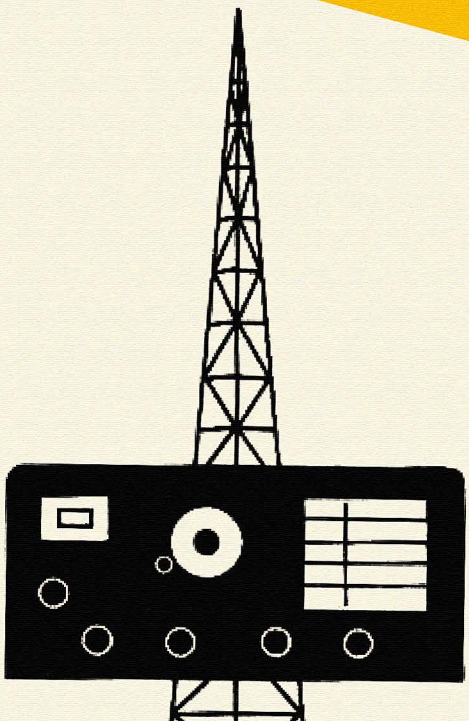
basic radio

by M. TEPPER



VOLUMES

1 to 6

a RIDER publication

basic radio

by MARVIN TEPPER

Electronic Services Division
Raytheon Company

Author of
FUNDAMENTALS OF RADIO TELEMETRY

VOL. 1

JOHN F. RIDER PUBLISHER, INC., NEW YORK

Copyright September 1961 by John F. Rider Publisher, Inc.

All rights reserved. This book or any parts thereof may not be reproduced in any form or in any language without permission of the publisher.

Library of Congress Catalog Card Number 61-11229

Printed in the United States of America

Fifth Printing, 1968

PREFACE

The purpose of this book is to fill a need for a text stating in plain, everyday language, what many people consider a complex technical subject. A technical subject need not be complex. Careful filling in of the background with essential information, and then leading step by step to the final explanation, provides a logical method of explaining the most difficult subject.

It would be impossible to cover in a single book or series of books, the immense scope implied in the word *electronics*. However, an understanding of radio circuits serves as a foundation for advanced study in all fields of electronics, such as television, radar, computers, etc. For teaching radio, the all-important basic tool of electronics, most available textbooks are woefully inadequate. One type contains information so brief as to acquaint rather than instruct. Another type is based on the premise that teaching a student to design a circuit is the best method of having him understand that circuit's operation.

Basic Radio represents the neglected middle ground. It is a course in radio communications, as distinct from a general course in electronics. The text deals with the circuitry and techniques used for the transmission and reception of intelligence via radio energy. Assuming no prior knowledge of electricity or electronics, the six volumes of this course "begin at the beginning" and carry the reader in logical steps through a study of electricity and electronics as required for a clear understanding of radio receivers and transmitters. Illustrations are used on every page to reinforce the highlights of that page. All examples given are based on actual or typical circuitry to make the course as practical and realistic as possible. Most important, the text provides a solid foundation upon which the reader can build his further, more advanced knowledge of electronics.

The sequence of *Basic Radio* first establishes a knowledge of d-c electricity. Upon this is built an understanding of the slightly more involved a-c electricity. Equipped with this information the reader is ready to study the operation of electron tubes and electron tube circuits, including power supplies, amplifiers, oscillators, etc. Having covered the components of electronic circuitry in Volumes 1 through 3, we assemble these components

in Volume 4, and develop the complete radio receiver, AM and FM. In Volume 5 we recognize the development of the transistor, and devote the entire volume to the theory and circuitry of transistor receivers and semiconductors. The last volume of the course, Volume 6, covers the long-neglected subject of transmitters, antennas, and transmission lines.

No prior knowledge of algebra, electricity, or any associated subject is required for the understanding of this series; it is self-contained. Embracing a vast amount of information, it cannot be read like a novel, skimming through for the high points. Each page contains a carefully selected thought or group of thoughts. Readers should take advantage of this, and study each individual page as a separate subject.

Whenever someone is presented with an award he gives thanks and acknowledgement to those "without whose help . . ." etc. It is no different here. The most patient, and long-suffering was my wife Celia, who typed, and typed, and typed. To her, the editorial staff of John F. Rider, and others in the "background", my heartfelt thanks and gratitude for their assistance and understanding patience.

MARVIN TEPPER

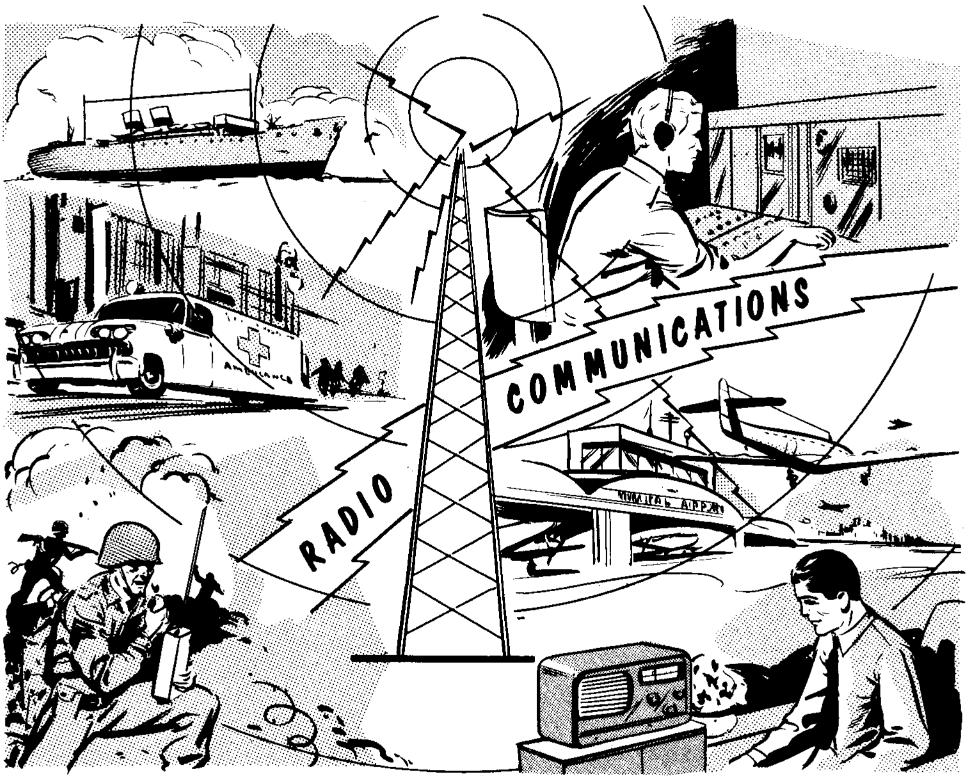
Malden, Mass.
September 1961

CONTENTS

VOL. I — BASIC RADIO

Introduction to Radio

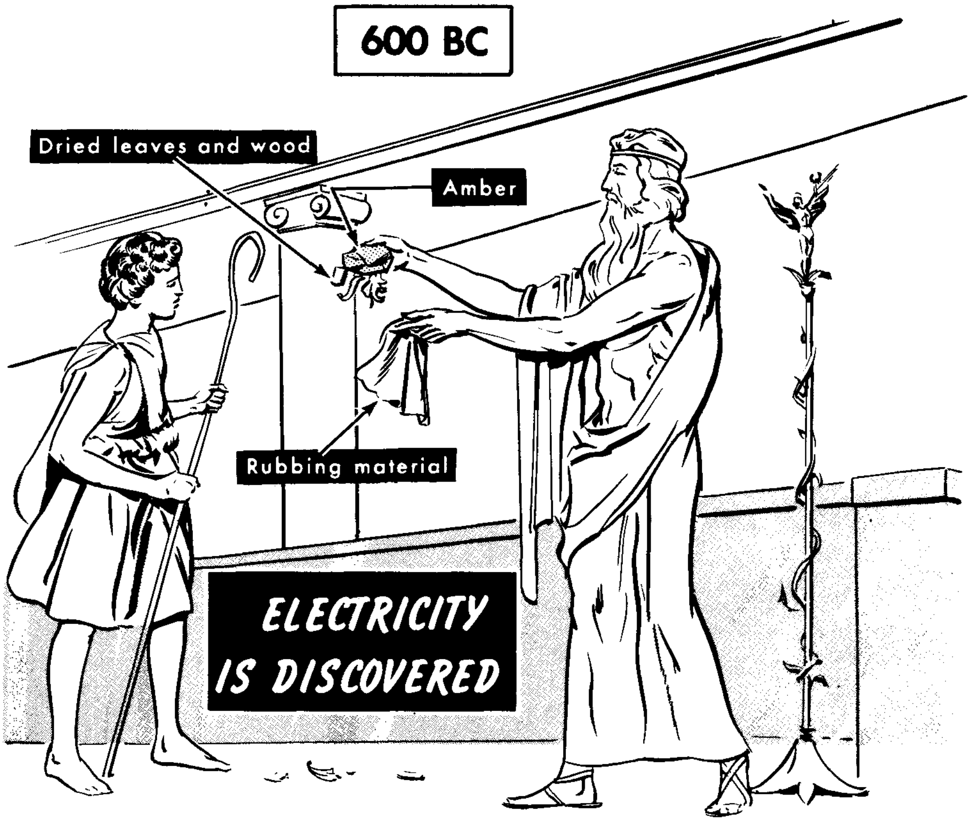
Radio is the name applied to the most successful and most frequently used facility for communicating with all parts of the world. Radio communication is speedy and reliable. An important news event -- a political incident, the death of a famous person, any significant occurrence anywhere in the world -- can be made known instantly everywhere else in the world through the medium of radio. The speed of the radio wave is approximately 186,000 miles per second, rapid enough to circle the globe at the equator slightly more than seven times per second. But the dissemination of news, culture and entertainment by radio broadcasting is but one of the many functions of radio. Police radio, marine radio, and aviation radio are equally important because they function to safeguard human life on land, on and beneath the sea, and in the air.



Without the benefits of radio, conquest of space would be impossible. The transmission of vital information from satellites to earth is one application of radio in this effort. Television is basically a radio system. The ability to detect an unseen hostile target and to locate its position and distance from a point of observation is another application of radio: it is called Radar -- RAdio Detection And Ranging. An understanding of the workings of radio is also the basis for understanding the widespread applications of electronics. And it all begins with a study of electricity.

The Early History of Electricity

Knowledge concerning electricity started with the Greeks. It is said that at about 600 BC, the Greek philosopher Thales of Miletus discovered that a certain substance (now known as amber), when rubbed with certain materials, displayed a peculiar force. It would attract tiny bits of dried leaves and wood to itself. Thales had no explanation for the action, but he gave the name electron to the substance. The word "electron" is in use today but its meaning is far different now, as you will learn.

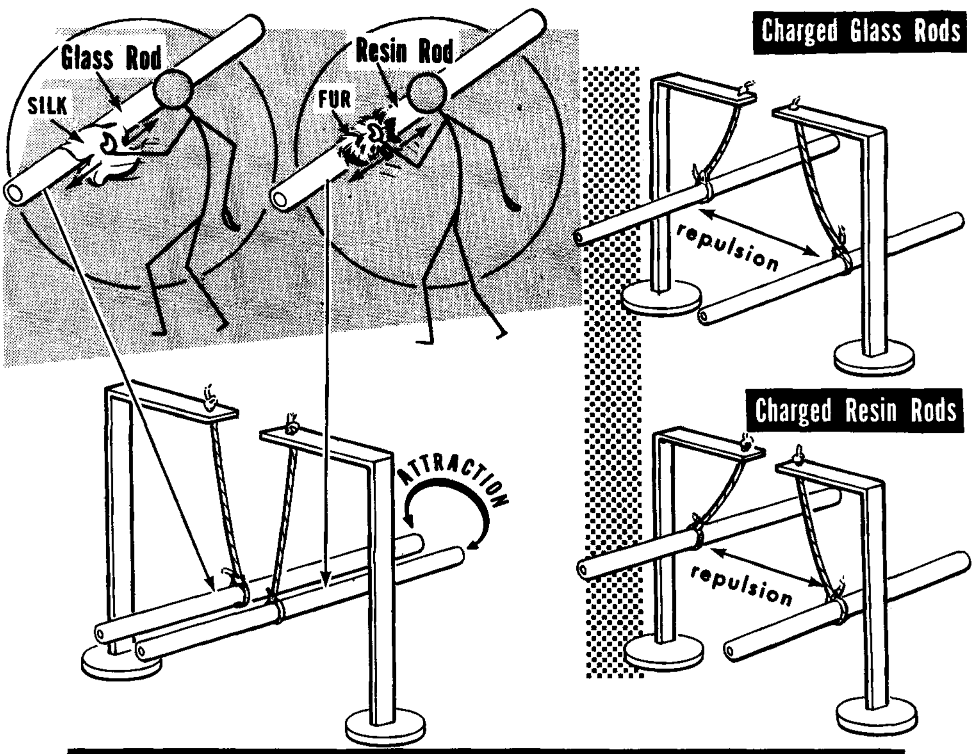


The behavior of amber remained a mystery for about a thousand years. But as time passed, more and more substances which behaved like amber were discovered. About the year 1600, an English scientist named William Gilbert compiled a list of so-called electriks, or substances that could be electrified or "charged with electricity", by rubbing (friction). When electrified, they could attract tiny objects such as bits of paper and threads of cloth. Then, in the early 18th century, experimenters with electriks discovered that many materials, when rubbed with other materials such as fur or wool, not only would attract tiny objects, but would attract or repel each other. The action was not understood and it was declared to be the display of electrical force due to a mysterious something known as electricity.

A Famous Experiment in Electricity

One of the far-reaching experiments performed in the field of electricity was the rubbing of a glass rod with a piece of silk, and the rubbing of a resin rod with a piece of fur, after which the two electrically charged or, simply, charged rods were suspended near each other and allowed to demonstrate their electrical effects on each other.

The charged resin rod and the charged glass rod attracted each other! The charged resin rod repelled another resin rod that had been similarly rubbed with the fur! The charged glass rod repelled another glass rod that had been rubbed with the silk!

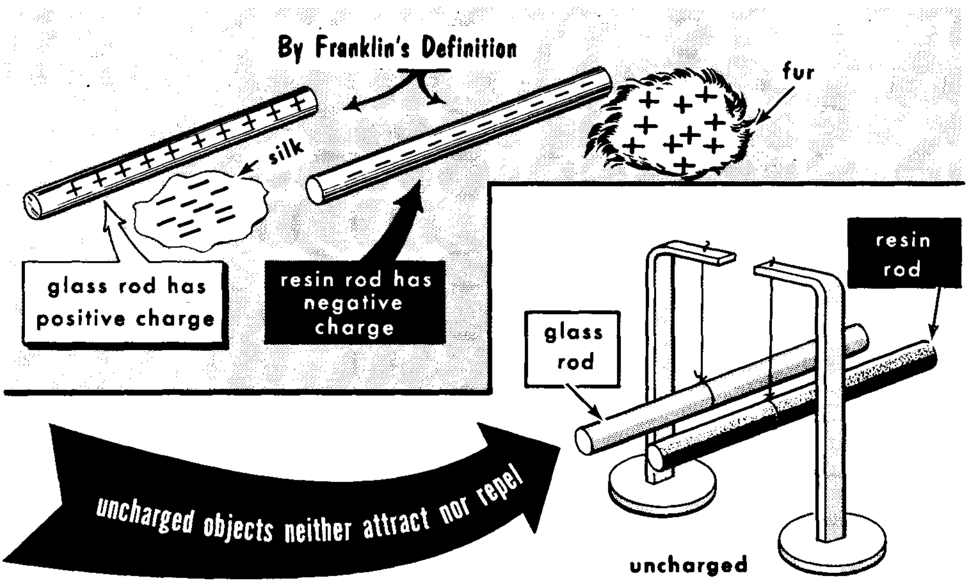


EARLY OBSERVATIONS MADE IN RUBBING GLASS AND RESIN RODS

It was also found that the action of rubbing the resin rod with the fur also had charged the fur. But the kind of charge was such that the fur and the resin rod were attracted to each other, whereas the charged glass rod and the fur repelled one another. The silk used for rubbing also displayed electrical effects. It too had acquired a charge when it was used to rub the glass rod, and its charge was such that it was attracted to the glass rod as well as by the fur, but it was repelled by the electrified resin rod.

A Famous Experiment In Electricity (contd.)

Many ideas were advanced as explanations for the actions observed. Charles Dufay, a French chemist, suggested that the different behavior of the charged rods was due to the presence of two kinds of electricity -- resinous in the resin rod, and vitreous in the glass. Benjamin Franklin, one of America's founders, recommended a change in the names identifying the two kinds of electricity. He suggested the name positive (symbolized by the plus sign +) for the kind of charge on the glass, and negative (symbolized by the minus sign -) for the kind of charge on the resin rod. These names were accepted and have been in use ever since.



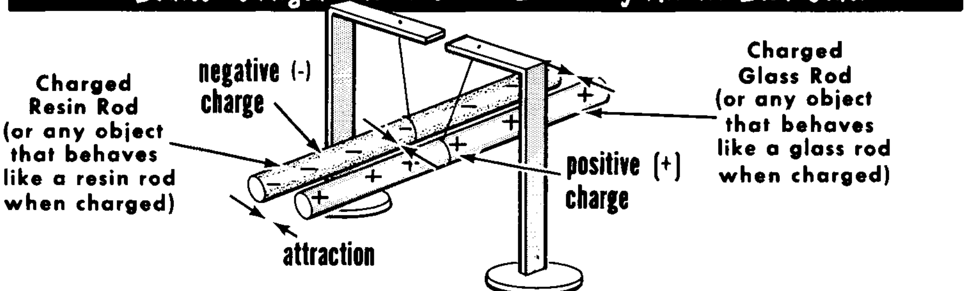
Franklin also suggested the idea that everything in its normal state, that is all objects which are not electrified, are made of equal amounts of positive and negative electricity. Such uncharged objects do not display electrical effects (attraction or repulsion) because the effects of one kind of electricity are offset by the effects of the other kind of electricity. He visualized the so-called electrified or charged condition of an object as being the presence of more of one kind of electricity than of the other kind. The part of the electrified resin rod which was rubbed was charged with negative electricity because it contained more negative electricity than positive electricity. The amount of negative charge was equal to the excess of negative electricity over positive electricity. In a practical sense, negative charge was produced in the resin rod by rubbing. The glass rod, on the other hand, was charged with positive electricity in the area where it was rubbed because it contained more positive electricity than negative electricity. The amount of positive charge was equal to the excess of positive electricity over negative electricity. Again, in a practical sense, positive charge was produced in the glass rod by rubbing.

A Famous Experiment in Electricity (contd.)

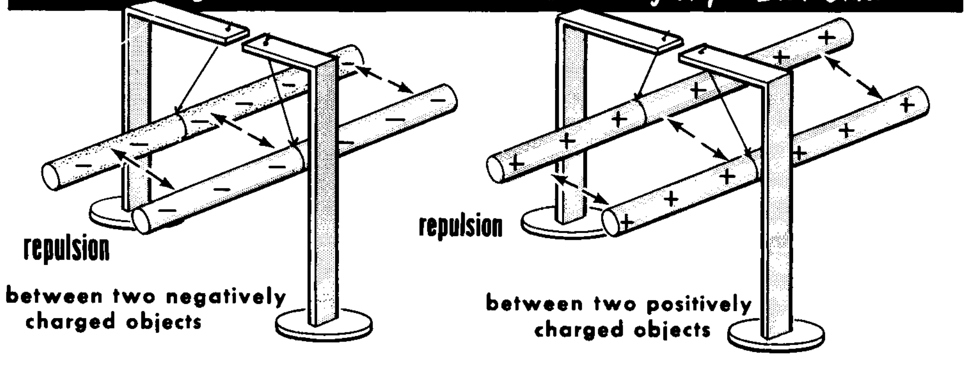
Franklin advanced several more fundamental ideas. He said that any object (body) which, when charged, behaved like the electrified resin rod, had a "negative charge." In other words, the glass rod was attracted to the resin rod and also the silk; therefore, the resin rod and the silk were charged with the same kind of electricity -- negative. By the same reasoning, any charged object (body) which behaved like the charged glass rod was charged with positive electricity. The resin rod was attracted to the glass rod and also to the fur; therefore, the glass rod and the fur were charged with the same kind of electricity -- positive.

It was observed that bodies which carried unlike charges -- the resin rod and the glass rod, the silk and the fur, the resin rod and the fur, the glass rod and the silk -- were attracted to each other. On the other hand, when the fur and the glass rod, or the resin rod and the silk (all of which carried the same kind of charge) were allowed to act on each other, it was noted that they repelled each other. From these findings came certain fundamental conclusions: namely, that bodies charged with unlike electricity attract each other, whereas bodies charged with the same kind of electricity repel each other. These experiments and observations did not explain what electricity was; they simply dealt with observable behavior of the assumed two kinds of electricity. Interestingly enough, these conclusions were correct and have become fundamental laws of electricity.

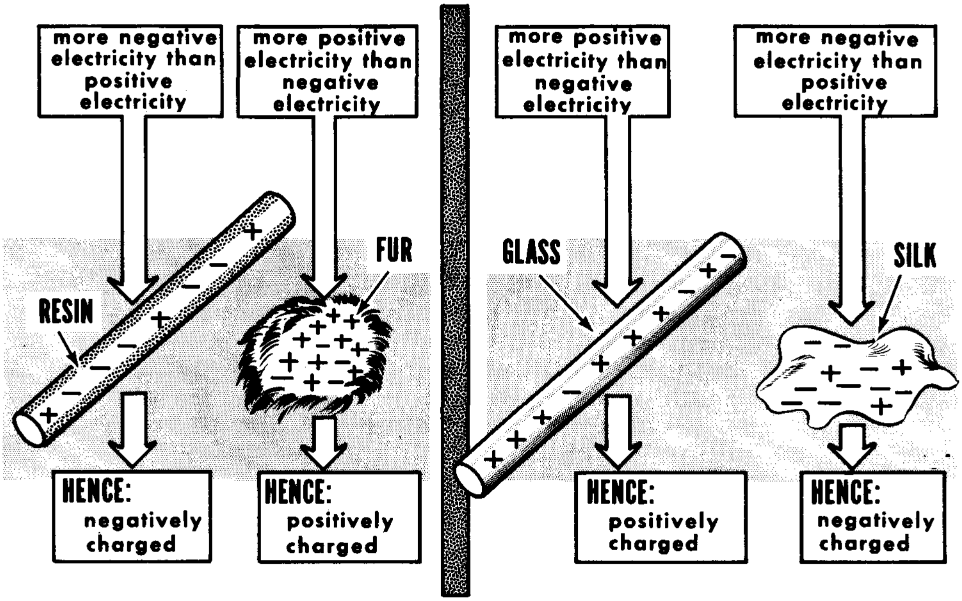
Bodies Charged With Unlike Electricity Attract Each Other



Bodies Charged With The Same Kind Of Electricity Repel Each Other



Electrical Charging By Rubbing



Early Explanation Of Negatively and Positively Charged Objects

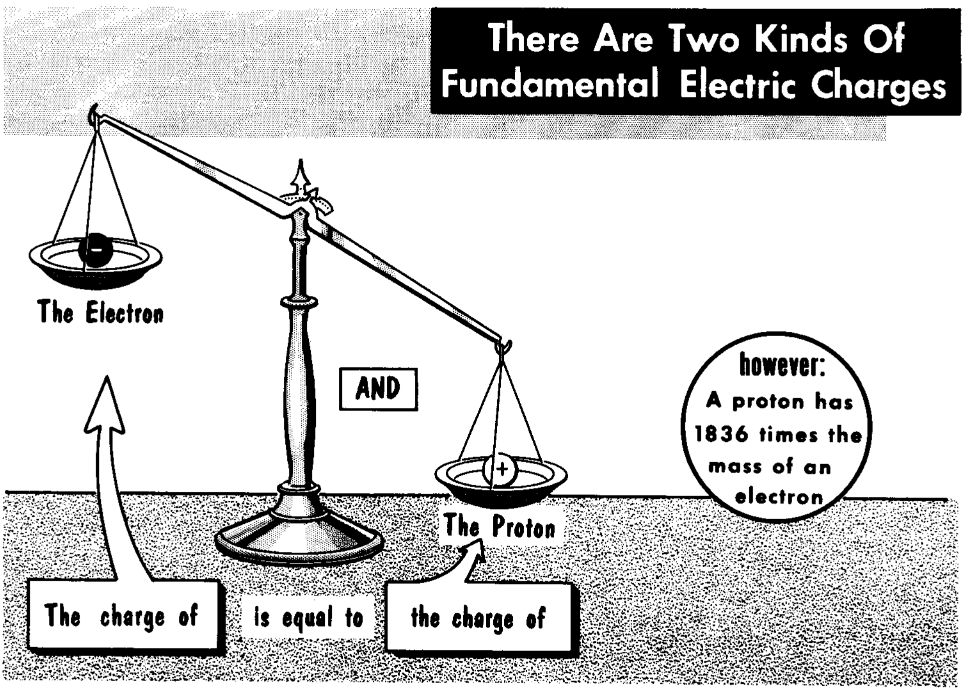
Franklin explained "charging by rubbing" in the following way. When the resin rod was rubbed with the fur, the surface friction caused some of the positive electricity in the resin to go to the fur. Now the rubbed part of the resin rod had more negative electricity than it had positive electricity; hence, it was negatively charged. On the other hand, the part of the fur used for rubbing now contained more positive electricity than negative electricity; therefore, it was positively charged. In the case of the glass rod rubbed with the silk, the silk gave up positive electricity which went to the glass, giving the rubbed portion of the glass rod a preponderance of positive electricity; therefore, a positive charge. Having given up positive electricity, the part of the silk used for rubbing was left with more negative electricity than positive electricity; hence, it was negatively charged.

It so happens that Franklin was not correct in his identification of which kind of electricity went where. Science has since learned that it was negative electricity which was displaced; from the fur to the resin rod, and from the glass to the silk. But the identities of the final charge created on the electrified resin rod and on the electrified glass rod as established then are in use today. Equally important, as you shall learn later on, is that several other concepts advanced at that time are in use today; namely that everything is made of two kinds of electricity and that one kind of electricity can be separated from the other kind. But before we can discuss the modern versions of these happenings, we must develop the atomic concept of matter and the electronic concept of electricity.

The Modern Concept of Electricity

Although our concepts of electricity differ substantially from the ideas of Franklin's time, the pattern of modern thinking, strangely enough, follows the explanation of electrical behavior as expressed then.

There are two kinds of electricity. Still using Franklin's terms, we speak of them as positive and negative. Whereas in Franklin's day electricity was visualized as a fluid, today we believe that electricity exists as minute, virtually weightless spherical specks or particles -- the positive electricity particles being called protons, and the negative electricity particles being called electrons. We don't know if the proton is, or just carries, a certain amount of positive electricity, or if the electron is, or just carries, a certain amount of negative electricity. Either concept is acceptable because it leads to the same thing.



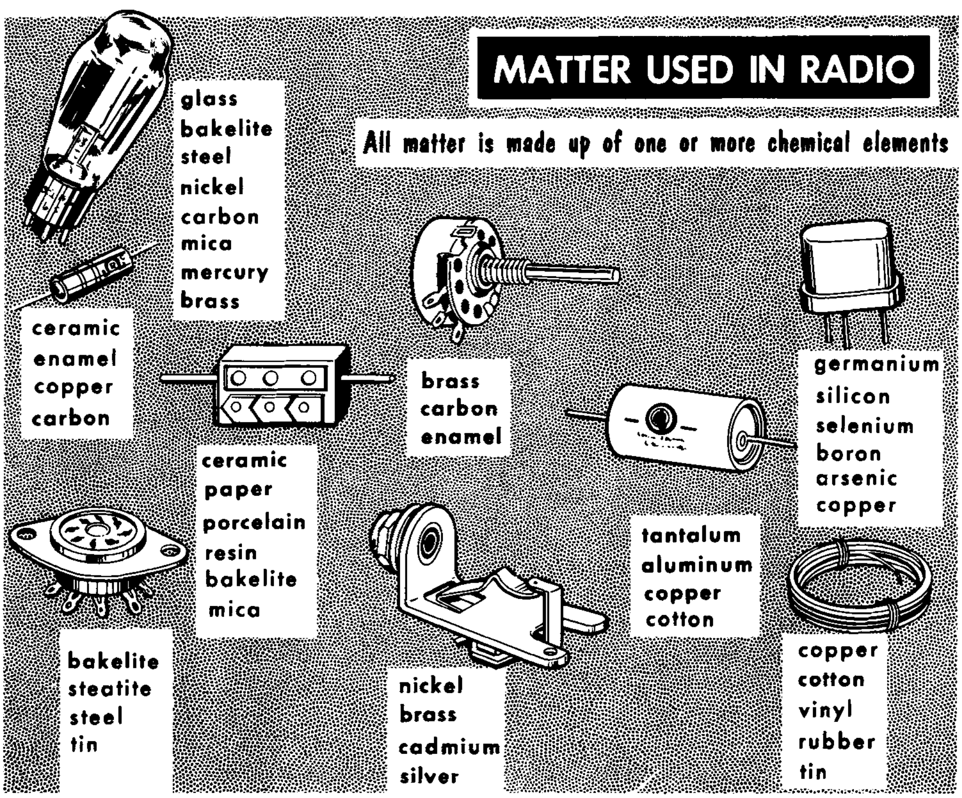
Man has never seen a proton or an electron because the particles are too tiny for even the greatest known magnification to make them visible. But this has not prevented the development of certain ideas about them. For instance, it has been established that the amount of positive electricity associated with a proton is exactly equal to the amount of negative electricity associated with an electron. Each is the smallest amount of electricity of its kind known. Therefore, the proton is the fundamental charge of positive electricity, and the electron is the fundamental charge of negative electricity. Each is an impractically small amount, but finite just the same. Operating electrical systems involve the motion of fantastic quantities of electrons, as you shall learn.

Matter and Chemical Elements

How do the specks of electricity -- protons and electrons -- fit into the scheme of things? The answer is that they are the main ingredients of everything.

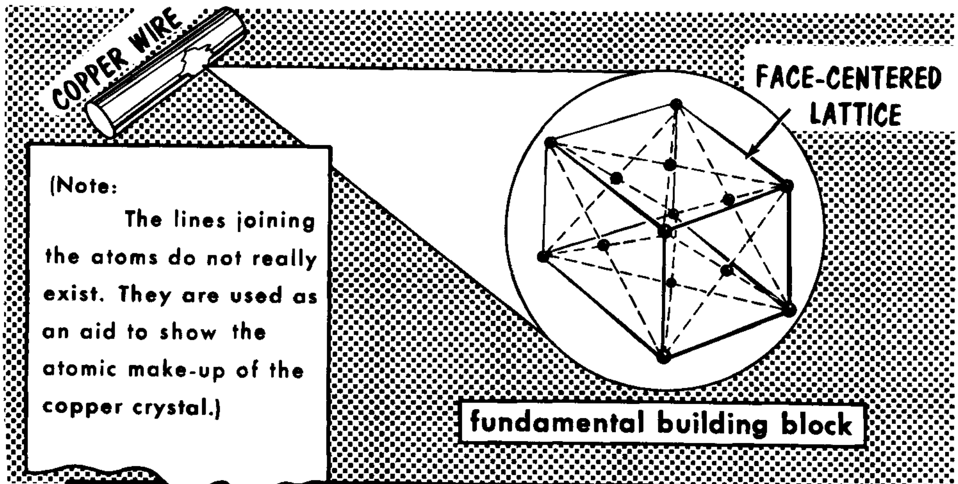
Everything in the world you can see or touch, and even the many things invisible to the naked eye but known to exist, make up the matter of the universe. Matter exists in solid, liquid, or gaseous states. Blood, skin, bone, steel, water, glass, rubber, powder, gold, smoke, copper, tobacco, and air are just some of the many examples of matter. Chemically speaking, all matter is made of one or more of the 102 "pure" substances identified as chemical elements. The word "pure" as applied to a chemical element is that it consists of only one substance. By definition, a chemical element is a substance which cannot be subdivided into two or more different substances by any known chemical means, nor can it be produced by the chemical combination of two or more different substances.

We are interested in chemical elements because so many of them are used as the ingredients that make up the various components that comprise electrical circuits and equipment. In some instances, elements such as copper, aluminum, and carbon are used directly; at other times, compounds (combinations of different elements) are used in the construction of electrical and electronic components.



The Atom

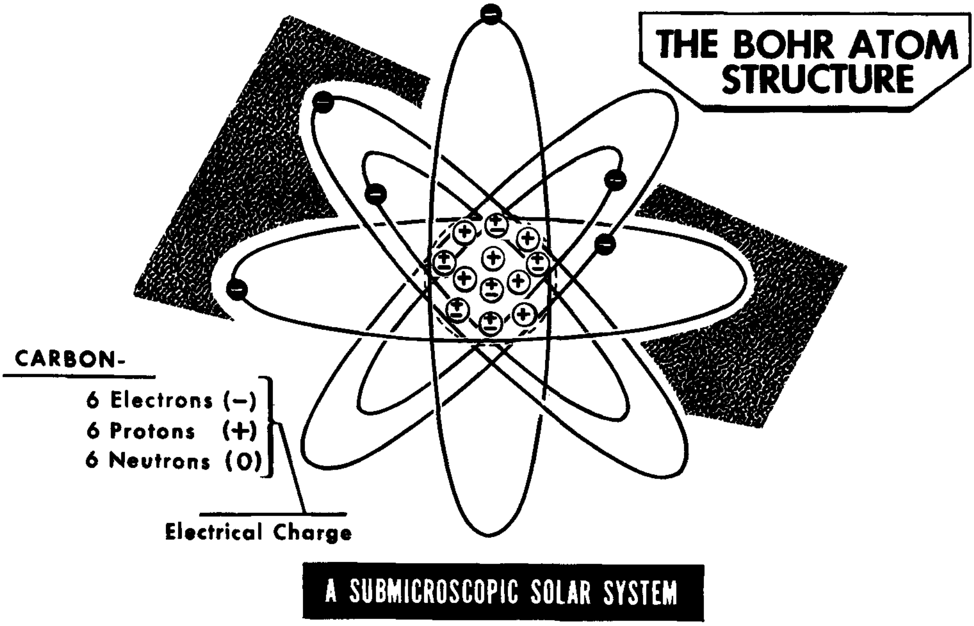
If a small quantity of a chemical element, solid, liquid or gas, could be reduced in amount by a continuous series of subdivisions, the fundamental building block of the element eventually would be reached. This is the atom. Science knows no way of performing such a subdivision, and consequently no one has ever seen an atom. But even though we lack this ability, we accept the idea that nature has built everything out of atoms. Atoms form the chemical elements, and chemical elements singly, or in combination make up everything else; therefore, in the fundamental sense, everything is made of atoms.



THE COPPER CRYSTAL -- THE TINIEST PIECE OF COPPER THAT BEHAVES LIKE COPPER. IT CONSISTS OF 14 ATOMS IN A FIXED FORMATION.

The smallest possible amount of some gaseous elements which display the properties of the gases is a molecule, made up of two atoms that are in an electrical bond. Oxygen is one of these, Hydrogen and Chlorine are others. Then there are gaseous elements for which the smallest quantity of the substance is a single atom. Helium, Argon and Neon are examples. In the case of the elemental metals, a different situation prevails. The smallest amount of the element is a single atom, as for instance copper, silver, gold, iron, lead, tin and others, but the smallest amount of the metal which displays the physical properties of the metal -- expansion, contraction, malleability -- is a crystal. (Sometimes, the crystal of a metal is called a giant molecule). A crystal of a metal is an organization of atoms of the element arranged in a particular formation. For instance, the tiniest amount of pure copper which will behave like the metal is 14 atoms of copper arranged in a geometric pattern called a face-centered lattice. A length of copper wire would be made of a tremendous amount of these crystals. Other elemental metals have their atoms arranged in different formations. Remembering the geometric pattern of the crystal of copper is not vital to understanding the workings of electricity.

Every Atom Is Made of Electricity

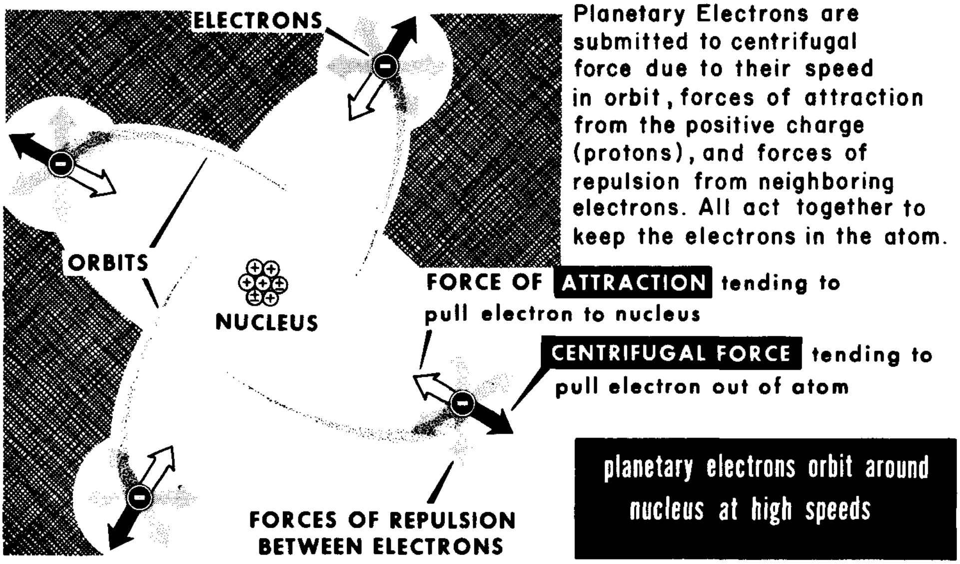


Every atom (regardless of kind) is an organized structure whose main parts are comprised of particles of positive electricity (protons) and particles of negative electricity (electrons). When we say that one atom is different from another (each being representative of a different kind of chemical element), we mean simply that the two atoms are made of different amounts of positive electricity, with each having as much negative electricity as positive electricity.

A popular concept of the atom was advanced by Niels Bohr, a Danish physicist. He visualized the atom as having a stationary center (or nucleus) in which was concentrated all the positive electricity of the atom (all the protons). Also present in the nucleus of all but the Hydrogen atom was still another kind of particle called a neutron. Each kind of atom contained a different number of neutrons, but since the neutron contributes nothing of electrical character to the atom, we need not discuss it any further. Revolving around the nucleus with very high velocity and at different distances from it are the specks of negative electricity -- planetary electrons -- which balance the positive electricity content of the atom. In any one kind of atom, there are associated with it as many planetary electrons as there are protons in the nucleus, or as much negative electricity as positive electricity. As a convenience in identifying the different kinds of atoms, each kind is associated with a number and with one or two letters. The numerical reference states the number of protons in the atom (hence, the number of planetary electrons too), while the letter symbol identifies the chemical element. As typical examples, the Hydrogen atom (H) is #1; the Carbon atom (C) is #6; the Copper atom (Cu) is #29; the Silver atom (Ag) is #47, and so on.

Electrical Forces In The Atom

What keeps the circling electrons from flying out of the atom under the influence of the centrifugal force that each planetary electron feels? The general concept is that the electrons are held in their orbits against the pull of centrifugal force by the electrical force of attraction between the protons in the nucleus and the orbiting electrons. The electrical force is manifested as a mechanical force; it pulls the electrons inwards towards the nucleus while the centrifugal force is pulling the electrons outward away from the center. The two forces are in exact balance; therefore, the electrons do not leave the atom nor do they "fall" into the nucleus. Each fundamental particle of electricity is inseparably endowed with the property of attracting an oppositely charged particle to itself, and repelling every similarly charged particle. The protons in the nucleus also feel a pull towards the electrons, but the much greater mass of the individual proton keeps it put where it is.



A force of repulsion exists between each orbiting electron and acts in all directions, thus keeping the planetary electrons in their positions in the orbits. A similar force of repulsion exists between individual protons which are separated, but when they are as densely packed as they are in the nucleus, there is a "something" which accounts for their not flying apart.

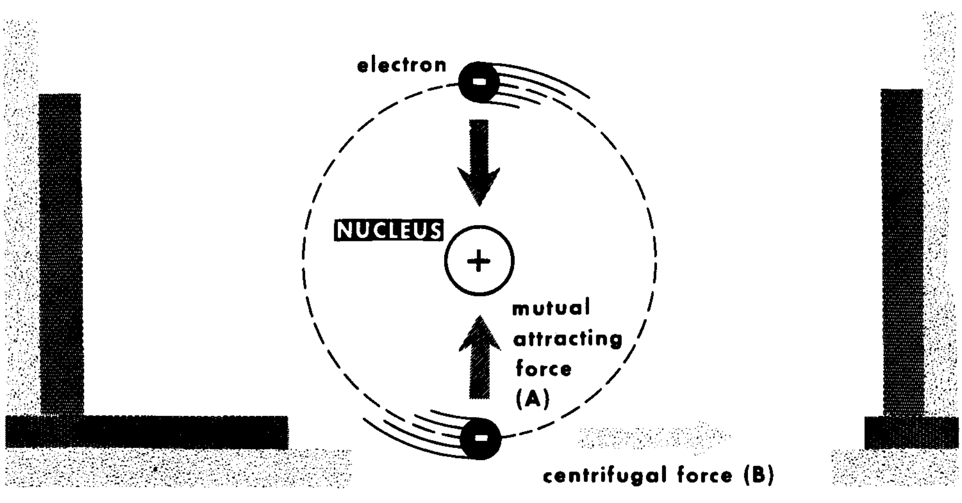
The electrical forces in the atom are fundamental and tremendous. Were this not so, the universe would fly apart. As long as the atom is in electrical equilibrium, that is, equal numbers of protons and electrons, the electrical and mechanical force condition is confined strictly to the inside of the atom. Neither the protons nor the electrons inside an electrically-neutral atom (equal numbers of protons and electrons) have any effect on other electrons and protons outside of the atom. This subject will receive more attention later, as we develop several related ideas.

Bound Electrons

The electrical forces referred to on the preceding page exist inside every atom but the behavior of these forces is not exactly the same in the different kinds of atoms. For our purposes, we need not consider the different kinds of atoms on an individual variety basis; it is sufficient if we deal with two main categories--nonmetals and metals. We shall consider each separately.

Under all normal conditions, the force of attraction between the protons in the nucleus and the planetary electrons in the atoms of nonmetals is sufficiently strong to keep all the electrons tightly "locked" inside the atomic structure. The electrons revolving in orbits close to the nucleus, as well as those revolving at the farthermost distances from the center of the atom, are "bound" to the atom. An occasional atom may let go of one of its outermost electrons, but by and large, we assume that the nonmetallic materials are made mainly of atoms which are in electrical balance (always have equal numbers of protons and electrons). They are electrically-neutral atoms.

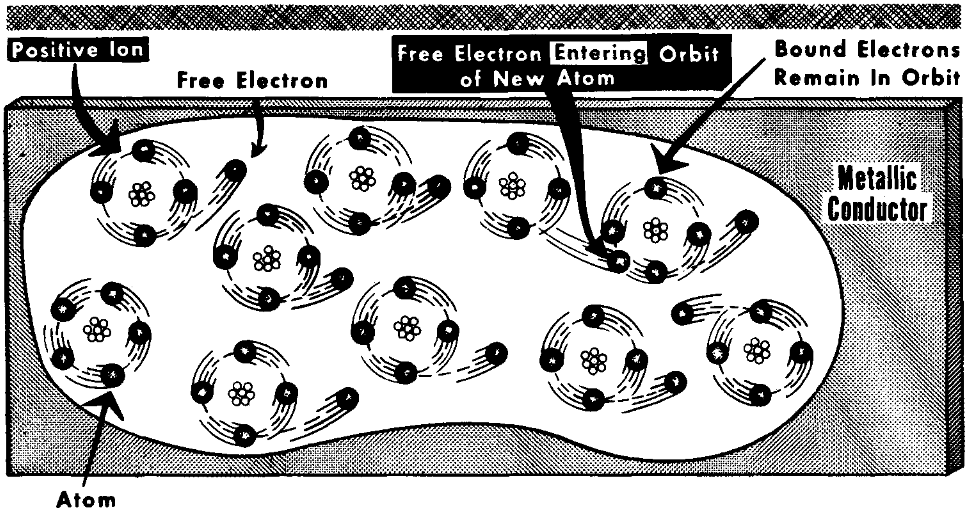
WHEN FORCE A EQUALS FORCE B, ELECTRON REMAINS IN ORBIT (IS BOUND)



This situation is not an unchangeable one Given a different set of conditions (which we might call "abnormal") such as the application of a sufficiently strong external force--one strong enough to overcome the binding forces between the protons and outermost electrons inside the atomic structure--it is possible to upset the electrical balance. It is possible to literally "tear" one or more electrons out of the atom, or even temporarily "add" an electron to the atom. Strangely enough, such external forces can be developed rather easily. Rubbing the resin rod with the fur and the glass rod with the silk are examples of "abnormal" conditions--that is, the application of such external force. The surface atoms of the fur and the glass rod released electrons, whereas the surface atoms of the resin rod and the silk temporarily accepted them.

Free Electrons In Metals

We have explained how the electrons in the atoms of nonmetals are bound to the atom. On the other hand, it is an accepted theory that the protons in the nucleus of the atom of a metal hold onto all but one of the normal complement of planetary electrons.



FREE ELECTRONS MOVE IN RANDOM MOTION THROUGHOUT METALLIC CONDUCTOR

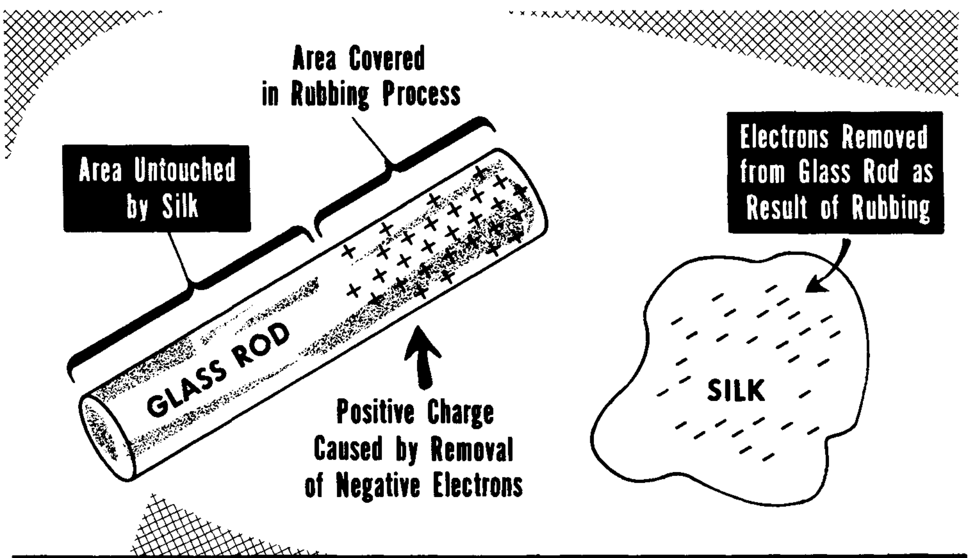
The situation is explained in the following way. The electron (or electrons) that orbits at the farthest distance from the nucleus in the atom of metals is believed to follow an elliptical path. (An ellipse is a geometrical figure which has the shape of a hoop that has been flattened slightly). The other orbits are assumed to be circular. As a case in point, let us assume an atom of copper. The outermost orbit is occupied by a single electron, this orbit being elliptical. At one point in the path of travel, the electron is very close to the nucleus; at another point, the electron is far removed from the nucleus. When the electron is farthest from the nucleus, it is released because the force of attraction between it and the cluster of protons is not sufficiently strong to keep the electron in its orbit. It is believed that this action occurs in the atoms of all metals, although not exactly to the same extent in each kind.

The liberated electrons are called "free electrons". They wander among the atoms of the metal throughout all parts of the metal in a random manner (presumably uniformly distributed throughout the metal), as many moving in one direction as in another. Every atom which has lost an electron now has a preponderance of positive charge, that amounting to one proton. These atoms are called "positive ions". However, for every positive ion in the metal there is a free electron so that the balance between positive and negative electricity is maintained and the metal as a whole remains electrically neutral.

The Modern Concept of Charging

The modern concept of "charging with electricity" or simply, "charging" is nothing more than disturbing either the equality of the electron and proton content in an object, or the uniform distributions of the negative and positive electricity content.

The four examples of nonmetallic substances used in the Franklin experiment became charged because electrons were torn from the surface atoms of one material (the fur and the glass rod) and transferred to the surface atoms of the other material (the resin rod and the silk) during the rubbing process. The emphasis on the surface atoms is made for a reason. The atoms of all solid substances are to all intents and purposes fixed in their locations. When electrons are transferred from the surface atoms of one nonmetallic object to surface atoms of another nonmetallic object, the negative charge is transferred from one particular place to another. Whatever charge is given up or acquired by rubbing, the action occurs over the area where the friction took place. The charged condition is therefore localized to certain places on the surface. This point is stressed because the action is somewhat different in objects made of metals.

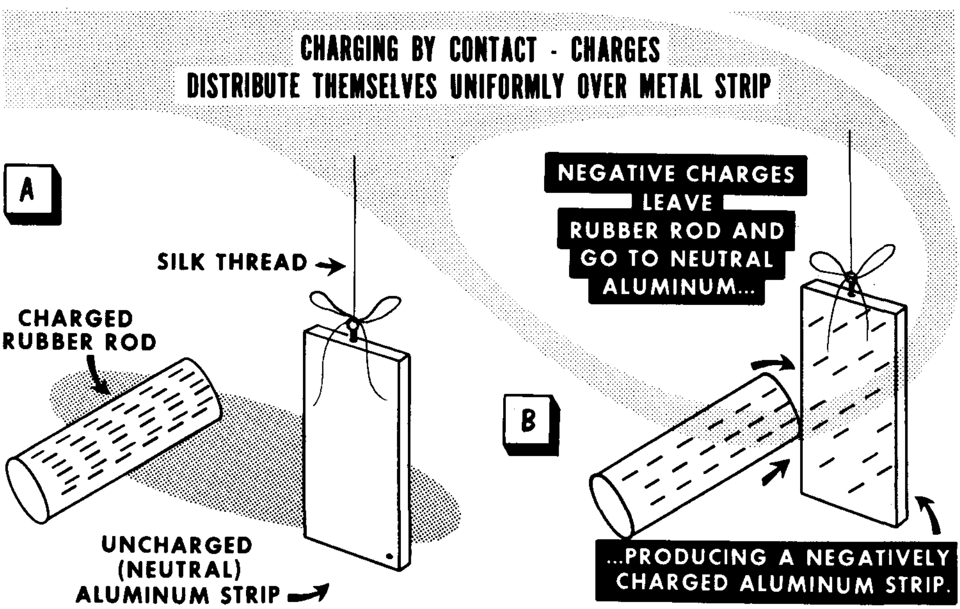


CHARGE ACQUIRED BY RUBBING OCCURS ONLY OVER AREA WHERE FRICTION TOOK PLACE

Charging By Contact

The frictional method of charging (rubbing) is applicable to nonmetals and metals alike, but not too successfully to the latter. A preferred method of charging metallic objects is by contact. We are concerned with the charging of metal objects because metals are used in radio equipment.

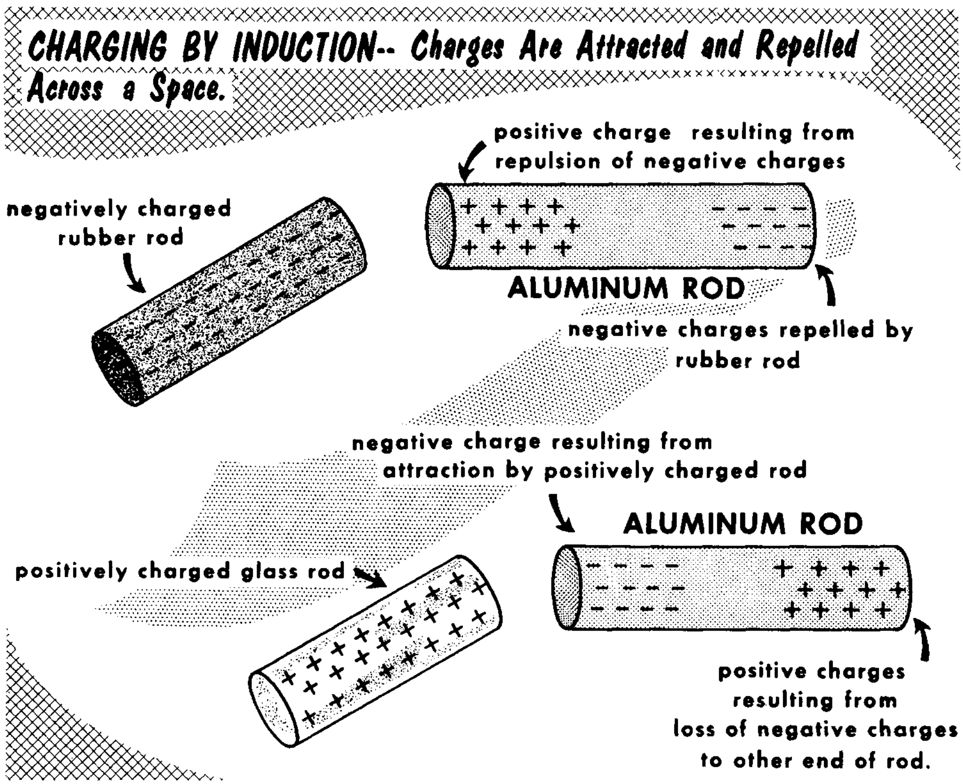
Let us assume we have a hard rubber rod which has been given a negative charge by some means. (Hard rubber behaves like resin). We also have a small strip of aluminum which is electrically neutral. The strip is suspended by a silk thread. We now make momentary contact between the charged rubber rod and the metal strip. Some of the electrons leave the rubber and go to the aluminum. Having given up some of its electrons, the rubber now has less negative charge; but now, the metal strip also has a negative charge. The total of the acquired electrons, plus the free electrons, plus the bound electrons in the atom of the metal exceeds the total of the protons in the metal. Thus, the total negative electricity content exceeds the total positive electricity content; hence, the net negative charge. However, unlike the behavior of the nonmetallic object, the electrons passed on to the aluminum distribute themselves uniformly over the entire surface of the strip, after which (to all intents and purposes) they are at rest.



In other words, the acquired negative charge is not localized to the point of contact with the rubber. It exists everywhere on the aluminum surface. If a positively-charged glass rod had been used as the charging body in place of the negatively-charged rubber rod, some of the free electrons moving along the surface of the aluminum would have gone to the glass, leaving a preponderance of positively-charged atoms (positive ions) in the metal. The metal strip then would have been charged positively.

Charging By Induction

There is still another method of charging an object. It is called charging by induction. We shall apply it to metallic objects rather than nonmetallic because the practical applications of the phenomenon involve metallic components.



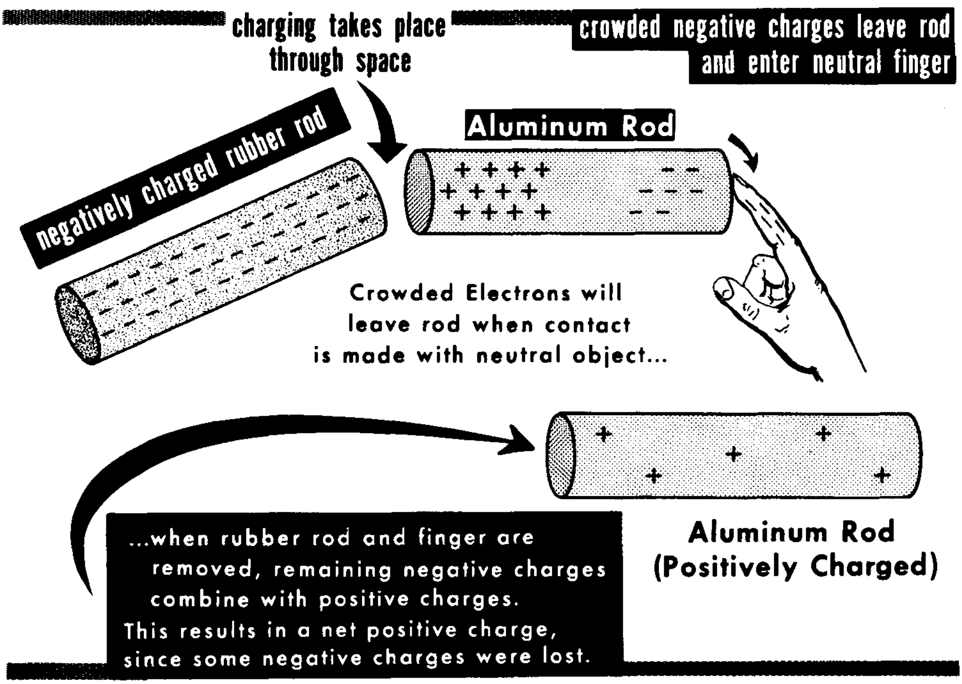
Imagine an uncharged aluminum rod suspended by a silk thread. (Any other metal could by used). Also a negatively-charged rubber rod. The negatively-charged rubber rod exerts an influence on the free electrons within the aluminum rod, acting across the space separating the two rods. Complying with the law that like charges repel, the rubber rod drives the free electrons wandering in the aluminum rod away from the end nearest the rubber rod. Now there is a decrease in the number of free electrons at that end of the aluminum rod which is nearest the charged rubber rod and a crowding of free electrons at the other end of the metal rod. This condition exists as long as the charged rubber rod is held near the metal rod.

If, instead of using a negatively-charged rod as the charging body, we used a positively-charged glass rod, the free electrons in the metal rod would be attracted towards the end which is nearest the glass rod, causing a crowding at this end and a shortage at the other end.

Charging By Induction (cont'd.)

Let us continue the experiment with the metal bar being acted upon by the negatively-charged rubber rod. As shown in the illustration, some of the free electrons are crowded at the end away from the rubber, having been repelled from the end near the rubber rod by the energy associated with the negatively-charged rubber.

Now we momentarily touch the end of the metal rod farthest from the rubber. The human body is not the best-known path for electricity, but it is good enough for this purpose. The crowded electrons tending to repel each other readily leave the metal and go to the finger. Then, we remove the charged rubber rod from the vicinity of the metal rod. The free electrons inside the metal redistribute themselves uniformly throughout the metal, but now, there are more atoms shy electrons than there are free electrons in the metal. The metal rod contains more positive electricity than negative electricity; hence, it has acquired a positive charge. It is to be noted that a negatively-charged charging body induces a positive charge in the original electrical neutral body. In other words, the charge induced is the opposite of the inducing charge.

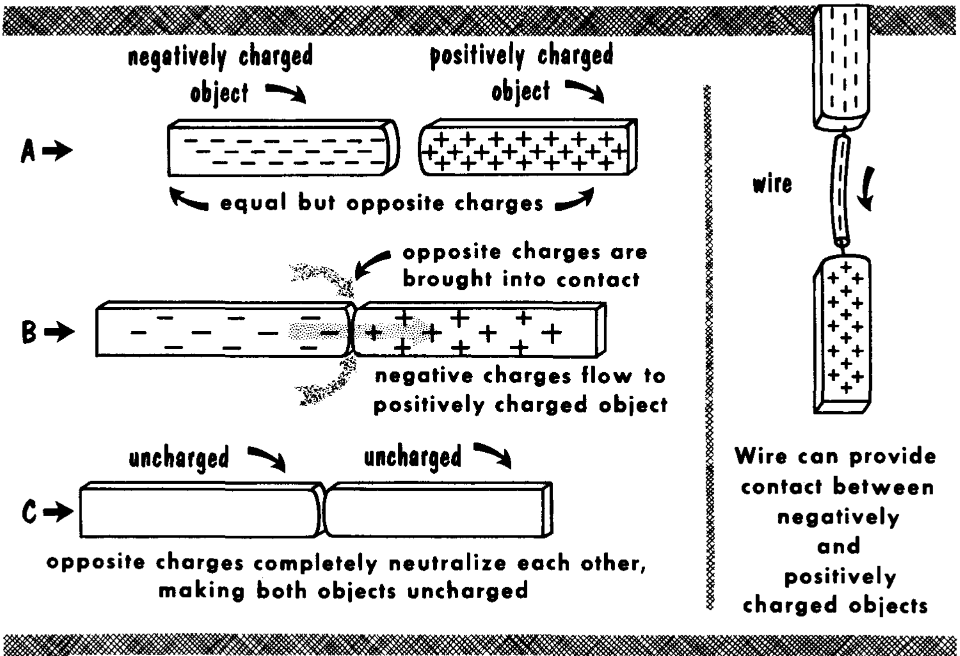


If the charging body had been a positively-charged glass rod, the metal would have had a shortage of free electrons at the end farthest from the glass rod. The positively-charged atoms at this end would pull electrons from the finger that touched the metal. Removing the charged glass rod from the vicinity of the metal would then leave the metal with more free electrons than positively-charged atoms -- or a net negative charge.

Discharging A Charged Object

You have learned that the electron and the proton are associated with electrical energy. If we think in terms of theory only, we say that a negative and positive charge are subject to a force which attracts each to the other; similar charges tend to repel each other. If we take the practical viewpoint, we realize that it is the electron, the very much lighter of the two fundamental charges of electricity, which performs the motion. (Each proton has about 1837 times the mass of an electron). Moreover, the positive electricity is locked inside the atom; it does not move under the influence of forces that act between charges. If we give a free electron the opportunity to move from one place to another under the influence of a nearby positive charge, it will do so, but the positive charge will not move.

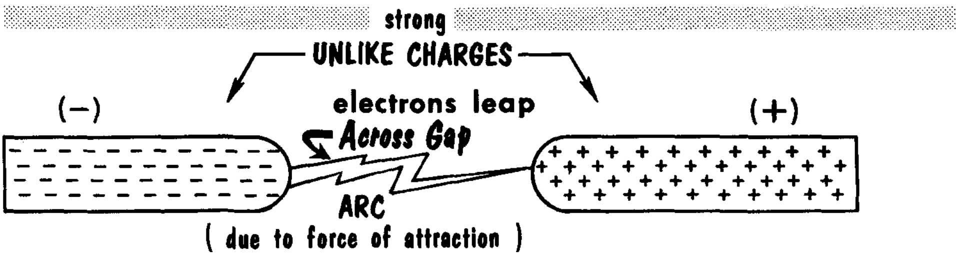
If we arrange for a negatively-charged object and a positively-charged one to make physical contact, the point of contact becomes the path over which the excess electrons on the negatively-charged object move to the positively-charged atoms on the positively-charged object. If both objects have equal amounts of unlike charge, all the electrons corresponding to the negative charge will flow to the positively-charged body, where they will neutralize the positive charge and create an electrically-balanced (neutral) condition in both objects. We have, in fact, discharged both charged objects by physical contact. Another kind of physical path could be a piece of wire which touches both objects at the same time. The flow of electrons through the wire is another story in itself.

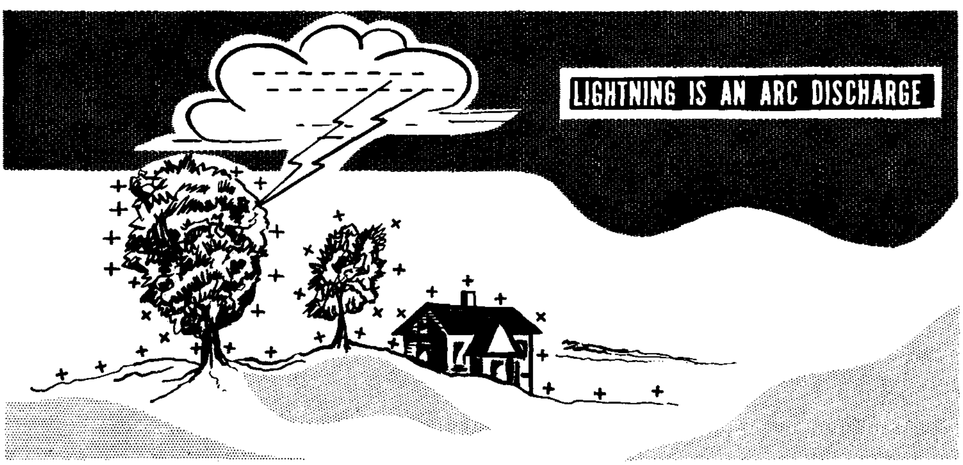


CHARGED OBJECTS CAN BE DISCHARGED BY CONTACT

Discharging A Charged Object (cont'd)

Objects bearing unlike charges can be discharged without using a physical path between them. Assume two objects, one with a very strong negative charge, the other with an equally strong positive charge. If the two charged objects are brought sufficiently close to each other but not touching, the excess electrons on the negative object will, under the influence of the force of attraction between unlike charges, leap across the gap to the positively-charged object. The movement of electrons through air produces a visible flash of light accompanied by an audible crackling sound. The visible phenomenon is called an arc. The greater the accumulation of the unlike charge on the two objects, the wider the separation that can be bridged by the arc. Lightning is an example of arc discharge between dissimilarly-charged clouds, or between a charged cloud and earth.



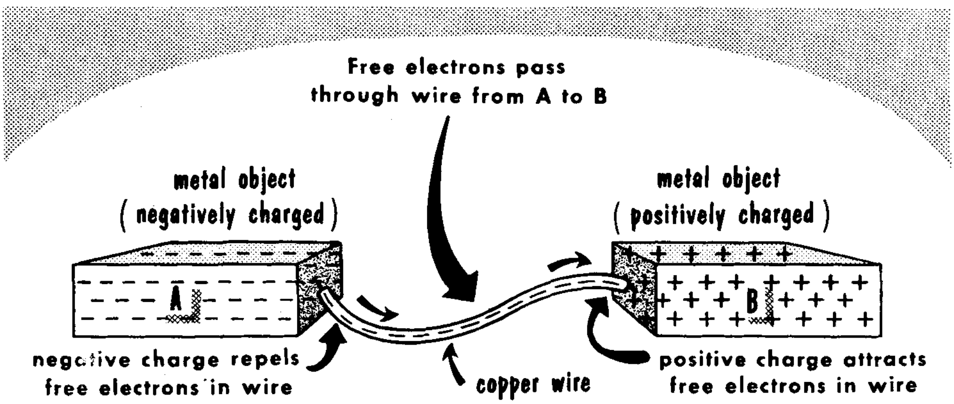


If all the surplus electrons on the negatively-charged object move over to the positively-charged object, both objects become electrically neutral. That is, the charged objects are discharged; the negatively-charged object because it has lost its surplus electrons, and the positively-charged object because it has acquired enough free electrons to neutralize the excess positively-charged atoms (positive ions) of the metal. It is conceivable that not all of the surplus electrons will move to the positively-charged object during the arc, in which case both objects will remain in a charged state, with each now bearing less charge than before the occurance of the arc.

Discharging Through a Wire

It is easy to visualize the action of discharging two charged objects by direct contact. The excess electrons on the negative object stream to the positive ions on the positive object. But how does the discharge action take place through a wire which joins two oppositely-charged metal objects? The answer is important to the study of electricity.

Let us assume the following conditions: The negatively-charged object (A) bears great numbers of excess free electrons along its surface. The positively-charged object (B) has an equivalent shortage of free electrons along its surface. Finally, a piece of copper wire is connected between the pieces of metal. When the copper wire joins A and B, the excess free electrons on A have free electrons which they can repel (the free electrons in the wire). Hence, the wire is a path into which the excess free electrons on A can move. Simultaneously, the excess positive ions on B now have free electrons (in the wire) on which they can exert a force of attraction. Thus, the free electrons in the wire feel a force of repulsion at one end and a force of attraction at the other. Actually, the electrons throughout the wire are subject to these forces. The net result is that the excess free electrons on A enter the wire at the same time the excess positive ions on B are pulling free electrons out of the wire. Of course, as excess electrons leave A, the amount of negative charge diminishes, and as they go to B the amount of positive charge diminishes.

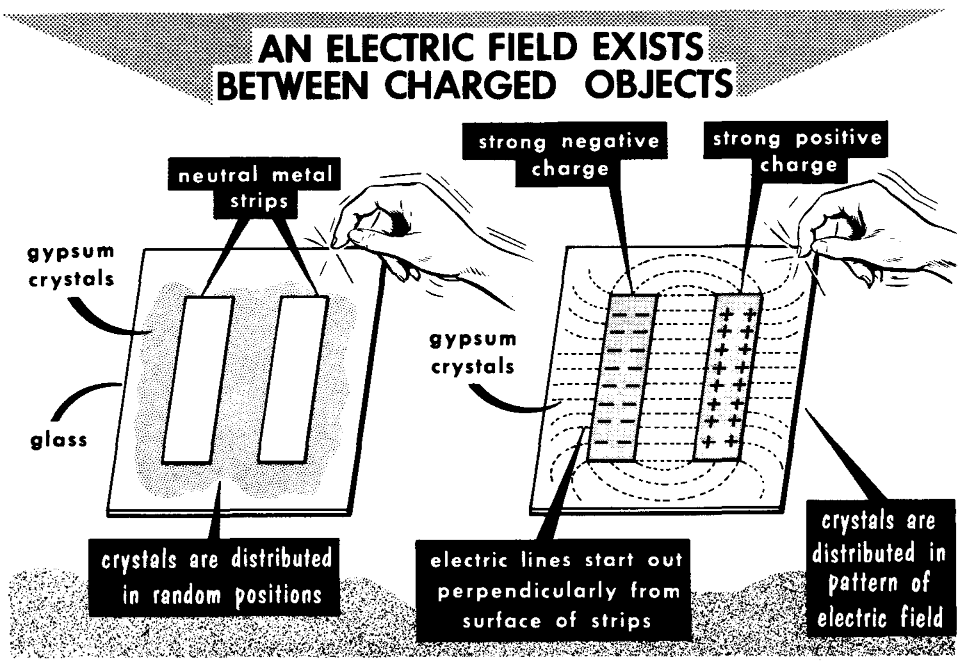


DISCHARGING CHARGED OBJECTS THROUGH WIRE
-- as many electrons leave wire and enter B as enter the wire at A

When all the excess electrons have left A and entered B, both strips are fully discharged and no further movement takes place between the two metal objects through the wire. Note that the movement of electrons through the wire is in a single direction -- from negative to positive -- and it happens without any change in the number of free electrons that are present in the wire at any one time, because as many leave the wire as enter it.

Demonstrating the Electric Field Concept

You have learned that the motion between charges or charged objects (attraction or repulsion) is caused by the presence of an electric force. We also said that the charging process was the equivalent of storing electrical energy. But where is the force which moves charged objects? And where does the storage of electrical energy take place? Strangely enough, both are found in the same place -- in the space between the charged objects. This is explained by the concept of the electric field of force, also known simply as the electric field. The space between and all around charged objects is filled with electrical energy--the energy of the electric field that is associated with the charge on the objects. This energy can do work.

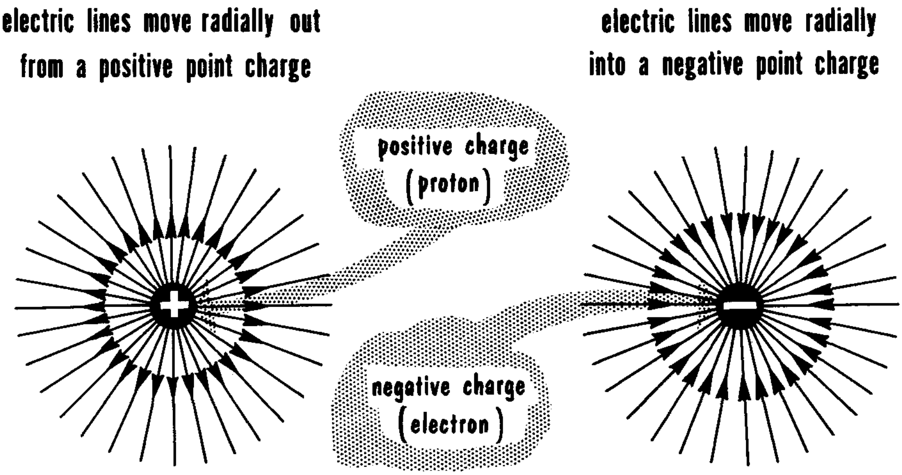


It can be demonstrated in the following way. Assume two electrically-neutral parallel-positioned metal strips cemented onto a pane of glass. Now we drop crystals of gypsum (or very tiny bristles of hair) onto the plate, and tap the glass slightly. The gypsum (or tiny pieces of hair) occupy random positions. That is, nothing happens. Now, remove the crystals (or hair) and give one metal plate an exceptionally strong negative charge and the other plate an equally strong positive charge. Then drop the crystals onto the plate and tap the glass. A distinct line pattern will be seen between the metal strips and around them. Something has happened! The unlike charge given to the two plates has created an electric field between and around the metal strips. The electrical energy that constitutes the field acted on the crystals (or the hair) and made them line themselves up in a special way. The field exerted mechanical force. If the two charged strips could move, they would move towards each other.

Electric Lines of Force

The electric field between the two charged metal strips shown on the preceding page has a distinct line pattern. One gathers the impression that the gypsum crystals acted on by the energy in the electric field aligned themselves along "lines" of energy. It is explained in the following way.

THE ELECTRIC FIELD AROUND POSITIVE AND NEGATIVE POINT CHARGES



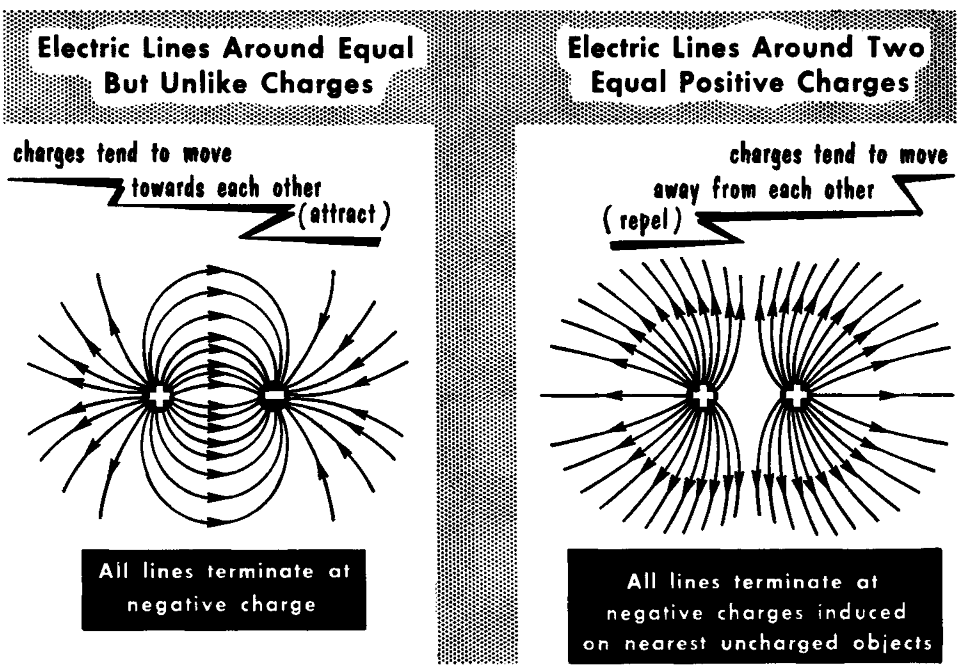
Every electron and proton has an electric field of its own. The energy of the fundamental charge is in this field. The electric field occupies the space all around the charge. Michael Faraday, an Irish scientist, pictured this field in a certain way. He visualized the field as being made up of "pencils of energy" which can conveniently be illustrated by innumerable straight lines which radiate outward in all directions from the center of the charge. He called these lines electric lines of force, or simply, lines of force.

Faraday said that a line of force behaved in certain ways. It had a direction of action; it exerted force in the direction in which it pointed, as indicated by an arrowhead drawn on the line. The direction was, and is (by arbitrary convention) that in which it would make an imaginary positive test charge move. So it pointed away from the proton and toward the electron. Another accepted form of behavior of lines of force is that those lines which advance in the same direction repel one another. A third form of behavior which he conceived was that lines of force which joined unlike charges behaved like stretched rubber bands that always wanted to contract. So, when we say that unlike charges attract, it is the action of the lines of force in the electric field between the charges that pulls them together. When we say that like charges repel, it is the force of repulsion between the lines of force going in the same direction that move facing charges apart.

Lines of Force (cont'd.)

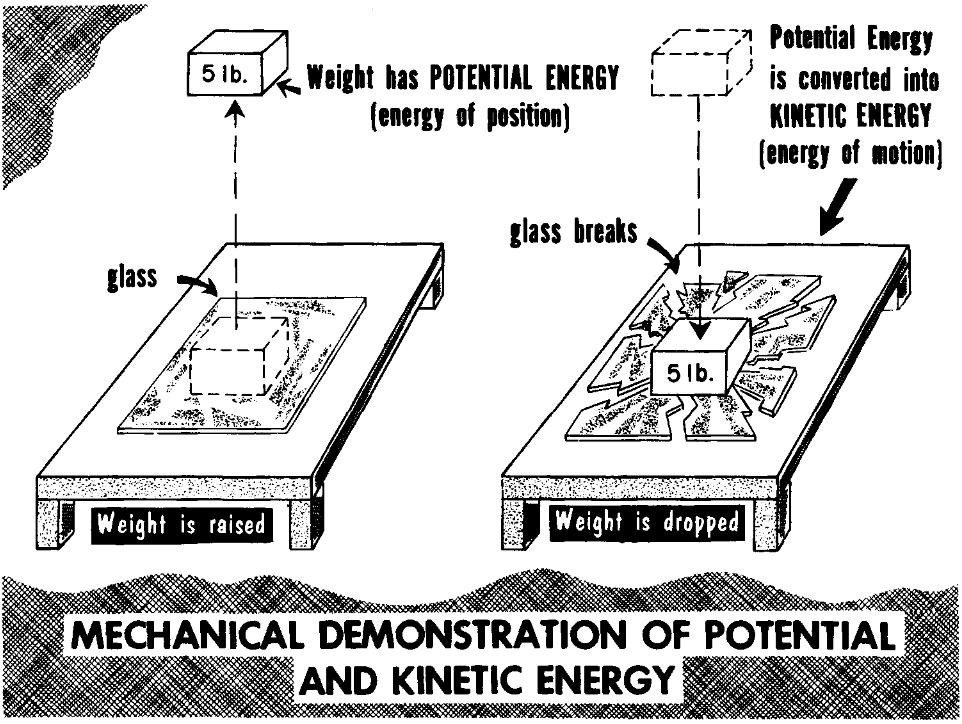
The lines of force concept is a ready means for explaining the apparent absence of electrical effects around electrically-neutral objects, and the electric field around charged objects. Interestingly, an electric field exists around the electrons and protons even in an uncharged object, but the fields have equal intensity and are opposite in direction of action at all points equidistant from the charge. One charge offsets the effects of the other everywhere; therefore, there is no detectable electric field surrounding the object. On this basis, an electrically-neutral atom is considered as having no external electric field. When the process of charging causes an accumulation of electrons on one object and positive ions on the other, each charged object has its own electric field, but two fields acting on each other produce a single net electric field between them.

In the case of the positive ions, each ion contains more protons than electrons; hence, it has a net electric field, that of the surplus protons. The energy of this field is detectable beyond the limits of the ion. The same condition is true in the case of the negative ion, except that now the electric field is that due to the excess electrons.



As to the line formation of the electric field between the parallel metal strips, it follows the reasoning of the lines of force concept. Actually, there are innumerable lines of force between the strips. The lines of force are substantially parallel between the strips, because the mutual force of attraction between them straightens them.

Potential and Kinetic Energy



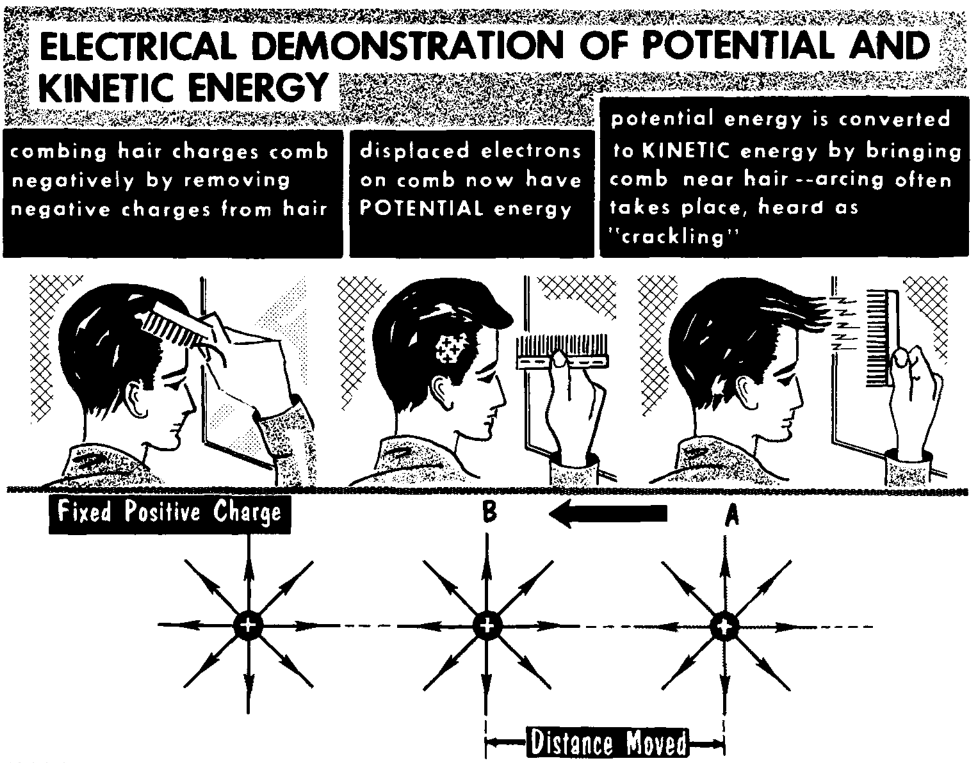
In the study of practical electricity, we are concerned mainly with the movement of free electrons and the conditions which govern their movement. In this connection, the subject of electricity has a language all its own wherein certain words identify particular conditions. One of these words is "potential". Potential is an abridgment of "potential energy", the energy associated with position. To illustrate this, consider the following. Assume a 5-pound weight resting on a pane of glass which in turn is resting on a table. Now you raise the weight above the glass to a height of say, 6 feet. Then you let it fall. The glass shatters. Where did the weight acquire the energy to break the glass? The answer is simple, even though you may not have thought much about it.

The energy to break the glass was acquired by the weight from your effort in lifting it against the gravitational pull of the earth. This pull accounts for everything falling to the ground, and if it were possible to do so, falling towards the center of the earth. You did work on the weight; that is, you gave energy to the weight when you raised it above the glass. You gave the weight potential energy--the energy of position. The higher you raise the weight above the table, the more work is done on it and the greater is the amount of potential energy stored in the weight. When you allowed the weight to fall, the potential energy was converted to kinetic energy--the energy of motion. When the weight struck the glass, the kinetic energy was changed into heat energy, sound energy, and mechanical energy, all of which are related to the actual physical breaking of the glass and the sound which accompanies the action.

Potential (cont'd.)

Let us now assume that we remove free electrons from an object by any one of the charging processes described, and pass these electrons on to another object. Whatever the method used, the displacement of the electrons requires the application of energy to make them move against either the force of attraction of the positive ions remaining after the electrons have been taken away, or against the force of repulsion from other free electrons which are on the surface to which the electrons are being transferred. The displaced electrons now have acquired potential energy. The greater the displacement, the greater is the amount of work that must be done, and the greater the potential energy acquired by the displaced electrons.

Now, if we arrange for the displaced electrons to go back to where they were, the potential energy is converted to kinetic energy (the energy of motion) and the moving charges can do work. The higher the potential of the charge at any point, the more work can be done when the potential energy is converted into kinetic energy. We save time by referring to the point or place where the charge or charges exist as being at a certain "potential" relative to some reference point, rather than describing it in terms of potential energy.

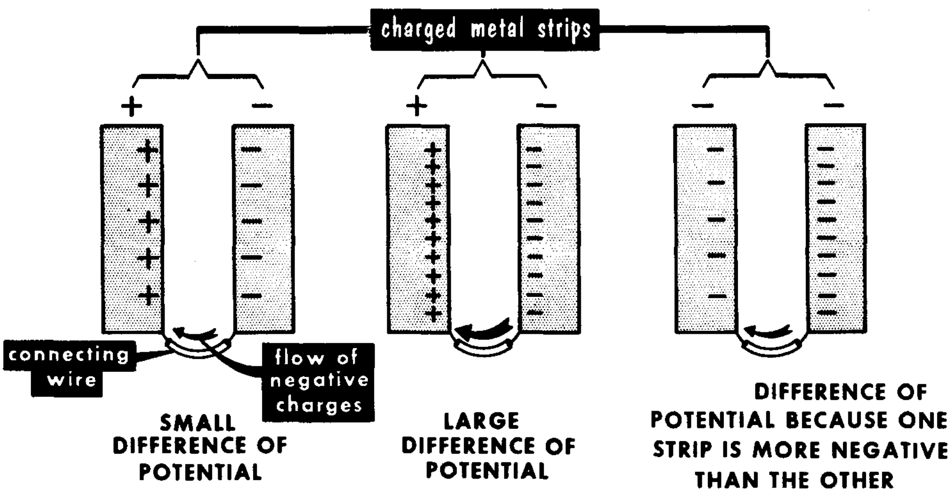


Positive charge at point B has potential equal to Work Done in moving it from point A to point B in opposition to repulsive force of fixed positive charge.

Difference of Potential

Another phrase encountered very often in electricity is "difference of potential" It, too, is related to the movement of free electrons. Difference of potential expresses a situation which determines the tendency of free electrons to move from one place to another. When we accumulate an excess of free electrons on an object and create a shortage of free electrons on another, we create a difference of potential between these two charged objects. You can consider the objects as charged metal strips, if you desire. Given a path between these two objects, free electrons would move from the "negative" strip to the "positive" strip. The greater the excess of free electrons on one strip and the more the shortage of free electrons on the other, the greater is the difference of potential between the two strips and the stronger would be the tendency of the free electrons on the negatively-charged strip to go to the positively-charged strip.

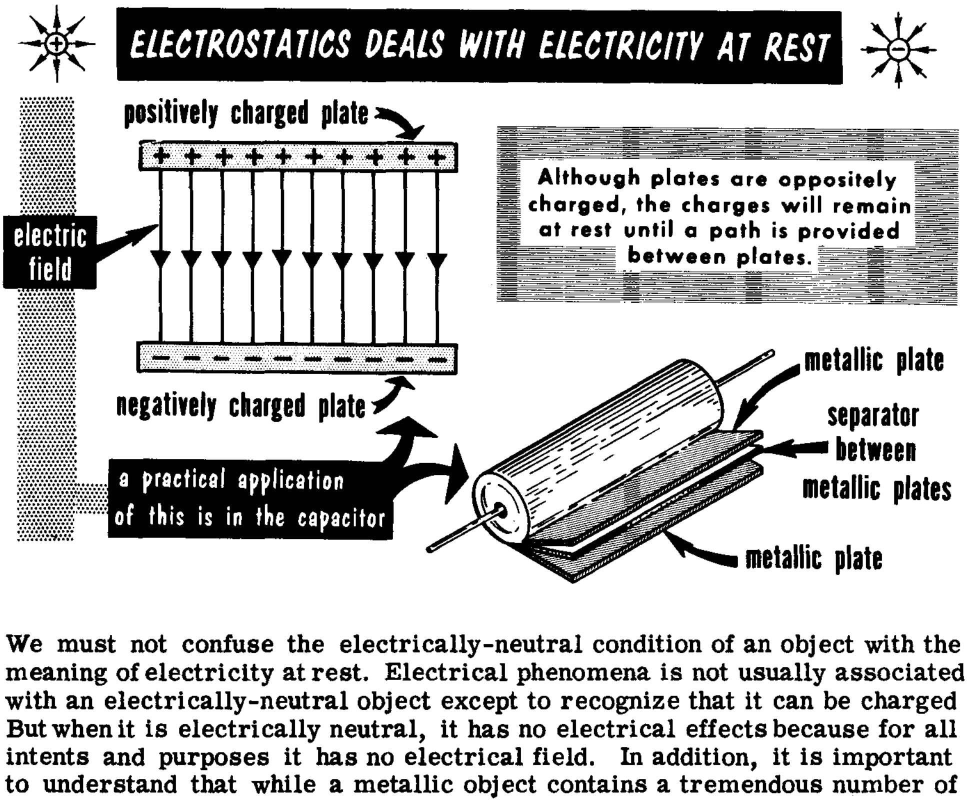
CHARGES TEND TO MOVE BETWEEN TWO POINTS THAT ARE AT DIFFERENT POTENTIALS



It is important however, to understand that a difference of potential exists between any two points or places where unequal amounts of free electrons exist. Conceivably, two metal strips (or any other objects) may be charged negatively, one less than the other. A difference of potential exists between these charged objects. Thus, there would be a tendency for electrons to move from the more negatively-charged object to the less negatively-charged object until both objects were at the same potential, after which there would be no further tendency of the electrons to move between the two charged objects. In other words, and this is a very important condition, movement of free electrons through the wire that connected the two metal strips described several pages back occurred while a difference of potential existed between them.

Static Electricity and Electrostatics

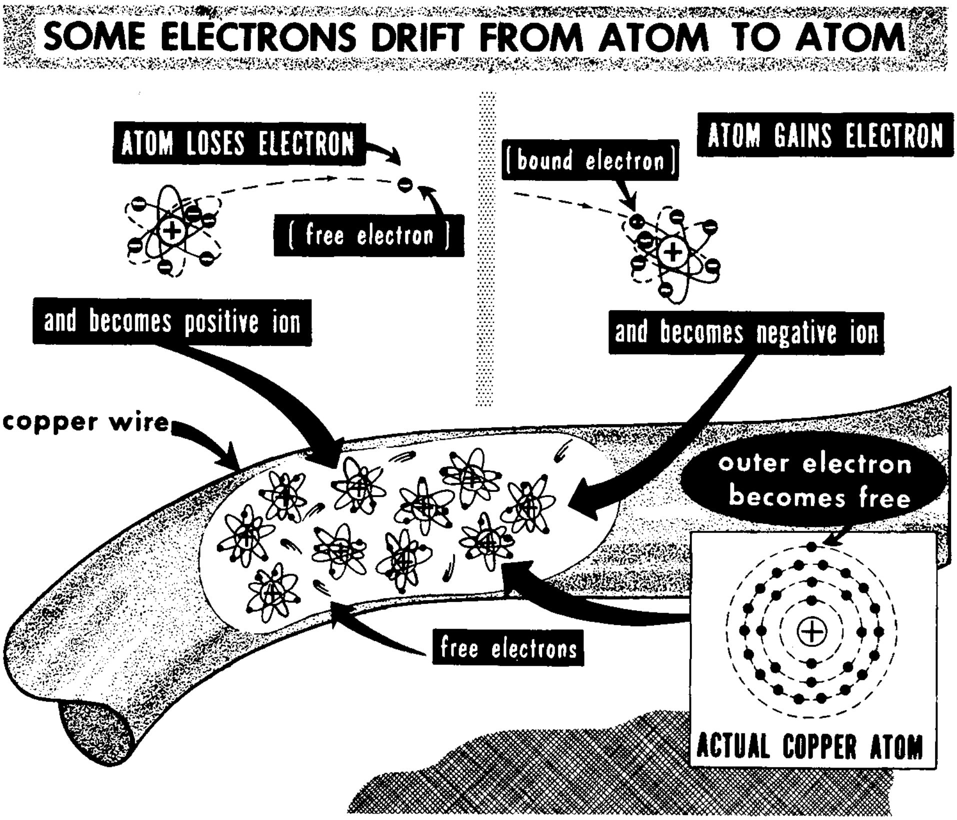
One area of the study of electricity is called electrostatics. We have discussed charging and discharging. Whether the method of charge was by friction, contact, or induction, electrons were displaced from one object to another, after which the electrons came to rest at a new location, creating a charged state. The name given to electricity at rest is static electricity, and the broad subject of charging and discharging (the behavior of static electricity) is called electrostatics. The study of electric fields between charged objects falls under the heading of electrostatics.



We must not confuse the electrically-neutral condition of an object with the meaning of electricity at rest. Electrical phenomena is not usually associated with an electrically-neutral object except to recognize that it can be charged But when it is electrically neutral, it has no electrical effects because for all intents and purposes it has no electrical field. In addition, it is important to understand that while a metallic object contains a tremendous number of free electrons in motion, when this object is charged by acquiring electrons, these surplus electrons are (to all intents and purposes) at rest along the surface of the metal.

There are practical reasons for having emphasized the phenomena of charging and discharging. Charging an object is the equivalent of storing electrical energy on the object; discharging an object is the equivalent of releasing the electrical energy. The practical device which does this in radio equipment, and about which you will learn more later, is called a capacitor (also known as a condenser). The purpose of a capacitor is to store electrical energy and to release it when it is needed. As we will learn, electrical charges also play an important role in the operation of tubes and transistors.

Free Electrons and Electric Current

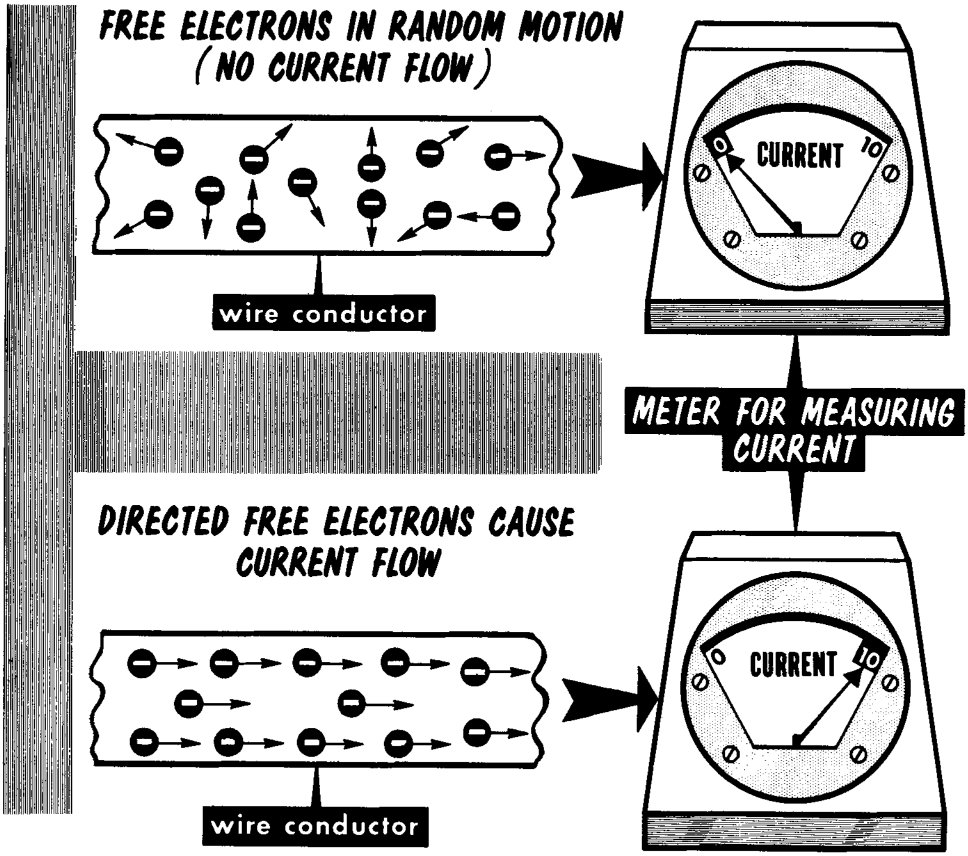


The study of the atom was intended to serve several purposes. First, to lay the foundation for the explanation of electrical charge and discharge. This we have done. Second, to be the basis for the explanation of electric current. This we shall do now.

Let us take an imaginary glimpse inside a piece of copper wire. What we see would not be too much different in any other kind of metal. The atoms shy electrons (more correctly known as positive ions) perform a to-and-fro motion over a very limited distance each side of a "fixed" position. Although describing a vibratory motion, we can, for all purposes, consider the ions as being fixed in location in the wire. In between, a host of free electrons drift slowly in all directions--as many moving in one direction as in the opposite direction. While performing this motion, the free electrons attach themselves momentarily to the ions, and even to atoms which have acquired the balancing electron--only to be freed again a moment later. The free electrons collide with each other, as well as with atoms. This is an important phenomenon, as you shall soon learn. Obviously, metal contains wandering negative electricity (wandering electrons) but it is not usable electricity as it is. We must change electron movement from random to controlled. Then it is electric current.

Free Electrons And Electric Current (cont'd.)

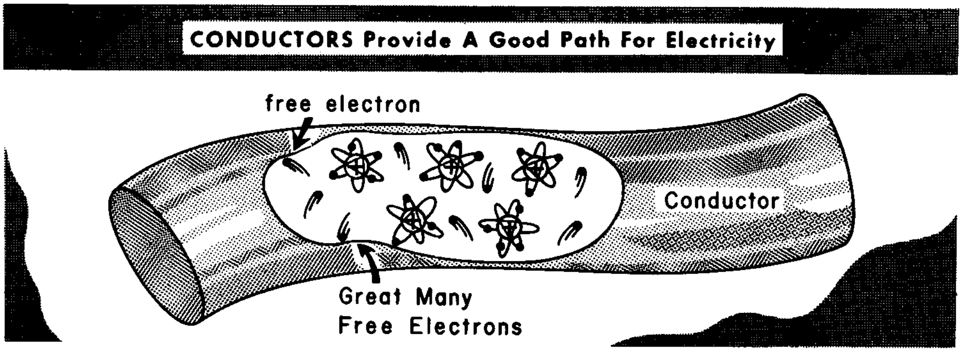
A few pages back, we described the motion of free electrons through a wire connected to two metal strips charged with unlike electricity. The direction of electron flow was from the negative to the positive strip. Such movement was a controlled motion--it had a definite direction--a singular direction throughout the wire. Although the flow of electrons did not last for long, it did constitute electric current while it lasted. We define electric current as free electrons performing a controlled motion.



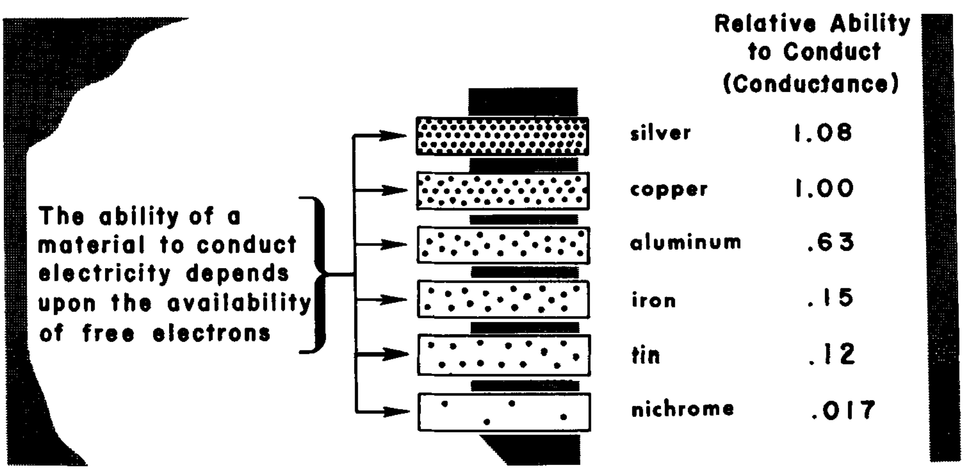
Let's consider electric current in a different way. When free electrons are drifting at random in a wire, as much charge (electricity) moves in one direction as in the opposite direction. There is, therefore, no continuous delivery of electricity from one point to another in a constant direction through the wire. But when the random motion of the electrons is changed to controlled motion, electricity is delivered continuously in a constant direction through the wire. Such delivery of charge is electric current. As you will learn later, two important considerations of electric current are the direction of flow, and the quantity of electricity that is transported in a period of time--the period being 1 second.

Conductors

Based on the atomic concept of matter and the electrical makeup of the atom, all things are made of electricity. Yet, all materials are not equally good as paths for electricity (electric current). Conductors is the name assigned to a broad category of materials through which a practical amount of electric current can flow under normal conditions. Most conductors are metallic.



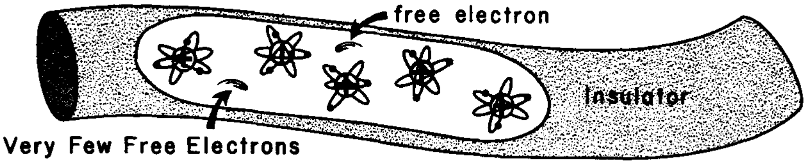
The usefulness of metal to conduct electric current stems from the abundance of free electrons in the material, the free electrons being electricity that is available to be moved through the metal when the material is part of a complete electrical system. However, not all metals are equally good conductors. The availability of free electrons for a given length and cross section is not the same in all metallic substances. The chart indicates the commonly-used metals in their order of suitability as paths for electric current. Silver is the best, but being expensive is used only in special cases. Copper, the second best conducting material, is plentiful and inexpensive; hence, is used most often. At the bottom of the list is Nichrome, a special alloy manufactured specifically to perform as an inferior conductor. It is one of numerous alloys that find special use in electrical systems.



Insulators

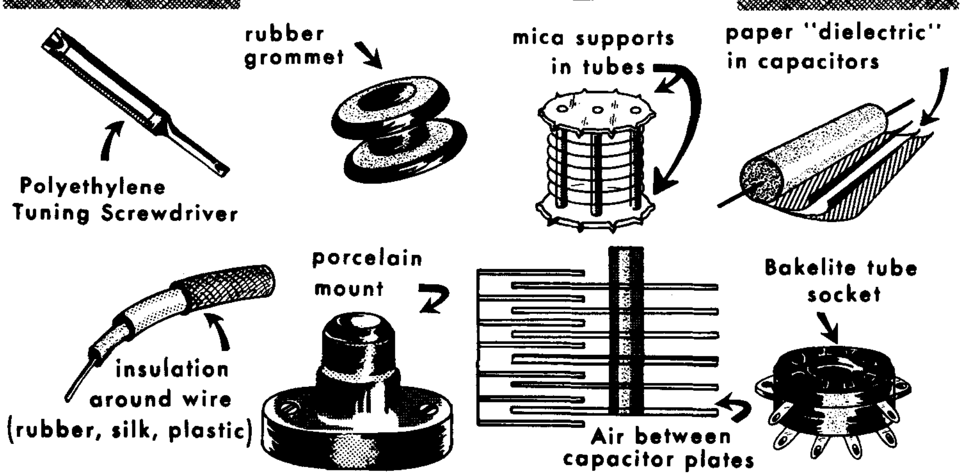
Insulators is the name given to a special group of materials that provide extremely poor paths for electric currents under all normal conditions. In fact, their function is to prevent the flow of electric current. To understand the basis of action of insulator materials, think of an atom whose electrons are bound to the atomic structure; in other words, the material contains very few free electrons. Since there are very few free electrons per given length and cross section of the material, there is very little charge which can be delivered from one point to another. No material is completely void of free electrons. But if they are so few in number, relatively speaking, that a practical and useful amount of electric current cannot flow under normal conditions, we consider the material to be an insulator. Examples of such substances are plastics, glass, mica, mineral oil, rubber, dry paper, dry wood, cotton, and ceramic substances. The above references are not in the order of their suitability as insulators.

INSULATORS Provide A Poor Path For Electricity



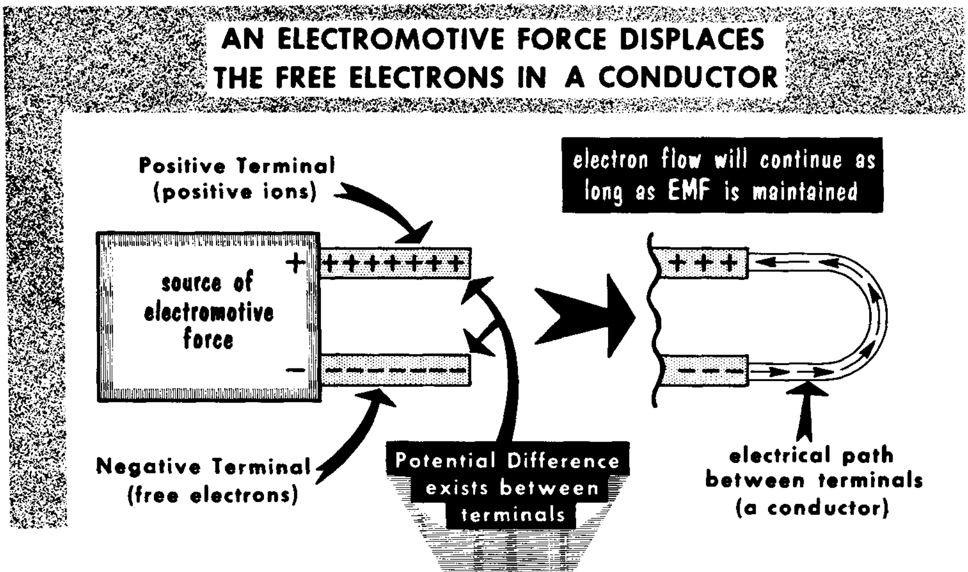
Let us clarify some points regarding the names associated with insulator materials. The material itself is called an insulating material, but when it is shaped into some form--large or small--and sometimes used as a support for wires carrying electric current, it is called an insulator. Another word for an insulator is dielectric. If the material is used as a covering around wires that carry electric current, it is called insulation.

ELECTRICAL INSULATION USED IN RADIO



Electrical Pressure and Electromotive Force

Q: How is the random motion of free electrons in a conductor changed to the controlled motion that is electric current? A: By the application of electrical pressure which is known as electromotive force. It is abbreviated emf, and pronounced ee-em-eff.



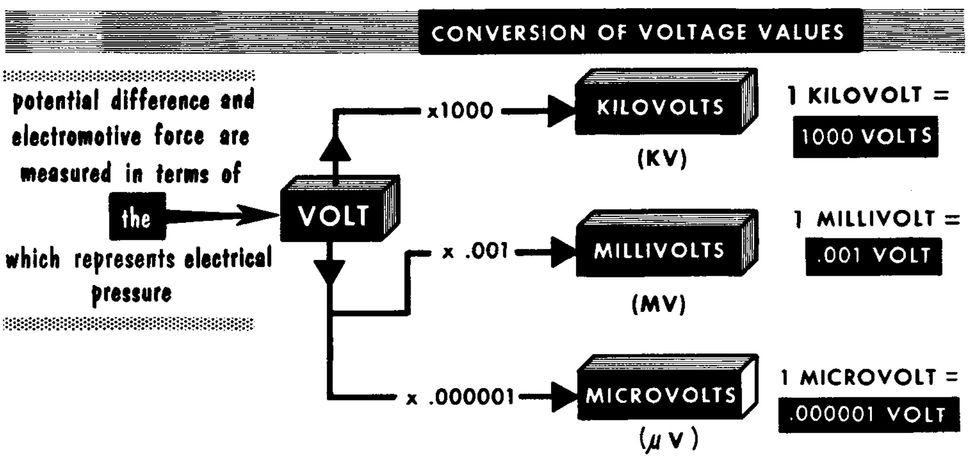
A source of electromotive force can be described as a device in which electrons are forcibly separated from atoms. This creates free electrons and atoms shy electrons or positive ions. The conventional cell or battery used in flashlights and portable radio receivers is an example of sources of electrical pressure or electromotive force. There are other sources about which you will study later.

The separating action referred to above takes place inside the device continuously while the device is functioning. The result corresponds to an accumulation of free electrons on the "negative" terminal (usually designated by the symbol -) and atoms shy electrons on the "positive" terminal (usually designated by the symbol +). The displaced free electrons are attracted to the atoms that are shy electrons (unlike charges attract each other), but they cannot do so by any path inside the device. The chemical action inside a battery prevents this. But if we provide some sort of electrical path outside the battery, connected between the negative and positive terminals, there will be a movement of free electrons between the battery terminals via the external path in a particular direction -- from the negative to the positive terminal. This directional quality of the action is called polarity. The electrical force acting on the electrons accumulated on the negative terminal urging them to the atoms that are shy electrons on the positive terminal is the electromotive force developed by the chemical action in the battery.

Unit of Electromotive Force (the Volt)

If we were speaking about water pressure or air pressure, we would refer to the amount of pressure in terms of "pounds per square inch". In the case of electrical pressure, or more correctly stated, electromotive force, the unit used is the volt. The actual amount of voltage is expressed by a number. Thus, the output of a flashlight cell is rated at an electromotive force of 1.5 volts; the modern automobile battery is rated at an electromotive force of 12 volts. Some other sources of electromotive force may be rated at 10 volts, 120 volts, or 5000 volts.

The difference of potential between two points in an electrical system can also be identified in terms of volts. It has become accepted practice to refer to the electromotive force of a device in terms of voltage output, the presence of electromotive force being understood when the voltage reference is made. Thus, by saying that the voltage output of a flashlight cell is 1.5 volts, we mean that it develops an electromotive force of 1.5 volts.



The voltages encountered in radio equipment involve a very wide range. Voltages may be as high as several thousand volts or as low as a fraction of a millionth of a volt. Sometimes, it is convenient to state them as a number and a word prefix having a numerical significance. For instance, it is quite common to use the prefix "kilo" to mean 1000. In other words, 1 kilovolt (abbreviated kv) means 1000 volts, and 10 kilovolts means 10,000 volts. When the voltage value is between 1 and 999, the amount is expressed as a simple number; beyond 999, the prefix kilo can be used.

Voltages less than 1 volt are sometimes stated as a decimal and sometimes by a number associated with a prefix. For instance, the prefix "milli" means one-thousandth; the prefix "micro" means one-millionth. To state three one-thousandths of a volt, it is customary to say 3 millivolts (3 mv) and, when writing the amount, the decimal .003 volt can be used. In like manner, 100 millionths of a volt would be stated as 100 microvolts (100 μv) and, when shown in a computation, it might appear as .0001 volt.

SUMMARY

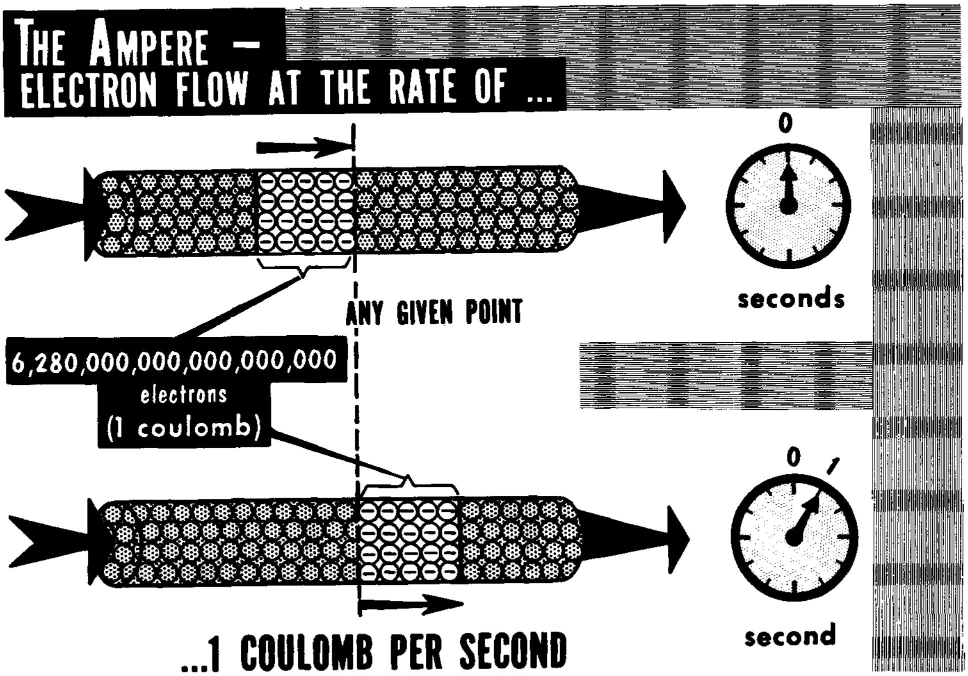
Many materials, when rubbed with other materials, not only attract tiny objects, but exert forces of attraction and repulsion on each other.
Bodies charged with unlike electricity attract each other; bodies charged with the same kind of electricity repel each other; positively-and-negatively-charged bodies attract each other.
Everything is made of two kinds of electricity--positive and negative--and one kind can be separated from the other.
Protons are the fundamental charge of positive electricity; electrons are the fundamental charge of negative electricity.
Electrons are held in their orbits against the pull of centrifugal force by the electrical force of attraction between the protons in the nucleus of the atom and the orbiting electrons.
All matter is composed of one or more of 102 substances known as chemical elements, which are comprised of atoms.
Matter or substances with an excess of electrons are said to be negatively-charged; matter or substances with a deficiency of electrons are said to be positively charged.
Atoms are comprised of positively-charged particles called protons, negatively-charged particles called electrons, and uncharged particles called neutrons.
Metallic objects can be charged by contact or by induction.
Objects bearing unlike charges can be discharged (made electrically neutral) through air (arc discharge), by direct contact, or through a wire.
Every electron and proton has an electric field of its own, which occupies the space around it. The energy of fundamental charge is in this field.
Potential energy is the energy of position; kinetic energy is the energy of motion.
Conductors are materials through which electric current flows under normal conditions; insulators prevent the flow of electric current.
An electromotive force changes the random movement of electrons through a wire to controlled movement (electric current). The unit of emf is the volt.
A difference of potential exists between two points or places in an electrical system and is identified in terms of voltage.
"Potential" and "voltage" are terms which are used interchangeably.

REVIEW QUESTIONS

1. What are two kinds of electricity called?
2. Define the fundamental laws of attraction and repulsion between two electrical charges.
3. How many elements are there? What is the smallest part of an element?
4. What are atoms comprised of?
5. What is the difference between charging by contact and by induction?
6. Explain what free electrons are.
7. What is an electric current?
8. What is the name given to electricity at rest?
9. Explain what is meant by the "lines of force concept".
10. Define a conductor. Give examples of good conductors.
11. Define an insulator. Give examples of good insulators.
12. Which is the better conductor, copper or silver? Which is more widely used and why?

Electric Current (Rate of Flow and Ampere)

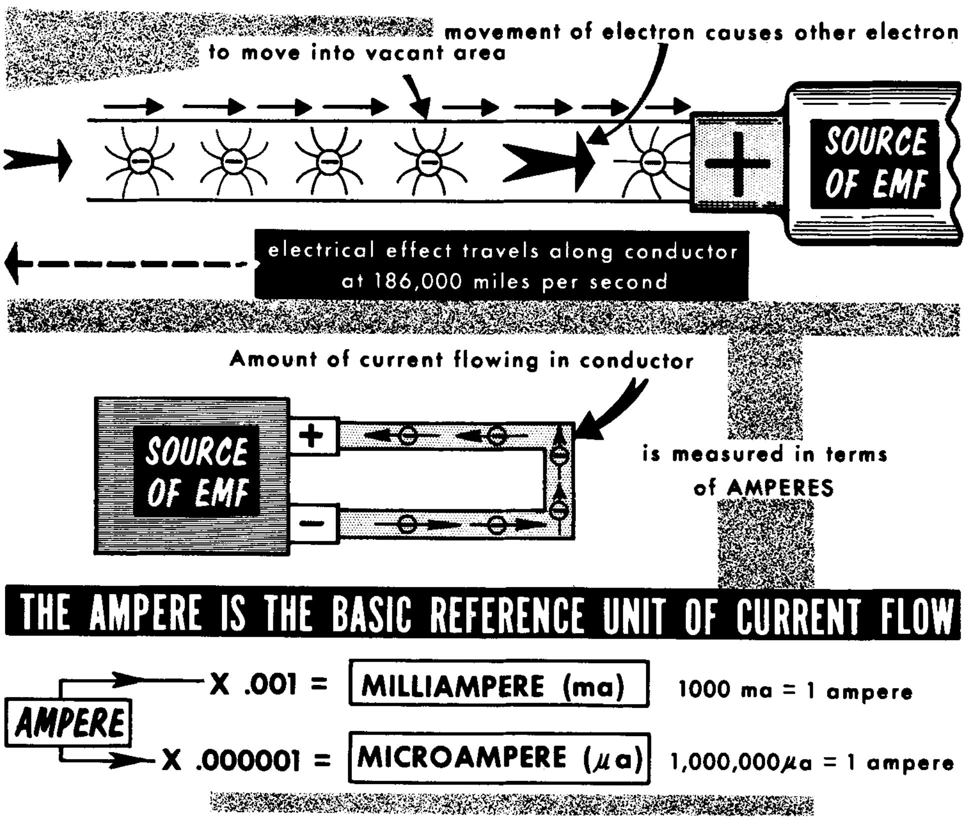
Most electrical devices function by virtue of the electric current which flows through them. An important characteristic of electric current is its rate of flow.



The unit of quantity for electric current is the coulomb, named after the scientist Charles Coulomb (1736-1806). It corresponds to the gallon in liquid measure. Although we seldom mention the equivalency between a gallon of fluid and the number of drops of the fluid that make up the quantity, there is a figure which expresses the amount of electricity contained in a coulomb in terms of the electricity carried by each electron. Specifically, it is the amount of electrical charge equal to 6.28×10^{18} or 6,280,000,000,000,000,000 electrons. (You need not bear this number in mind but it is useful to show that a coulomb is a specific amount of electricity). It is not sufficient to know the quantity of electricity passing through a device; we must know the rate of flow. In this connection, the unit time reference is 1 second; hence, the rate of flow of electric current is expressed as coulombs per second. This term corresponds to "gallons per minute". However, in the electrical language, we shorten "coulombs per second" to the simple word "ampere", named after the French physicist Andre Marie Ampere (1775-1836). We say that 1 ampere of electric current flows when 1 coulomb of electric charge moves past any given point in 1 second. Note that the word ampere already includes the quantity and the time references. Thus, a reference to 5 amperes means the flow of 5 coulombs of electricity past a given point every second. Another short cut used in engineering language is the letter 'I' to represent the symbol for current.

Velocity of Electrons in a Conductor

We have spoken of current flow as consisting of the movement of free electrons under the influence of an electromotive force. We learned that the individual free electrons moved from atom to atom in the general direction of the positive electromotive force charge, and away from the negative charge. Actually, the forward motion of the individual electrons is not very rapid, but their electrical effect is -- approximately 186,000 miles per second, the speed of light.

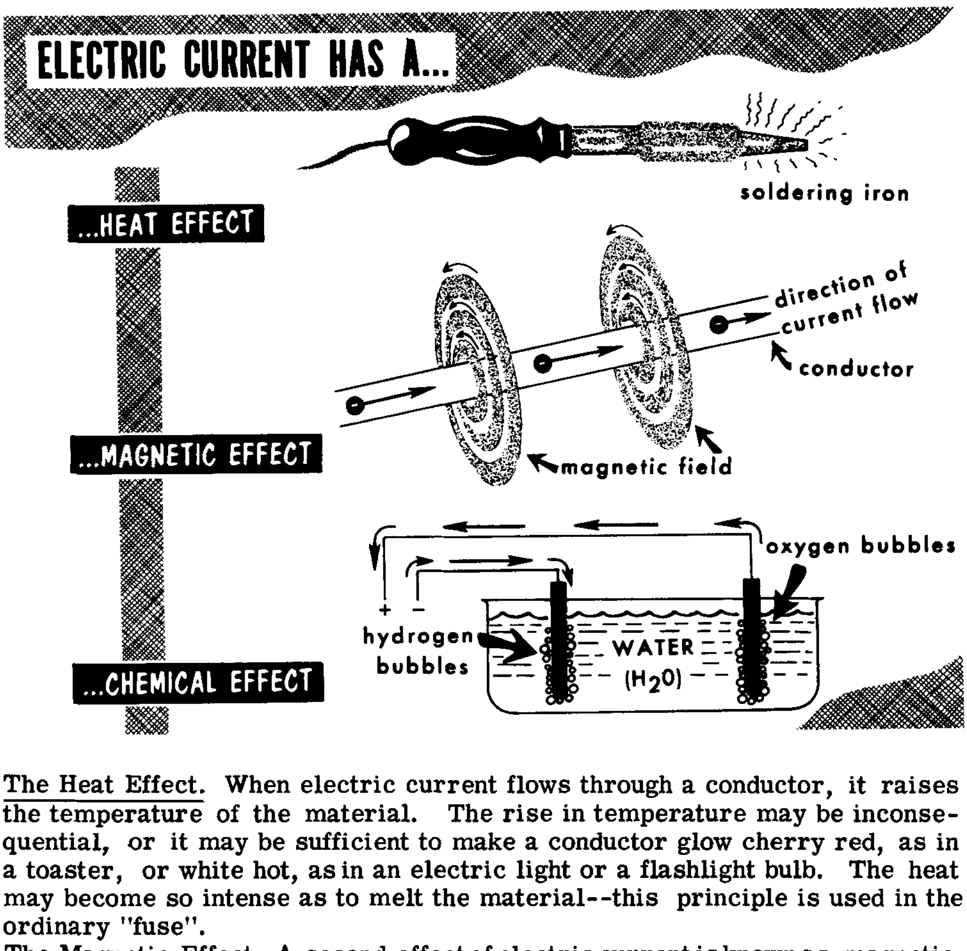


When an electron is attracted to the positive terminal, there is a deficiency of negative charge in the area just vacated. A new free electron is attracted to fill the vacancy. This effect travels along the length of the conductor, with the positive terminal attracting and the negative terminal repelling free electrons. Thus, despite the fact that the electrons themselves drift very slowly, the effect of the changes in positions of the electrons is transferred along the conductor, connecting the terminals of the source of emf almost instantly.

In radio, the amount of current flow may be hundreds of amperes in transmitting equipment; but most often, it is in fractions of an ampere. Quantities such as the milliampere (.001 ampere) and the microampere (.000001 ampere) are commonly used.

Effects of Electric Current

Although an electric current is invisible, its presence can be detected because of its effects. The greater the amount of current, the stronger the effects.



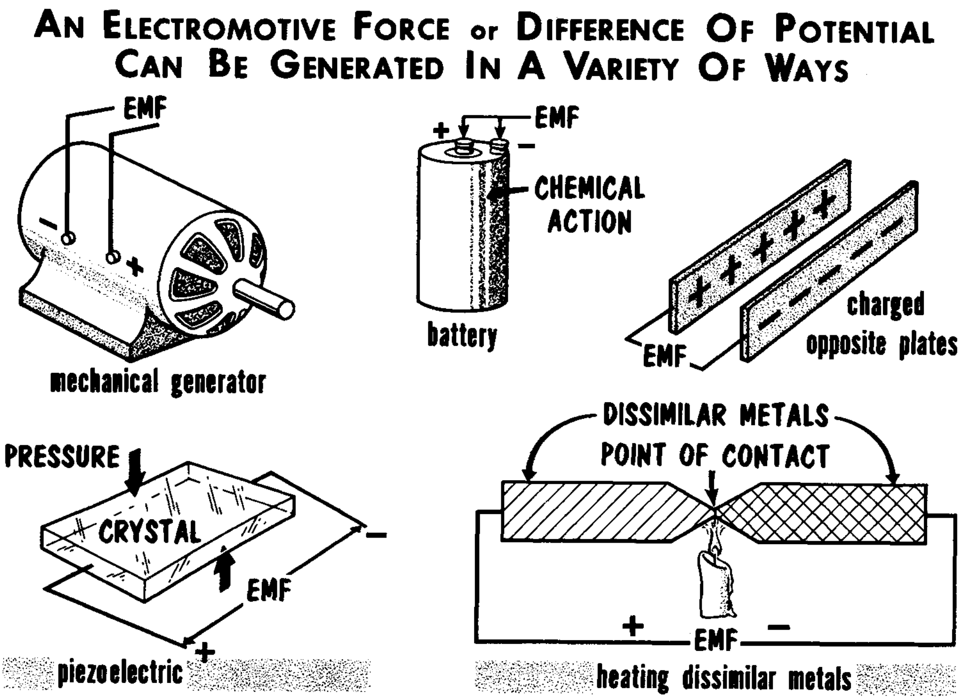
The Heat Effect. When electric current flows through a conductor, it raises the temperature of the material. The rise in temperature may be inconsequential, or it may be sufficient to make a conductor glow cherry red, as in a toaster, or white hot, as in an electric light or a flashlight bulb. The heat may become so intense as to melt the material--this principle is used in the ordinary "fuse".

The Magnetic Effect. A second effect of electric current is known as magnetic. Every conductor through which current is flowing is surrounded on all sides by an invisible area of energy known as a magnetic field. The origin of the field is the moving electron, and the intensity of the field depends upon the strength of the current. Of great importance is the fact that the magnetic field is capable of exerting mechanical forces on a variety of metals such as steel, iron, nickel and on materials known as "ferrites". Also, it is capable of exerting mechanical forces on electrons moving in nearby conductors.

The Chemical Effect. A third effect is known as the chemical effect. The passage of electric current through different fluids can cause the breakup (decomposition) of the fluid into its constituent atoms. Also, under certain conditions, electric current passed through acids in which certain metals are immersed can cause decomposition of the metal. This is the basis of electroplating.

The Generation of Electromotive Force

Electromotive force (emf or voltage) was described earlier as being the result of the forcible separation of electrons from atoms, causing an accumulation of electrons (negative charge) at one point, and an accumulation of atoms shy electrons (positive charge) at another point.



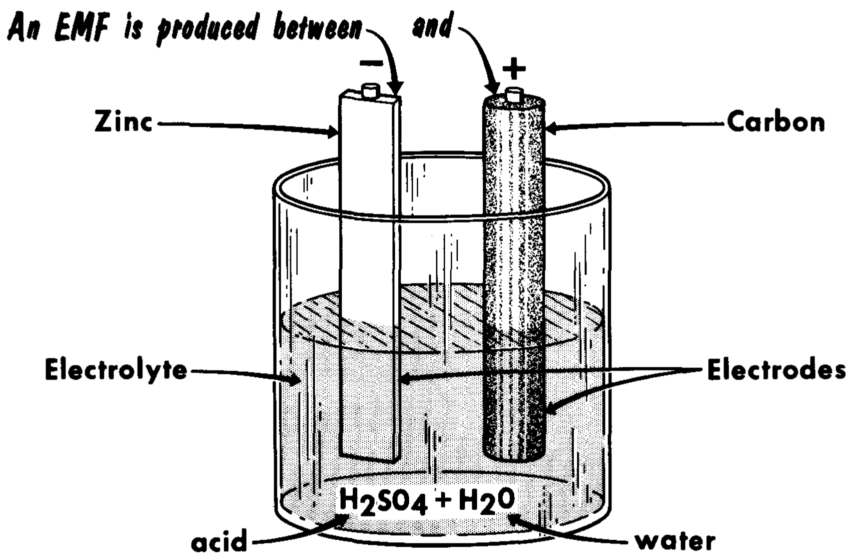
The condition can be developed in a variety of devices. For instance, two oppositely-charged plates can be a source of an electromotive force as long as the oppositely-charged condition exists on the plates. When these charges are neutralized by the transfer of electrons, the electromotive force between the plates ceases. The friction of rubbing resin with fur and glass with silk separated electrons from atoms and accounted for the generation of an electromotive force between these charged objects. Also, you will recall a previous reference to batteries. They generate an electromotive force by internal chemical action, continuing to do so as long as the chemical action continues.

Rotating machines know as generators constitute another category of devices that generate an electromotive force. Later, we shall discuss the method. Still other substances, like slabs of quartz and crystals of Rochelle salts, when subjected to pressure or physical deformation, result in the separation of electrons from atoms within them and so generate electromotive force between their faces. This is known as the "piezoelectric" effect. Two dissimilar metals, placed in contact with each other and with their point of contact heated, will generate an electromotive force at the open ends.

The Primary Cell

Experiments conducted during the 18th century disclosed that when two dissimilar materials like zinc and carbon (or copper used in place of carbon) were immersed in a bath of acid solution (sulphuric acid diluted in water), chemical action produced an electromotive force between the zinc and the carbon (or copper). The materials immersed in the acid were called the electrodes, with the zinc being the negative electrode and the other (carbon or copper) being the positive electrode. The diluted acid solution was called the electrolyte; the whole assembly was called the cell. Often, cells are mistakenly referred to as batteries. Actually, the cell is the basic unit; two or more cells used together comprise a battery. We shall show examples of these later.

THE BASIC CELL

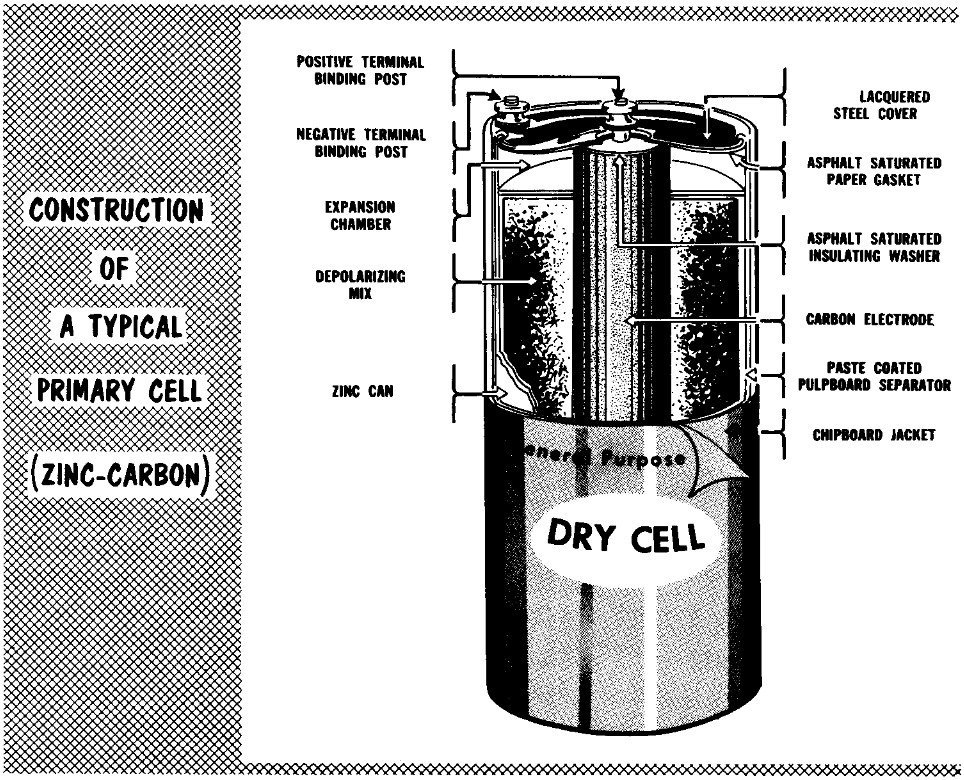


When the + and — terminals are connected externally, current will flow

As to the broad significance of a "primary" cell, it has a particular meaning, this being that the zinc dissolves slowly while the cell is functioning. Eventually, the zinc is eaten away to the extent that it prevents further operation of the cell. When this stage is reached, the cell has exhausted its useful life. In other words, a primary cell is of the kind which has a limited life, after which its operation cannot be restored and the cell as a whole must be disposed of. There are other kinds of cells which will be dealt with separately in this section of the course.

The Zinc-Carbon (Le Clanche) Primary Cell

By far, the most popular and commonly-used primary cell is the zinc-carbon type, sometimes referred to as the Le Clanche cell. In this cell, the positive electrode is made of carbon (C) and the negative electrode is made of zinc (Zn). The electrolyte is a chemical known as ammonium chloride (NH_4Cl), often called sal ammoniac. The negative electrode is in the form of a container and holds the entire cell. The positive element is in the form of a carbon rod located at the center of the cell. The electrolyte is mixed with cornstarch or flour to form a paste. Thus, a dry cell is not really "dry". In fact, when the electrolyte drys out, the cell becomes useless. A mixture of finely-ground manganese dioxide (MnO_2) is packed around the carbon electrode to act as a depolarizer.

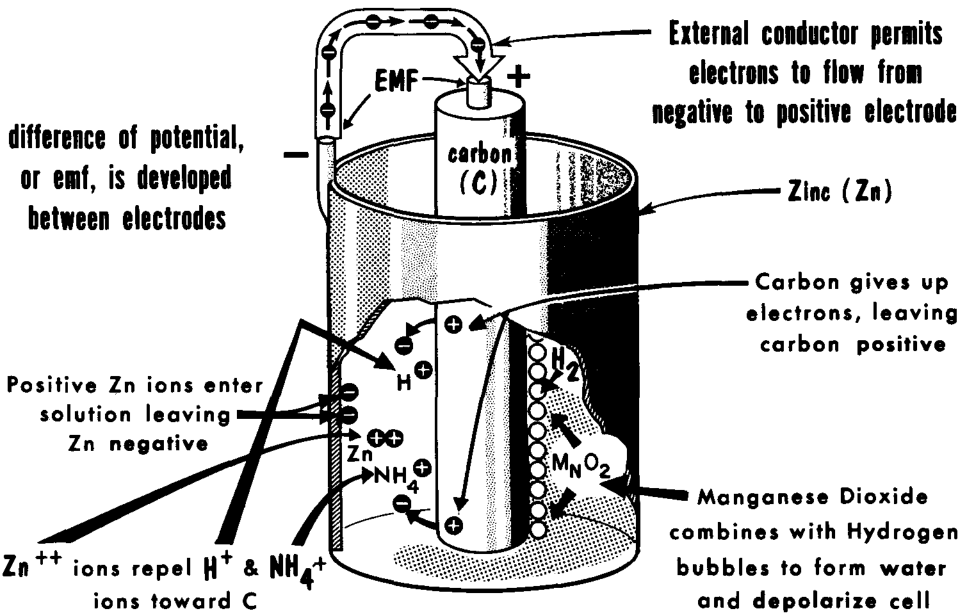


When the cell is operating properly, a difference of potential (or voltage) of about 1.6 volts (fully charged) is developed across the positive and negative terminals. As the cell becomes "used up", either by the electrolyte drying out or the zinc electrode being "eaten" away, the terminal voltage will fall off. At about 1.1 volts (discharged), this type of cell becomes useless for most applications -- it cannot be recharged, and must be discarded. Different combinations of dissimilar metals and electrolytes will produce different terminal voltages.

Action in a Primary Cell

Let us connect a conducting wire between the terminals of a zinc-carbon dry cell. In the electrolyte (NH_4Cl), ammonia ions (NH_4^+) and chlorine ions (Cl^-) are present. When the zinc makes contact with the electrolyte, zinc ions (Zn^{++}) enter the solution, each leaving two electrons on the negative electrode. The accumulation of negative electrons constitutes a negative charge on the zinc plate. The Zn^{++} ions in the solution repel the positive hydrogen ions (H^+) and ammonia (NH_4^+) ions that are already there toward the ammonia atoms which collect on the surface of the carbon electrode in the form of gaseous bubbles. The loss of electrons leaves the carbon electrode with a positive charge. The Zn^{++} ions combine chemically with the Cl^- ions to form zinc chloride ($ZnCl_2$), a white substance. You may have seen this substance when you removed "dead" cells from a flashlight case. Therefore, the zinc plate is gradually used up to form zinc chloride ($ZnCl_2$) during the operation of the cell.

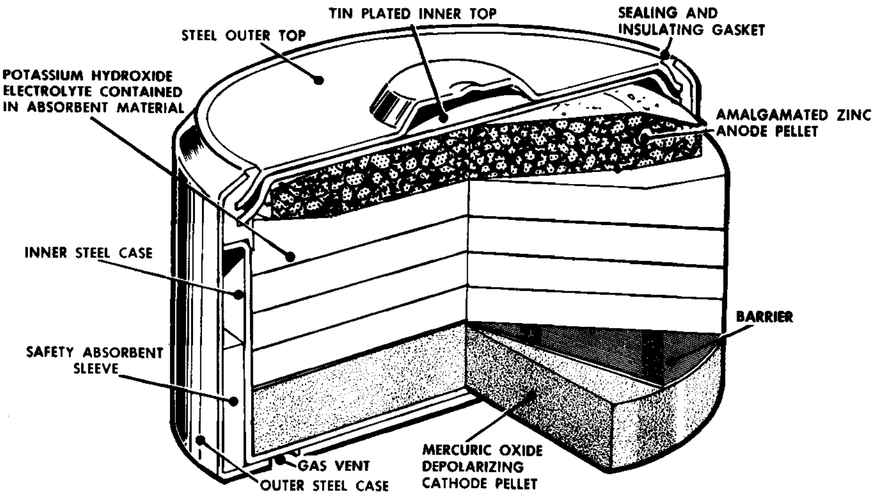
OPERATION OF THE ZINC-CARBON DRY CELL



The overcrowded electrons on the zinc plate repel each other. This repulsion, plus the attraction of the positive charge on the carbon plate, results in the creation of the emf of the cell. This emf causes electron, hence current flow, through the external conducting path between the electrodes. The accumulation of hydrogen bubbles over the carbon electrode adversely affects the operation of the cell by blocking normal chemical action. This is called polarization. The manganese dioxide in the electrolyte prevents polarization by combining with the hydrogen bubbles to form water.

Primary Cells (Zinc-Mercury Oxide Type)

Another kind of primary cell is the zinc-mercury oxide type invented during World War II. The cell consists of an amalgamated zinc negative electrode either in powder form or in coiled corrugated strips. The positive electrode consists of a mixture of mercuric oxide and graphite, molded under pressure in a steel cup and assembled into the cell. The electrolyte is a solution of potassium hydroxide and zinc oxide. Cellulose material is used as a separator and maintains the electrolyte in paste form to prevent spilling.



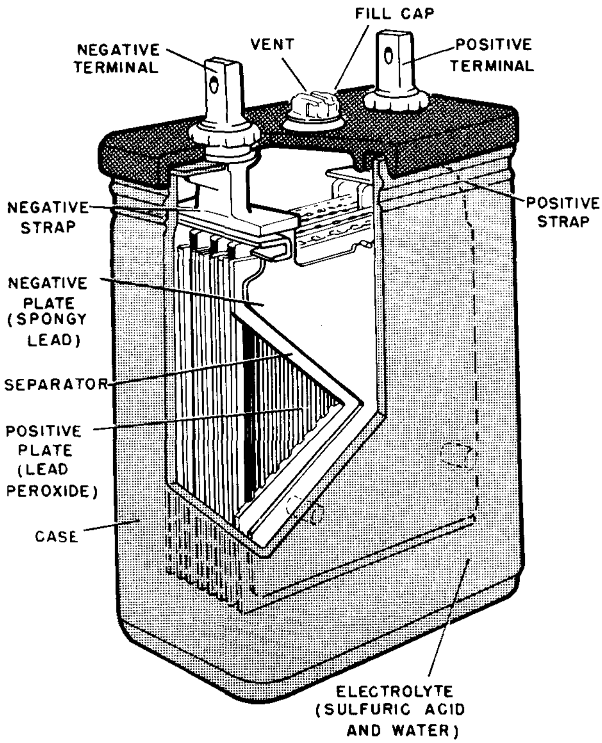
CONSTRUCTION OF THE ZINC-MERCURY PRIMARY CELL

The voltage of this cell when not in use is 1.34 volts, but under normal current drain, the voltage will drop to between 1.31 and 1.24 volts. Compared with most other types of primary cells, zinc-mercury oxide cells have several very desirable features. The voltage over the operating life (discharge cycle) is nearly constant, and its ability to store electrons on the negative electrode is superior to that of the zinc-carbon cell. It has a comparatively high current output which it can maintain for a considerable period of time. Also, storage and shelf life are maintained at higher temperatures. However, zinc-mercury cells are more expensive, and have been in demand mainly where their reduced size (1/2 to 1 inch in diameter and almost 1 inch long) makes their use feasible. Some of these applications are: hearing aids, small portable radios, portable communication equipment, electrical test equipment, scientific instruments, and in some applications as a voltage reference.

An interesting feature of this cell is its ability to maintain a relatively constant voltage over its useful lifetime. In many instances, the voltage of this cell is used as a standard with which to check voltage-measuring instruments.

The Secondary Cell (Lead-Acid)

The primary difference between the secondary cell and the primary cell is in the fact that the secondary cell is rechargeable. That is, after the secondary cell has been in use and has discharged, its chemical action can be reversed and the cell recharged. The most popular and widely-used secondary cell is the lead-acid "storage" cell, found in automobiles, trucks and a wide variety of vehicles. When fully charged, this cell has an output voltage of approximately 2. 2 volts. Automobile batteries are generally made containing three or six of these cells. The lead-acid cell is capable of delivering extremely high currents, running into the hundreds of amperes.



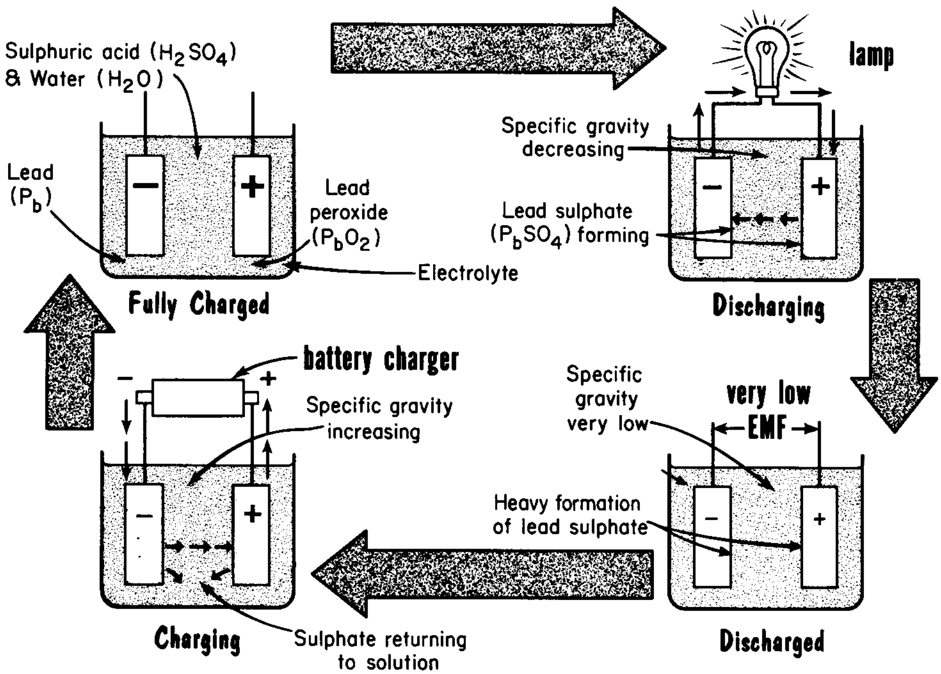
CUTAWAY VIEW OF A SINGLE LEAD-ACID STORAGE CELL

The two dissimilar metals in the lead-acid cell are finely divided or spongy lead (Pb) in the form of a plate, and lead peroxide (PbO_2). The lead is the negative electrode and the lead peroxide the positive electrode. These materials are relatively soft, and are passed into the openings of a grid to form a rugged plate. The electrolyte is sulphuric acid (H_2SO_4) mixed with distilled water (H_2O). A cell consists of several negative and several positive plates interleaved and fitted into a container, together with the electrolyte. The condition of this cell can be checked with a hydrometer which measures the specific gravity of the electrolyte (the weight of the electrolyte as compared to the weight of the water). When fully charged, the specific gravity may be 1. 25; when discharged, about 1. 1.

Charge And Discharge in the Lead-Acid Cell

Let us look at the chemical action in a lead-acid cell. When fully charged, the negative plates (electrode) consist of lead, and the positive plates, of lead peroxide. The electrolyte is sulphuric acid and water. If we connect a conductor between the positive and negative terminals, current will flow and the cell begins to discharge. During discharge, the acid content of the electrolyte decreases, and lead sulphate ($PbSO_4$) deposits on both the positive and negative plates. The amount of water in the electrolyte increases. This process continues until both the electrodes contain a maximum of lead sulphate and the specific gravity of the electrolyte is very low. At this time, since the two electrodes are no longer dissimilar, the difference of potential across them is at a minimum.

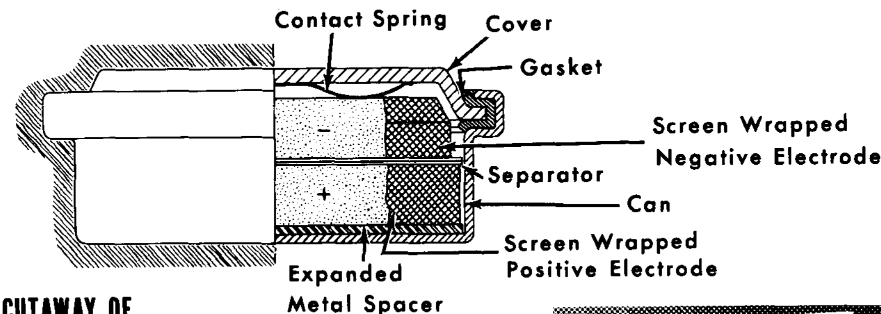
CHARGE AND DISCHARGE ACTION IN A LEAD-ACID CELL



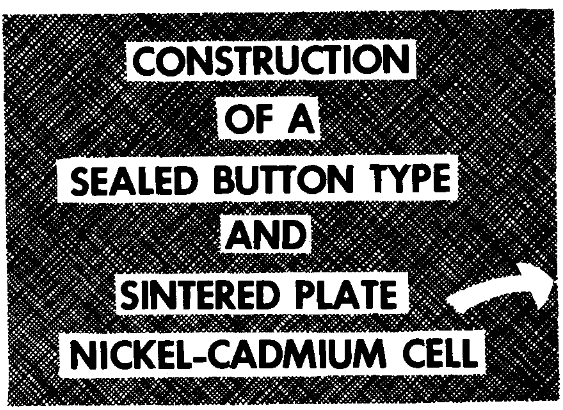
The lead-acid cell can then be recharged by reversing the direction of the discharge current. This is done by connecting the positive terminal of the battery to the positive terminal of a battery charger, and the negative terminal of the battery to the negative terminal of the charger. During the charging process, the negative plate returns to lead and the positive plate to lead peroxide. The sulphate returns to the electrolyte and its specific gravity increases. During charge, hydrogen and oxygen are given off, and some water must be added to the electrolyte to replace the water lost. This is the reason you have to add water to your automobile battery on the average of two or three times a year.

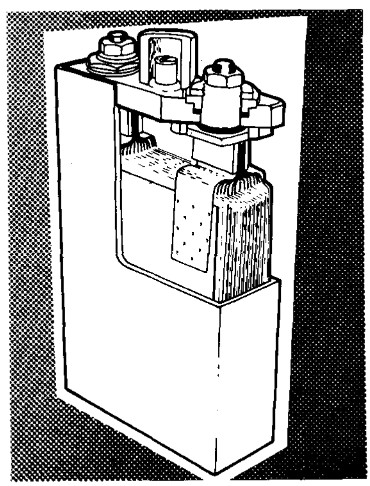
The Nickel-Cadmium Cell

The nickel-cadmium cell is a comparatively recent development that has found considerable use in portable and mobile electronic equipment. It is a mechanically-rugged cell that can withstand overcharging, overdischarging, or standing idle in a discharged condition for a long time. Being a secondary cell, this unit can be recharged after having been discharged. In the charged condition, the positive electrode of a nickel-cadmium cell is a nickelic hydroxide; the negative electrode is metallic cadmium. The electrolyte is potassium hydroxide. The average operating voltage of the cell under normal discharge conditions is about 1.2 volts.



CUTAWAY OF TYPICAL BUTTON CELL

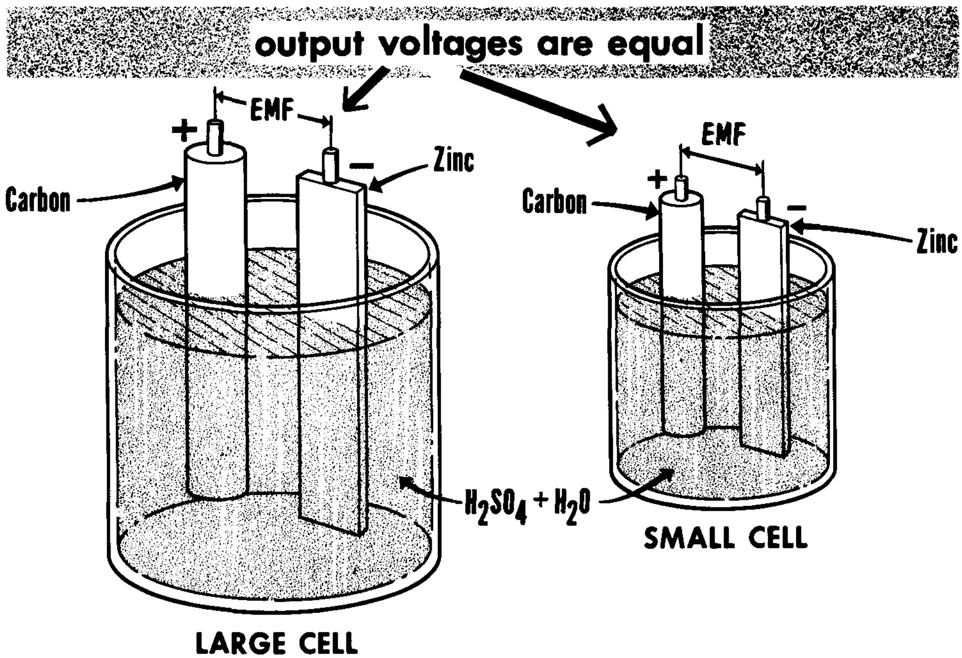




Nickel-cadmium cells come in a wide variety of shapes and types, the most popular being the hermetically-sealed variety available in rectangular, cylindrical and button form, and the sintered-plate construction requiring a vent. In the sintered-plate type, the plates are arranged in groups connected by welded group straps and are separated by layers of fabric. Positive and negative plate groups are intermeshed and placed in plastic containers. During charge and discharge of a nickel-cadmium cell, there is practically no change in the specific gravity of the electrolyte. The electrolyte acts only as a conductor for the transfer of hydroxide ions from one electrode to the other, depending upon whether the cell is being charged or discharged.

Output Voltage Rating of Cells

An interesting fact about the electromotive force (emf) generated in cells is that the output voltage of a cell depends upon the kinds of materials used in the cell, and not upon the dimensions of the materials. For instance, all zinc-carbon cells use zinc and carbon electrodes, with an electrolyte of ammonium chloride. This is true whether we consider the tiny "pen" type cells or the very large No. 6 cells used for bells and electric trains. The output voltage of both is exactly the same -- 1.6 volts for a new fully-charged cell. As we will learn later, the big difference between these cells is in the available current output. The same is true of the common lead-acid storage cell. A small cell of this type having a few plates will have an output voltage of 2.2 volts, the same as a large lead-acid cell having a great many large plates. There are many other different combinations of chemicals that give various voltages, but the quantity of chemicals involved do not affect the output voltage.

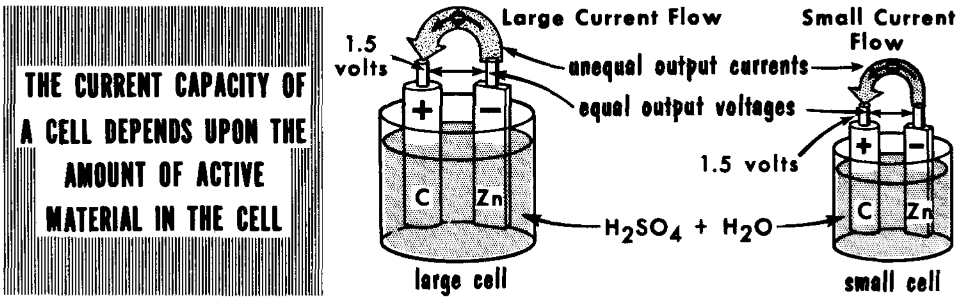


THE OUTPUT VOLTAGE OF A CELL DEPENDS UPON THE KIND OF CHEMICALS USED AND NOT UPON THE QUANTITY OF CHEMICALS USED

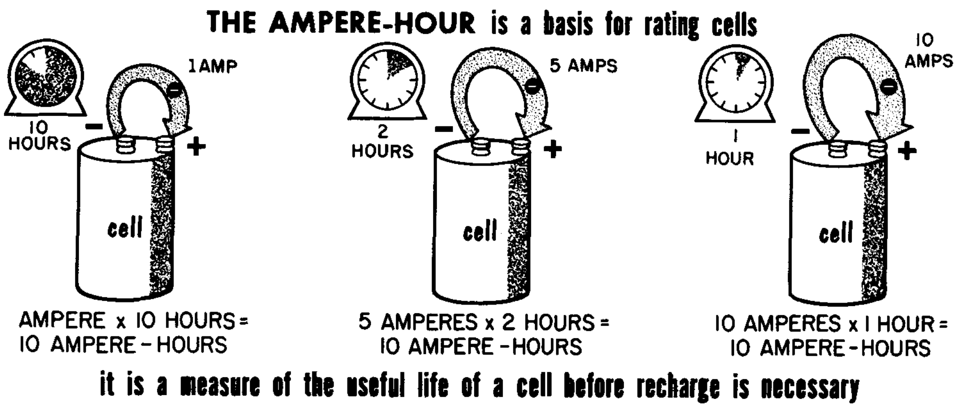
To measure the output voltage of a cell, a "voltmeter" is used. This instrument will be covered in detail later in this course. The voltmeter is connected across the terminals of the primary cell and the voltage is read on the voltmeter. The same is true of secondary cells; however, to get a more accurate check of the quality of such a cell, it is common to make the voltage measurement while the cell is discharging, or "under load".

The Current Capacity Rating of Cells

You have learned that if a wire is connected between the negative and positive terminals of a cell, a current will flow through the wire. The fact that electrons leave the cell and enter the wire is the basis for viewing the cell (or a battery of cells) as being a source of current. As long as the proper chemical action continues, the supply of electrons continues. The ability of the cell to supply electrons at a certain rate is referred to as its current capacity. The maximum rate of supply of electrons depends upon the size or amount of active material in the electrodes, assuming proper condition of the electrolyte. This explains why the large No. 6 dry cell can supply more current than the much smaller cell.



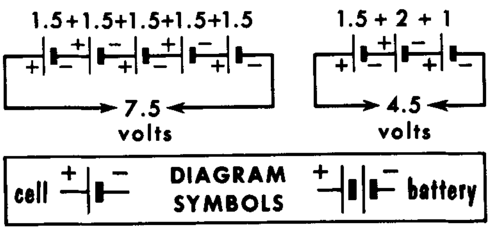
When the current capacity of a cell is expressed, (a practice most common with secondary cells), it is done by stating the maximum number of amperes which the cell will supply for 1 hour. Thus, a 20 ampere-hour cell rating means that the cell will furnish 20 amperes of current for 1 hour, after which it will be in a discharged state. If, however, the rate of discharge is less than the maximum capability, the cell can supply current for more than 1 hour. For example, the 20 ampere-hour cell could supply 1 ampere for 20 hours. Working in the other direction, the current capacity could be proportionately greater for a shorter period, such as 100 amperes for one-fifth of an hour, or 12 minutes. The product of the current in amperes and the time in hours cannot exceed the ampere-hour rating of a particular cell.

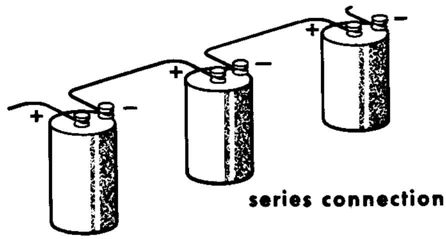


Connecting Cells in Series to Form a Battery

Under certain circumstances, the voltage available from a single cell may be sufficient, as in some small flashlights. Under other circumstances, higher values of voltage may be required. This can be achieved by connecting as many cells (primary or secondary) in series as are needed to achieve the necessary voltage. Such a bank of cells forms a battery.

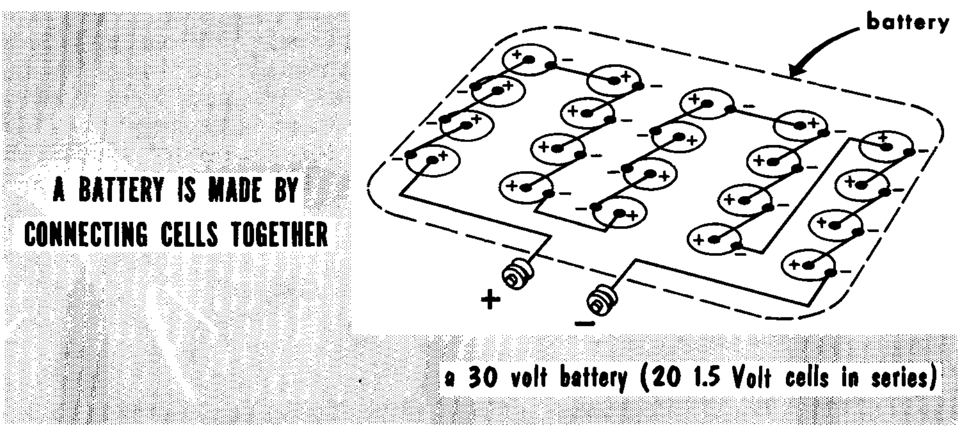
THE OUTPUT VOLTAGE OF CELLS CONNECTED IN SERIES IS EQUAL TO THE SUM OF ALL THEIR INDIVIDUAL VOLTAGES.





WHEN CONNECTING CELLS IN SERIES, THE POSITIVE TERMINAL OF ONE CELL IS CONNECTED TO THE NEGATIVE TERMINAL OF THE OTHER.

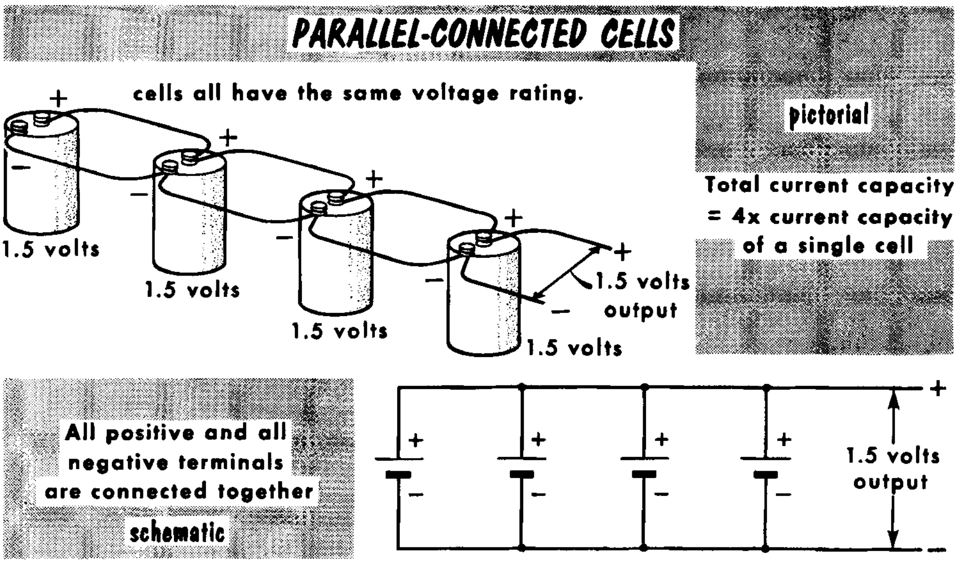
The total voltage of series-connected cells is the sum of the voltage output of each of the cells. Thus, if four 1.5-volt dry cells are series connected, the total voltage is 1.5 + 1.5 + 1.5 + 1.5 + or 6 volts. If 30 such cells are connected in series; the output voltage is 30 x 1.5 or 45 volts. The 6-volt lead-acid storage battery consists of three 2-volt cells series connected. The 12 volt lead-acid storage battery consists of six 2-volt cells series connected.



When cells are connected in series, the positive terminal of one is connected to the negative terminal of the other. By doing this, all the individual potentials or voltages aid each other, and add. The above examples involved cells rated at the same voltage. This need not be so; cells of any voltage output can be connected in series. Similarly, batteries of like or unlike voltage rating can be connected in series to increase the voltage. However, each cell (or battery) in a series arrangement should have the same current capacity.

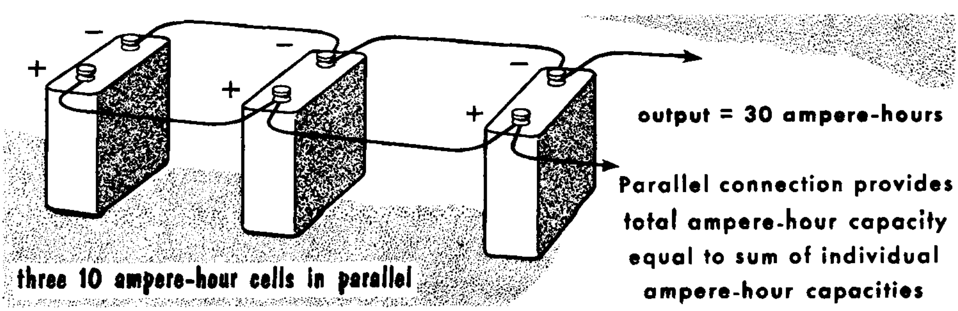
Connecting Cells in Parallel to Form a Battery

A battery can also be formed by connecting cells in parallel. Only cells of identical output voltage rating should be connected in parallel. The purpose of a parallel connection is to increase the current capacity of the battery beyond that of a single cell. The parallel connection creates the equivalent of an increase in the physical size of the electrodes and of the amount of electrolyte, and thus increases the amount of current available.

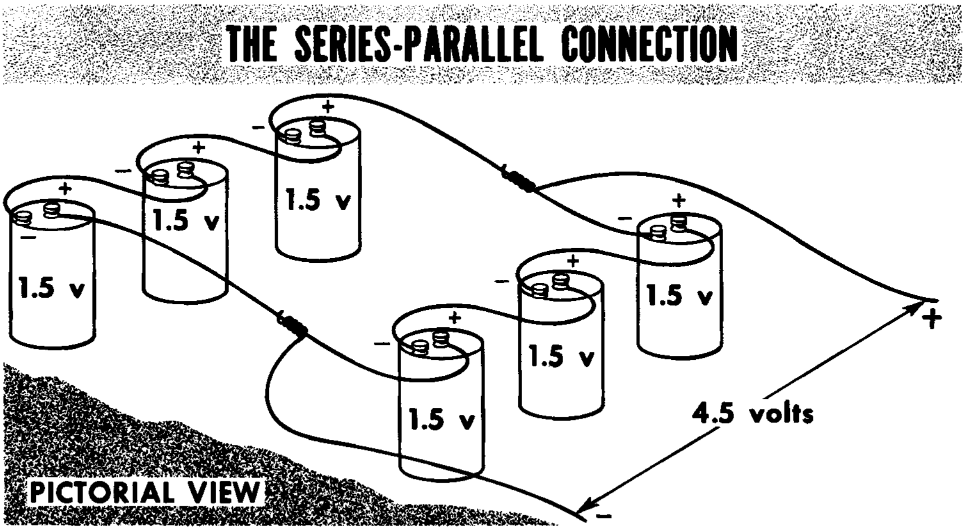


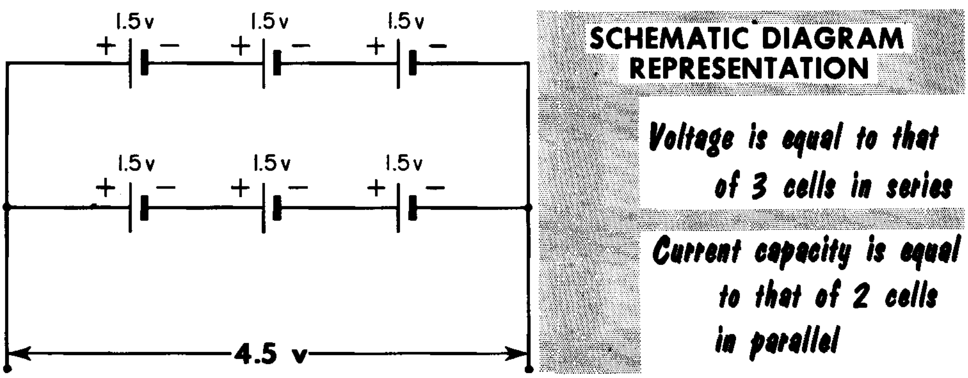
For example, if three cells are connected in parallel, the current capacity of the battery equals three times that of a single cell. Therefore, each cell is contributing one-third of the total current.

Connecting cells in parallel does not change the voltage rating. The final voltage rating of the parallel cells is the same as that of single cell. When cells of unlike output voltage are connected in parallel, current flows between the cells and constitutes wasted electrical energy. Also, there is a possibility that the cells would be damaged.



Series-Parallel Connected Cells





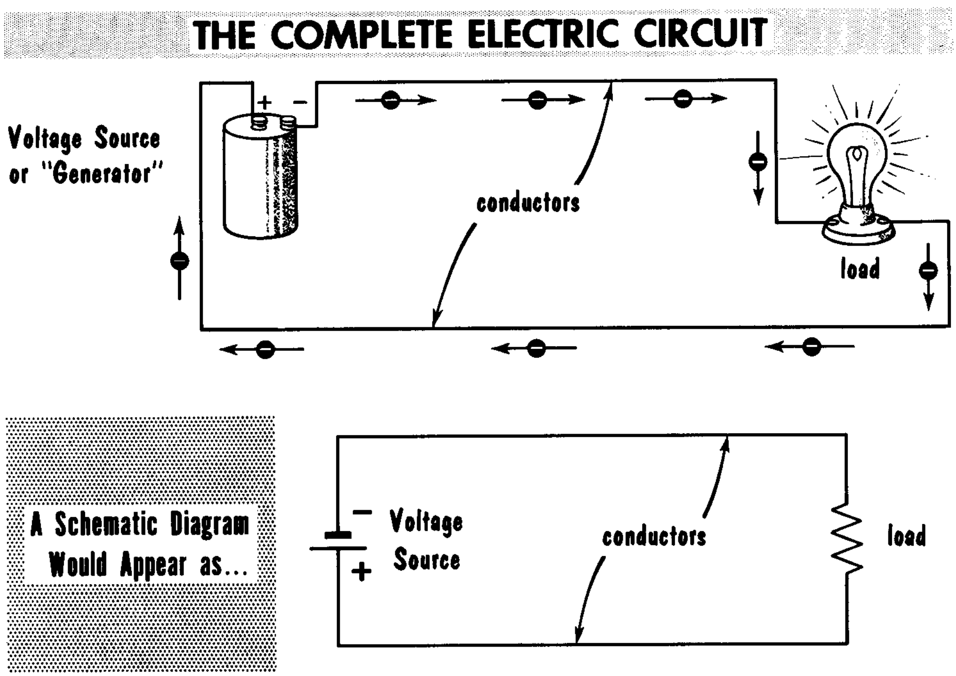
The features of series connection and parallel connection can be combined in a circuit called series-parallel. This connection gives the higher output voltage afforded by series connection and the increased current capacity afforded by the parallel connection, simultaneously. As in the previous examples in parallel connection, it is desirable that the current and voltage ratings of the cells be similar. If a high voltage cell is connected across a lower voltage cell, the higher voltage will cause current to flow through the lower voltage cell, possibly damaging it. Generally, this type of connection is seldom used since higher current capacity can be obtained by using larger cells. However, there may be emergency instances where the series-parallel circuit is the only practical method of obtaining the voltage and current combination desired.

When making a series-parallel connection, the usual rules of polarity must be observed: in series circuits, connect positive to negative; in parallel circuits, connect positive to positive and negative to negative.

The Electric Circuit

Electricity can be put to work through the use of an electric circuit. This is an organized combination of electrical components that permit some specific function to be performed. Basically, all electrical circuits consist of three main components:

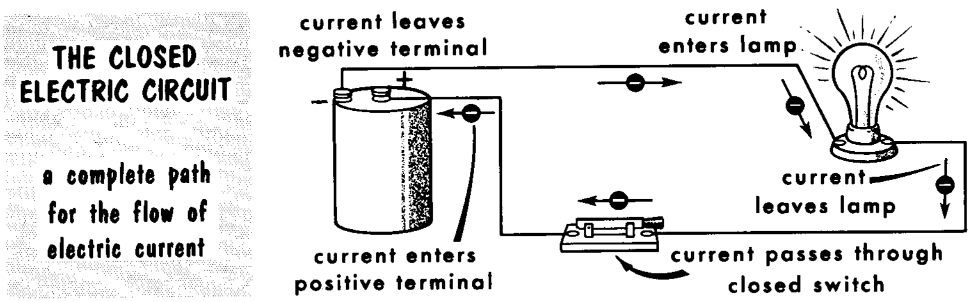
1. A source of voltage -- thus far, the only voltage source discussed has been the cell or battery. As we will study later, there are many other voltage sources, such as generators and crystals.
2. A load -- this is the component for which the entire circuit is constructed. By having voltage applied to it and current flow through it, the load will perform some specific task. The load may be a lamp, toaster, electric motor, bell or any other device that operates from a voltage applied to it.
3. A conductor -- the load may be close to the voltage source, as in a flashlight, or quite far away, as in the case of an electric power company having to deliver a voltage many miles from the generator. In either case, there must be a connection between the voltage source and the load. As we will learn in electrical circuits, the conductor that connects the load with the voltage source is invariably a copper wire.



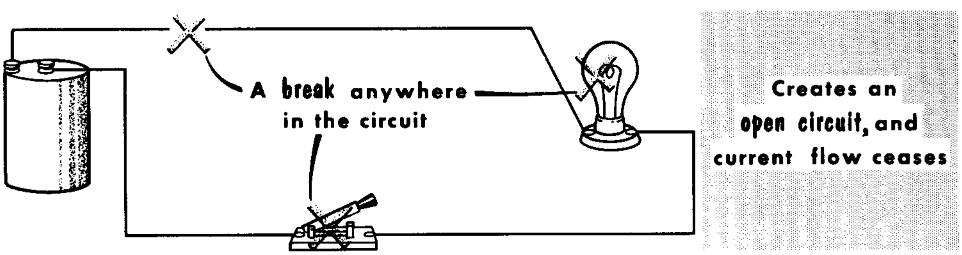
In an electric circuit, there must be a complete path for current flow. That is, there must be a conductor from the negative terminal of the voltage source to the load, and a "return" conductor from the load back to the positive side of the voltage source. Should any one of the above components be missing, we do not have an electric circuit. It is common to refer to the voltage source as the "generator" regardless of what type of voltage source is used.

The Closed Electric Circuit

Consider a very simple electric circuit consisting of a flashlight battery, a flashlight bulb, a socket for the bulb, a switch that can close or open the circuit as desired, and the wires that interconnect the elements of the circuit. We show these elements pictorially. Each of these circuit devices has two connecting points which correspond to the two openings of a piece of lead pipe. One connecting point allows the electrical energy to enter the device, the other end permits the energy to leave the device. It is to these connecting points on each device that the conducting paths (wires) are joined when forming the circuit. Each piece of conducting wire also has two connecting points, the two ends of the wire.



We have interconnected the elements of the circuit so that there is only one conducting path from one terminal of the battery to the switch -- from the switch to the bulb, and from the bulb to the other terminal of the battery. Let us also assume that we have depressed the movable blade of the switch so that there is an uninterrupted path through the switch for the current. We now have a closed circuit. It is considered closed because all the elements of the circuit are joined to each other and a continuous uninterrupted path exists for the flow of electric current from the battery through the devices and back to the battery.

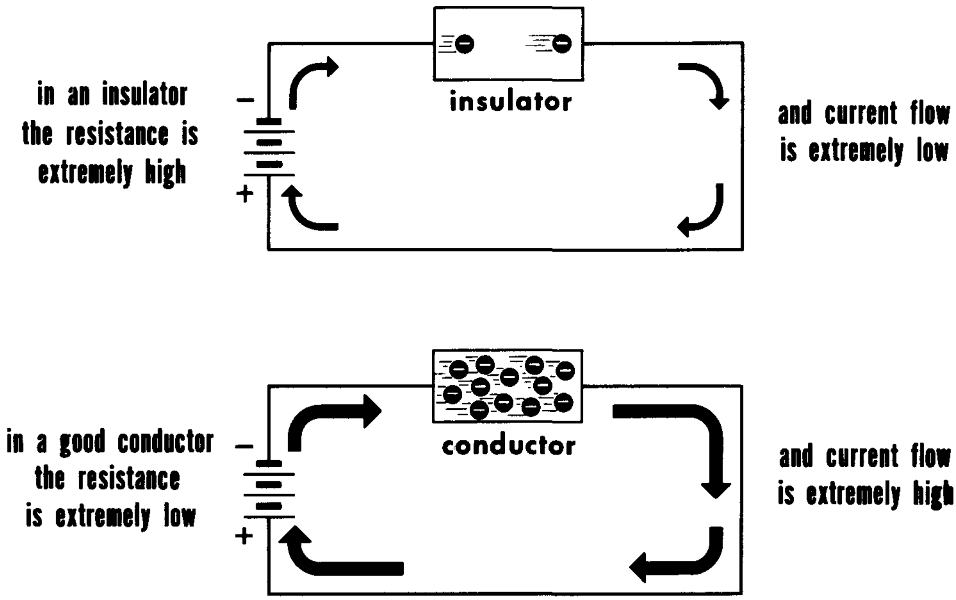


Should this closed circuit be broken (interrupted) at any point, we have what is called an "open" circuit. The break may be unintentional, such as a broken conductor, or a burnt out filament in the lamp, or it may be intentional, such as opening the switch in order to turn off the lamp. In any case, breaking the circuit causes the current to stop flowing, and we have an open circuit.

Resistance

The word resistance as used in everyday language means opposition to an action or an idea. In the language of electricity, resistance is defined as "opposition to the flow of electric current". Now, if you will recall that electric current was described as being the directed movement of free electrons, then resistance can be looked upon as being the opposition to the motion of free electrons. From this, we can conclude that wherever free electrons are in motion, there is resistance to their movement. This is called d-c (direct current) resistance. (Direct current describes current flow in one direction. Later in this course, we will study a-c or alternating current, where current flows alternately in both directions). If we relate this action to metals, all metals regardless of their shape or purpose, offer electrical resistance.

RESISTANCE IS THE OPPOSITION TO THE FLOW OF ELECTRIC CURRENT

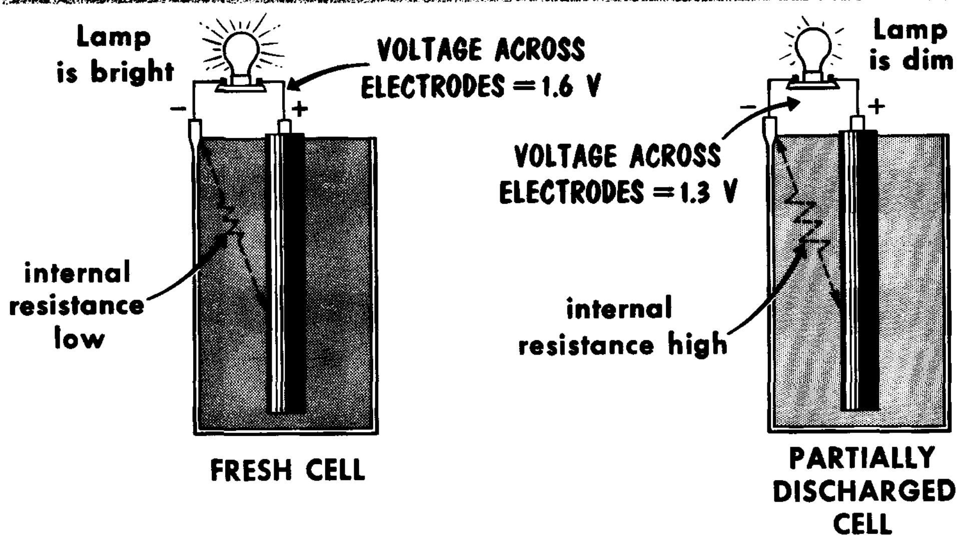


The importance of d-c resistance stems from its controlling effect on the amount of current which flows under the influence of an applied voltage. For instance, if a high voltage is applied across the ends of a glass rod, there is negligible current flow through the rod. This is so because few free electrons exist in the glass. Glass is called an insulator because it has few free electrons to perform as electric current; an insulator does not conduct electricity well. If, now, we interpret poor conductivity due to insufficient free electrons as being the equivalent of very high d-c resistance, we have still another basis of d-c resistance. So, although several different conditions underlie d-c resistance, the end result is the same -- a limitation on the amount of current flow.

D-C Resistance of Voltage Sources (Internal Resistance)

The chemical action inside primary and secondary cells involves the movement of electricity through the electrolyte in which the active electrodes are immersed. In both the dry and the wet cell, the electricity in motion within the cell is in the form of positive and negative ions -- atoms shy electrons and atoms with excess electrons. The generation of voltage is accounted for in part by the decomposition of the electrodes, as well as by a change in the makeup of the materials as the two kinds of ions enter or leave the electrodes. It is this action which ultimately leads to the supply of free electrons from the negative terminal.

CELLS HAVE INTERNAL RESISTANCE

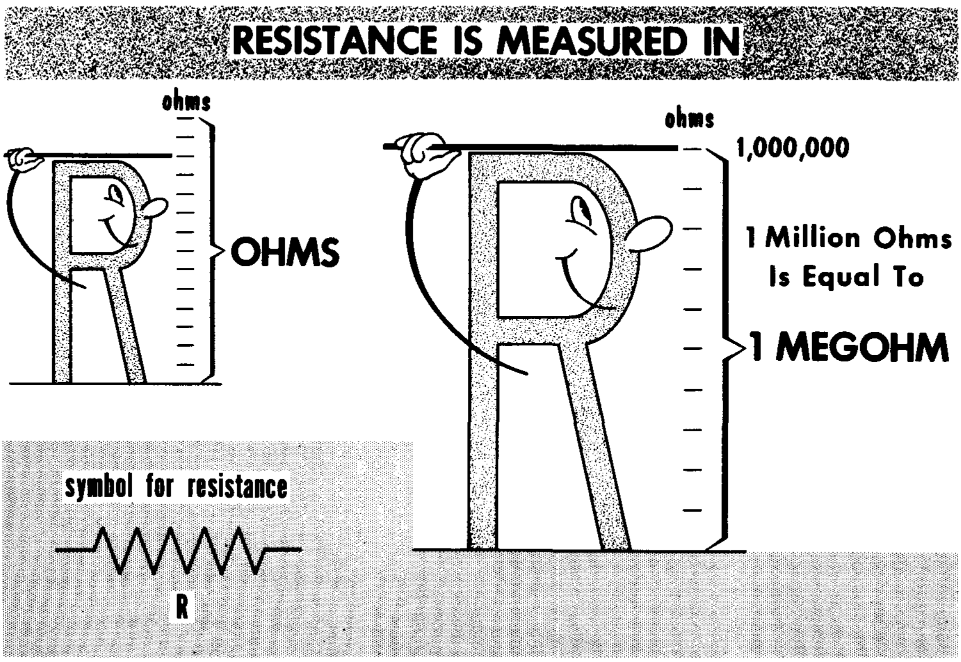


AS ELECTRODE MATERIALS ARE USED UP AND ELECTROLYTE DRIES UP, IT BECOMES DIFFICULT FOR ELECTRONS TO PASS THROUGH CELL

However, the conversion of chemical energy into electrical energy does not occur with 100% efficiency; some waste usually takes place. We attribute the loss to the presence of resistance inside the cell -- resistance that is inseparably associated with the action. It is called internal resistance. It is least when the chemicals are fresh and the action is strong. As the cells function and discharge, the internal resistance gradually increases, becoming higher when the cell is discharged. In the secondary cell, the internal resistance is high when the battery is discharged but reduces to its normally-low value as the battery is recharged. When cells are part of an electrical system and are supplying the circuit current, the internal resistance of each cell of the battery is considered as being in the path of the circuit current.

Unit of Resistance (The Ohm)

D-c resistance is indicated by the capital letter R and, when symbolized, it appears as a zigzag line. Because the amount of resistance presented to current flow is not always the same and it is necessary to express and compare different amounts of resistance, science has established a unit of resistance. It is the ohm. The origin of the word ohm is the name George Simon Ohm. He was a German scientist who advanced the concept that a definite relationship existed between the amount of current flow, the amount of voltage applied, and the resistance of the electrical system. The concept defined the unit ohm as being that amount of resistance which limits current to 1 ampere when the applied voltage is 1 volt.

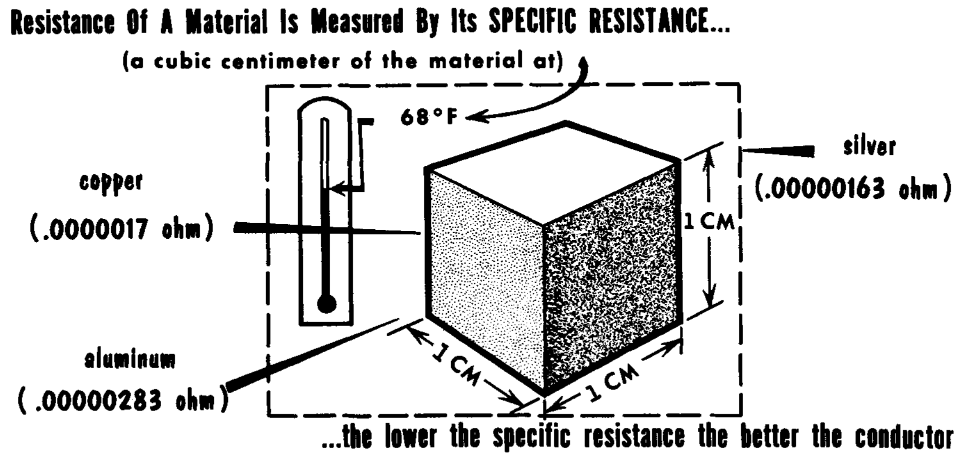


Practical amounts of resistance encountered in radio circuits range from a fraction of 1 ohm to several million ohms. As a matter of convenience, the words "milli" meaning one-thousandth, "kilo" meaning thousand, "meg" meaning one million, and "micro" meaning one-millionth are used in connection with resistance. Thus, .001 ohm is 1 milliohm, 1000 ohms is 1 kilohm, 1,000,000 ohms is one megohm, and 1 micro-ohm is 1/1,000,000 or .000001 ohm. You will recall these words being used in connection with voltage and current values. In addition to prefixes being used to express values of resistance, several letter symbols indicate resistance.

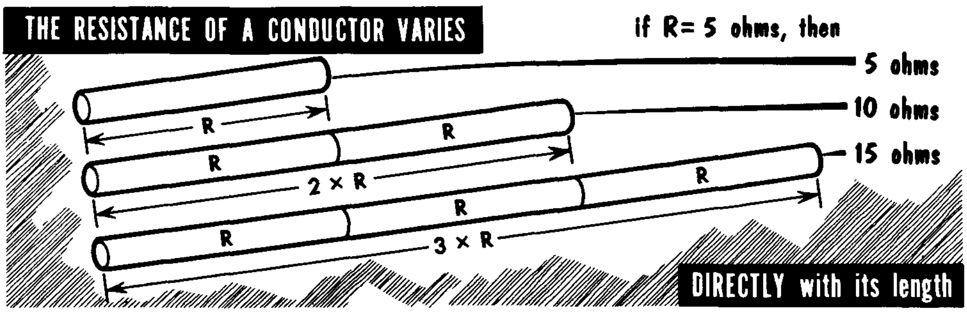
OMEGA Ω	=ohms	100 Ω	=100 ohms
KILO or K	=1000	4 K	=4000 ohms
MEG or M	=million	5 MEG	=5,000,000 ohms

Factors That Determine Resistance in Metals (Material and Length)

Several physical features of conductors determine the amount of resistance they present to the flow of current. They are: (1) kind of material; (2) length; (3) cross-section; (4) temperature.



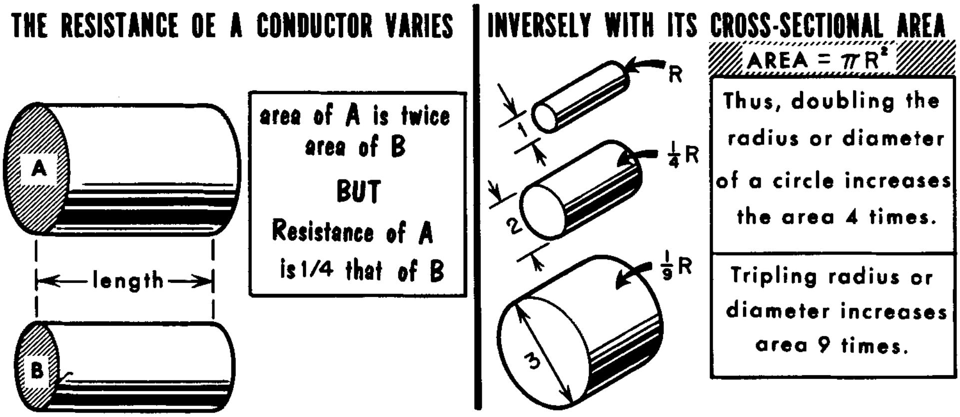
A basis for comparing the usefulness of metals as conductors of electricity is by specific resistance. This is the ohmic value of a standard amount (1 cubic centimeter) of the substance measured at a particular temperature (68° F). If the specific resistance of a substance is high, it is a "high-resistance" material; if the specific resistance is low, it is a "low-resistance" material. The lower the specific resistance, the better the material performs as a conductor of electric current.



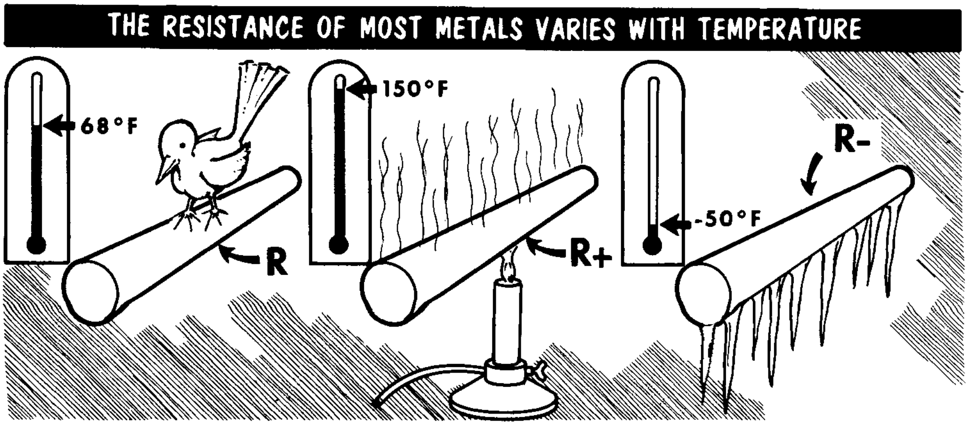
Given any one particular conductor material having a specific resistance, the longer the conductor, the more will be its resistance. If, for instance, the conductor is pure copper and a piece 1 foot long and 3/8 inch in diameter has a resistance of .00005 ohm, a piece 100 feet long will have a resistance 100 times as great or .00005 x 100 or .005 ohm. Each foot of length contributes its share of resistance to the total. This leads to the basic rule "the resistance of a conductor is directly proportional to its length".

Factors That Determine Resistance in Metals (Cross Section and Temperature)

The "thickness" or cross section of a wire made of a given material has a great effect on its resistance. The greater the cross section of a conductor (the thicker the wire), the less its resistance per unit length -- the reason being that the greater the cross section, the more numerous the free electrons to be moved by a given applied voltage. Given two equal lengths of the same wire A and B, with A having twice the cross-sectional area of B, wire A will present one-half the resistance of B. (Do not confuse cross-sectional area with diameter.) We can state this as a rule: the d-c resistance of a given conductor is inversely proportional to the cross-sectional area.



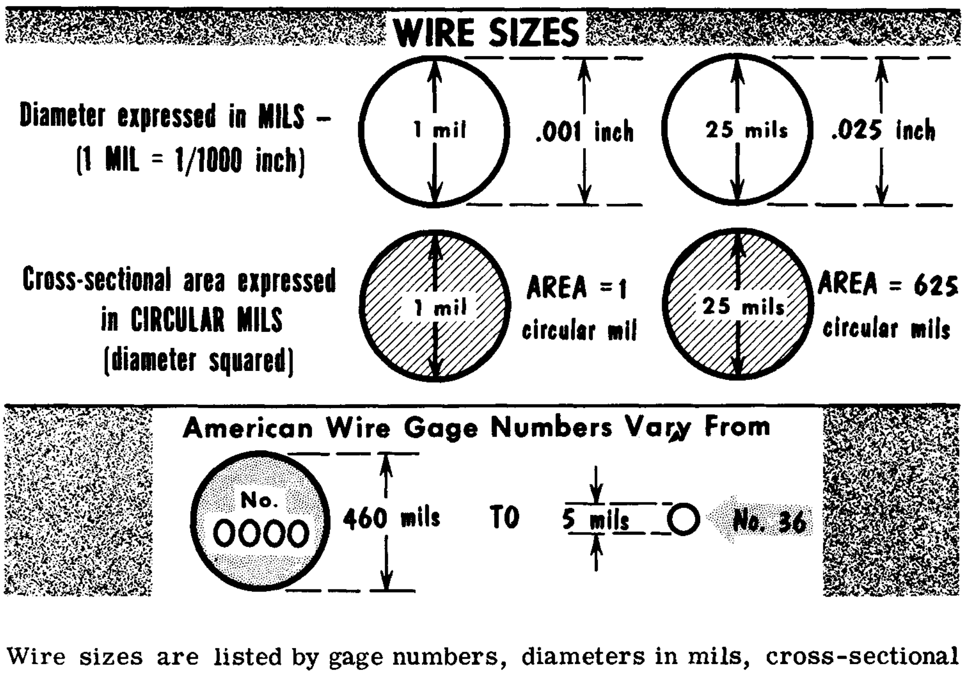
The higher the temperature of ordinary metal conductors, the higher the resistance per unit length and cross section. The reason for this is simply that an increase in temperature causes more violent motion of the free electrons and atoms inside the wire, thus increasing the number of collisions between electrons, and between electrons and atoms. The result is that the free electrons are retarded more in their advance through hot wire than through cold wire.



Practical Units of Wire Sizes — American Wire Gage

It is standard American practice to use different units to express length and cross-sectional area. Length can be expressed in a variety of units--inches, feet, or miles (depending on the circumstances). Units of 1000 feet are commonly used. To express cross-sectional area, it is customary to use circular mils, in which case, the diameter is stated in terms of the mil. A mil equals 1/1000 inch (.001 inch). One circular mil is the area of a circle with a diameter of 1 mil. This follows from the formula for the area of a circle, which is the area proportional to the square of the diameter, or d^2. Thus, a round wire with a diameter of 100 mils has a cross-sectional area of d^2 or 100 x 100 = 10,000 circular mils. If the diameter is 400 mils, the cross-sectional area is 400 x 400 or 160,000 circular mils.

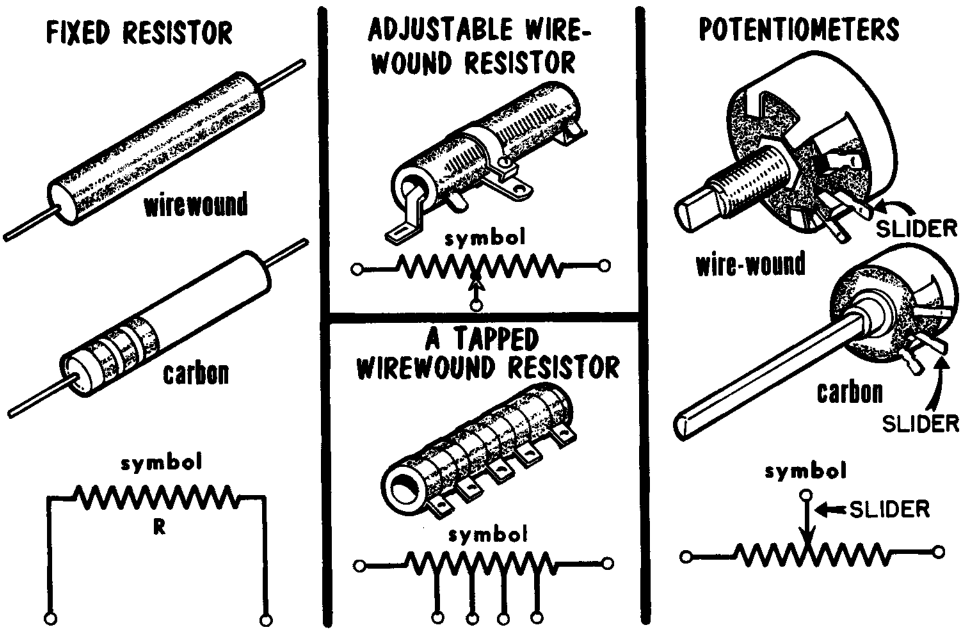
If the wire is rectangular, the cross-sectional area is expressed in square mils, and is equal to the width multiplied by the height, each dimension being expressed in mils. A bus wire 1 inch square has a width of 1000 mils and a height of 1000 mils; therefore, it has a cross-sectional area of 1,000,000 square mils.



Wire sizes are listed by gage numbers, diameters in mils, cross-sectional areas in circular mils, resistance in ohms per thousand feet, and weight in pounds of conductor material per thousand feet. Although there are many types of gages, the most frequently used is the American Wire Gage (AWG), formerly called the Brown and Sharpe (B & S) Gage. The AWG is so constructed that the ratio of any one diameter to the next in order is constant. The largest size, No. 0000 or 4/0, has an arbitrary diameter of 460 mils and the smallest, No. 36, a diameter of 5 mils. Between these two sizes there are 39 other sizes. A wire table of the AWG is given in the Appendix.

Resistors

The manufactured version of resistance is the resistor. A resistor is an electrical component which, when made part of an electrical circuit, is intended to introduce a definite amount of d-c resistance in very compact form. In many applications, the purpose of the added resistance is to limit the amount of circuit current to a predetermined value. As we will learn, there are other functions as well.



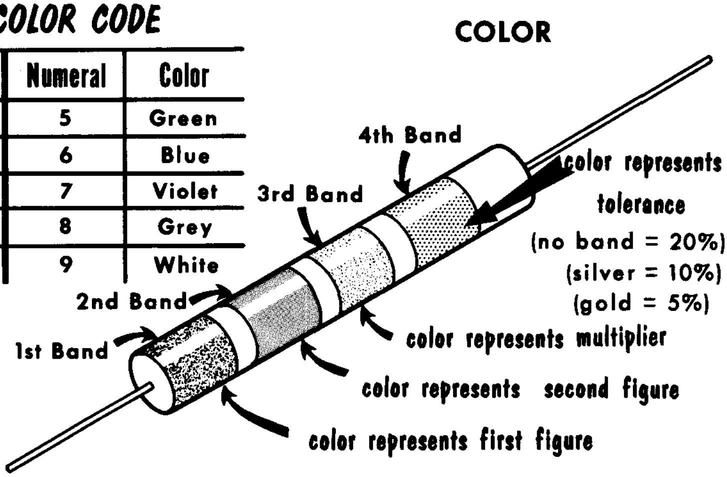
Resistors are made in many sizes and shapes, and in a variety of materials. The so-called "wire-wound" type makes use of a special alloy wire or ribbon as the resistance element, and is wound on an insulating form, with or without a ceramic covering. Because of the current carrying capabilities of wire, this type is used when the circuit currents are relatively high. They are available in fixed, adjustable, or continuously-variable types. Resistance values range from a fraction of an ohm to about 100,000 ohms, in power (wattage) ratings from 1/2 watt up to 200 watts. Another type is made of graphite or carbon powder that is formed into rods and cut to length according to the resistance desired. They are fixed in their ohmic value and in their power-handling capabilities. Another type of fixed resistor consists of a thin film of metal deposited on an insulating form. Both the carbon and the deposited-metal types are low-current units and are availabe in resistance values from several ohms to as high as 50,000,000 ohms (50 megohms); and in wattage ratings from 1/2 watt to 2 watts. Connection to the wire resistor is by means of terminals. Carbon and deposited-metal types are connected by means of wire leads called pigtails. Continuously-variable resistors (potentiometers) provide any amount of resistance to the maximum. A rheostat is a kind of variable resistor designed to handle large amounts of current.

Resistor Ratings and Color Codes

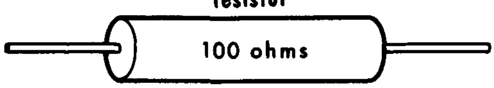
The usefulness of a resistor is measured by its electrical rating. There are three factors used to determine this rating: the ohmic value; the heat dissipating capability; and the resistance tolerance. The tolerance of a resistor is the permitted variation in percent from its rated ohmic value. The heat dissipating capability is also known by another name -- wattage rating -- which we shall learn in detail later. The ohmic value cannot be recognized by merely looking at the resistor; hence, the resistor is labeled in some way. Fixed, adjustable, or variable wirewound resistors often have their values stamped right on them. In the case of fixed carbon and deposited-metal resistors, a color code consisting of three or four colored bands or dots is used. The color code indicates both ohmic value and resistance tolerance. To "read" the code correctly, the resistor must be held in such a position that the colored bands or dots appear at the left, and the value is then read from left to right. The length and thickness of the resistor is an indication of its wattage rating. If no tolerance indication is given, a tolerance of 20% can then be assumed.

RESISTOR COLOR CODE

Numeral	Color	Numeral	Color
0	Black	5	Green
1	Brown	6	Blue
2	Red	7	Violet
3	Orange	8	Grey
4	Yellow	9	White



RESISTOR TOLERANCE

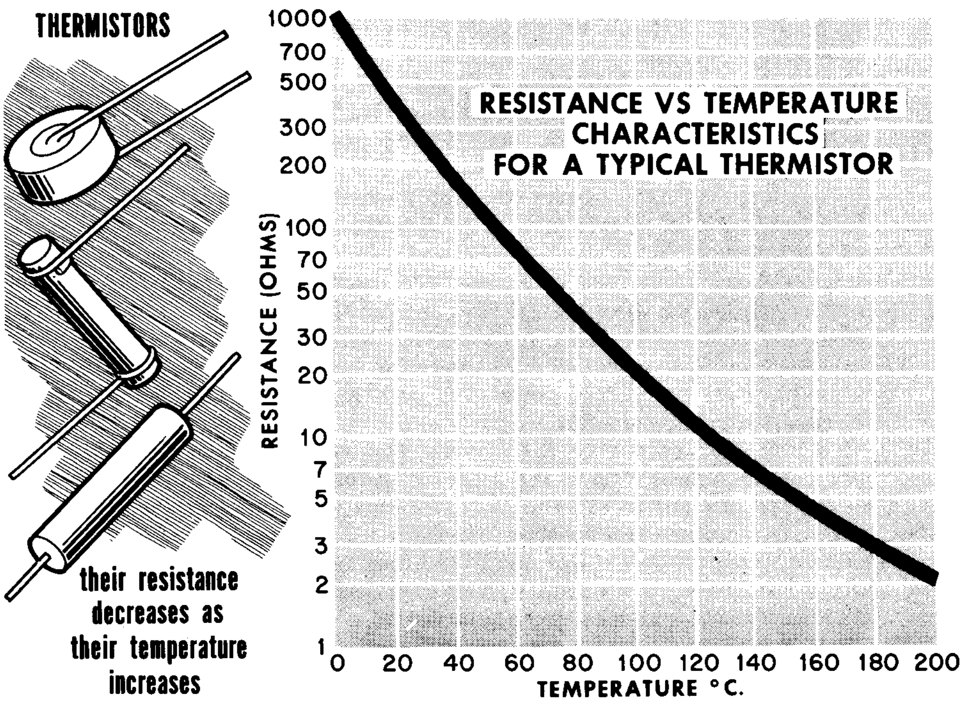


If Tolerance is:	Resistance may be anywhere from:
20%	80 ohms to 120 ohms
10%	90 ohms to 110 ohms
5%	95 ohms to 105 ohms
1%	99 ohms to 101 ohms

Assume a fixed carbon resistor that is color-coded yellow, violet, orange and silver. It would read "four, seven, three zeros, and 10%", meaning 47,000 ohms rating with a resistance tolerance of plus or minus(±) 10%. An earlier code now obsolete consisted of a body color, an end color and a dot. The body color represented the first significant figure, the end color represented the second significant figure, and the dot represented the number of zeros.

Special Resistors

Sometimes, it is necessary to use resistors having special characteristics or functions. Such a unit is the ballast resistor. This is a resistor whose resistance increases as the amount of current flowing through it increases, or as its temperature increases. While this characteristic is true of all conventional resistors, the ballast resistor is placed in certain circuits for the sole purpose of maintaining a high resistance while current flow is high, and to exhibit a lower resistance when the current flow is low. In some instances, it is used to provide nothing more than a "voltage drop" in a circuit.



Another special resistor is the thermistor. Unlike conventional resistors whose resistances increase as their temperatures increase, the thermistor exhibits the opposite properties -- its resistance decreases when its temperature increases (has a negative-temperature coefficient). Thus, an increase in current flow through a thermistor would cause its temperature to increase and its resistance to decrease. As we will see, this property of the thermistor is used in special "temperature-compensating" circuits.

The third special resistor is the voltage-sensitive varistor. The varistor exhibits a significant and non-linear change of resistance with applied voltage. An increase in the applied voltage causes an increase in current which is much higher than is the case with a standard resistor. Because of the change of resistance with voltage, no resistance rating is listed regarding varistors. Instead, a voltage is listed for a specified current.

The unit of quantity for electric current is the coulomb; a coulomb per second is defined as an ampere.

Electric current has three effects: heat; magnetic; and chemical.

Resistance is the opposition to the flow of electric current.

Electrical circuits consist of a source of voltage, a load, and a conductor.

The resistance of any material depends on its length, cross-section, and temperature.

A primary cell is not rechargeable and has a limited life; a secondary cell is rechargeable and has a longer life.

Dry cells are rated in ampere-hours at a specified rate of current flow.

The most commonly-used primary cell is the zinc-carbon type.

The most widely-used secondary cell is the lead-acid "storage" type.

The lead-acid cell can be recharged by reversing the direction of the discharge current.

Batteries should be tested "under load" (while discharging).

The current capacity rating for storage batteries is usually based on a steady 20-hour discharge. A 100 ampere-hour lead-acid battery will furnish 5 amperes for 20 hours.

A hydrometer is used to measure the specific gravity of the electrolyte in a lead-acid cell, which is a good indication of the charge condition of the battery.

When cells are connected in series, the positive terminal of one is connected to the negative terminal of the other.

Connecting cells in series increases the voltage available.

Connecting cells in parallel increases the current capacity of a battery.

Series-parallel connected cells give both higher voltage output and increased current capacity.

Active material refers to the spongy lead and lead peroxide which are pressed into the grids of the plates of a lead-acid cell.

The ohm is the unit of electrical resistance; a resistor is the component which offers this resistance.

The usefulness of a resistor is determined by its electrical rating.

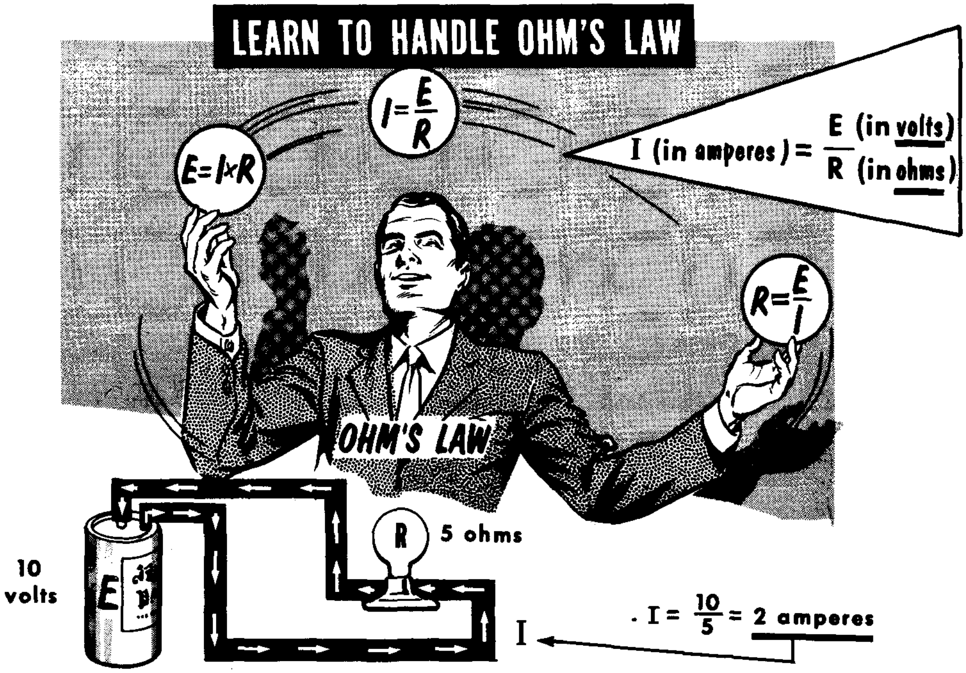
Ballast resistors, thermistors, and varistors are resistors which have special characteristics and applications.

REVIEW QUESTIONS

1. What is the unit of measurement for resistance? For electric current? For electromotive force?
2. Name three types of resistors having special characteristics and applications. What are these characteristics?
3. Define resistance. What physical factors influence resistance?
4. What colors are used to indicate a resistance of 4700 ohms, 10% tolerance?
5. How are dry cells rated?
6. Electrical circuits consist of three main components. What are they?
7. What is the difference between a primary cell and a secondary cell?
8. Upon what does the voltage of a primary cell depend?
9. How should batteries be properly tested?
10. How does connecting cells in parallel affect the current capacity of a battery?
11. What method of combining cells gives both increased current capacity and higher output voltage?

Ohm's Law For Finding Current

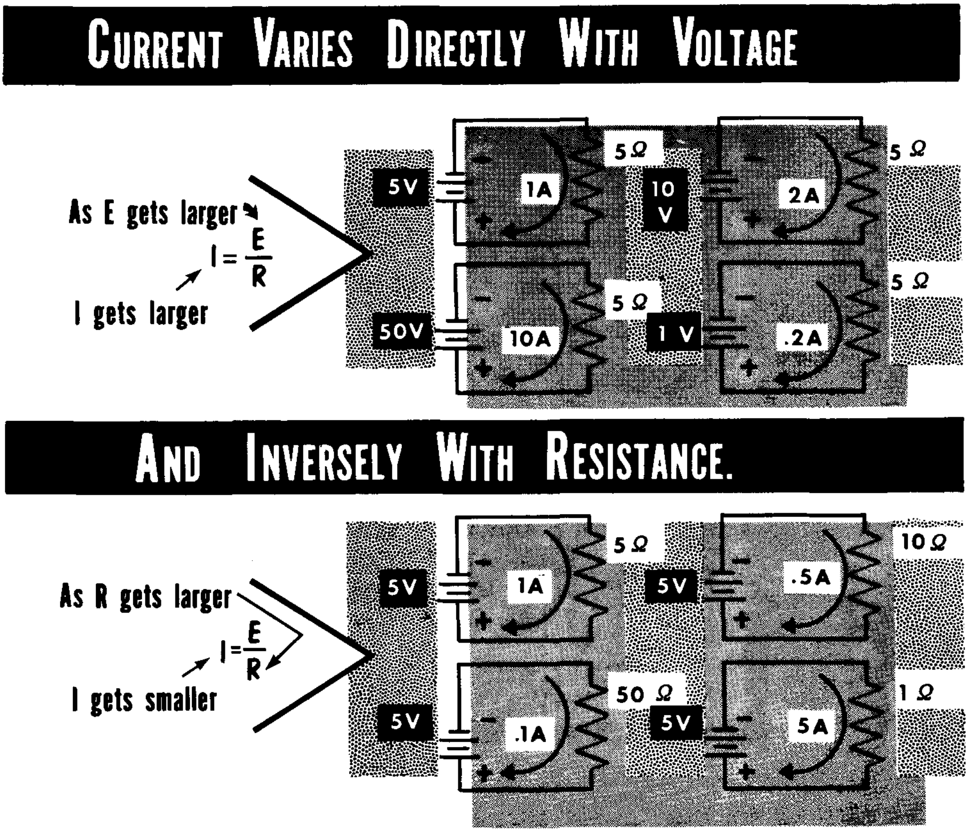
Applying a voltage to a circuit causes current flow; the presence of resistance opposes the flow of current. How can we reconcile the seemingly opposite actions of voltage and resistance? George Simon Ohm, a German scientist, investigated the behavior of electric current more than 100 years ago and summed it up in a single, simple formula called Ohm's law. This law, in its various forms, is one of the foundations on which the study of electricity rests.



Let us examine the formula by studying its application to a very simple d-c circuit. The battery supplies a voltage we call E to a flashlight bulb filament whose operating resistance we call R. We neglect the resistance of the circuit-connecting wire, the internal resistance of the battery, and the resistance of the switch. When the switch closes the circuit, current flows. But how much current? To answer the question, we must know the values for E and R, or how much voltage is pushing and how much resistance is opposing the current. So we shall say that E = 1.5 volts and R = 3 ohms. Then I = 1.5/3 or .5 ampere, which also equals 500 milliamperes. Let's try another set of values. Suppose that E = 3 volts (or twice as much as before) but R is 3 ohms (the same as before). How much current would flow in the circuit? Substituting the numbers in the equation, I = 3/3, or 1 ampere. Doubling the voltage doubled the current. The significant fact to remember here is that the current varies directly with the applied voltage when the resistance remains constant. This is one of the basic relationships between current and voltage when resistance is kept constant. The same relationship would hold true if the voltage were reduced while the resistance remained constant.

Ohm's Law (Cont'd)

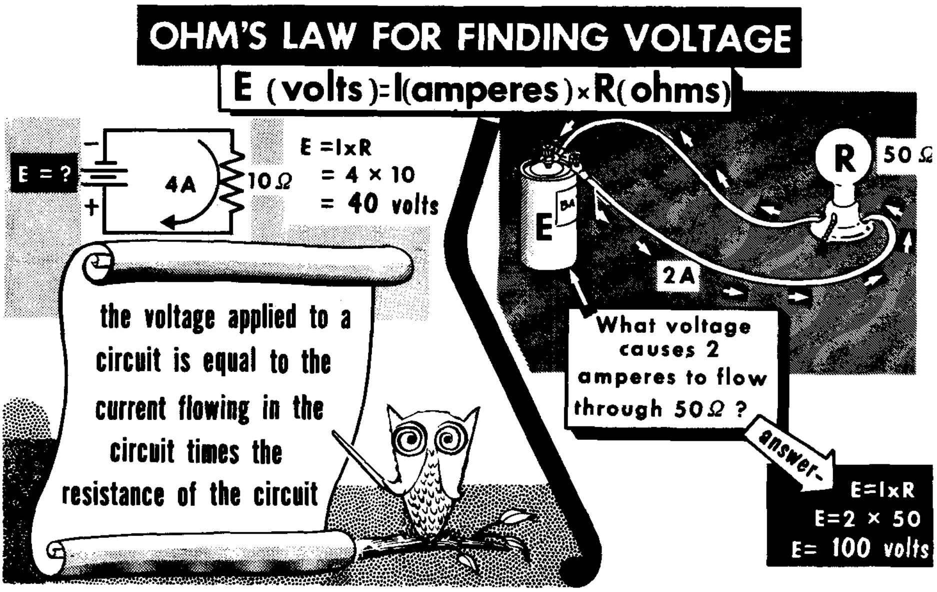
We have seen that current varies directly with voltage when resistance is constant, but what happens to the current when the resistance varies? Since current is still the unknown quantity, we work with the same equation ($I = E/R$).



Assume that E = 1.5 volts, but now, we shall substitute a different filament for R. Now the operating resistance is 10 ohms. How much current will flow in the circuit? $I = 1.5/10 = .15$ ampere (or 150 milliamperes). It is readily evident that with the voltage constant, any increase in resistance reduces the current. In fact, we can be specific and say that current varies inversely with resistance. Let's hold E at 1.5 volts and now assume that R is 30 ohms, or three times the value used before. Now $I = 1.5/30$ or .05 ampere (or 50 milliamperes). Thus, the resistance was increased three-fold and the current decreased to one-third its previous value. In other words, when the voltage is held constant, the current varies inversely with the resistance. This too is a basic relationship. Make another computation with E = 1.5 volts and R = 5 ohms. Your answer is .3 ampere (or 300 milliamperes). It shows that decreasing resistance increases the current -- always inversely -- with the change in resistance.

Ohm's Law for Finding Voltage

There is still another form of Ohm's law. It enables the determination of the voltage E when the current I and the resistance R are known. This equation is a tool for determining the answer to two possible questions that are frequently encountered in electrical circuits. One is: What value of voltage causes the flow of a known current I through a known resistance R? The other is: How much voltage must be applied in order for a given current I to flow through a known resistance R?

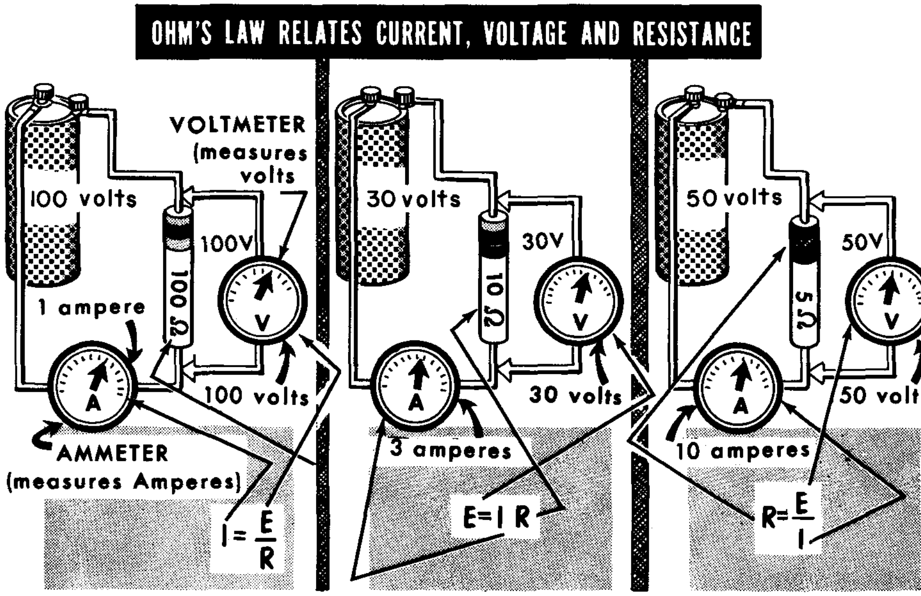


For instance, a flashlight bulb requires .5 ampere (or 500 milliamperes) current for proper operation, and its operating resistance is 3 ohms. How much voltage must be applied for this value of current to flow through the bulb filament? Since the voltage is unknown, the equation $E = I \times R$ is used. Substituting the known electrical quantities for the letters, we have $E = .5 \times 3 = 1.5$, or $E = 1.5$ volts. Let's consider another case. A different flashlight bulb is used. This one is rated at .5 ampere current (I) flow but its operating resistance (R) is 30 ohms. How much voltage must be applied so that the filament will glow with the proper intensity? Now $E = .5 \times 30 = 15$, or $E = 15$ volts. With the current kept constant, an increase in resistance requires an increase in applied voltage. In other words, the change in required voltage varies directly with the change in resistance.

If you will examine the equation for voltage (E) closely, you will note that E is the product of two quantities, I and R. If either of these quantities remains constant, the voltage E changes in direct proportion to the change in the other quantity. Suppose $I = .25$ ampere and $R = 3$ ohms. The voltage then equals $.25 \times 3$ or .75 volts. Compare this with the first problem on this page. Note that halving the current reduced the required voltage exactly in half.

Ohm's Law For Finding Resistance

There is still another important form of the basic Ohm's Law. It is used when the resistance R is the unknown quantity and current I and voltage E are known. This formula is used to determine the amount of resistance that is present in a circuit when a given amount of current is being driven through a circuit by a known amount of voltage. It is also used to determine the value of resistance required to limit the current to a known value when a voltage of a known value is applied.



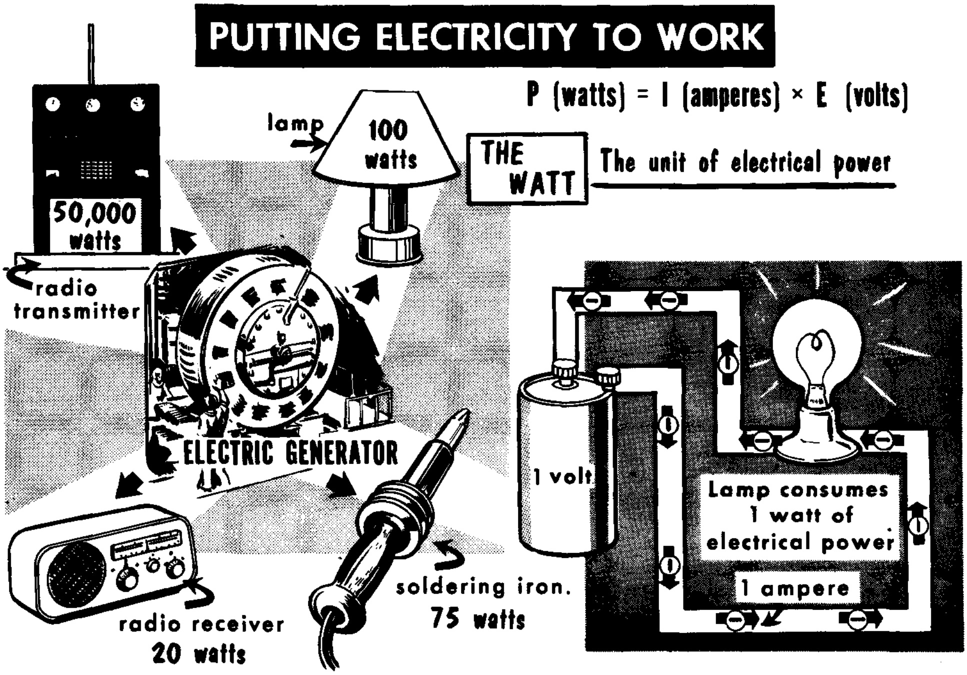
The formula shows that R varies in direct proportion to E and inversely with I. If E increases and I is constant, R must increase. If E is constant and I increases, R must be less; if E is constant and I decreases, R must increase. In other words, R changes in the same direction as E, but in the opposite direction to I.

Assume E to be 3 volts and the flashlight bulb filament to be rated at .15 ampere. What is the operating resistance of the filament? Solving for R = 3/.15 = 20 ohms. Consider this case: E remains at 3 volts, but now we use a bulb whose filament requires .3 ampere for proper operation. What is its resistance? Solving R = 3/.3 = 10 ohms. Note that with the voltage constant, doubling the current required a reduction of the resistance to one-half the original value. This is the inverse proportion relationship. Let's double the voltage and keep the current constant at .3 ampere. Then R = 6/.3 = 20 ohms. Doubling the voltage with the current constant required a two-fold increase in resistance. This is the direct proportion relationship.

Try solving this problem: E = 150 volts and I = 20 milliamperes (20 milliamperes = .02 ampere). What is the value of R? Is your answer 7500 ohms?

Electrical Power and the Watt

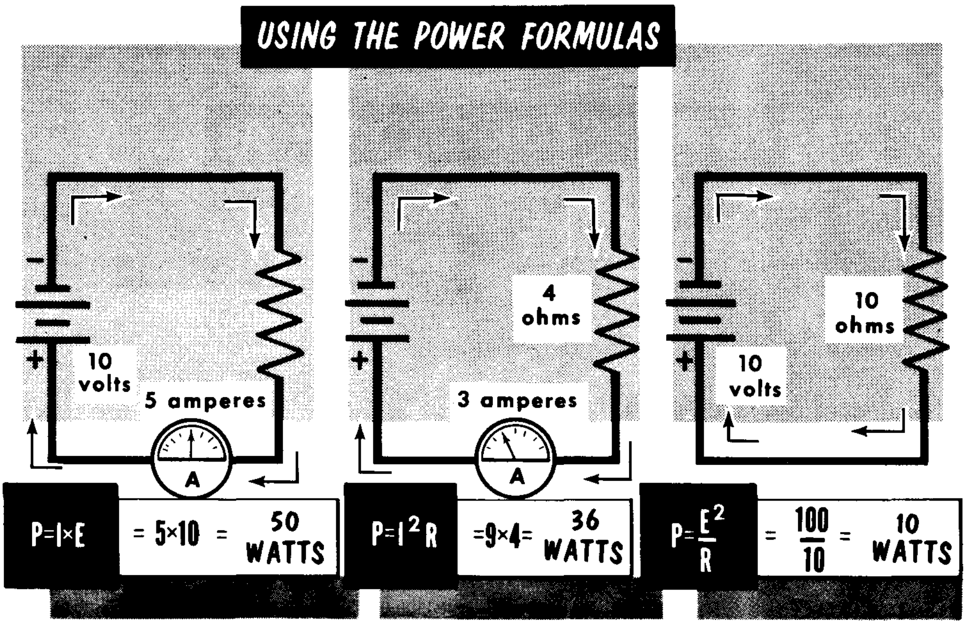
When a voltage forces current to flow through a resistance, it does so against the opposition of the resistance. Overcoming the resistance of the circuit uses up electrical energy. We can compare this to pushing a large box across the floor. In this case, the opposition is the friction of the floor to the box. This friction causes a certain amount of heat to be produced, with the heat representing a loss, or wasted energy. In a simple electrical circuit containing a battery, conductor, lamp, and switch, there will be an electrical loss in the form of heat throughout the circuit. For all practical purposes, however, the heat loss everywhere but in the lamp is insignificant. In the lamp, the filament glows white hot, and radiates light. The rate at which electrical energy is consumed or used up is called electrical power, designated by the letter P.



The unit used to express electrical power is the watt, named after James Watt, the inventor of the steam engine. Using our electrical standards, we can say that 1 watt of electrical energy is used up forcing 1 ampere of current through a circuit under the influence of 1 volt. It is important to remember that power represents the rate of doing work, and not the amount of work done. A man can do as much work as a steamshovel, but in far more time -- hence, a man has far less power. Electrical power is equal to the amount of current in a circuit times the voltage applied to the circuit ($P = I \times E$), and is measured in watts. To compare electrical to mechanical power, we use the figure 746 watts (electrical horsepower). This amount of electrical power will do as much work as a mechanical device capable of performing 550 foot-pounds of work per second (1 mechanical horsepower).

Power Formulas

Being able to determine the power conditions in an electrical system is very important. It underlies the selection of the correct component to do a job, as well as recognizing if the operating conditions are right or wrong. Three related equations enable the calculation of electrical power (P) in watts. Each has a particular application, depending on which two of the three quantities -- volts (E), current (I) and resistance (R) -- are known. Under any circumstances, either E and I, I and R, or E and R must be known.



Let's solve some power problems using the battery and the flashlight bulb. Assume that E is 1.5 volts and the bulb filament requires .5 ampere (therefore I is .5 ampere). The applicable equation is $P = E \times I$. Using numbers instead of the letters, $P = 1.5 \times .5 = .75$ watt or 750 milliwatts.

Suppose we do not know the voltage, but we do know that the filament current is .5 ampere and the operating resistance is 3 ohms. Then $P = I^2 \times R$ or $(.5 \times .5) \times 3 = .25 \times 3 = .75$ watt.

As a third case, assume that we know the voltage to be 1.5 volts and the resistance to be 3 ohms. We do not know the current. The applicable formula is $P = E^2/R$ or $P = (1.5 \times 1.5)/3 = 2.25/3 = .75$ watt.

Using the above examples of power formulas, try solving the power for the following conditions: E = 120 volts, I = 10 amperes and R = 12 ohms. Use each of the three formulas, using only those quantities which apply. Is your answer 1200 watts? (1200 watts = 1.2 kilowatts). It should be!

Power Rating of Resistors

All electrical devices and circuits bear some association to electrical power ratings, but resistors in electrical and communication equipment are most prominent in this connection. The wide variety of resistors have many kinds of uses, the foremost of which is the control of current. A prime requirement, therefore, of all resistors, is the ability to dissipate the electrical energy delivered to them; i.e., to be able to dissipate the consequent heat generated within them. The power rating of the resistor, in watts, expresses this capability, assuming the resistor has adequate circulation of air around it.

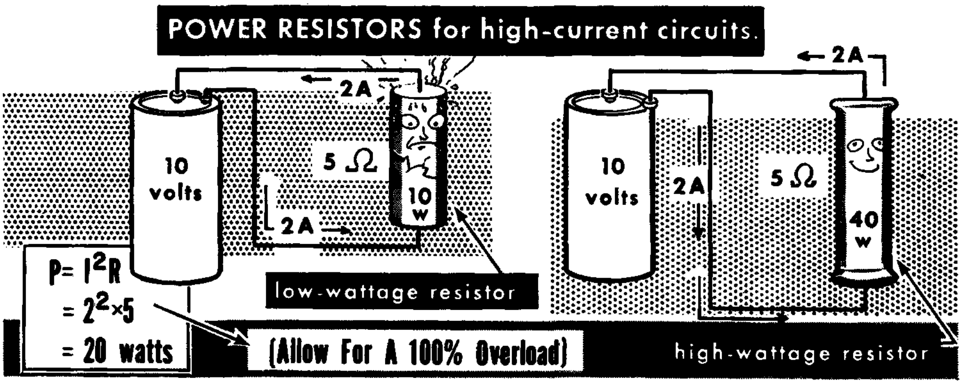
The power rating sets the limit on the maximum amount of current that can flow through the resistor without damaging the resistor element. This is so for all kinds of resistors -- fixed and variable, carbon, deposited metal or any other. In all cases, the power rating applies to the whole resistor and not to a part.

How can we calculate the maximum amount of current which can be permitted to flow in a resistor of a given wattage rating without damaging it? The equation is: $I = \sqrt{\frac{P}{R}}$

Assume a wirewound resistor rated at 20 watts and 5 ohms. What is the maximum amount of current that can be passed through the resistor? Substituting these values in the equation:

$$\text{maximum current} = \sqrt{\frac{20}{5}} = \sqrt{4} = 2 \text{ amperes}$$

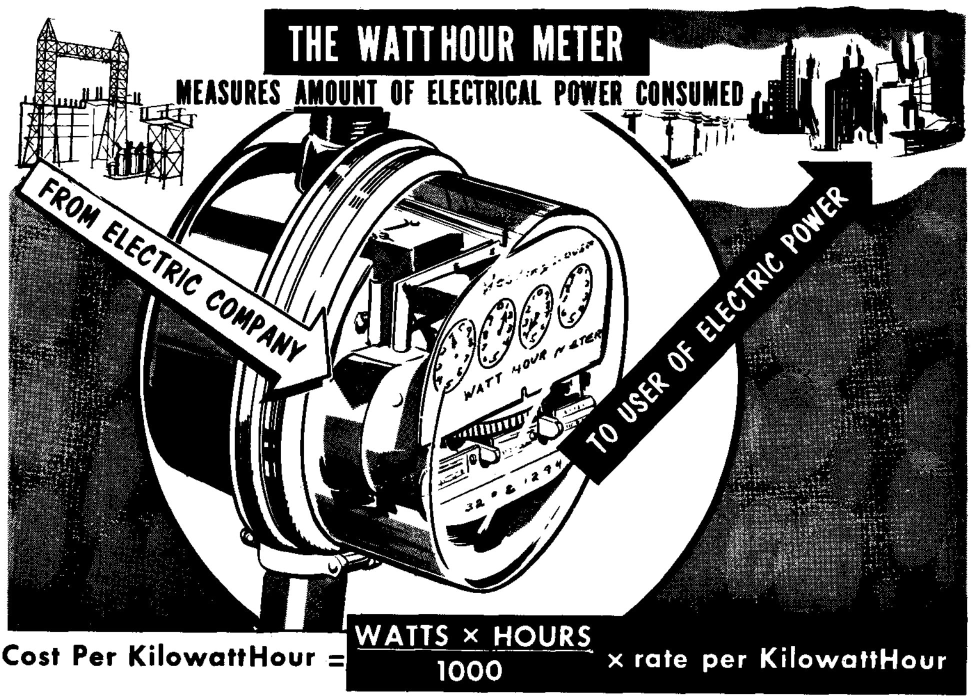
Try to solve the following: The resistor is rated at 0.5 watt and 5000 ohms. What is the maximum current? Your answer should be .01 ampere or 10 ma.



It is important to note the power formula $P = I^2R$. As the current through a resistor doubles, the power dissipated by the resistor increases four times. As the current triples, the power dissipation requirement increases nine times. From this, we can see the importance of current flow with regard to power -- it becomes the most important consideration.

The Kilowatthour

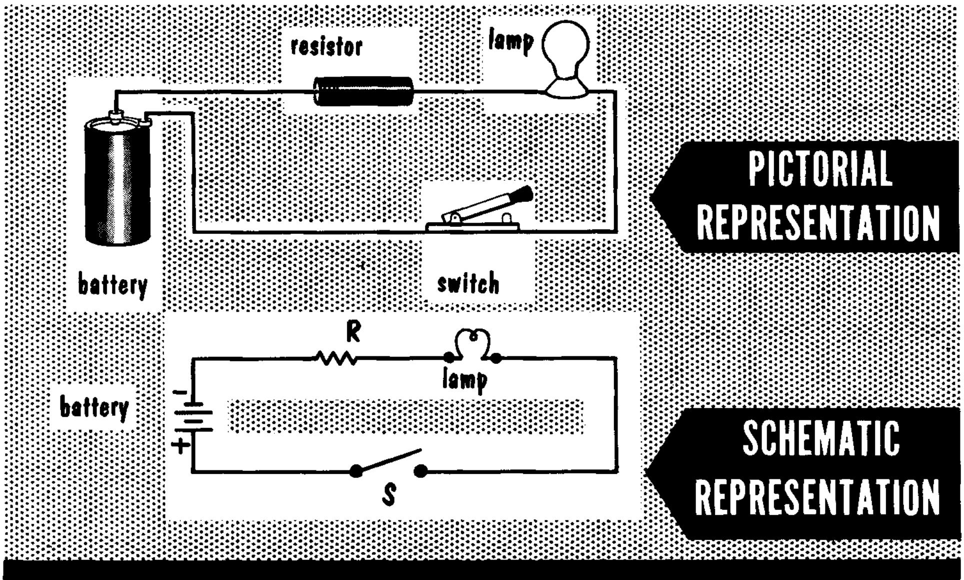
We have learned that power represents the rate at which electrical energy is being consumed, or used up. A 50-watt lamp uses electrical power at twice the rate of a 25-watt lamp. Electric utility companies are in the business of selling electricity. Through huge electrical generators turned mostly by water or steam power, the electric companies deliver voltage to homes, offices and industrial plants. No power is consumed, however, until a circuit is closed and current is drawn. Electric companies charge for the use of this power on the basis of how many watts are consumed and for how many hours. From this comes the term watthour. The watthour is obtained by multiplying the number of watts consumed by the number of hours it was consumed. For instance, a 100-watt lamp burning for 10 hours gives us a figure of 1000 watt-hours. Since the amount of watthours used by a home or industry during the course of a month (a billing period) is usually a very large figure, the kilowatthour (KWA) is used. A kilowatthour equals 1000 watthours.

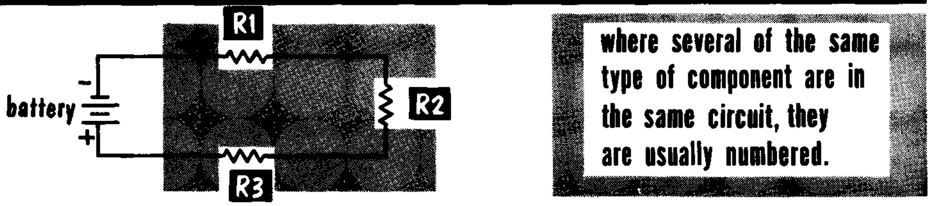


The electric company places a watthour meter at every electrical input. At specified periods, the amount of watthours consumed is read and subtracted from the previous total. A charge is made from this, based on the cost per kilowatthour. For example, an office uses 1000 kilowatthours during a month. At 5 cents per kilowatthour, the cost of electricity would be 5 × 1000, or 5000 cents ($50.00). A special meter is used to measure watts, which we will discuss later.

The Organization of D-C Circuits (Schematic Representation)

We have discussed the meaning of voltage, current, resistance, and power as individual electrical quantities and the units in which these quantities are expressed and compared. To use this information we must apply it to circuits.



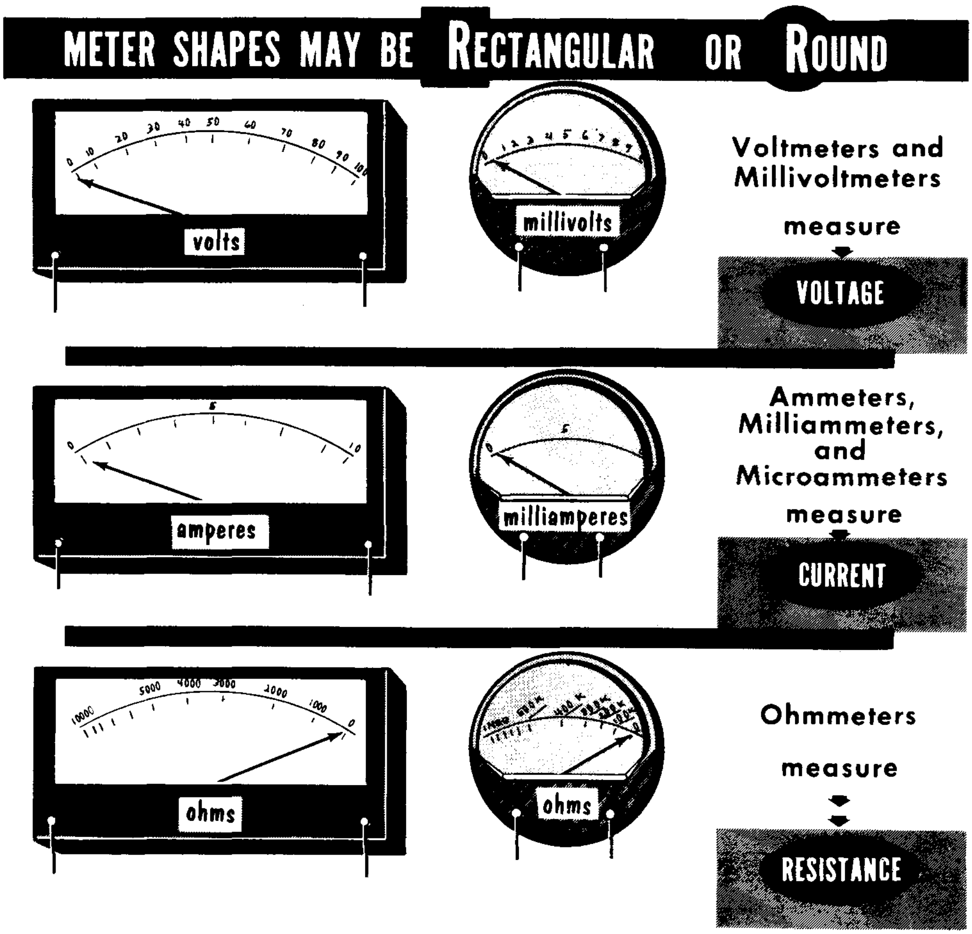


To understand the principles of electricity, we must be able to recognize and "read" a circuit. The circuits shown so far have been mostly in pictorial form. Another form, known as schematic uses symbols to show the components of a circuit. Schematic representation is the usual way of showing circuitry. In a sense it corresponds to the electrical language presented graphically. We shall start with very few symbols, increasing them as the variety of circuits and electrical components is increased.

Symbolization on the schematic is usually accompanied by two forms of identification: a letter with or without an accompanying numeral, (such as R3, indicating the third resistor in the circuit), and numbers expressing the electrical value of the component (for example, the ohmic value if it is a resistor and possibly the power rating and tolerance). Sometimes a word will appear next to the symbol for purposes of identification (such as LAMP).

Introduction to Measuring Devices

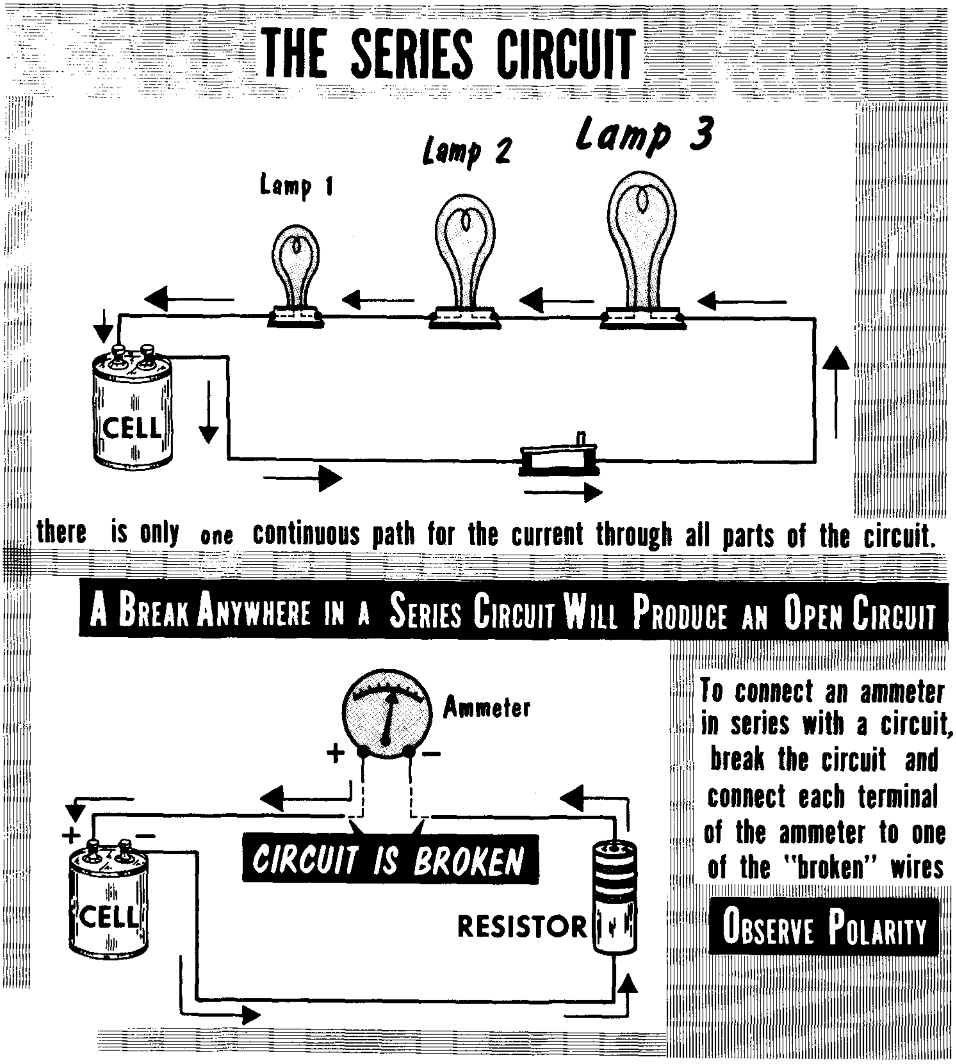
Voltage, current, resistance and electrical power are measurable quantities. Such measurement is done with electrical measuring instruments known as "meters". Although the principles underlying the operation of meters have not yet been discussed, it is nevertheless possible to become familiar with voltmeters that measure voltage, ammeters that measure current, ohmmeters that measure resistance, and wattmeters that measure electrical power.



Meters have calibrated scales on the "face" of the instrument. These scales are marked off (calibrated) in fractions or multiples of the units of the quantity being measured. When the meter is connected to the circuit under measurement, a pointer moves across the scale and comes to rest at the position which corresponds to the magnitude of the quantity being measured at the particular location in the system where the measurement is made. The complete details concerning the theory of electrical meters are given later on.

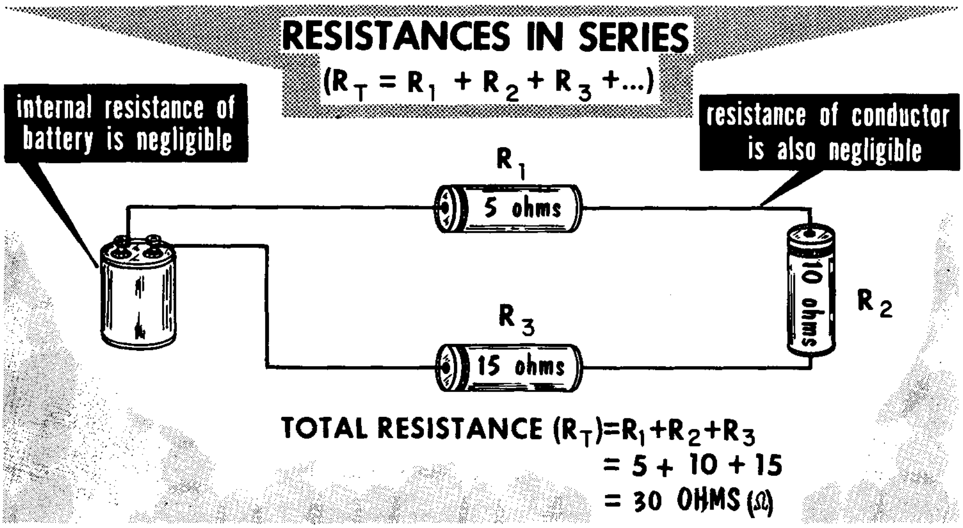
The Series-Connected or Series Circuit

The terms series-connected or series circuit applies to the arrangement of an electrical circuit in which there is only one continuous path for the flow of the current. A series circuit may contain many individual devices--a battery, numerous resistances, switches, et. al., --but they are so connected that the current must flow through each one. An important characteristic of the series circuit is that a break or opening in any part of this circuit will prevent any current from flowing, and make the circuit completely inoperative. Thus, to insert any additional device into the series circuit, the circuit must be broken, and the new device so connected that all the current in the circuit must flow through it.



Determining Resistances in Series (Calculation)

It is not enough to say that resistance exerts control on the amount of current that flows in a circuit. There are many ways in which the control of current by resistance is used.



THE TOTAL RESISTANCE OF A SERIES CIRCUIT IS EQUAL TO THE SUM OF THE INDIVIDUAL RESISTANCES

Consider a circuit consisting of a battery (a voltage source), a switch, and a flashlight bulb. The complete circuit is a series circuit because there is only one path for the current. For the moment, all we know about the circuit is that the operating resistance (the "hot" resistance) of the filament is 3 ohms. Everything in the circuit appears normal as judged by the normal brightness of the filament. Now, we raise the bar of the switch to open the circuit; also, we open the circuit at one terminal of the flashlight bulb socket and wire a suitable 3-ohm resistor in series with the circuit. We close the switch and immediately note that the filament glows less brightly. The reason is simply that less current is flowing in the circuit, thereby reducing the circuit current. In other words, resistances in series are additive in their effects. Each resistance contributes opposition to the current. This rule can be stated:

$$R \text{ total or } R_t = R1 + R2 + R3$$

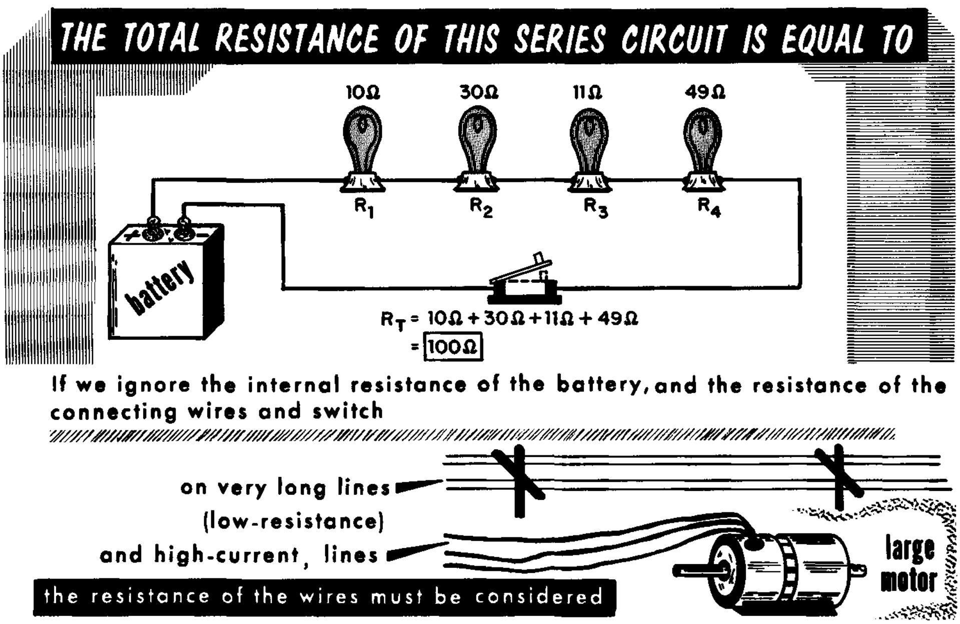
where R1, R2, and R3 are the resistances in series. Applying this formula to the above example, $R_t = 3 + 3 = 6$ ohms total resistance.

Solve this problem. Four resistances are connected in series: R1 = 11 ohms; R2 = 7.3 ohms; R3 = 6.8 ohms; and R4 = 116 ohms. What is the total resistance? Your answer should be 141.1 ohms.

Connecting-Wire Resistance

Now that you know the basic rule for calculating the total resistance of resistances connected in series, we can consider a more advanced case to illustrate an important point. Four different imaginary electric light bulbs are connected in series. Arbitrarily, we say that their filament resistances are: R1 = 10 ohms; R2 = 30 ohms; R3 = 11 ohms; and R4 = 49 ohms. They are connected to a battery through a switch.

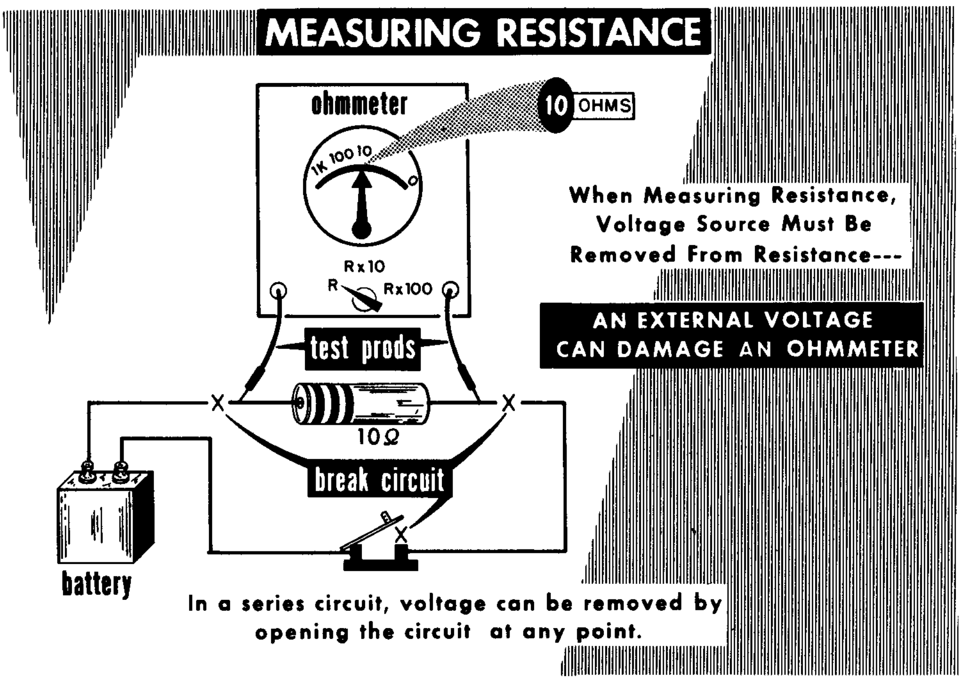
Now you are asked, "What is the total circuit resistance?" You get your answer by totaling the four resistance values, or 10 + 30 + 11 + 49 = 100 ohms. For all practical purposes, the answer of 100 ohms is correct. But it is not correct if you are to be technically accurate. We did not consider several other sources of resistance in the circuit: the internal resistance of the battery; the resistance of the metal comprising the switch; and the resistance of the wire used to interconnect all parts of the circuit. These resistances are present and they are in series with the circuit; therefore, they contribute to the total resistance of the circuit.



Fortunately, however, the internal resistance of a good battery is very low, usually much less than .1 ohm. Similarly, the switch and the interconnecting wire resistances are very low in the usual radio circuit. The total of these resistances may be .4 to .5 ohm, certainly an insignificant figure relative to the total resistance of 100 ohms of the devices that make up the load on the voltage source. When such a relative resistance situation prevails, it is permissible to forget about the battery, switch, and connecting-wire resistances as factors that contribute to the total circuit resistance, and to think only of the resistance made up by the load devices. We shall do this in all future discussions except where noted.

Resistances in Series (Measurement)

The total resistance of resistances in series, or the ohmic value of the individual resistances in a series circuit, can be determined by measurement using an ohmmeter. Because of the electrical characteristics of the ohmmeter, the resistance being measured cannot be part of an electrical circuit in which current is flowing while the measurement is being made. The resistor or device whose resistance is being measured must be disconnected from whatever source of voltage may be acting on the circuit.

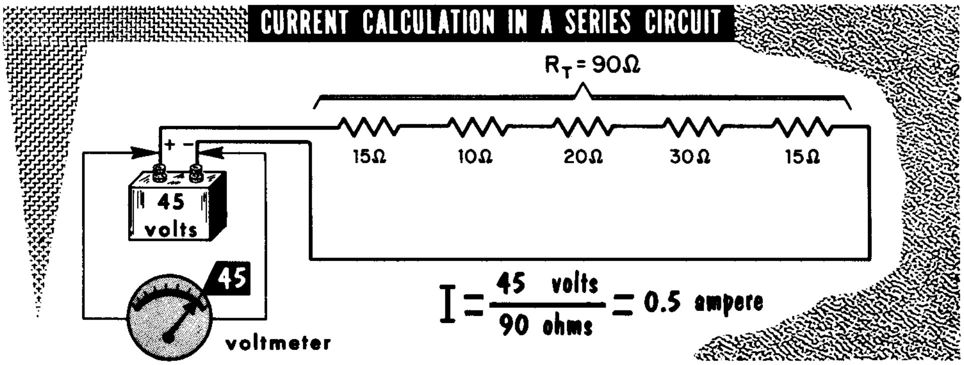


The measurement is made by connecting one test prod of the ohmmeter to one of the terminals of the device or resistor (alone or a series chain), and the other test prod to the other terminal of the resistance, and reading the indication on the scale. Although more will be said later, all ohmmeters afford different ranges of resistance measurement by manipulating a selector switch. The ideal condition for accurate measurement is when the indication of the resistance is at the middle to low end of the scale.

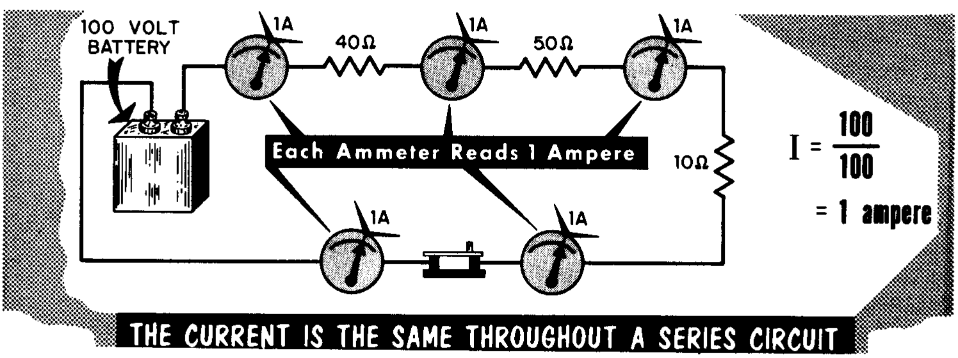
A very important fact to bear in mind when measuring the resistance of lamp filaments is that there is a major change in the resistance when the filament is cold and when it is hot. Hot means that the filament is drawing current. The resistance of a hot filament is higher. More important, the ohmmeter cannot be used to measure the resistance of a glowing lamp filament. Damage to the instrument is sure to result. Every device whose function is to convert electric energy to heat energy has a "cold" and a "hot" resistance.

Current In A Series Circuit (Calculation and Measurement)

There are two ways of determining the amount of current in a series circuit. One is by calculation, using Ohm's Law for current ($I = E/R$). Two quantities must be known -- the total resistance (R) of the series circuit and the applied voltage (E). The second method is by measurement, using an appropriate current meter. As a typical case, consider four resistors -- R1 = 10 ohms, R2 = 30 ohms, R3 = 11 ohms and R4 = 49 ohms -- joined in series and connected across a 12-volt source. The total circuit resistance is 100 ohms. How much current is flowing in the circuit?



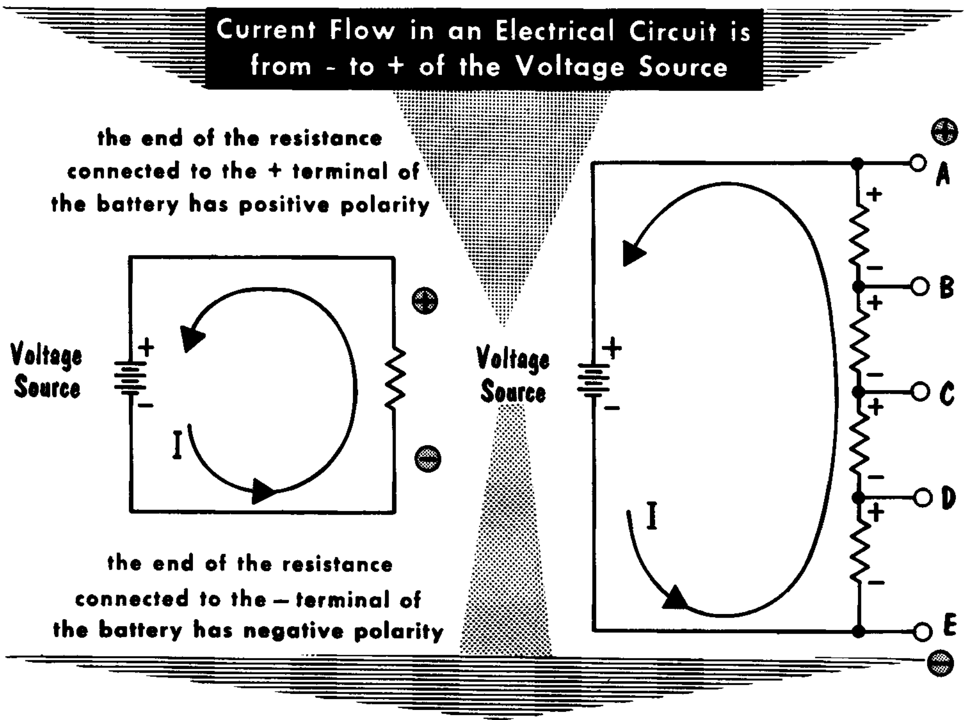
The current is measured by inserting the current-reading meter anywhere in the series circuit. The current being measured must flow through the meter. Because the series circuit offers only one path for the current, the current is the same throughout the circuit. This is an important rule. Having calculated the current to be 0.12 ampere, we know that a d-c ammeter (or a d-c milliammeter) with a maximum range of 0.5 ampere (500 milliamperes) is ideal for the measurement. A meter with a higher maximum range could be used, but it would make reading the indicator more difficult.



Try this exercise problem. Four filaments totaling 90-ohms resistance are connected in series. How much resistance must be added in series with the filament string to limit the current to 0.12 ampere when the battery voltage is 12 volts? ($R = \frac{E}{I} = \frac{12}{.12} = 100$ ohms, total R. Since we already have 90 ohms, We must add 10 ohms.) Is your answer 10 ohms?

Polarity In Series Circuits

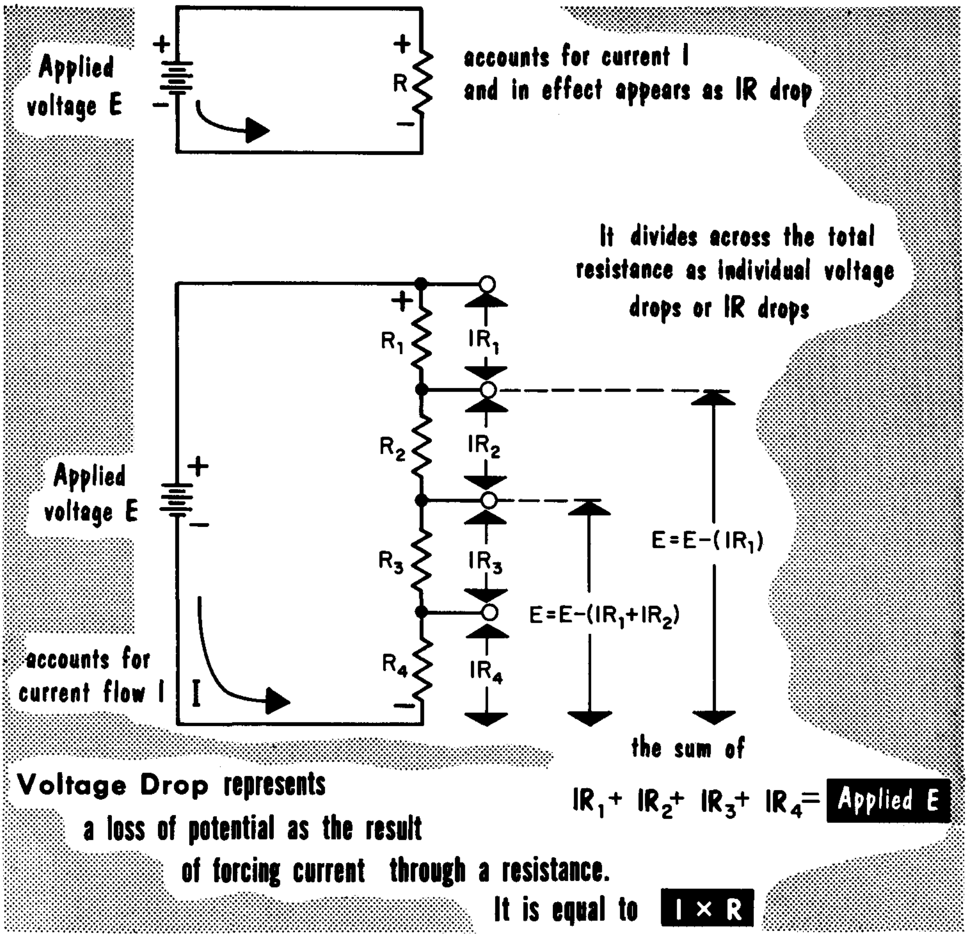
Voltage polarity in series circuits determines the proper usage of a d-c measuring device. A series circuit consisting of a single resistor connected across a voltage source presents a simple set of polarity conditions. That end of the resistor connected to the positive terminal of the voltage source is the positive end of the resistor; that end of the resistor connected to the negative terminal of the voltage source is the negative end of the resistor. If we use the direction of current flow as indicated, the same polarities prevail -- the end at which the current presumably enters the resistance load is the negative end; where it leaves the resistance load is the positive end.



A somewhat more complicated set of conditions prevails in a circuit containing a string of series-connected resistors. Although the end terminals of the series string have positive and negative polarity just as in the single-resistance circuit, points along the series string (other than the two extremes) have dual polarity. Point (A), for example, has a positive polarity relative to any point along the series chain. Point (B), on the other hand, is negative relative to point (A) but positive relative to points (C), (D) and (E). Relative to point (B), the current enters at (C), but leaves at (B). On the other hand, relative to point (A), the current enters at (B), thereby making this point negative while it leaves at (A), making this point positive. If we take point (E) as the most negative reference point, points (D), (C), (B) and (A) are positive with respect to point (E).

Voltage Drop In Series Circuits

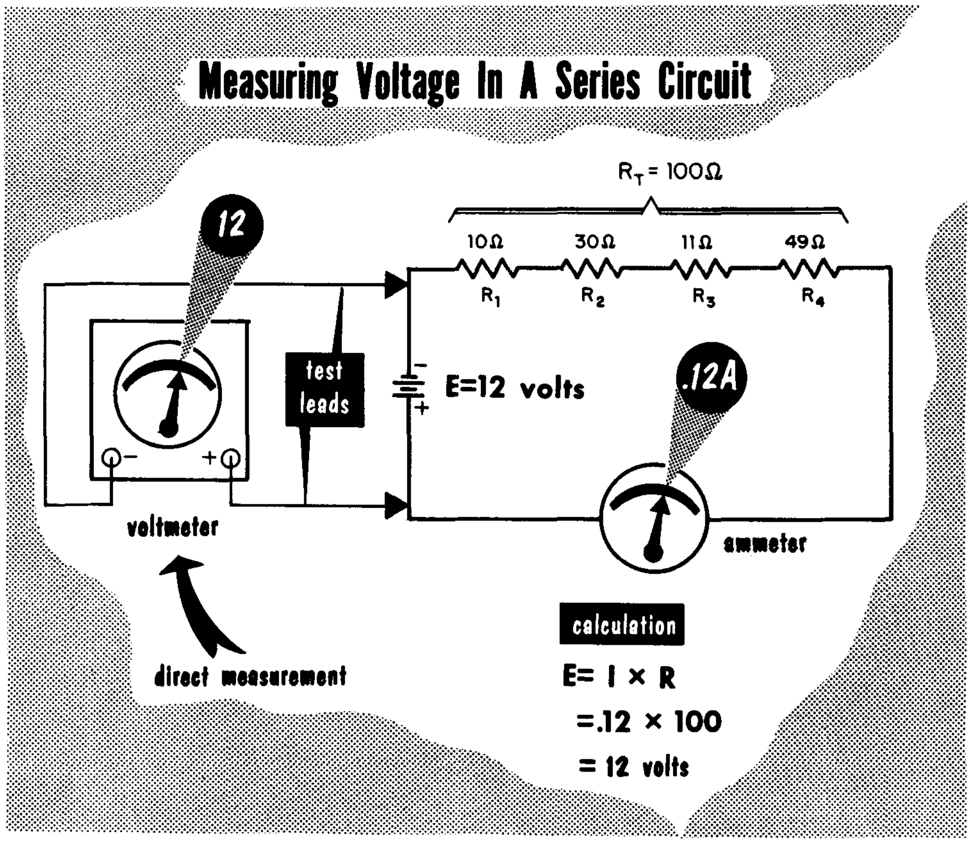
Voltage drop , or IR drop as it is also known, is interpreted in two ways. In one sense, it means the voltage that appears across any resistor R through which a current I is flowing. The same thought is conveyed when the product of IR is referred to as simply voltage without adding the word drop .



The second meaning of voltage drop (or IR drop) relates expressly to series circuits; meaning that it refers to the voltage which appears across each resistance R in a series chain when a current I is flowing through the chain. Since the sum of the individual resistances in a series circuit is the total resistance of the circuit (hence, determines the circuit current I for a given applied voltage), the sum of the individual voltage drops equals the applied voltage. Utility-wise, voltage drop is the means whereby a d-c voltage that is less than the applied voltage can be made available at a particular point in a series circuit. This function of the series circuit is the reason for its being called a voltage divider.

Determining Applied Voltage In A Series D-C Circuit

The applied voltage in a series d-c circuit can be calculated when circuit current and total resistance are known, or it can be measured directly with a d-c voltmeter. When I and total R are known, Ohm's Law for voltage, E = IR, is used. It is applicable to all d-c circuits.



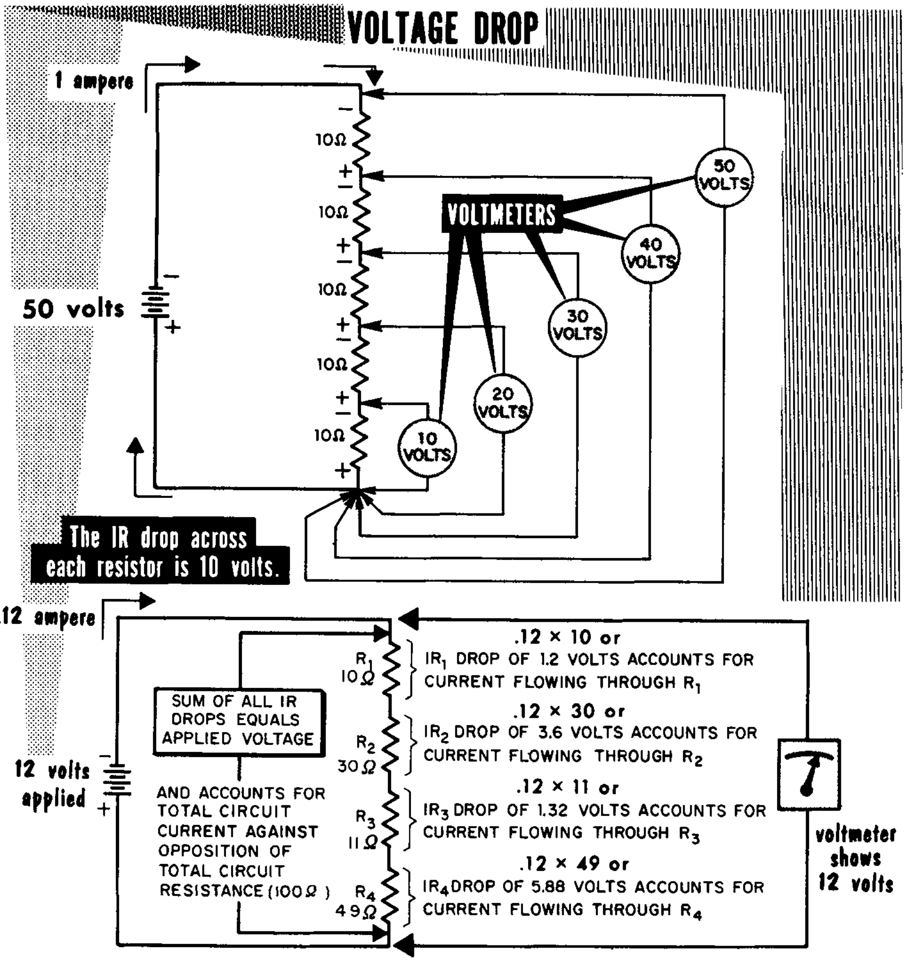
In the circuit used for the example, the current is 0.12 ampere (120 milliamperes) and the total circuit resistance made up of R1, R2, R3 and R4 equals 100 ohms. The applied voltage is:

$$E = I \times R, \text{ or } E = .12 \times 100 = 12 \text{ volts.}$$

The 12 volts applied causes 0.12 ampere to flow through the series circuit. We can reaffirm this by connecting a suitable d-c voltmeter across the circuit at the terminals of the voltage source. The positive test lead from the voltmeter is connected to the positive-polarity side of the circuit, and the negative test lead (sometimes marked common) is connected to the negative-polarity side of the circuit. A suitable voltmeter would be one with a full-scale range of from 30 to perhaps 50 volts, so that the 12 volts indication would be easily readable.

Calculating and Measuring Voltage Drop in Series Circuits

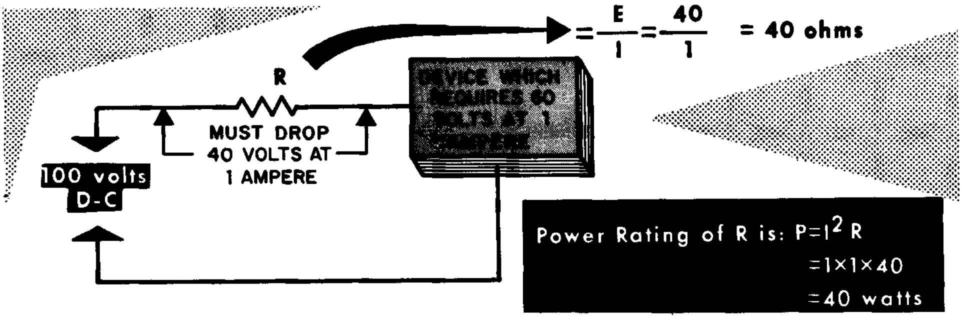
Voltage drop or IR drop can be calculated, or measured with a suitable voltmeter. To calculate voltage drop, the current and the resistance must be known. Ohm's Law for voltage ($E = I \times R$) is then used, I being the circuit current and R representing the ohmic value of each resistance.



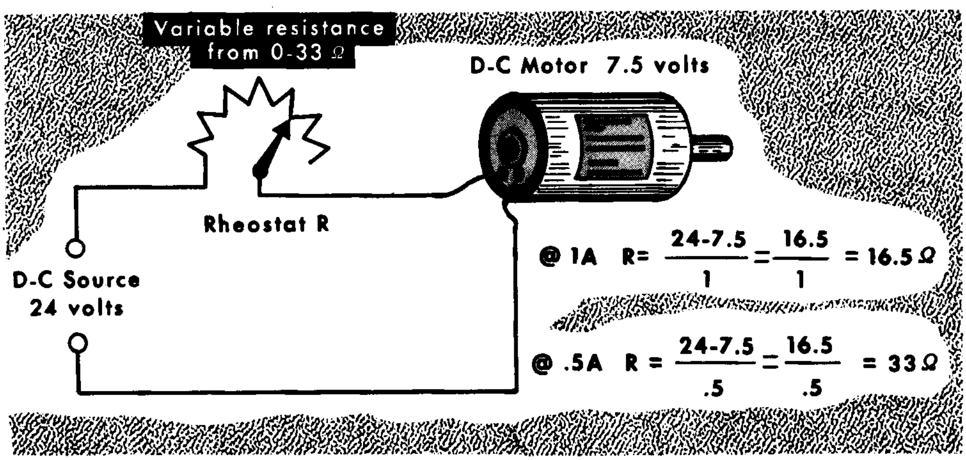
To measure voltage drop, a suitable voltmeter is connected across the resistance (or resistances) through which the current is flowing. Across means that the two leads from the voltmeter are connected to the limits of the circuit or that part of the circuit where the voltage is to be measured. In the examples shown, the voltage drop across the entire series train, as well as across its individual elements, is being measured. The voltage developed across two or more adjacent resistors also is measurable. Note that the polarity of the voltage differs at different points along the series-connected resistances, several points having dual polarity.

Applying Voltage Drop

How is voltage drop used? Imagine the following situation: a d-c device that requires 60 volts and 1 ampere for its operation must be used with a d-c voltage source which delivers 100 volts. How can this be done? The answer is to connect the device in series with a voltage-dropping resistor which would reduce the voltage amount to the difference between the available voltage (100) and the required voltage (60).

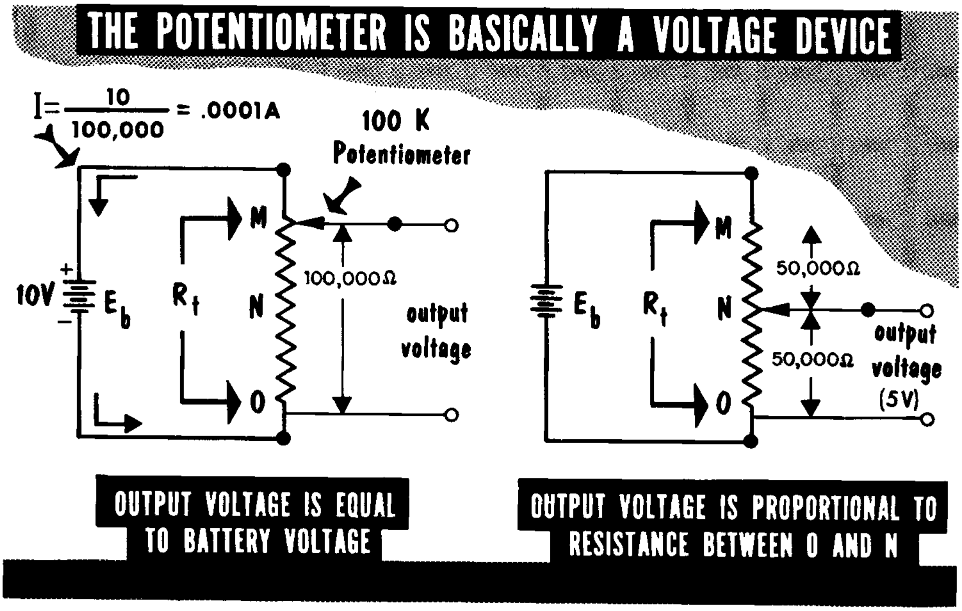


Consider another case. A d-c motor requires that the current through it be varied at irregular intervals to suit a variety of needs. The required range of current is a maximum of 1 ampere and a minimum of 0.5 ampere. The voltage rating at 1 ampere is 7.5 volts, whereas the voltage source is 24 volts d-c. How do we arrange the current needs? By connecting a wirewound rheostat in series with the device. At full current (1 ampere), a voltage drop equal to 24 - 7.5, = 16.5 volts is needed. The resistance required is $R = E/I = 16.5/1 = 16.5$ ohms. To reduce the current from 1 ampere to .5 ampere requires that the amount of resistance be doubled. Thus, the maximum resistance required is 33 ohms. Since a rheostat affords variable resistance from zero to a maximum, a rheostat rated at 33 ohms maximum and capable of passing 1 ampere will be adequate. The power rating is determined by the highest current flow or $P = I^2R = (1 \times 1) \times 33 = 33$ watts.



Applying Voltage Drop (Cont'd) (Potentiometer Voltage Divider)

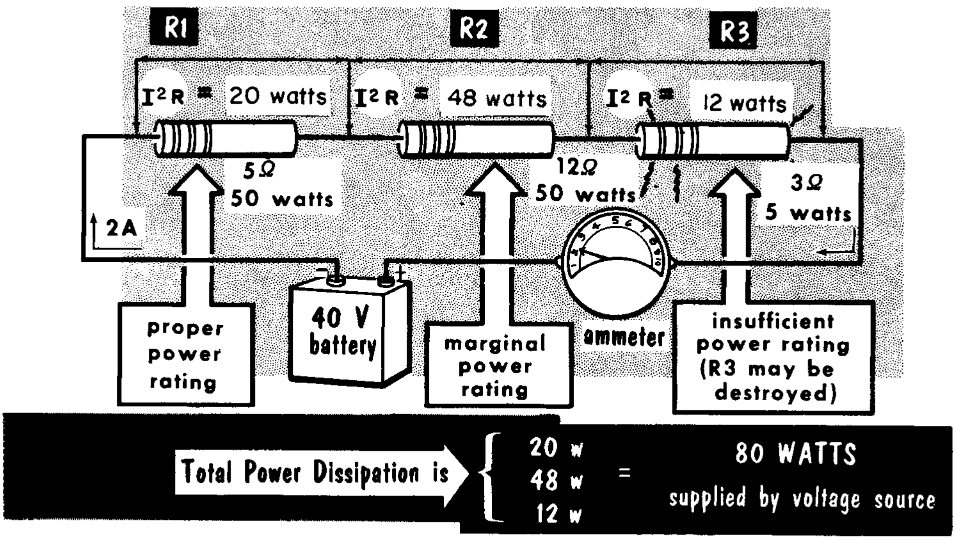
A frequently used arrangement found in radio systems makes use of a potentiometer as a voltage divider. It is a means of obtaining a variable voltage from a constant voltage source. Assume a voltage source E_b (although it need not always be a battery) and a potentiometer with a total resistance R_t of 100,000 ohms. The circuit current is $I = E/R_t = 10/100,000 = .0001$ ampere, or 100 microamperes. The entire 10 volts applied appears across the potentiometer element as a fixed voltage drop. By moving the slider along the resistance element, we can pick off any percentage of the total applied voltage and it becomes the output voltage. Assume that the slider is positioned at the top of the element (position M). The output voltage M-O is the full voltage (E_b) existing across the element.



Now assume that the slider is moved half-way down the resistance element to position N, so that 50,000 ohms are above the slider position and 50,000 ohms below the slider position. The current through the potentiometer is still .0001 ampere, so that the voltage drop (IR) across M-N is $.0001 \times 50,000 =$ 5 volts, and the voltage drop (IR) across N-O is $.0001 \times 50,000 = 5$ volts. If these two voltage drops are added, they total the applied voltage. The voltage available as output is the voltage drop which appears across the resistance element between the slider location and the 0 reference point. The voltage drop across the resistance element between location M and the slider is, you might say, wasted. By moving the slider anywhere between 0 and M, any desired value of output voltage between zero and the value of the applied voltage is made available as output. Usually, the total resistance of the potentiometer chosen is high, so that the current through it is low. Potentiometer voltage dividers are usually used when the current drawn through the slider is very low.

Power in the Series Circuit

As current flows through a series circuit, it flows through each resistance in that circuit. As we have seen, there is a voltage drop, or IR drop, as the result of current flowing through that resistance. In the process of forcing electrons through a resistance, a certain amount of power is "wasted", or consumed. This power is equal to I^2R, and is often referred to as the <u>I^2R loss</u>. Since power is measured in terms of <u>watts</u>, virtually all electrical equipment is rated in terms of how many watts it consumes or dissipates. Resistors of 1 watt or more usually have their power ratings printed right on the body of the resistor.

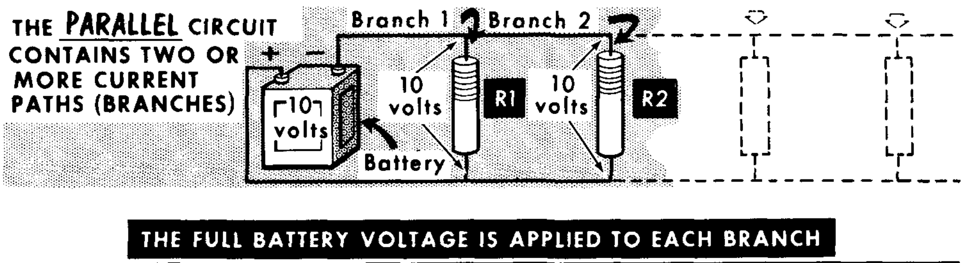


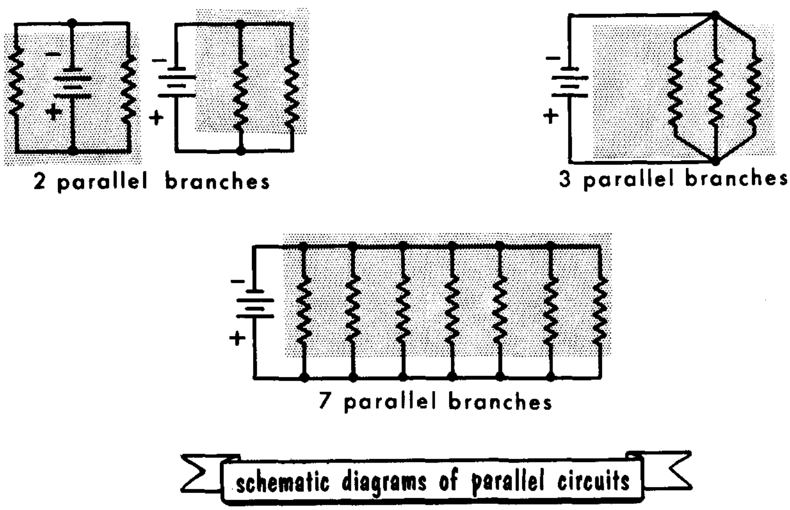
In the illustration, we have a simple series circuit using three resistors in series. With two amperes flowing through the circuit (I = 40/20), the power dissipated in R1 is 20 watts, R2 48 watts, and R3 12 watts. A resistor must be of sufficient size and proper design to dissipate the heat produced within itself, and to be safe, circuit designers usually use a resistor having a wattage rating of <u>twice</u> that which is necessary. R1, a 50-watt resistor, will easily handle the 20 watts dissipated in it, and may feel almost cool to the touch. R2 is a 50-watt resistor that must handle 48 watts dissipation. This resistor can run warm to very hot, depending upon the circulation of air around it. R3, a 5-watt resistor having 12 watts dissipated in it, will become extremely hot and will most likely be destroyed if this condition is maintained for any length of time.

The total power consumed by the three resistors is 80 watts, and this power must be supplied by the battery or other source of power. High wattage resistors, of 2 watts or more, should generally be located away from other components that may be damaged from the heat. In a series circuit, the total power supplied by the voltage source is equal to the sum of all the I^2R losses in the circuit.

Parallel Circuits

The parallel-connected or parallel circuit is a second type of circuit arrangement. The term shunt-connected or shunt circuit is also used to describe this circuit arrangement. In the parallel circuit, the total battery or source voltage is applied across each and every electrical device in the circuit. The connection is such that each electrical device connected to the source acts independently of each other. The circuit that each device is connected into is called a branch, and the full source voltage is applied to each of these.

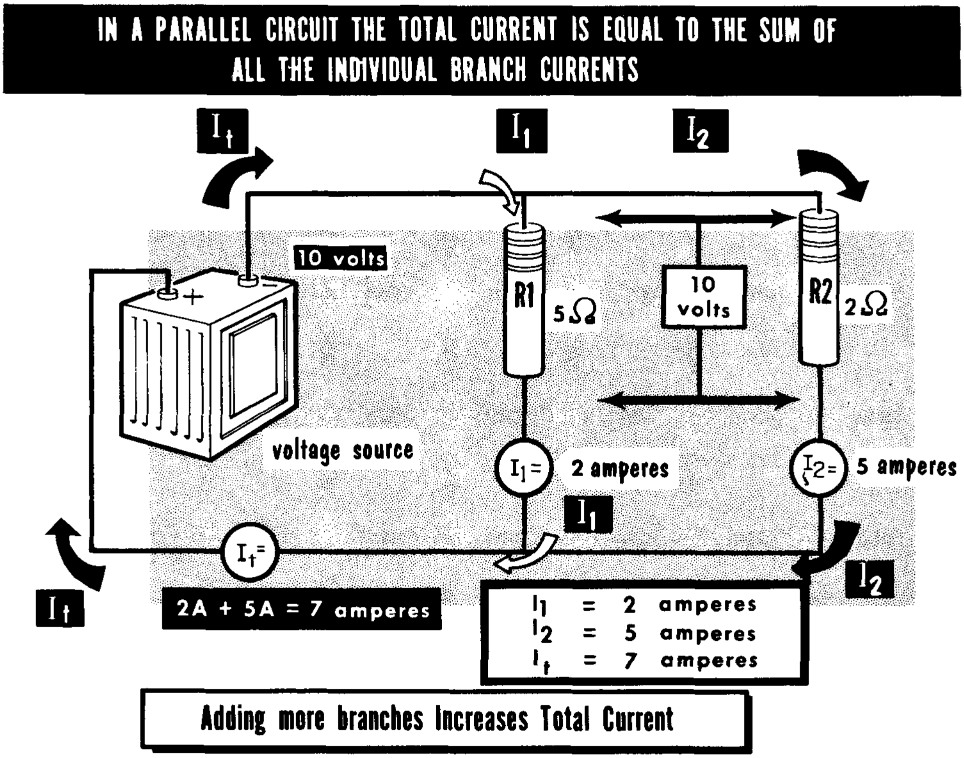




The parallel circuit is used in all homes, offices, and industry. When you turn on a lamp or other electrical device in the home, the full line voltage is applied to that device. In addition, when a lamp is turned off, it has no effect on other lamps or electrical devices. There is no limit to how many branches a parallel circuit might contain. Of course, if there are too many branches, or if the current drain from some branches becomes too great, the voltage source may be unable to supply the necessary power, and the result will be a lowering of the line voltage. Schematically, we show the branches of a parallel circuit tied together to a common point, or tied to two parallel lines drawn from the positive and negative terminals of the voltage-source symbol.

Current in a Parallel Circuit

We have seen that in the parallel circuit, two or more electrical devices are connected across a voltage source and act independently of each other. Since the total voltage is applied to each branch, the current flow in each branch would be determined by the resistance of each branch. If the voltage source is 10 volts and the resistance of one branch is 5 ohms, the current through that branch would be 2 amperes. If the second branch had a resistance of 2 ohms, the current through it would be 5 amperes. Although each branch operates completely independent of the other branches, there is one point they have in common -- their individual branch currents combine at the terminals of the voltage source to form a total current (I_t) for that circuit. In our example, the total current would be 7 amperes, or the sum of all the individual branch currents.

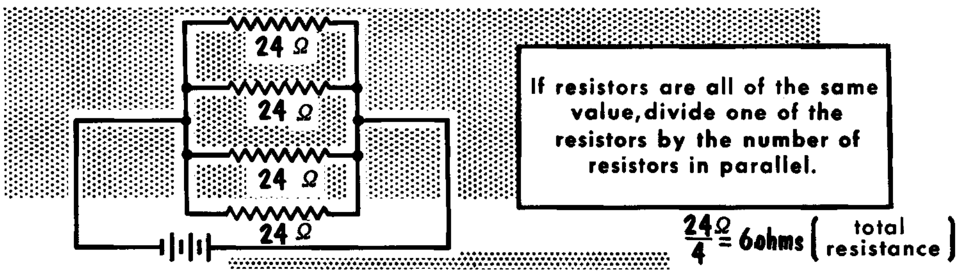


The amount of branches that can be used in a parallel circuit depends upon the amount of current drawn by the devices and upon the ability of the wires to conduct the current without overheating. Circuits containing high-current devices such as motors and heating units have very few branches; those containing low-current devices such as lamps, small radios, etc., may have many branches. The wiring in most homes permits a total circuit current of from 15 to 20 amperes, and contains a fuse which will "blow out" and break the circuit should the sum of the branch currents exceed the rating of the fuse.

Computing Resistances in Parallel

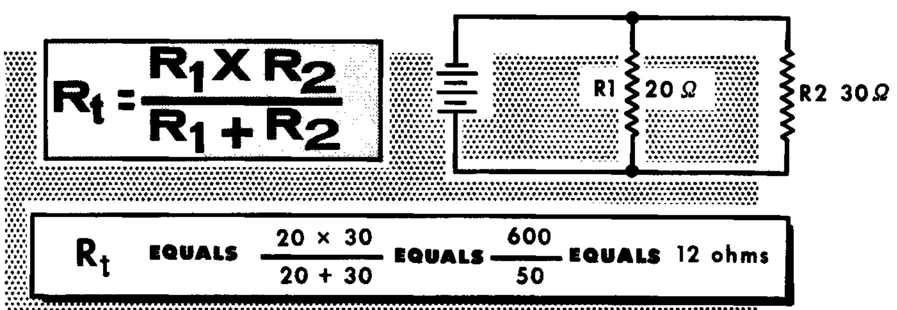
Current in parallel circuits can be measured or computed. Familiarity with the latter process is essential to be able to analyze a parallel circuit. To compute current, circuit resistance must be known. For the purposes of this explanation, we shall neglect the connecting-wire and battery resistances. Several rules related to parallel resistances help to make the calculation easier. They should be remembered.

FINDING THE TOTAL RESISTANCE OF PARALLEL RESISTORS



Rule 1. When equal values of resistance are connected in parallel, the resultant resistance is equal to the ohmic value of any one of the resistances divided by the number in parallel. The method of solution applies to any number of equal resistances in parallel.

To find the total resistance of two resistors in parallel, use this formula.



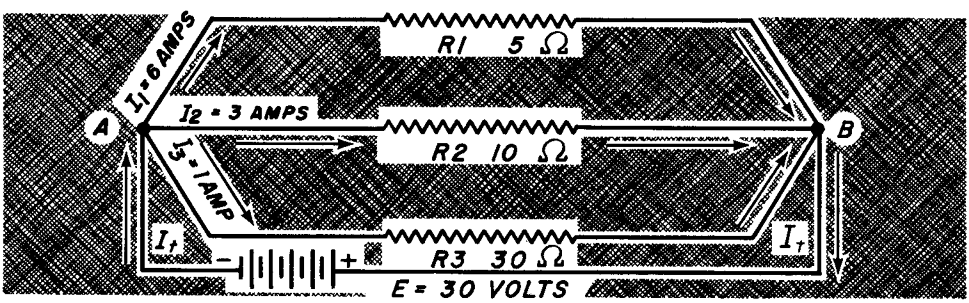
Rule 2. Another rule involves the use of multiplication, addition and division, and applies to any two unequal values of resistance in parallel. It states that the result of two unequal resistances in parallel is equal to the product of their ohmic values divided by the sum of their ohmic value. Rule 1 is, of course, much more convenient to use when the resistances happen to be of equal ohmic value.

Calculating Current in a Parallel Circuit

In a parallel circuit, the same voltage is applied across each of the parallel resistors. In the illustration, the voltage applied across R1, R2, and R3 is 30 volts, the same as the source voltage, E. Current flows from the negative terminal of the source to point A, where it divides and passes through R1, R2, and R3 to point B and back to the positive terminal of the source voltage. The amount of current flowing through each branch depends on the source voltage and on the resistance of that branch. The lower the resistance of the branch, the higher the current will be through that branch. The individual currents can be found by applying Ohm's law to the individual resistors. Thus,

$I1 = \frac{E}{R1} = \frac{30}{5} = 6$ amperes; $I2 = \frac{E}{R2} = \frac{30}{10} = 3$ amperes; $I3 = \frac{E}{R3} = \frac{30}{30} = 1$ ampere.

The total current, I_t, is equal to the sum of the currents through the individual branches, which is 10 amperes.



To find the total resistance R_t in the circuit, Ohm's law is first used to find each of the branch currents. Using the formula above, $\frac{E}{R_t} = \frac{E}{R1} + \frac{E}{R2} + \frac{E}{R3}$; and,

$$\frac{E}{R_t} = E\left(\frac{1}{R1} + \frac{1}{R2} + \frac{1}{R3}\right)$$

Both sides of this equation may be divided by E without changing the value of the equation, therefore: $\frac{1}{R_t} = \frac{1}{R1} + \frac{1}{R2} + \frac{1}{R3}$

The fraction $\frac{1}{R_t}$ is the reciprocal of the total resistance of the parallel circuit.

The reciprocal of a number is the number divided into 1, (e.g., the reciprocal of 5 is $\frac{1}{5}$.) The reciprocal of the total resistance of a parallel circuit is equal to the sum of the reciprocals of all the branches.

Using the above to find the total resistance of the parallel circuit,
$\frac{1}{R_t} = \frac{1}{5} + \frac{1}{10} + \frac{1}{30}$; and with the least common denominator of 30, you have,

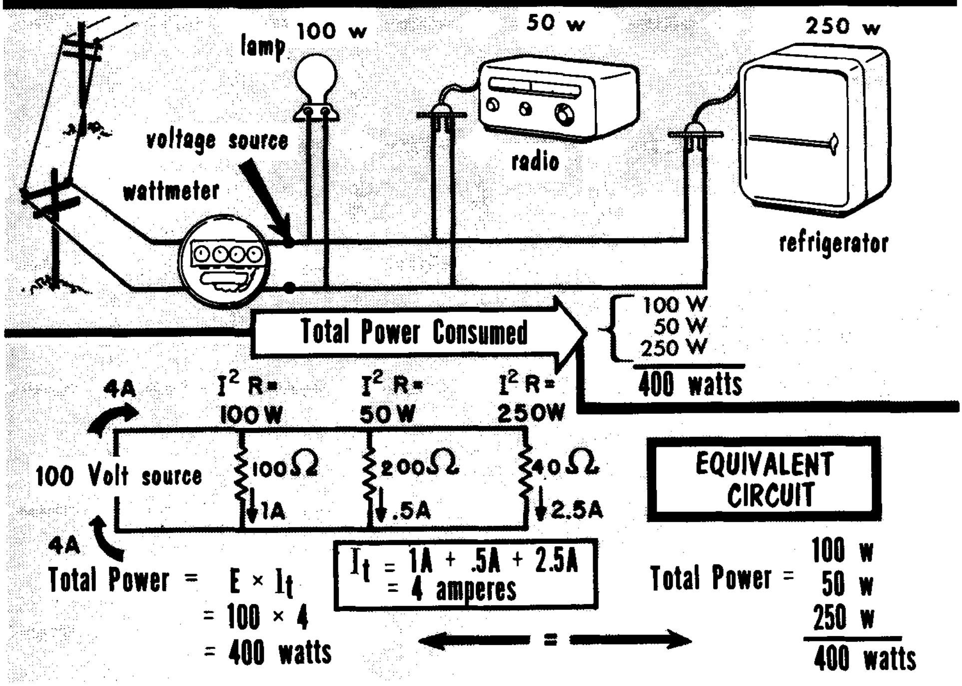
$$\frac{1}{R_t} = \frac{6}{30} + \frac{3}{30} + \frac{1}{30} = \frac{10}{30}$$

Taking the reciprocals of both sides (dividing them into 1), $R_t = \frac{30}{10} = 3$ ohms.

Power in Parallel D-C Circuits

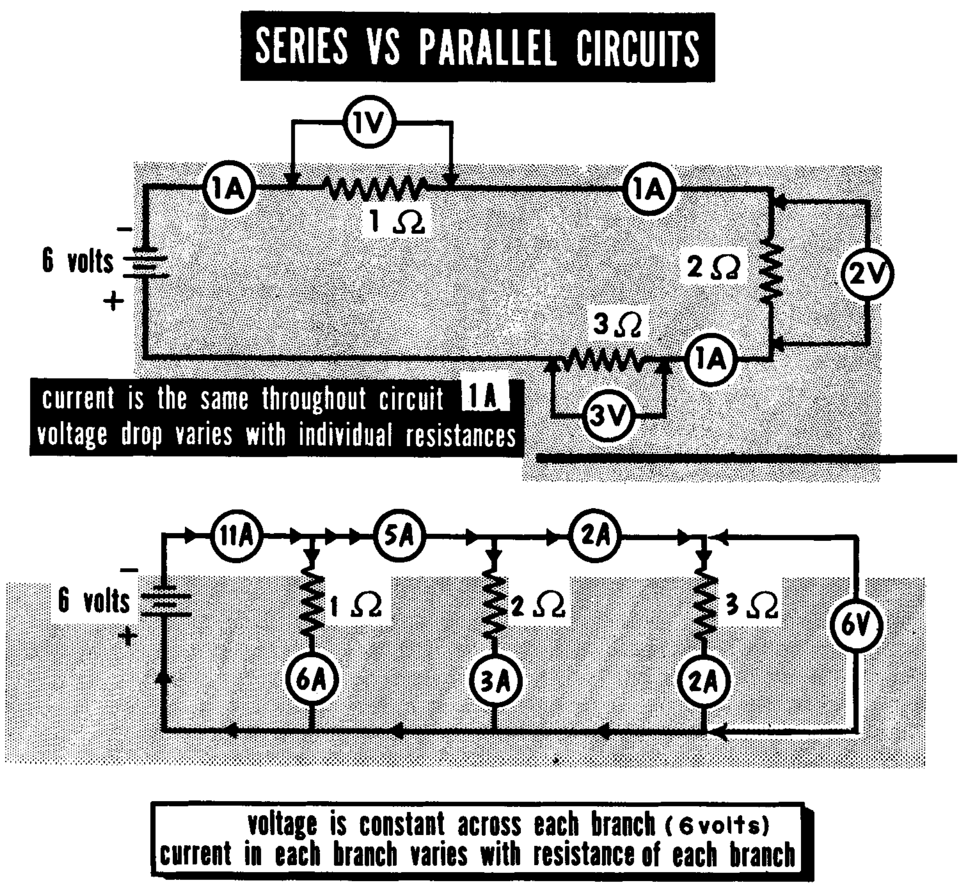
As we studied in series circuits, a power loss takes place in a resistance as the result of current flowing through it. This power loss is equal to I^2R, and refers to each and every resistor in the parallel circuit. Each branch of a parallel circuit contains a resistance, and each of the resistances has the same voltage (the source voltage) applied across it. Thus, the current through each branch would be determined by the formula $I = E/R$. Knowing the current in each branch, we can now compute the power consumed by each branch. We have also seen that each branch circuit operates independently of the others. Thus, the total power consumed by a parallel circuit is equal to the power consumed by all the individual branches.

TOTAL POWER CONSUMED BY A PARALLEL CIRCUIT IS EQUAL TO THE SUM OF THE POWER CONSUMED BY THE INDIVIDUAL BRANCHES

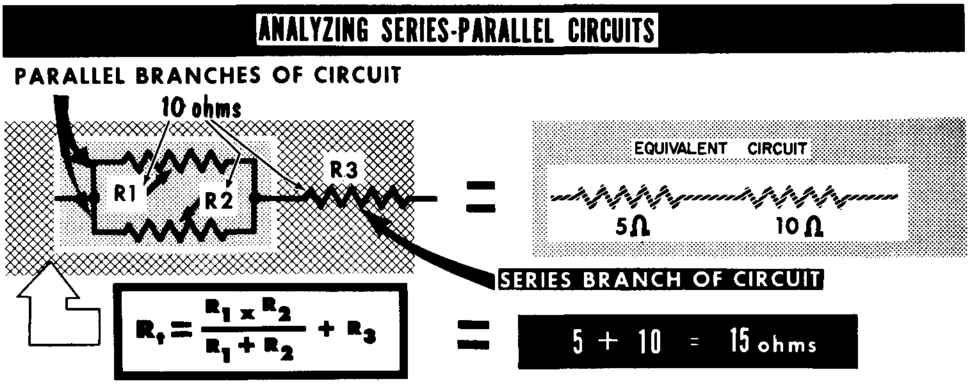


In our homes, each time we turn on an electrical device or appliance, we close a branch of a parallel circuit and draw additional current from our power supply (which are the power lines drawn into our homes by the electric utility company.) The utility company places a wattmeter in the incoming circuit to measure the total power consumed by the various devices in use. Every time a branch is closed, more current is drawn and more power used. As branches are opened, less power is consumed. Since the total current in a parallel circuit increases as additional branches are added, we see that the total resistance of a parallel circuit decreases as resistances are placed in parallel.

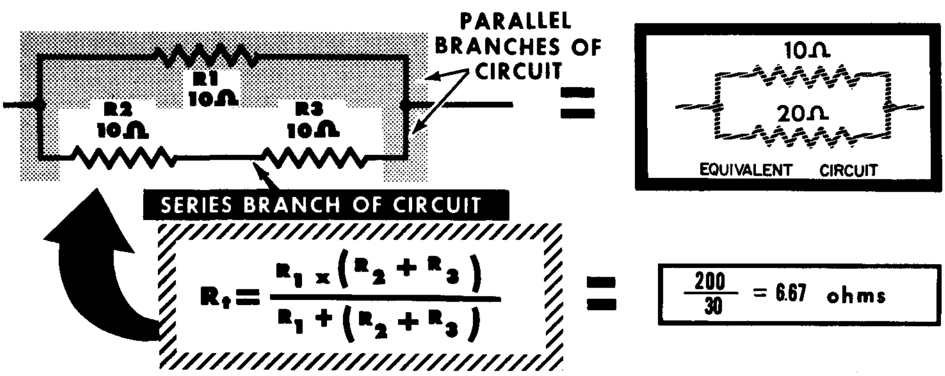
SERIES CIRCUIT	PARALLEL CIRCUIT
1. The sum of the individual voltage drops equals the applied voltage.	The applied voltage is the same across each branch.
2. The total resistance is equal to the sum of the individual resistances that make up the circuit. $R_t = R1 + R2 + R3 + \text{etc.}\ldots$	The total resistance equals the reciprocal of the sum of the reciprocals of the resistances. Resultant resistance is less than the smallest resistance of the parallel combination.
3. Current is the same in all parts of the circuit.	The current divides in each branch according to the resistance of each branch.
4. Total power is equal to the sum of the power dissipated by the individual resistances.	(Same as series circuit) Total power is equal to the sum of the power dissipated by the individual resistances.



Series-Parallel Circuits



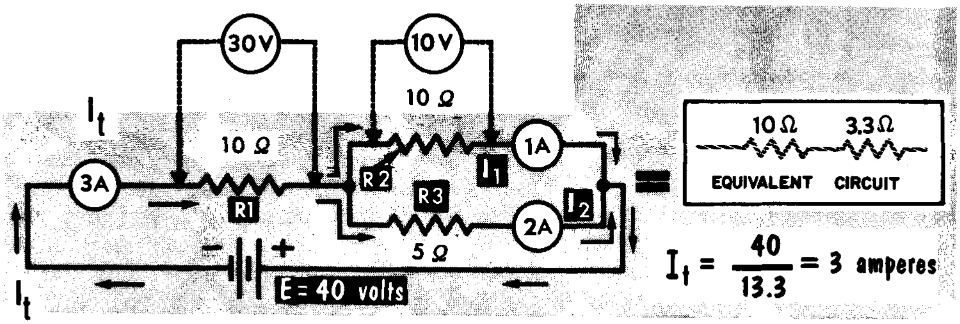
The third type of circuit arrangement is the series-parallel circuit. In this circuit, there is at least one resistance connected in series and two connected in parallel. The two basic arrangements of the series-parallel circuit are shown here. In one, resistors R1 and R2 are connected in parallel and this parallel connection, in turn, is connected in series with resistance R3. Thus, R1 and R2 form the parallel component, and R3 the series component of a series-parallel circuit. We can find the total resistance of any series-parallel circuit by merely reducing it into a simple series circuit. For example, the parallel portion of R1 and R2 can be reduced to an equivalent 5-ohm resistor (two 10-ohm resistors in parallel). We then have an equivalent circuit of a 5-ohm resistor in series with the 10-ohm resistor (R3), giving a total resistance of 15 ohms for the series-parallel combination.



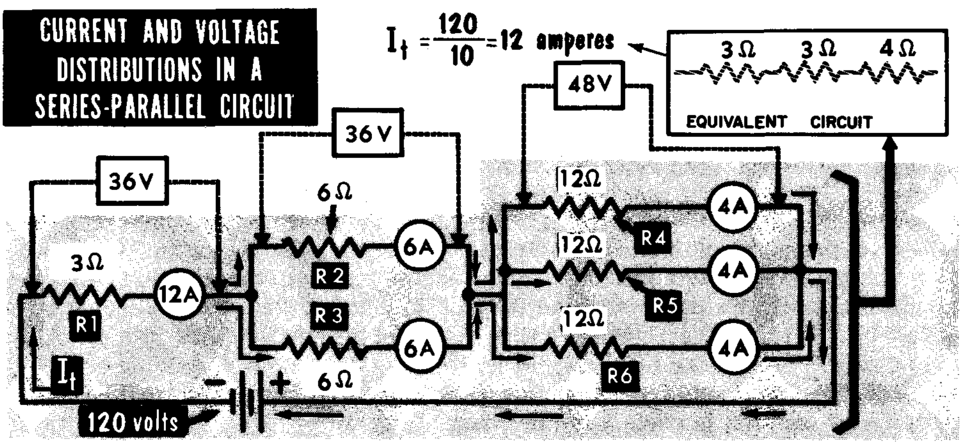
A second basic series-parallel arrangement is shown where basically we have two branches of a parallel circuit. However, in one of the branches we have two resistances in series. To find the total resistance of this series-parallel circuit, we first combine R2 and R3 into an equivalent 20-ohm resistance. The total resistance is then 20 ohms in parallel with 10 ohms, or 6. 67 ohms. From these circuits, we see the basic scheme in solving series-parallel circuits: combine all series resistances and add them to the resistance of all parallel branches in the circuit.

Current and Voltage in Series-Parallel Circuits

In the series-parallel circuit shown, we see the distribution of current and voltage in such a circuit. Some things become quite apparent. For instance, since R2 and R3 are in parallel, the voltage across them would be the same. In addition, since R2 is twice the resistance of R3, the current flow through it would be only half as great as the current through R3. The current flow through R1 would be equal to the combined current through R2 and R3. Whenever the total current flows into a parallel branch, it splits up, with the greater amount flowing through the smaller resistor. The voltage drop, of course, across each branch of a parallel circuit would be equal.

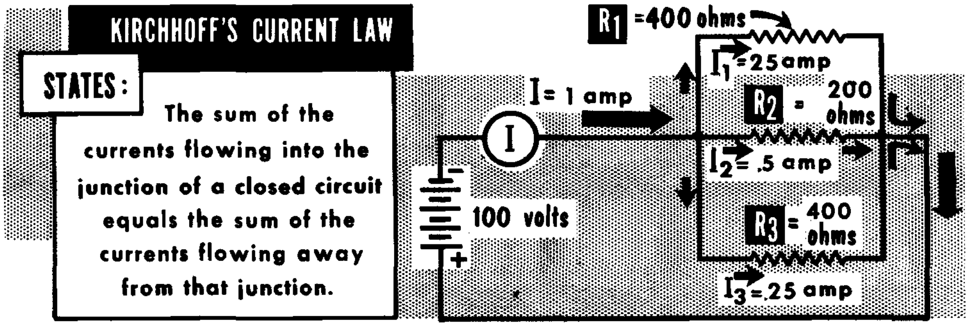


In the second circuit shown, we have a series-parallel circuit consisting of three sections. First, there is a series resistance R1. In series with this is a parallel branch R2 and R3 and, in series with this, is a three-branch parallel circuit consisting of R4, R5, and R6. The total current of 12 amperes flows through R1. It then breaks into two 6-ampere sections through R2-R3. Finally, it breaks into three 4-ampere sections through R4-R5-R6. Note that the voltage across each parallel section is equal to the IR or voltage drop in that section, and that the sum of the currents in each parallel section is equal to the total current. Keeping these general rules in mind, any kind of series-parallel circuit can be solved.

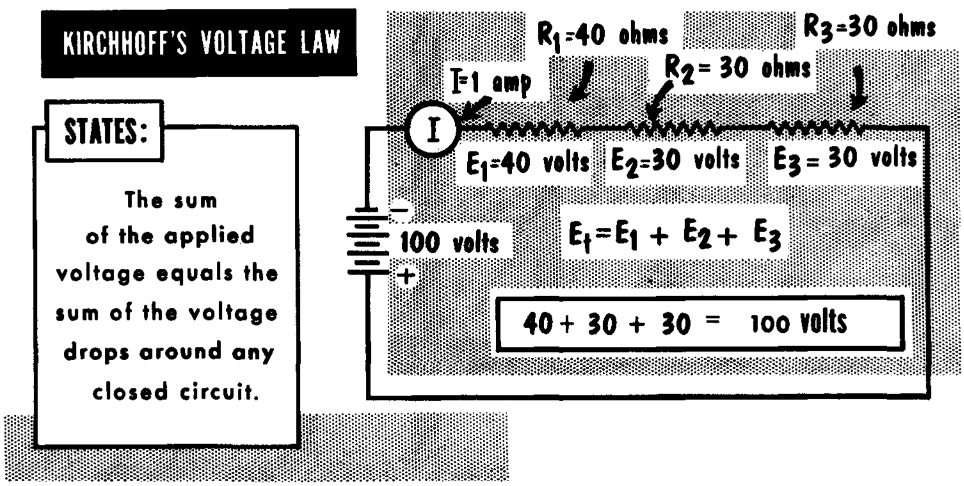


Kirchhoff's Laws

An interesting set of laws often used in circuit analysis are known as Kirchhoff's laws, named after Gustav Robert Kirchhoff (1824-1887). The principles they state have already been made known to you, although not as laws associated with a particular name.



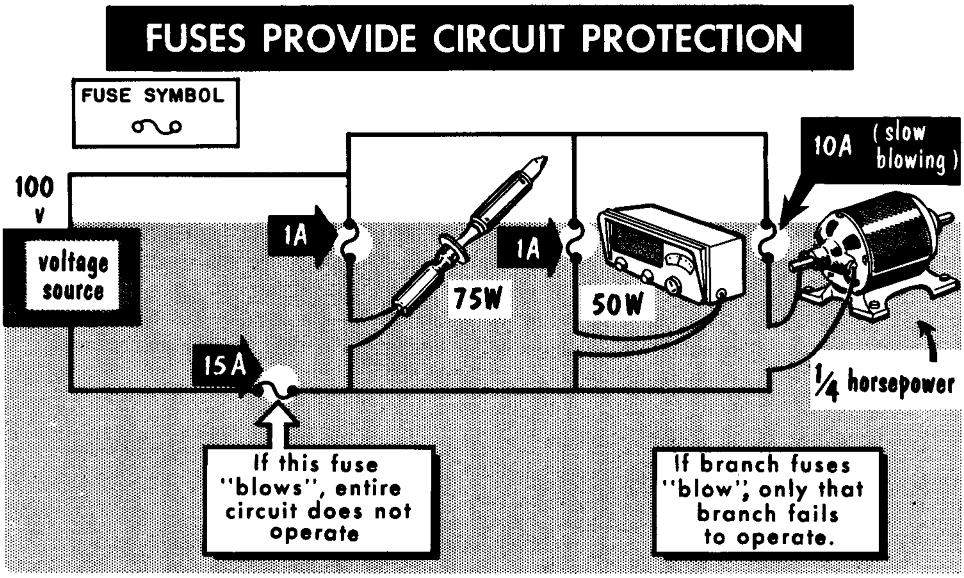
Kirchhoff's first law deals with the distribution of currents in a circuit. It states that the sum of all the currents flowing into a point or junction in a circuit is equal to the sum of all the currents flowing away from that point or junction. Thus, if 1 ampere flows into a junction, 1 ampere must flow away from that junction, whether in a single path or in many paths. In other words, current cannot accumulate anywhere in a circuit.



Kirchhoff's second law deals with the distribution of voltage in a closed circuit. It states that the sum of all the individual voltage drops in a closed circuit is equal to the applied voltage. Thus, if the sum of all the voltage drops in a circuit should total 100 volts, this should be equal to the applied voltage. If there is any difference in the two amounts, it indicates that an error has been made in the calculations.

Fuses

All electrical devices--soldering irons, motors, radios, etc., are designed to draw a certain amount of current when plugged into their rated voltage source. Occasionally, there is an electrical defect, the resistance of the electrical device falls off, and the amount of current drawn by the device increases greatly. This would cause the wires, often located inside walls, to overheat and possibly start a fire. To prevent such an occurrence, fuses are used. A fuse is connected in series with the electrical device it is to protect so that the total current flowing through the device will also flow through the fuse.



We can say then that a fuse is basically an overcurrent device. It consists usually of a short length of wire or metal ribbon within a suitable enclosed container. The ribbon, or link, is usually made of an alloy that has a low melting point and of a size which will carry a given current indefinitely. A current larger than that at which the fuse is rated will cause the fuse metal to heat and melt, opening the circuit being protected. Fuse metal generally consists of an alloy of tin and bismuth, but copper, aluminum, German silver, and iron alloys are also used.

Fuses may be further classified as instantaneous or time-delay types. The instantaneous fuse carries its rated current indefinitely, but quickly melts when its rated current capacity is exceeded by about 25%. Time-delay fuses are designed to have a time delay for overloads. This feature is necessary to keep short-time surges (such as high starting currents for motors) from melting ("blowing") the fuse. This time delay permits momentary high current without damaging the fuse, while continuous excessive current produces a melted fuse. Some fuses are "sealed" and must be discarded when open; others have provisions for replacing the melted ribbon.

Ohm's law defines the relationship between voltage, current, and resistance.

When resistance is held constant, current varies directly with voltage.

When voltage is held constant, current varies inversely with resistance.

Ohm's law for voltage is $E = IR$; for current, is $I = \frac{E}{R}$; for resistance, is $R = \frac{E}{I}$.

The unit used to express electrical power is the watt.

A watthour meter measures the amount of electrical power consumed.

Voltmeters measure voltage, ammeters measure current, ohmmeters measure resistance, and wattmeters measure power.

There are three basic types of electrical circuits: series, parallel, and series-parallel.

In a series circuit, there is only one continuous path for current to flow.

The current in a series circuit is the same throughout the circuit.

The basic rule for calculating the total resistance of resistances connected in series is $R_t = R1 + R2 + R3 +$ etc.

The voltage in a parallel circuit is the same across each branch of the circuit; the current divides according to the resistance of each branch.

The sum of the voltage drops in a series circuit is equal to the applied voltage.

When measuring voltage drop, a voltmeter is connected across the resistance through which current is flowing.

A potentiometer is used as a means of obtaining a variable voltage from a constant voltage source.

The amount of power wasted in the process of forcing electrons through a resistance is called the I^2R loss.

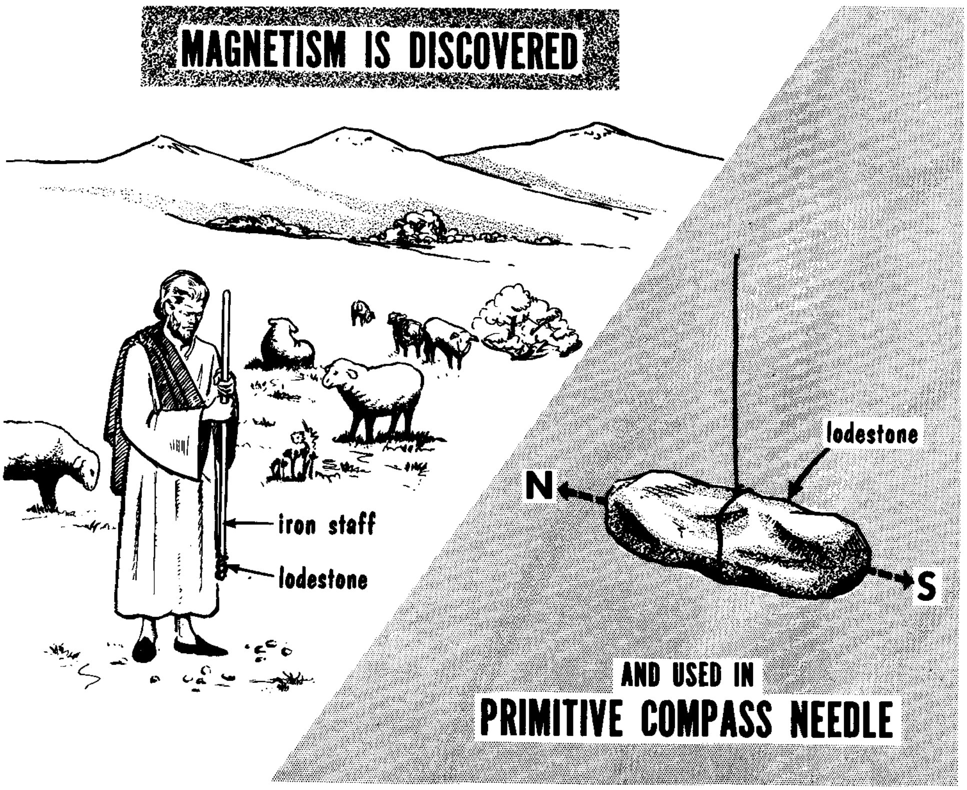
Kirchhoff's law for current states that the sum of all the currents flowing into a point is equal to the sum of all the currents flowing away from that point.

REVIEW QUESTIONS

1. What are Ohm's law's formulas for finding voltage, current, and resistance?
2. What formula is used to find the total resistance of two resistors connected in parallel?
3. What happens to the current in a circuit when the voltage is doubled and the resistance is held constant?
4. How does the current change when the voltage is held constant and the resistance is doubled?
5. What is the difference between current flow in a series circuit and in a parallel circuit?
6. How is a voltmeter connected when measuring voltage drop?
7. What voltage is required to light a 150-watt lamp drawing 3 amperes of current?
8. What is the wattage rating of a television receiver connected to a source of 120 volts and drawing 5 amperes?
9. How much power does an electric toaster consume if it operates from a source of 110 volts and draws 7.5 amperes?
10. What is the total resistance of a circuit having one 5-ohm and two 10-ohm resistors connected in parallel?

Magnetism and Electromagnetism

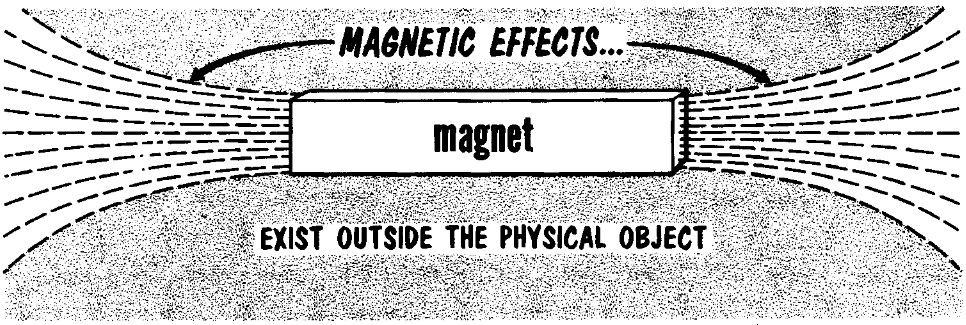
More than 20 centuries ago, an ore called magnetite was discovered. The ancients called it "lodestone", or leading stone. It displayed the peculiar ability to attract tiny bits of iron to itself. Although the action was not understood, it was attributed to an invisible effect called magnetism, named after Magnesia, the area in ancient Greece where this type of "rock" was found. The lodestone, in turn, was given the related name of magnet.

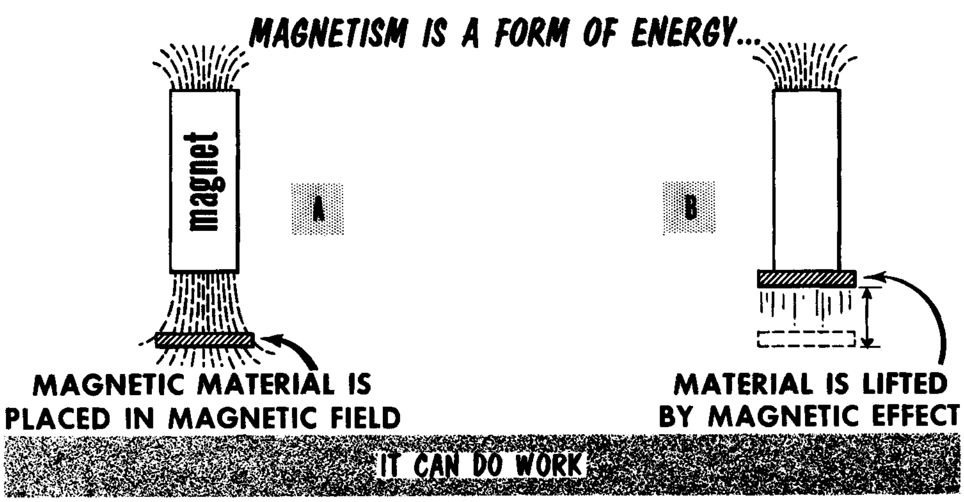


For a long time, the magnet was regarded merely as a curiosity. Then, it was discovered that if an elongated piece of lodestone was suspended horizontally so that it could turn freely, one end always pointed towards the North and the other end always pointed towards the South. This was the birth of the magnetic directional compass which has become invaluable to travelers on land, sea, and in the air. Much later it was found that the earth too behaved as a huge magnet; also, that the interaction between the magnetic effects of the magnet and that of the earth accounted for the "directional pointing" of the lodestone. It was concluded that what were called "magnetic effects" were the result of magnetic energy--magnetism was a form of energy that could do work. In time, it was realized that magnetism and electricity were related. It was found that effects associated with magnets could produce electric current in wires and that electric current gives rise to magnetic effects.

Magnetism Is A Form of Energy

If we sprinkle iron filings onto a sheet of paper and then hold a magnet slightly above the paper, filings will "leap" upward to the magnet. Iron filings are light and tiny, but as light as they may be, each filing is a physical entity and therefore, has some weight. The ability of whatever phenomenon is associated with the magnet to make the filings leap upward from the paper provides observable evidence of two conditions: first, magnetism exists outside of the physical body of the magnet; second, magnetism is a form of energy, as shown by the work done on the iron filings when they are raised against the force of gravity.

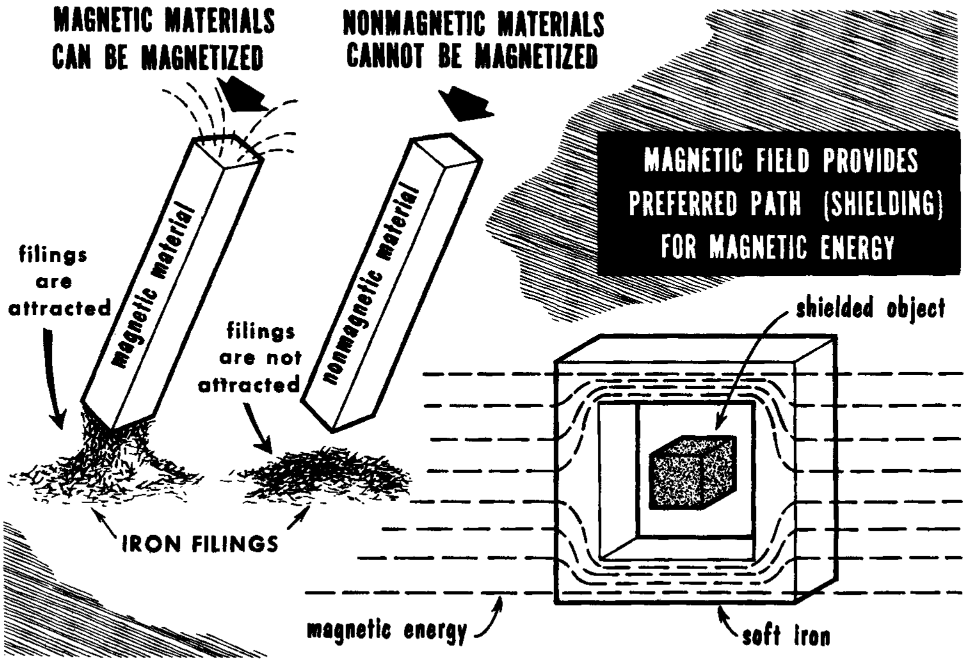




Magnetic energy normally produces no response from the human senses. The area around the most powerful magnet is filled with magnetic energy, yet we cannot see, hear, taste, touch or smell it. We recognize its presence by its effects--magnetic effects. Magnetic energy can do much work. It can make another object, made of magnetizable material, into a magnet. Magnetic energy can raise very heavy objects made of magnetizable materials; it can exert mechanical force on other magnets; it contributes to the generation of emf by mechanical means; and it has many other applications.

Magnetic and Nonmagnetic Materials

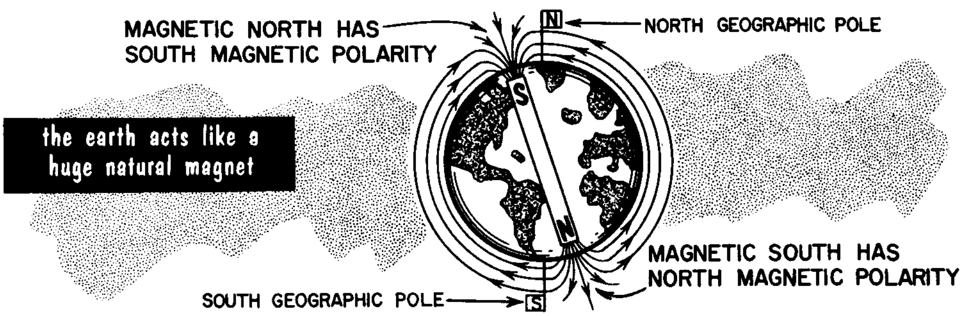
Magnetic materials can be magnetized and thus made to attract iron and certain other metals. Foremost among magnetic materials are soft iron and alloys of steel such as Alnico, containing one or more other metals including nickel, cobalt, and aluminum. Some recently developed "ceramics" make excellent magnets, among these being Indox, which contains barium-oxide and Vectolite, an iron-content compound. Temporary magnets are made from iron. They behave as magnets only as long as the magnetizing force which makes them magnets acts on the iron. Iron gives up its magnetic properties almost immediately after the magnetizing force is removed. Steel and its alloys, on the other hand, are readily made into permanent magnets; i. e., magnets which hold on to their magnetic properties indefinitely.



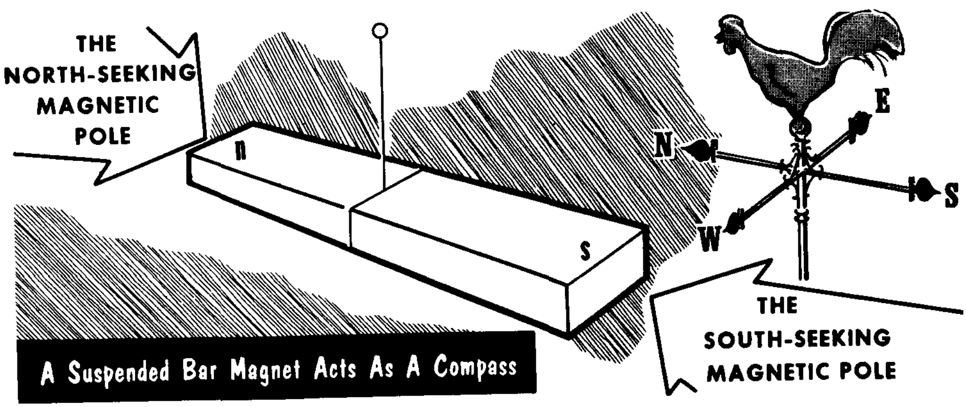
In contrast to the relatively few magnetic materials, there are a host of nonmagnetic materials; these cannot be magnetized. For our purposes, we need mention but a few: glass, paper, wood, rubber, plastic, cotton, tin, and copper. Although nonmagnetic materials cannot be magnetized, magnetic energy will pass through them; in fact, they are completely "transparent" to this kind of energy. Magnetic materials can be penetrated by magnetic energy also--but once inside the substance, the energy will use the material as a preferred path for as long as possible, rather than leave it and travel through air or through some other nonmagnetic substance. Thus, a magnetic substance is sort of a confining path for magnetic energy. It is this form of behavior that makes soft iron an excellent "shield" around an object from which we want to keep out magnetic energy.

The Poles Of A Magnet

If a permanent bar magnet is suspended horizontally so that it is free to rotate, it will come to rest with one end pointing towards the North pole of the earth, and the opposite end pointing towards the South pole. It has become conventional to refer to the ends of a magnet as the "poles" of the magnet, the end pointing towards the North being called the "north-seeking" pole, and conversely, the end pointing towards the South called the "south-seeking" pole.



Generally speaking, every magnet has two poles; a North (or N pole) and a South (or S pole). There are exceptions to this rule wherein a magnet may have several N poles and an equal number of S poles. These are, however, not common. No magnet can have only one pole. Wherever there is an N pole, there is a corresponding S pole and vice versa. Sometimes, the N pole is labeled with a plus (+) sign and the S pole with a minus (-) sign.

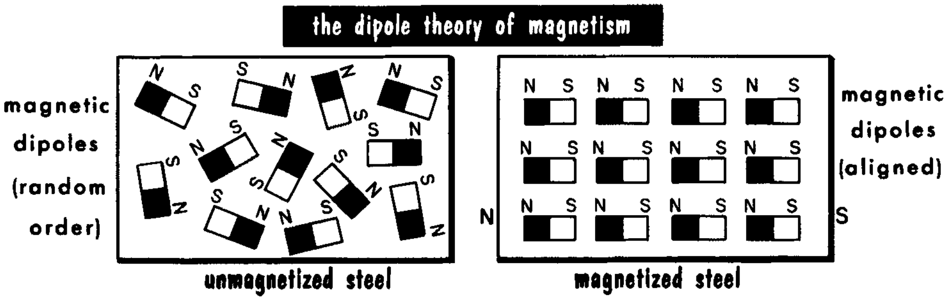


A Suspended Bar Magnet Acts As A Compass

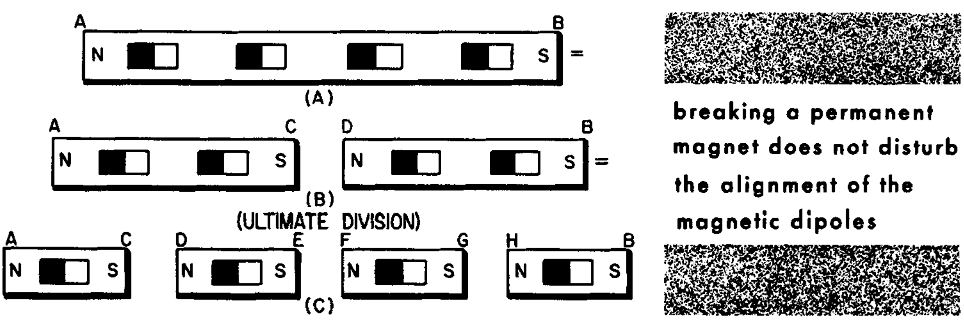
The earth behaves like a huge natural magnet, having a so-called North and South magnetic poles in the approximate direction (but not in the immediate vicinity) of their respective geographic poles. Based on the accepted behavior of a magnet and the laws of magnetism, the earth's North magnetic pole actually has south magnetic polarity, whereas the South magnetic pole actually has north magnetic polarity.

The Basis of Magnetism

How do we explain magnetism? The details are not exactly known but it is believed that magnetism is associated with the electrons in the atoms of which magnetic substances are made. These atoms are believed to be tiny magnets or "magnetic dipoles", each having a North pole and a South pole. Presumably, each atom accounts for a certain amount of magnetism, some more and some less. When a magnetizable material is in an unmagnetized state, the tiny atomic magnets are oriented in random fashion; that is, the magnetic poles point in all directions. When a substance contains randomly-oriented atomic magnets, it is assumed that the magnetic effects of one atom offset or cancel the magnetic effects of another. Thus, the material as a whole does not display any magnetic characteristics.

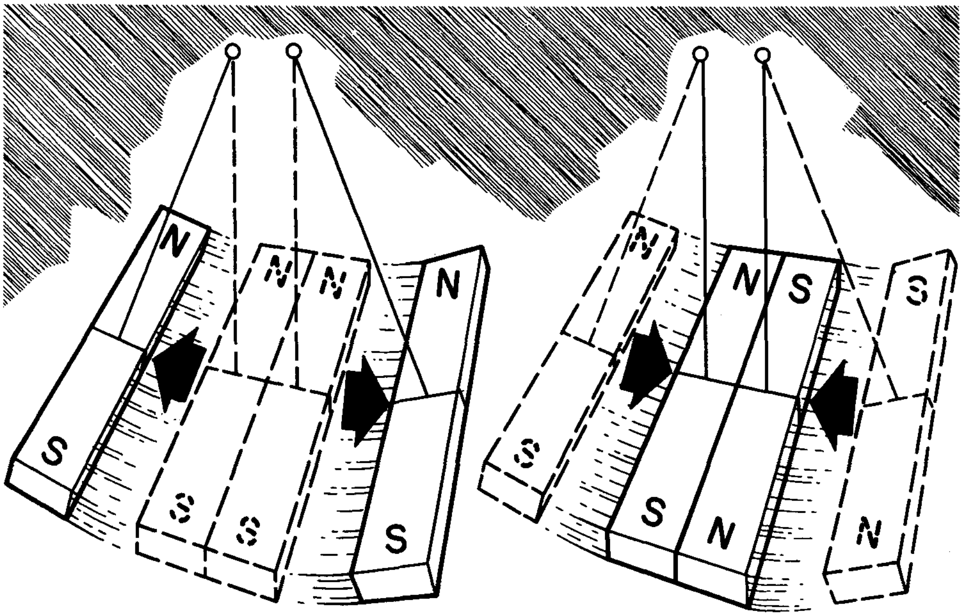


The process of "inducing" magnetism is believed to be the reorientation of the randomly-positioned atomic magnets so that their poles line up in horizontal rows--the N pole of one facing the S pole of the other. Each horizontal row of atomic magnets forms an extremely thin magnet, and the great many parallel horizontal rows of magnets form a single composite magnet with a common N pole at one end and a common S pole at the other end. This theory explains why it is possible to break a permanent magnet into any number of parts, each of which remains a magnet with an N and an S pole. When a material "loses" its magnetism, the atomic magnets change their "magnetized" orientation to a random one. In a permanent magnet, once it has been changed from random orientation to "magnetized" orientation, the atomic magnets in it stay that way until disrupted by heat or physical shock.



Attraction and Repulsion Between The Poles of Magnets

Much of what is known about magnetism has been learned by experiment and observation of its effects. If we suspend two bar magnets with the two N poles (or S poles) near to and facing each other, a mutual force of repulsion is demonstrated between the like poles. The magnets move away from each other, indicating the presence of energy that can exert mechanical force. The action is described by a basic law of magnetism; namely, like magnetic poles repel each other.



LIKE MAGNETIC POLES REPEL EACH OTHER

UNLIKE MAGNETIC POLES ATTRACT EACH OTHER

If the magnets are arranged so that the N pole of one faces the S pole of the other, the magnets will move towards each other. If the magnets are not perfectly in line, they turn on their axes so as to bring the unlike poles as close together as possible, again indicating the presence of energy around the magnet. Such behavior demonstrates another basic rule of magnetism, this being that unlike magnetic poles attract each other.

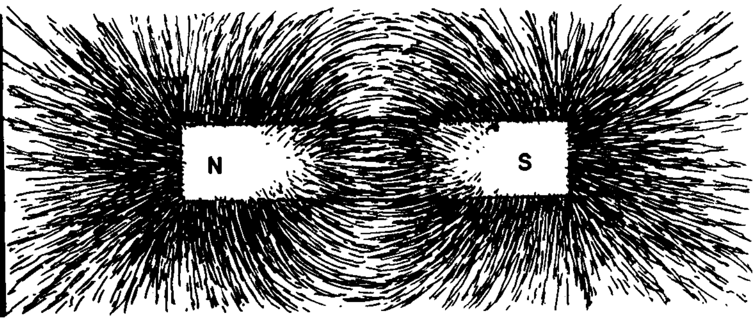
The force of repulsion between like poles, or attraction between unlike poles, varies inversely as the square of the separation between the poles. If the separation is doubled, the force decreases by a factor of 4 or to 1/4; if it is tripled, the force decreases by a factor of 9, or to 1/9, and so on. If the separation is decreased by a factor of two, the force increases by the square of 2 or 4 times; if the separation is decreased by a factor of 3, the force increases by the square of 3 or 9 times, and so on.

The Magnetic Field (Lines of Force)

How do we describe the energy around a magnet? Michael Faraday advanced the concept that the region in all directions around a magnet is a "zone of magnetic influence"; i. e., it is occupied by an invisible "magnetic field of force" or simply, a "magnetic field." Any magnetizable object placed within the magnetic field is made into a magnet--temporary or permanent, depending on the material--by the energy contained in the magnetic field.

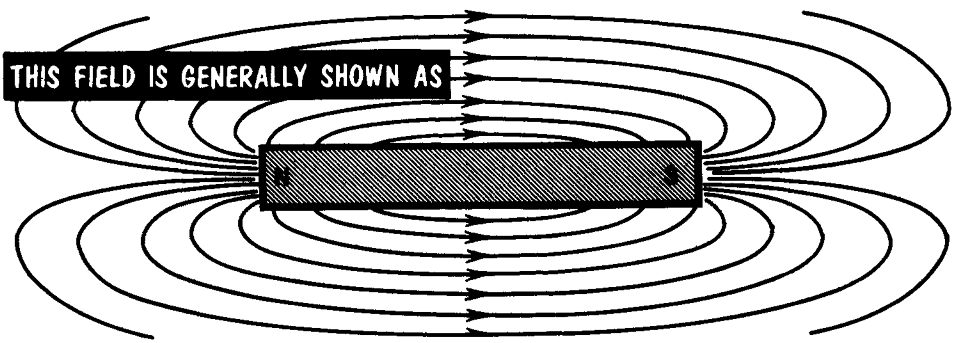
MAGNETIC FIELD AROUND A BAR MAGNET

iron filings indicate lines of force



Faraday visualized the magnetic field as being made up of imaginary "magnetic lines of force." Each line of force may be thought of as a thin, thread-like region where magnetic energy exists. To explain magnetic behavior, we assume that the line of force is real. We say that each line of force "leaves" the magnet at the N pole, completes an elliptical path outside the magnet to the S pole, then "re-enters" the magnet and advances to the N pole inside the magnet. Thus, each magnetic line of force is a complete loop of magnetic energy, part of which is outside the magnet, and part of which is inside the magnet. Moreover, we say the field has a "directional quality". Specifically, the direction of the magnetic field, externally, is from the North to the South pole.

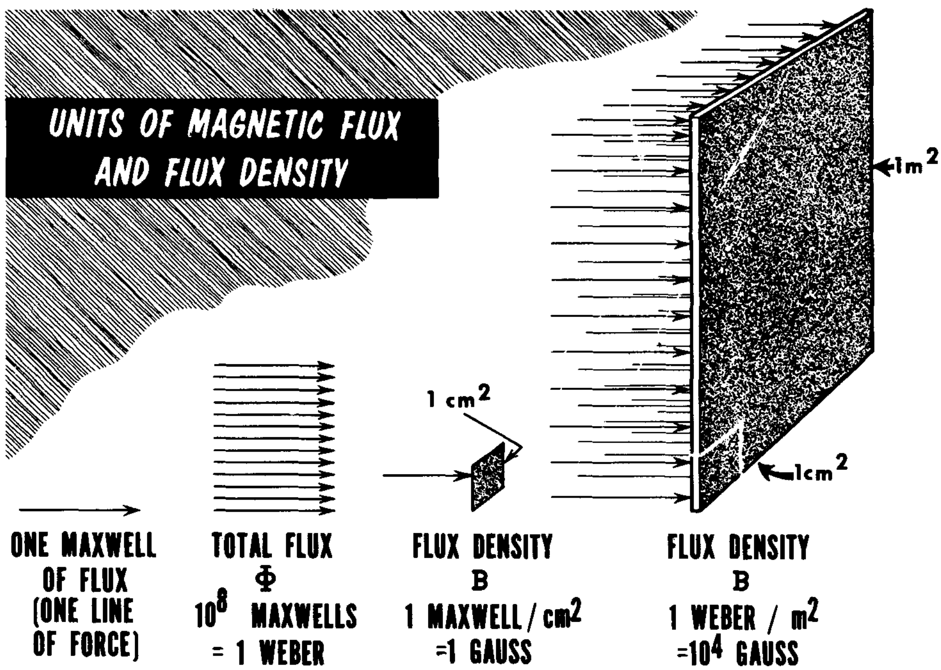
THIS FIELD IS GENERALLY SHOWN AS



We can place a sheet of cardboard over a bar magnet, and sprinkle some fine iron filings. If we tap the cardboard lightly, the filings will arrange themselves in the "line" pattern shown. The pattern indicates the organization of the lines of force.

Lines of Force and Flux Density (Units)

To express the strength of the magnetic field at any given point around the magnet, or to compare the magnetic strength, we refer to the number of magnetic lines of force that pass through a unit area in the magnetic field. The unit area used for this purpose is a cross-section 1 centimeter on each side of 1 square centimeter (or 1 cm^2).



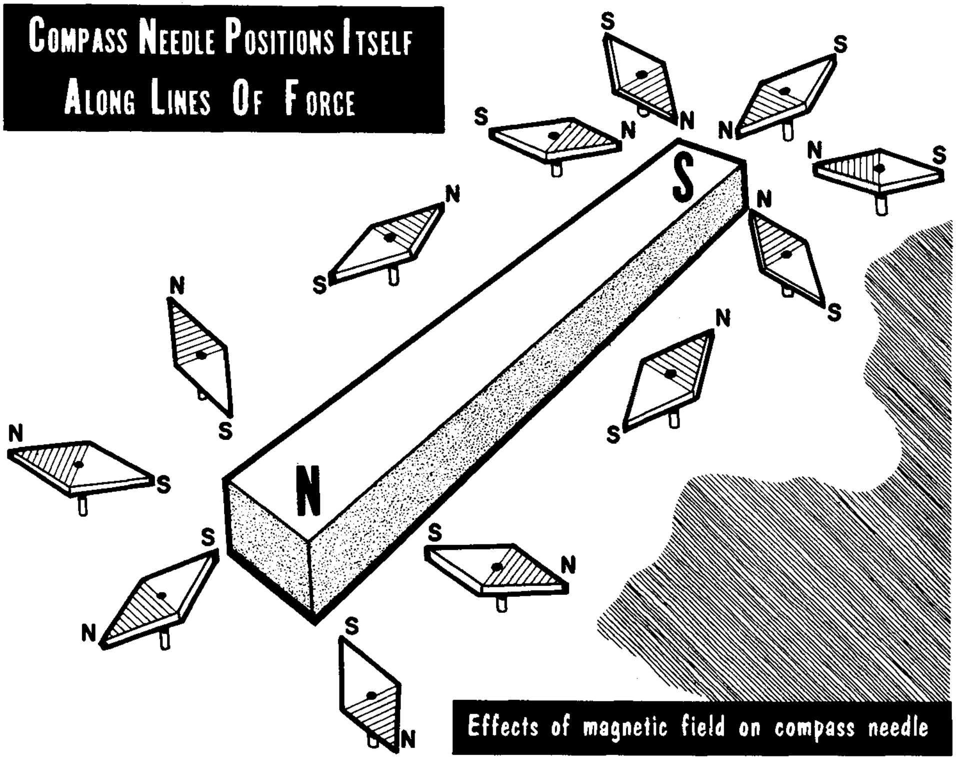
When speaking about the magnetic field strength, it is customary to refer to a single magnetic line of force as a "maxwell of flux". If 1 maxwell of flux (1 magnetic line of force) passes through a cross-sectional area of 1 square centimeter, the strength of the magnetic field at that location is 1 gauss. Associating the number of lines of force with the unit area (or the number of flux lines per square cm) expresses flux density (measured in gausses). Thus, two magnets or two magnetic fields may be compared by referring to the flux density at corresponding points in the two fields. If magnet A is rated at a flux density of 100 gausses and magnet B at a flux density of 500 gausses at the same distance from the N poles, magnet B is by far the stronger magnet. Electrical devices which make use of magnetic lines of force created by electric current may function with flux densities of many hundreds and even thousands of gausses.

If reference is made to the total number of magnetic lines of force of a magnetic field, it is called total flux and is indicated by the symbol Ø. The unit used is 100, 000, 000 or 10^8 maxwells (lines of force). It is called "weber"; i. e., 10^8 maxwells = 1 weber.

Properties of Magnetic Lines Of Force

Magnetic lines of force have numerous properties. Only two are discussed here; the rest are explained in connection with the magnetic effects of current.

Compass Needle Positions Itself Along Lines Of Force



Effects of magnetic field on compass needle

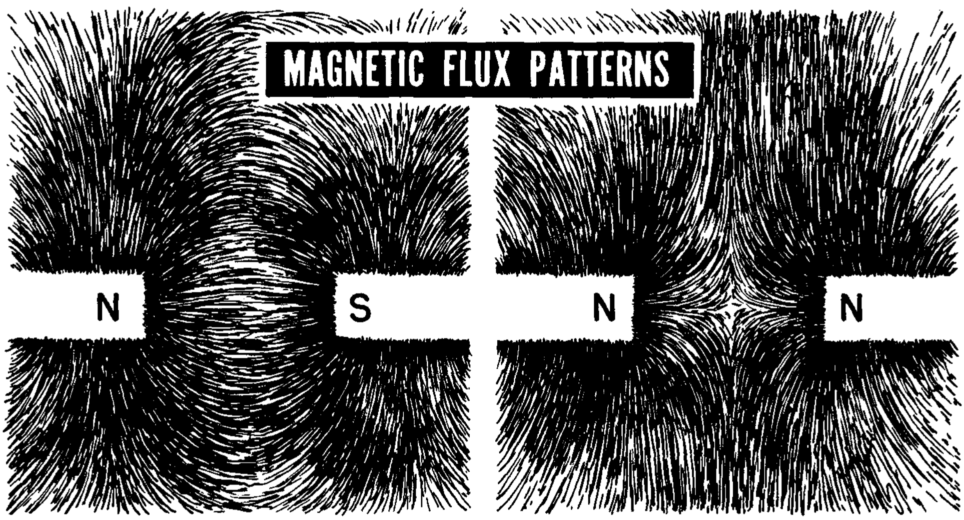
The formation of the "line" pattern by the iron filings, or by the directional indications of the compass needle, is due to the action of the magnetic energy in the lines of force. The energy magnetizes each filing and urges it (and the compass needle--already a magnet) to take positions along the lines of force. In the process of magnetizing a filing (or some other magnetizable object), the energy in the magnetic field induces magnetism in the filing. The end of the filing entered by the lines of force becomes the South pole of the magnetized bit of iron; the end where the lines of force leave the filing is the North pole. Thus, the magnetizing lines of force pass lengthwise through each magnet they create.

Being positioned along the lines of force, each tiny magnet in the field is threaded by the greatest number of flux lines. This is a rule of magnetic behavior. A magnetizable object located in a magnetic field will position itself so that it will be threaded by the greatest number of lines of force. Given sufficient energy in the magnetic field, the lines of force exert turning, pulling, or pushing force on other magnets on which they act, so that the magnets will be threaded by the greatest number of flux lines.

Properties of Magnetic Lines of Force (Cont'd)

Two magnets placed near each other have separate magnetic fields which occupy an area common to both. One might gather from this that the lines of force cross or intersect. This does not happen; lines of force never cross. When like poles face each other, the lines of force of each field remain separate entities, but each field becomes distorted. When unlike poles face each other, the lines of force of each field interact with the other and produce a strong resultant field.

Wherever the lines of force of the two fields are parallel and have the same direction, a strengthened field is created. This follows from the condition that a greater number of lines of force acting in the same direction are present per unit area than if only one magnet and one magnetic field were present. Where the lines of force of the two fields act in opposite directions, the energy of one field tends to offset the energy in the other field. The result is a weakened field in that area. This condition is equivalent to fewer lines of force traversing the area.



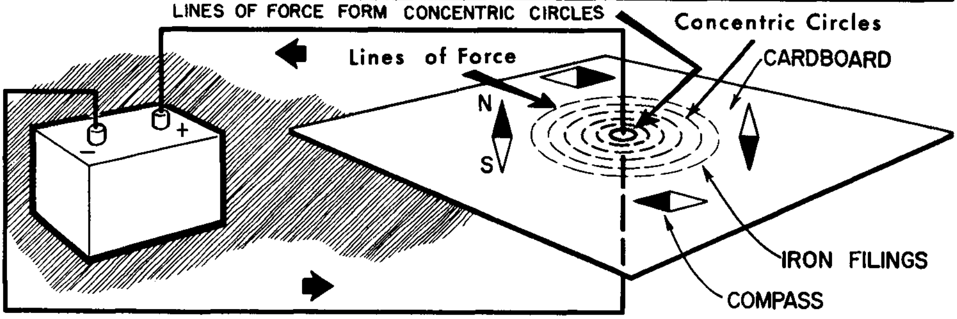
opposite poles facing
(mutual attraction exists)

like poles facing
(mutual repulsion exists)

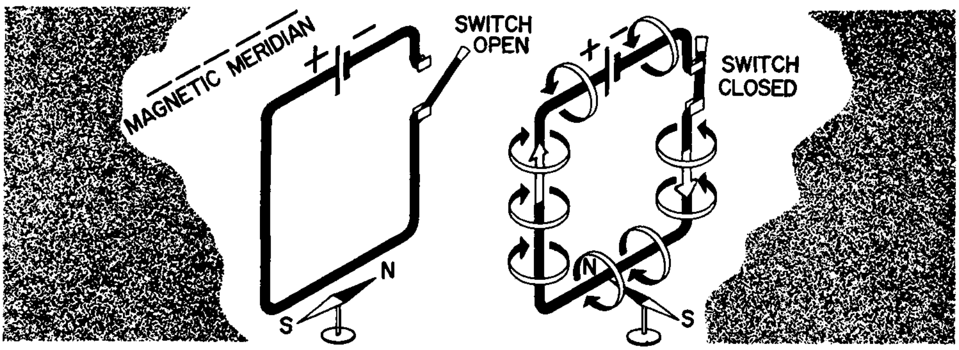
Still another important behavior of magnetic lines of force is that each magnetic loop behaves as though it were a rubber band stretched, which, if given the opportunity, would contract. This explains why unlike poles attract each other. The loops of magnetic force passing through the magnets behave as though they shrink, thus pulling the magnets toward each other. This is the simplified version of the condition whereby when one portion of a resultant of two fields is strengthened and another is weakened, the movable sources of the field move from the stronger to the weaker portion of the field. Strengthened and weakened magnetic fields in an area are illustrated by more or fewer lines of force.

Electromagnetism

MAGNETIC FIELD IS PRODUCED AROUND CURRENT-CARRYING CONDUCTOR



In 1819, Hans Christian Oersted, a Danish physicist, observed that the directional indications of a small magnetic compass located near a wire through which electric current (d-c) was flowing were influenced by the current. The presence or absence of the current, as well as its direction of flow in the wire, had a major effect on the behavior of the compass needle. Later experiments by other scientists resulted in the conclusion that electric current produced a magnetic field or a zone of magnetic energy around itself. This was the beginning of electromagnetism, the term used to denote magnetic effects that result from electric current.

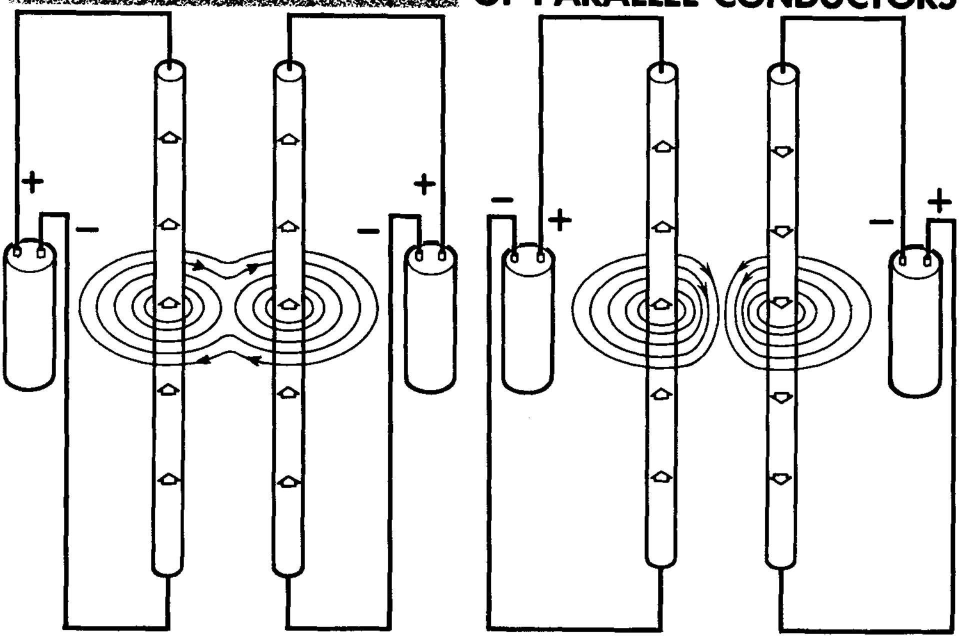


The magnetic field surrounding a current-carrying conductor is most dense nearest the wire (greatest field strength), and thins out at increasing distances from the wire (reduced field strength). If the above experiment is repeated as the cardboard is moved up and down along the wire, the same pattern appears at every point, indicating that the loops of force (concentric circles) surround the current all along its path. If a small compass is moved around the current-carrying conductor, the poles of the needle line themselves up with the magnetic lines of force and change their direction as the compass is moved around the wire. Thus, the lines of force created by the current have direction, just as the lines of force associated with a magnet have direction, except that the direction is now always at right angles (perpendicular) to the axis of the wire.

Action of Magnetic Field Around Two Parallel Current-Carrying Wires

We showed that two magnets with facing poles demonstrate mechanical forces. Two parallel wires carrying current also demonstrate mechanical forces of attraction or repulsion, depending on the direction of the current in the wires. The energy is in the magnetic field surrounding the wires. When the currents in the two wires move in the same direction, the lines of force between the wires move in opposite directions. One offsets the other; hence, the resultant field between the wires is weakened. At the same time, the directions of the lines of force on the outside of the wires are in the same direction. Each field strengthens the other at these points. The overall result is a strengthened field on the outside of the wires and a weakened field between the wires. The contracting behavior of the "outside" magnetic lines of force now comes into play, and cause the two wires to move physically towards each other if motion is possible. Although the action is attraction between two wires, a much more useful description is to say that the strengthened portions of the field tend to push the conductors toward the weaker part of the field; hence, towards each other.

ACTION BETWEEN MAGNETIC FIELDS OF PARALLEL CONDUCTORS

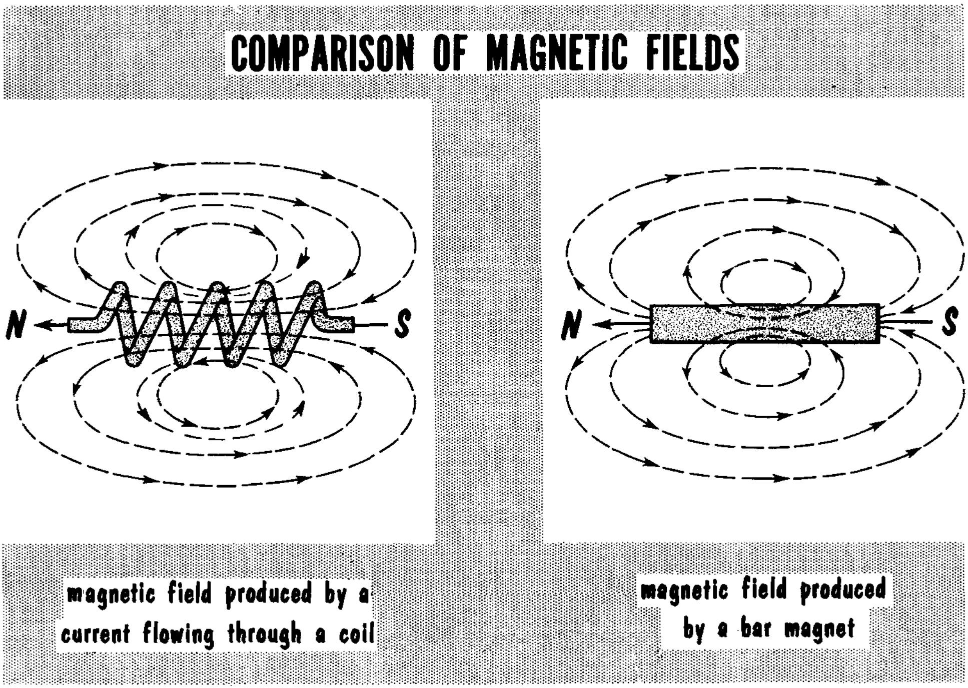


Currents Flowing In The Same Direction **Currents Flowing In The Opposite Direction**

When the currents in the two parallel wires flow in opposite directions, the resultant field is strongest between the wires and weakest on the outside. The two wires therefore move apart, again moving from the stronger portion of the field (inside) towards the weaker portion (outside).

The Magnetic Field Around A Current-Carrying Coil

When a straight conductor is bent into a series of loops, forming a coil (also known as a solenoid), a very much stronger magnetic field is created within the space inside the winding as well as outside the coil, but especially inside. The explanation is that each magnetic loop of force that encircles each turn passes through the inside of each turn--hence, through the inside of the coil as a whole--and all of them have the same direction. Because of this, the number of lines of force (or flux density) per unit area inside the coil is greater than anywhere else for the same area outside the turns. The magnetic field of a coiled wire is therefore very much stronger than if the wire had not been coiled.

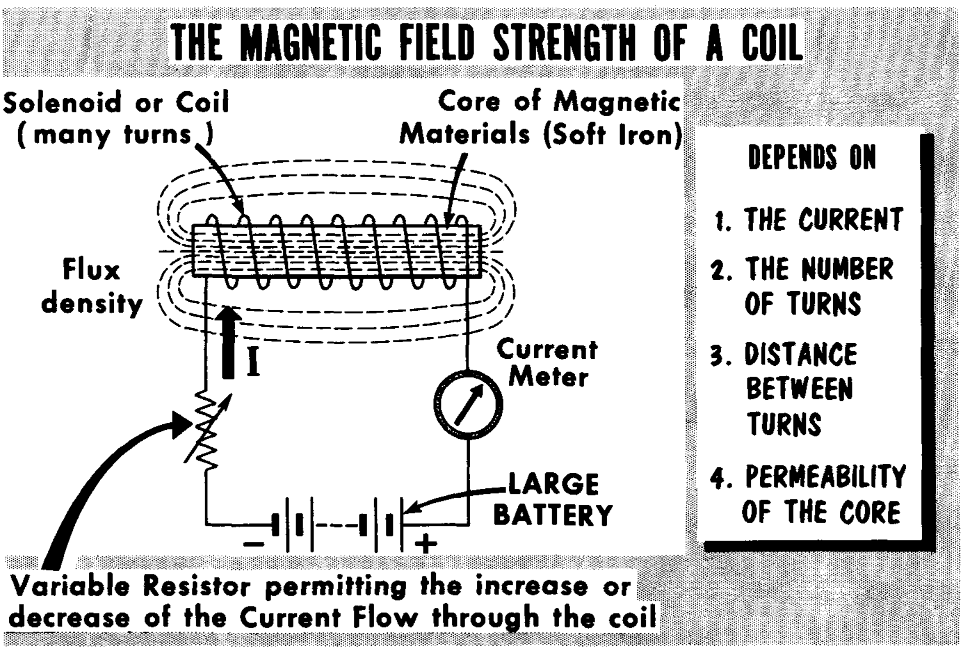


When current flows in a solenoid, the magnetic lines of force that surround the winding enter the winding at one end and leave at the other. Where the lines of force enter the winding is called the South end, or South pole, and the end where they leave is called the North end, or North pole. This organization of the magnetic field is like that of the conventional bar magnet; hence, a solenoid with current flowing in it is an electromagnet

The term electromagnet denotes a magnet created by virtue of the magnetic lines of force that issue from an electric current. It is a "temporary" magnet because it behaves as such only while current is flowing in the wire. When the current ceases, the magnetic behavior ceases. Coils of few turns make weak magnets; to make them strong magnets requires many turns.

Strength of the Magnetic Field Around A Solenoid (Ampere-Turns)

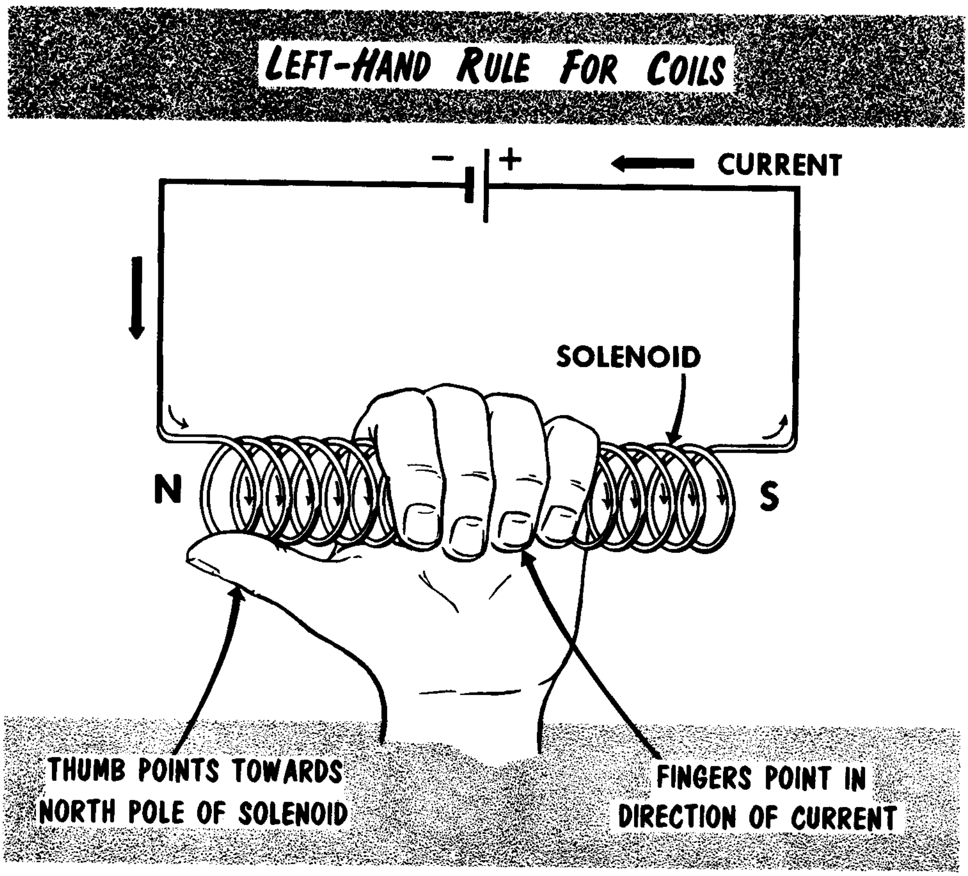
The strength of the magnetic field created by current in a solenoid is determined by several factors: the amount of current; the number of turns; the separation between the turns; and the core material. The more current there is in the winding, the greater the number of flux loops there are that surround the coil and the more energy there is in the magnetic field around the coil. Each turn through which current flows contributes lines of force to the total current; hence, the more the number of turns, the greater the number of flux loops everywhere around the coil. These two factors are combined into a single term called "ampere-turns". It is the product of the current in amperes and the number of turns. A coil with 100 turns and .5 ampere of current has the same ampere-turns as a coil with 500 turns and .1 ampere of current, for $100 \times .5 = 500 \times .1$. The closer the turns are to each other, the greater the linkage between the flux lines; i.e., the more readily the flux loops around the individual turns link with other turns, and the stronger the magnetic field is around the coil.



A coil wound on an insulated form with an air-filled core will, for a given current, have a weaker magnetic field than a coil wound on a core made of a magnetizable material such as soft iron. The soft iron core becomes magnetized and, in this state, creates additional lines of force which add to the total due to the current. The core material is described in terms of its permeability or magnetic conductivity. Permeability expresses the ease with which magnetic lines of force are established in a material relative to that in air or in a vacuum, both of which have a permeability of 1. Magnetic materials with a permeability exceeding 50,000 are available.

The Left-Hand Rule For a Coil

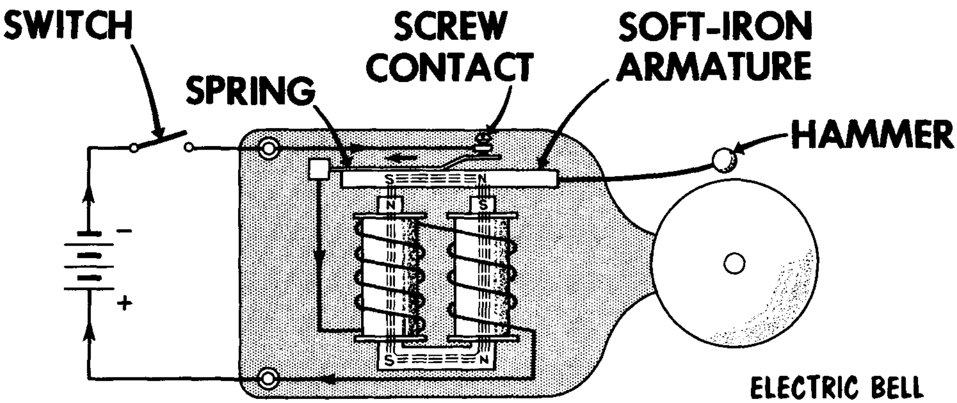
The magnetic polarity produced by the flow of current through a coil is determined by the direction of the current and the direction of the coil winding. We should not confuse the voltage drop across a coil with the polarity magnetic produced in the coil. North and South are terms applied to opposite magnetic poles; minus and plus are terms used to describe a difference of potential.

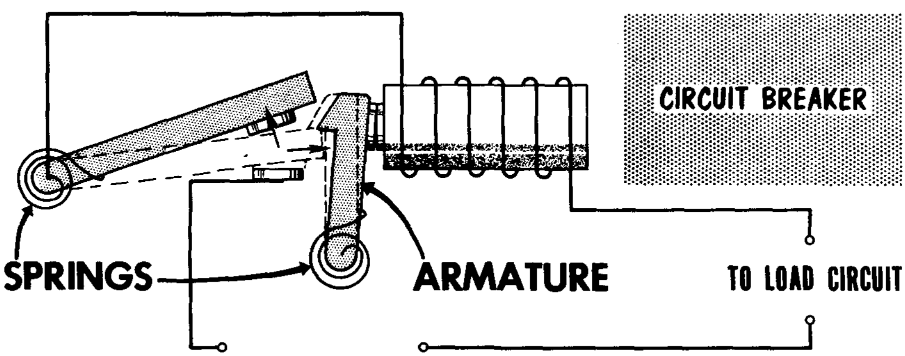


Based on the electronic concept of current, a method known as the "left-hand rule for a coil" determines the relationship between the direction of current and the direction of magnetic flux through the coil. If the coil is grasped in the left hand with the curled fingers pointing in the direction of current flow through the turns of the coil, the extended thumb will be pointing in the direction of the flux and the North pole of the coil. Obviously, the other end of the coil is then the South end, or South "pole." As you will see later, the interaction between the magnetic field that surrounds a conductor in which current is flowing and a permanent magnetic field that exists between magnetic poles is the basis for operation of a wide variety of electrical devices.

The Electromagnet

No other electromechanical device is so widely used or representative of magnetism and electricity as the electromagnet. This unit is composed of a coil or wire wound around a soft iron core. When direct current flows through the coil, the core becomes magnetized with the same polarity that the coil (or solenoid) would have taken without the iron core. When the current through the coil is reversed, the polarity of the electromagnet is reversed. The polarity follows the left-hand rule the same as the solenoid. The addition of the soft-iron cores serves two functions: first, the magnetic flux is increased because the soft-iron core is more permeable than the air core; second, the flux is much more highly concentrated. The permeability of soft iron is many times that of air; therefore, the flux density is increased considerably when a soft-iron core is inserted in the coil.





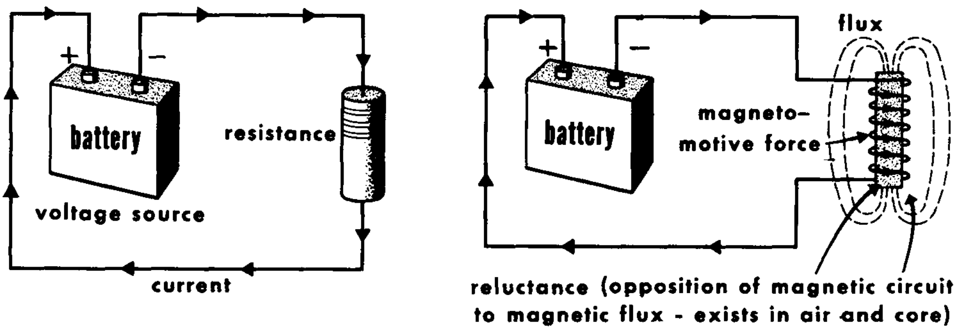
Popular applications of the electromagnet are in bells and buzzers, in circuit breakers used to protect circuits, and in relays, used to open and close circuits. Note that in the devices shown, the fundamental operation depends upon the electromagnet becoming energized and attracting a movable piece of metal called an armature.

The Magnetic Circuit

Lines of magnetic flux do not end at magnetic poles, but are continuous or closed on themselves like the lines of electron flow in an electric circuit. In fact, a magnetic circuit is, in many ways, similar to an electric circuit. To produce an electric current, a voltage called an electromotive force is required. Similarly, to produce a magnetic flux, a force called a magnetomotive force is required. In an electric circuit, for a given amount of voltage or electromotive force, the current is determined by the circuit resistance. Similarly, in a magnetic circuit, for a given amount of magnetomotive force, the flux density depends upon the amount of opposition or reluctance. Thus, in the magnetic circuit, as in the electric circuit, we can say: the result produced is directly proportional to the force that produces it and inversely proportional to the opposition encountered.

COMPARING THE ELECTRIC AND MAGNETIC CIRCUIT

$$\text{I (current)} = \frac{\text{E (electromotive force)}}{\text{R (resistance)}} \qquad \phi\ \text{(flux or lines of force)} = \frac{\text{F (magnetomotive force)}}{\text{R (reluctance)}}$$



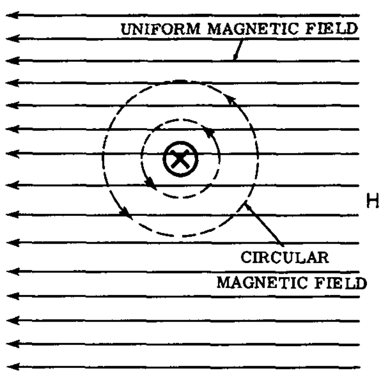
There are two important differences in the relationship of the electric circuit to that of the magnetic circuit. The first difference is that in the electric circuit, the resistance is a constant value and can be determined by measuring the ratio of voltage to current. In a magnetic circuit, however, the reluctance is not a constant but depends upon the flux, or field strength. The second difference is that in electric circuits, current actually flows (electron flow) from one point to another. In the magnetic circuit, there is no actual flow of flux, but merely an indication of the intensity and direction of the magnetic field.

The magnetizing force set up due to current flowing in a coil or solenoid is known as the magnetomotive force (abbreviated mmf). The practical unit of magnetomotive force is the ampere-turn. We can then say that the magnetomotive force is proportional to the current (in amperes) in the circuit and to the number of turns of the coil. The magnetic flux (Ø) comprises the total number of lines of force in the magnetic circuit. The maxwell is the unit of flux (1 line of force is equal to 1 maxwell).

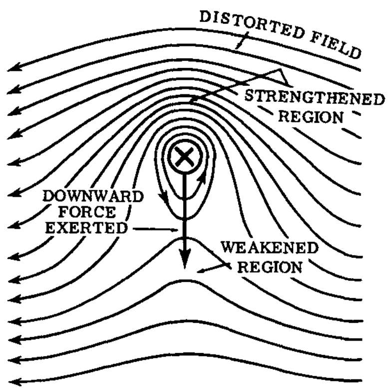
Motion of A Current-Carrying Conductor In A Magnetic Field

You have learned that a current in a conductor is surrounded by a magnetic field. If this conductor is placed within a stationary magnetic field, the two fields combine to produce a single resultant distorted field which exerts force on the conductor. The force is attributable to the creation of a strengthened and a weakened region in the resultant field around the conductor. For instance, the lines of force of the individual fields have the same direction above the wire; hence, a strengthened field is created above the wire. Below the wire, the field produced by the current has a direction which is opposite to the field of the magnets, so the two fields buck each other and create a weakened region. The conductor will be moved downward because it is "pushed" from the strengthened region to the weakened region of the field.

MOTION OF A CURRENT-CARRYING CONDUCTOR PLACED IN A MAGNETIC FIELD



CIRCULAR MAGNETIC FIELD SUPERIMPOSED ON UNIFORM MAGNETIC FIELD

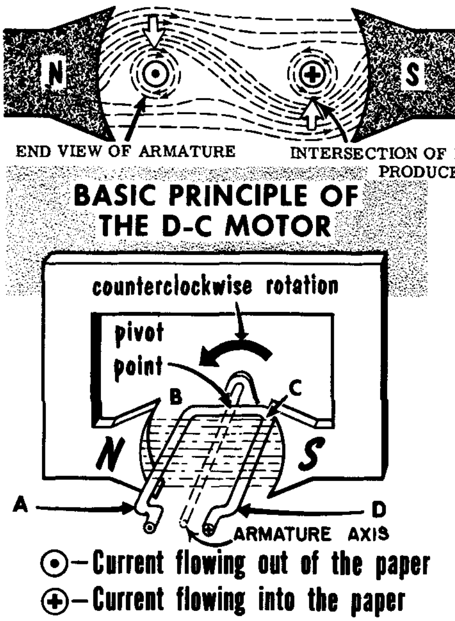


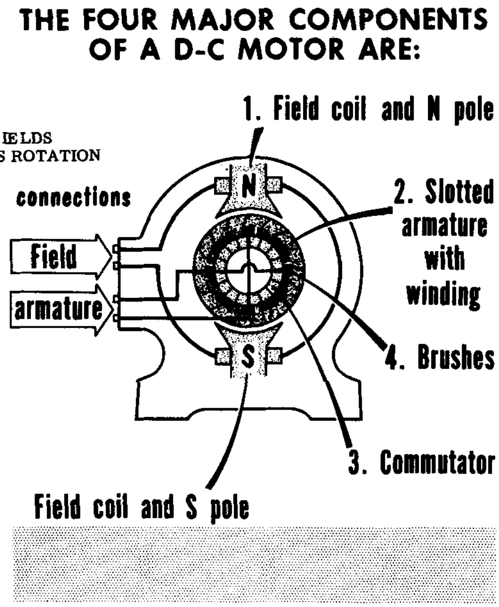
RESULTANT OF THE TWO SUPERIMPOSED FIELDS

By changing the direction of the current in the wire, we can transpose the locations of the strengthened and weakened regions of the resultant field around the wire. The former is now located below the wire and the latter above it. The wire now is "pushed" upward. The greater the amount of current in the wire, the stronger are the aiding and bucking effects; therefore, the greater the difference in magnitude of the strengthened and the weakened parts of the field. This results in greater force being exerted on the wire and more rapid motion. The foregoing description explains a fundamental condition which has many applications. Two prominent ones are related to the electric motor and the "moving-coil" type of electric meter. Both are discussed in more detail on the pages to follow.

The D-C Motor

A d-c motor is a rotating machine that transforms electrical energy into mechanical energy. The basic principle was explained on the preceding page. The stationary magnetic field can be created by a d-c current in a winding on a suitably-shaped core of soft iron, as well as by the use of a permanent magnet.





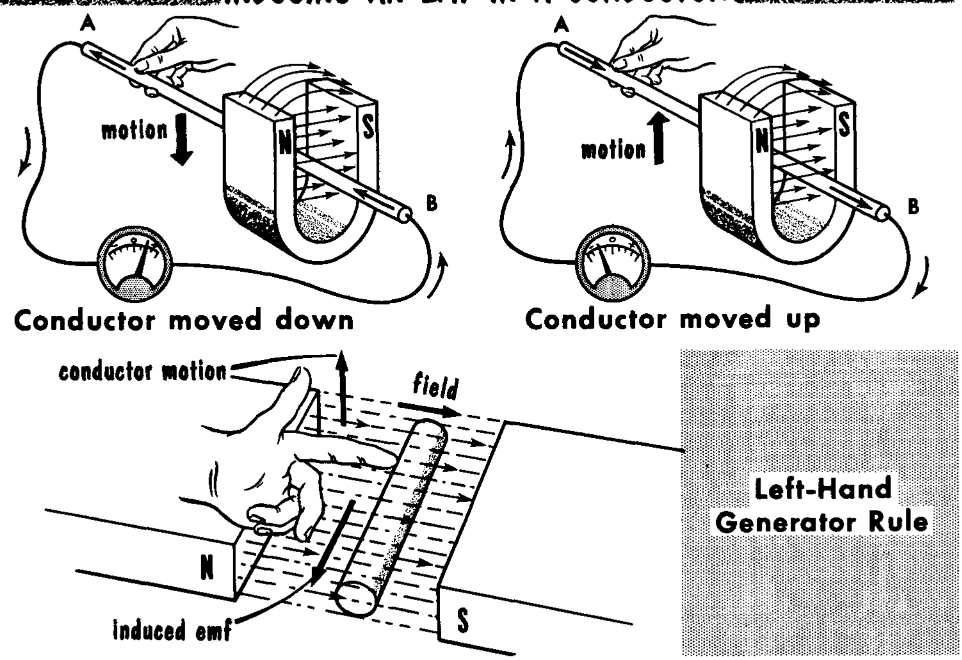
The two sides of the pivoted conductor (the armature) create two conditions of current direction within it. The left side (A-B) carries current towards the reader, whereas the right side (C-D) carries current away from the reader. Side B-C is just a conducting link. The magnetic field of the current in side A-B, acting in conjunction with the stationary field of the magnet, causes the resultant field to be stronger above the conductor and weaker below it. In the case of the right side, C-D, the direction of the current causes the resultant field to be stronger below the conductor and weaker above it. Thus, a downward push is felt by A-B and an upward push by C-D. With the loop of wire pivoted at two points, the mechanical forces produce a torque (a force which tends to turn things), causing rotary motion on the part of the loop as a whole.

The practical motor has many turns for the armature. Physically, it consists of a soft steel frame with slots into which the turns are wound. The turns are divided into separate groups which terminate on the commutator, a device used for making electrical contact with the conducting loops as they rotate. The commutator and brushes enable the current always to be passed into those turns which face the pole pieces, thereby generating continuous torque to keep the armature rotating.

Induced Electromotive Force (emf); The Moving Conductor Generator

In the d-c motor, electrical energy is transformed into mechanical energy by making use of magnetic effects. The reverse is also true; mechanical energy can be transformed into electrical energy by utilizing magnetic effects. This occurs in the electric generator. A simple experiment demonstrates the process. A short length of stiff copper wire is connected to a sensitive zero-center galvanometer by means of two lengths of flexible copper wire. Also required is a strong "U"-shaped ("horseshoe") permanent magnet with pole pieces which are close together, thus affording a concentrated stationary magnetic field between them.

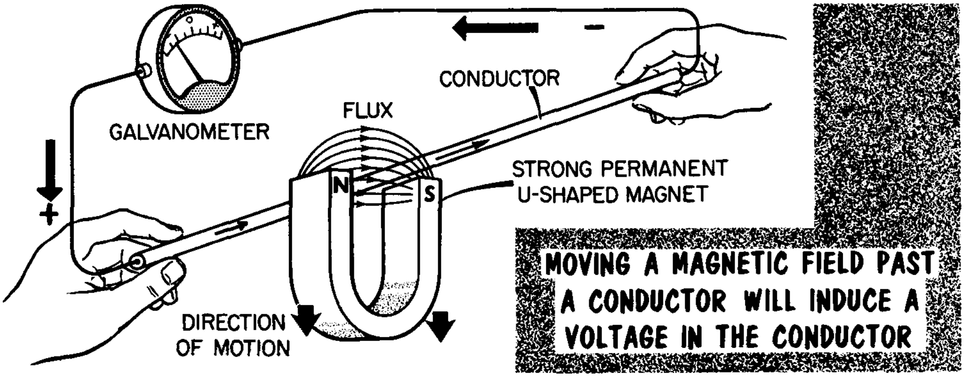
INDUCING AN EMF IN A CONDUCTOR



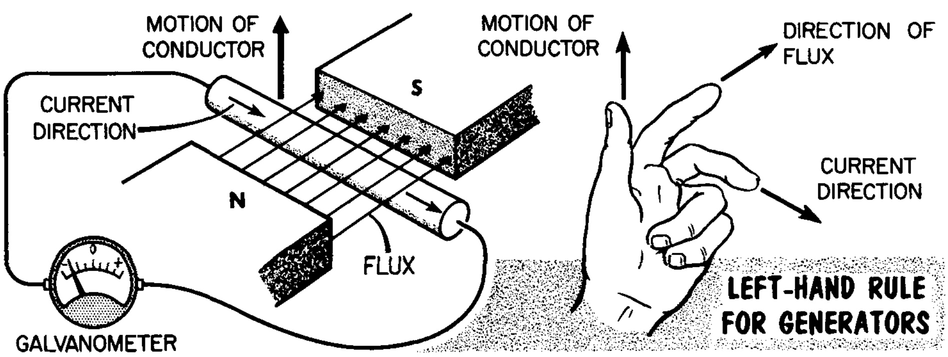
When the stiff copper conductor is moved downward between the pole pieces with a rapid motion, thereby cutting across the stationary lines of force, the galvanometer needle "kicks" away from zero in one direction, and returns to zero when motion of the conductor ceases. Then, when the conductor is moved upward between the pole pieces in a rapid motion, the meter needle kicks in the opposite direction. The two directions of deflection of the needle indicate current flow in opposite directions. Both are the result of an electromotive force (voltage) being induced in the stiff copper wire while it is cutting across the magnetic lines of force between the pole pieces. If the wire is moved horizontally (i.e., parallel to the lines of force in either direction between the pole pieces), no current is evident on the galvanometer. The overall action is described by saying that "when a conductor cuts magnetic lines of force, an electromotive force is induced in the conductor."

Induced Electromotive Force; Moving Magnetic Field

Instead of moving the conductor so that it will cut stationary magnetic lines of force, we reverse the process and move the magnetic field so that moving lines of force cut the stationary conductor. As with the case of the moving conductor, maximum emf is induced when the lines of force cut the conductor at right angles, and minimum emf (effectively zero) is induced when the moving lines of force have a direction parallel to the axis of the conductor.

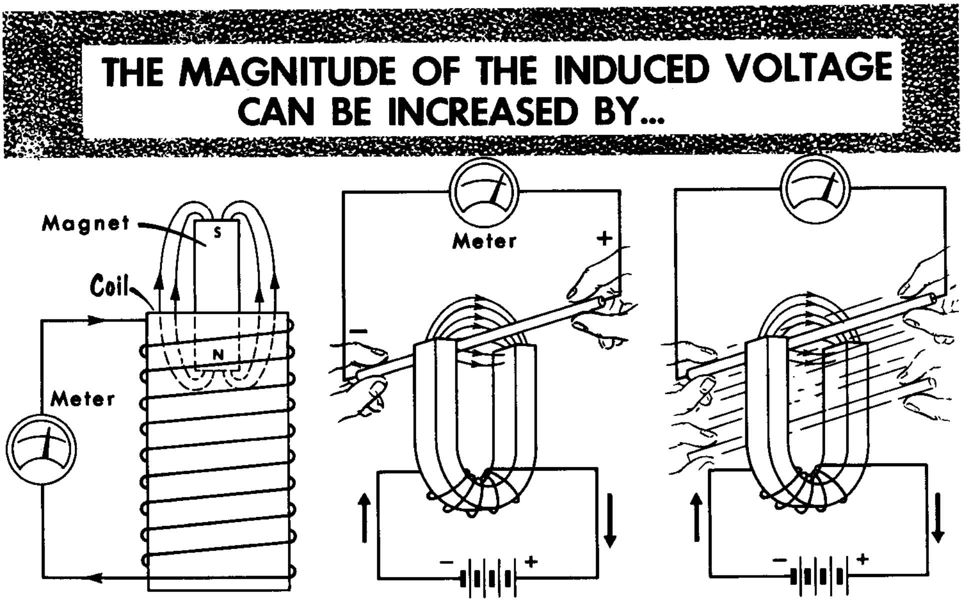


The electrical polarity of the voltage induced in the conductor, hence, the direction of the resultant current in the galvanometer circuit, is a function of the relative direction of motion between the moving magnetic lines of force and the stationary conductor. When the magnetic field moves downward, it is the same as though the conductor moves upward. When the magnetic field moves upward, it is the same as though the conductor moves downward through the field. As to the polarity of the voltage induced in the conductor by the relative motion between the conductor and the magnetic lines of force, it is indicated by the "left-hand rule for generators." This rule states that if you hold the thumb, first and middle fingers of the left hand at right angles to one another, with the first finger pointing in the direction of the lines of force and the thumb pointing in the direction of motion of the conductor, then the middle finger will point in the direction of the induced emf (the direction in which current will flow).



Factors That Control The Amount of Induced EMF

The amount (magnitude) of the emf (voltage) induced by relative motion between magnetic lines of force and a conductor is determined by several factors. In this connection, do not try to associate every magnetic field with a permanent magnet; think just of the magnetic field regardless of its origin.



1 winding the conductor to form a coil of many turns

2 replacing the magnet with a strong electromagnet

3 increasing the relative speed of the movement between conductor and field

The angular relationship between the magnetic lines of force and the conductor already has been mentioned. Another is the strength of the magnetic field which is cut, or which is doing the cutting. The greater the field strength (that is, the greater the flux density), or the more the number of lines of force per unit area which are cut or are doing the cutting per unit time, the greater will be the induced emf.

Another way of increasing the induced emf is by increasing the velocity of the relative motion between the conductor and the field. The faster the armature of a generator turns, the greater is the induced emf because more lines are cut every second. The faster the motion of the bar magnet inside a solenoid, the more rapid the rate of cutting; hence, the greater is the induced emf. The longer the conductor which is cut or is doing the cutting, the greater the induced emf. Still another way of increasing the induced emf is to increase the number of conductors which are cut by the flux lines or which themselves cut the flux lines. The emf induced in each conductor is added to that induced in the others. This happens when a conductor is formed into a solenoid.

Pieces of iron ore called magnetite were found to possess magnetic properties which attracted bits of iron. These natural magnets were called lodestones.

Any magnet that retains its magnetism over a long period of time is a permanent magnet; if it loses its magnetism rapidly, it is a temporary magnet.

The space surrounding a magnet is called a magnetic field. Magnetic lines of force are concentrated around the poles of a magnet.

Pieces of iron or steel become magnetized by induction when they are brought close to, or in contact with, a magnet.

Every magnet has at least one north pole and one south pole.

Artificial magnets can be made by stroking an unmagnetized piece of iron or steel with a magnet, or by electrical means.

Residual magnetism is the amount of magnetism retained in a material after the magnetizing force has been removed.

Like magnetic poles repel each other; unlike magnetic poles attract each other.

A magnetic field surrounding a current-carrying conductor increases when the current increases and decreases when the current decreases.

Magnetic lines of force take the form of concentric circles around a current-carrying conductor.

A coil through which current is flowing has two polarity identities--voltage and magnetic.

The "left-hand rule" for a coil determines the relationship between the direction of current and the direction of magnetic flux through the coil.

Ampere-turns is the product of the current in amperes and the number of turns of the coil. Magnetomotive force is expressed in ampere-turns.

An electromagnet is composed of a coil or wire wound around a soft iron core.

The maxwell is the unit of magnetic flux; 1 line of force is equal to 1 maxwell.

When the current through the coil of an electromagnet is reversed, the polarity of the electromagnet is reversed.

Moving a magnetic field past a conductor will induce a voltage in the conductor.

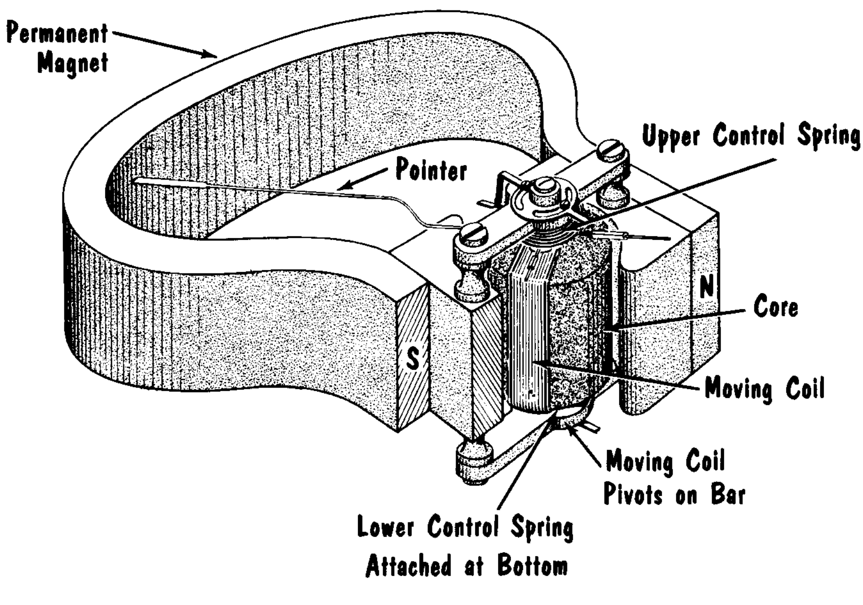
REVIEW QUESTIONS

1. State the "left-hand rule" for a coil.
2. What is an electromagnet?
3. What forms do the magnetic lines of force take around a current-carrying conductor?
4. What is the practical unit of magnetomotive force?
5. How does an increase or decrease in current through a conductor affect its magnetic field?
6. How do the magnetic fields of an electromagnet and a permanent magnet differ?
7. Define reluctance. Define permeability.
8. Give two methods for making artificial magnets.
9. How does a permanent magnet differ from a temporary magnet?
10. What is the maxwell? What is 1 maxwell equal to?
11. What happens in a conductor when a magnetic field is moved past it?

Electrical Measuring Devices: The D'Arsonval Meter

Electrical measuring instruments are used for checking the operation of electrical and radio equipment. The most commonly-measured electrical quantities are current, voltage, and resistance. The basic instrument used in d-c ammeters, voltmeters, and ohmmeters is the D'Arsonval permanent magnet, moving-coil d-c meter (this mechanism is also known as the Weston meter). Its principle of operation is like that of the d-c motor. Because of the torque or rotational force developed in the d-c moving coil is directly proportional to the current flowing through the coil, the scale graduations of d-c current meters are uniformly divided (linear).

COMPONENT PARTS OF THE BASIC D'ARSONVAL PERMANENT-MAGNET MOVING-COIL INSTRUMENT

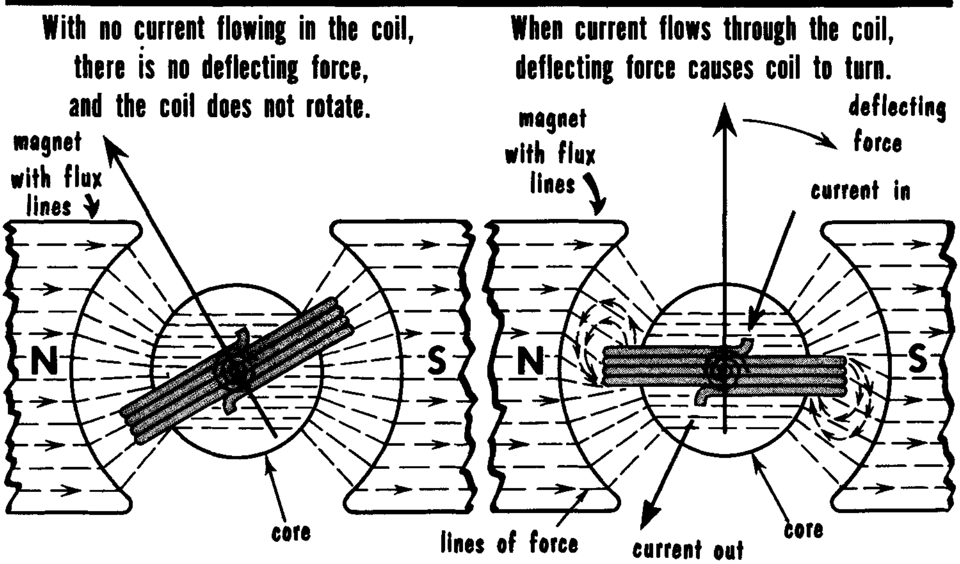


As to the physical makeup of the device, a current-carrying loop (the moving coil) is wound on an aluminum frame or bobbin mounted around a circular core of magnetic material. The entire coil assembly is positioned between the pole pieces of a permanent horseshoe magnet. The mounting pins of the assembly pivot on sapphire bearings. A top and bottom spring, mechanically attached to the coil assembly, conduct the meter current to and from the coil; load the coil and prevent free spinning when magnetic reactions develop the turning force; and supply the restoring forces which return the coil to its normal position when meter current ceases flowing. A pointer attached to the coil swings across the meter scale as the coil turns between the pole pieces. Currents induced in the aluminum bobbin "damp" the moving coil and prevent it from moving back and forth around the point of deflection.

Operation of the Moving-Coil D-C Meter As a Current Meter

Although the moving-coil d-c meter is used to measure voltage and resistance, as well as current, it is basically a current-operated device. The current in the moving coil develops the magnetic field that reacts with the stationary magnetic field of the permanent magnet (produces rotary motion of the coil) and moves the meter pointer across the scale of the instrument. The direction of rotation of the coil is from the strengthened portion to the weakened portion of the resultant magnetic field, just as in the d-c motor. Depending on the particular design of a current meter, a certain maximum current in the coil causes maximum turning force, or full-scale deflection. Exceeding the maximum current rating "slams" the pointer off scale against the end stop. For example, passing 2 amperes through a 0.5 ampere (500 milliamperes) meter movement can easily damage the meter by burning out the coil or by bending the thin pointer. Less than full-scale current flowing in the coil produces a proportionately reduced deflection. Thus, 50 milliamperes of current flowing through a 100-milliamperes full-scale meter causes half-scale deflection.

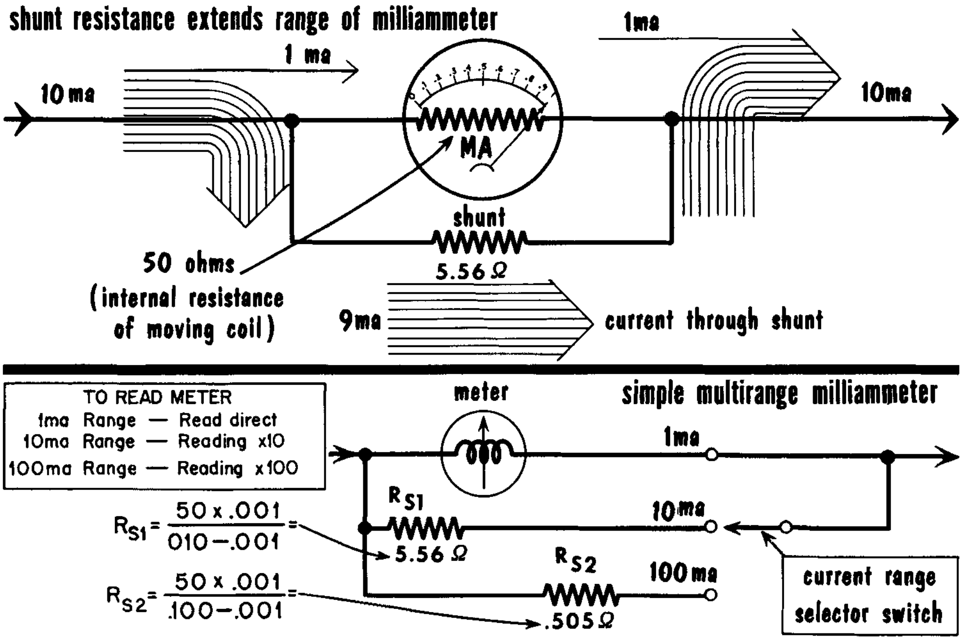
OPERATING PRINCIPLE OF A PERMANENT MAGNET, MOVING-COIL INSTRUMENT



Some current meter designs incorporate a moving coil in which as little as 10 microamperes of current causes full-scale deflection; other coils may require 50 microamperes of current and still others may be designed for 1 milliampere or higher. How can these meters be used to measure more current than can safely flow through the moving coil? The problem is solved by using a parallel current path, or shunt, around the coil, either inside or outside the meter. This arrangement is explained on the next page.

How to Calculate The Resistance of a Shunt

Although it is not conventional practice for users of radio test equipment to build shunts for current meters, it is still valuable to know how the resistance of a meter shunt is calculated. The general philosophy of the meter shunt is as follows: if a meter is designed to indicate the flow of 1 milliampere (ma) of current at full-scale deflection and it is desired to measure 10 ma, the ohmic value of the shunt must be such(relative to the ohmic value of the moving coil in the meter) that 9/10 of the total current (9 ma) passes through the shunt and 1/10 of the current (1 ma) passes through the meter.



Assume that we wish to use a 0-1-ma meter (milliameter) having a moving-coil resistance of 50 ohms in a circuit that may have as much as 10 ma flowing through it. To use the 1-ma meter to measure up to 10 ma, we must add a shunt to the meter movement, connecting the shunt directly to the outside terminals of the meter. To calculate the shunt, the following formula is used:

$$R_s = \frac{R_m \times I_m}{I_t - I_m}$$

where R_s is the resistance of the shunt in ohms, R_m is the resistance of the meter movement in ohms, I_m is the full-scale current rating of the meter movement, and I_t is the total current in amperes to be carried by the meter and shunt. Substituting the values, we get:

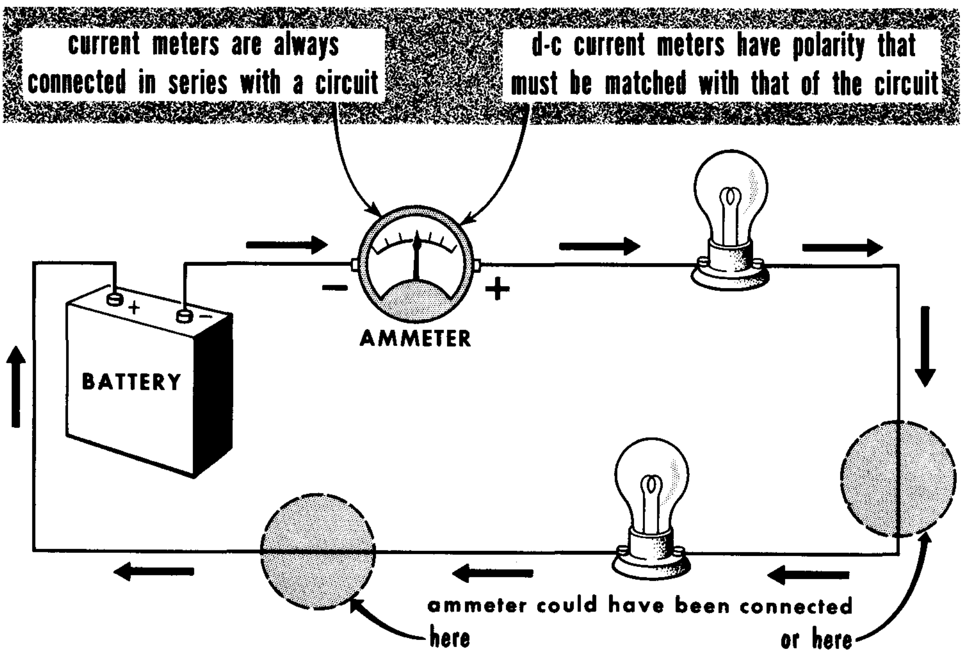
$$R_s = \frac{50 \times .001}{.010 - .001} = \frac{.05}{.009} = 5.56 \text{ ohms}$$

The construction of accurate shunts can be done using precise resistance-measuring devices. To calculate other shunts, substitute the new values in the above formula.

Using the Current Meter

Reference to d-c current meters includes ammeters, milliammeters, and microammeters. Meters suitable for measuring full-scale currents up to 500 microamperes (μa) are considered to be microammeters; from 500 μa (0.5 ma) up to 500 ma full scale are milliammeters; and from 500 ma (0.5 ampere) or more full scale are ammeters.

The first rule for proper use of current meters is that they must be connected in series with the circuit where the current is to be measured. The current must flow through the meter. Usually, this means "breaking" into the circuit to insert the instrument. However, some circuits provide jacks for insertion of current meters where measurements are to be made. As a safety precaution, always disconnect the voltage source when inserting a current meter into the circuit.

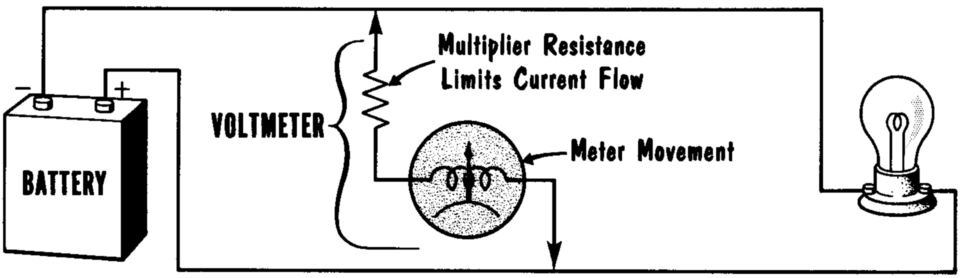


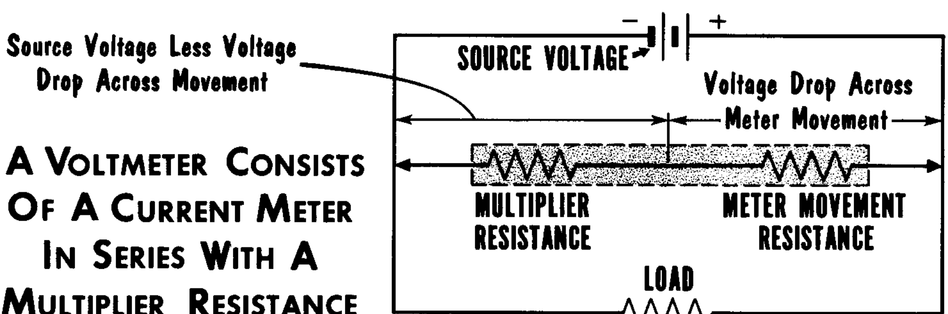
The second rule for proper use is that the polarity of the meter be correct. Using the polarity of the voltage source as a reference, the current must flow into the meter at its negative terminal and out of the meter at its positive terminal. If the polarity is wrong, the meter pointer will move in the opposite direction and possibly damage the meter.

The third rule is that whenever uncertain as to how much current is in a circuit, always start at the highest current range of a meter. If the pointer barely moves away from the zero point, turn the meter range switch to the next lower current range. Continue doing this until there is sufficient deflection to obtain an accurate reading.

The D-C Moving-Coil Voltmeter

The d-c moving-coil current meter is the basis of the d-c moving-coil voltmeter. A d-c current meter can be used as a d-c voltmeter in the following way: Assume that a current meter has a moving coil rated at 100 μa (.0001 ampere) full-scale current and a d-c resistance of 1000 ohms. Full-scale coil current means, therefore, a voltage drop (E = IR) of .0001 × 1000 = 0.1 volt across the moving coil of this instrument. In fact, every current meter bears a fixed, internal voltage-drop rating equal to the product of the full-scale coil current and the d-c resistance of the coil. Thus, although the ability to measure voltage is inherent in every current meter, its usefulness is limited because its maximum range and resistance are both very low.



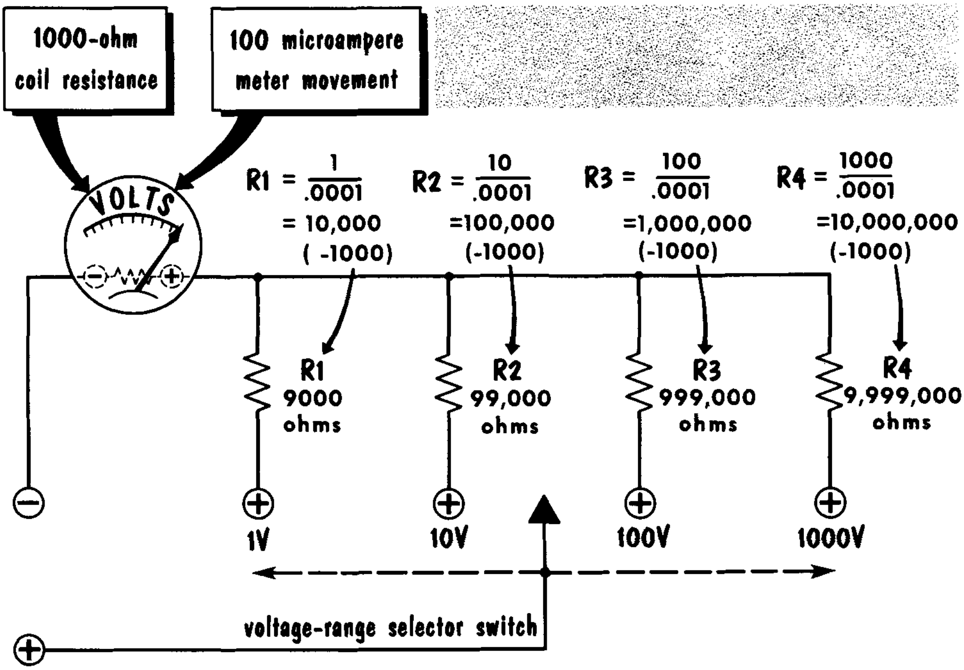


A VOLTMETER CONSISTS OF A CURRENT METER IN SERIES WITH A MULTIPLIER RESISTANCE

How is a current meter used as a voltmeter to measure voltages in excess of the internal voltage drop across the moving coil? By making the current meter part of a circuit in which a multiplier resistance is placed in series with the meter coil! The ohmic value of the multiplier resistance is such that when added to the meter coil resistance, the total resistance limits the circuit current to the full-scale current rating of the meter for any given applied voltage. In this way, the applied voltage divides between the multiplier resistance and the meter resistance in direct proportion to their respective resistances; the voltage drop across the meter coil never exceeds the internal voltage rating for full-scale deflection. The function of the multiplier resistance is to develop a voltage drop equal to the excess between the applied voltage (the voltage being measured) and the internal voltage drop across the meter coil. This proportioning of voltage drops becomes automatic when the multiplier resistance limits the circuit current as described.

Calculating the Multiplier

Let us assume that we have a 100-microampere ammeter (microammeter) having a d-c resistance of 1000 ohms. We want to convert this meter movement into a voltmeter that reads 25 volts full scale. When the 100-μa full-scale current flows through the meter, there is a voltage drop of 0.1 volt across the meter coil. Our task then is to calculate what resistance must be placed in series with the meter movement so that when a current of 100 microamperes flows through the meter and resistance, there will be a 25-volt drop across the meter resistance and multiplier combination. We can find the total resistance of the combination at 25 volts by using Ohm's law ($R = E/I = 25/.0001$, or 250,000 ohms). Of course, we must not lose sight of the internal meter resistance of 1000 ohms. Thus, the series multiplier resistance is equal to 250,000 ohms less the meter resistance of 1000 ohms, or 249,000 ohms.



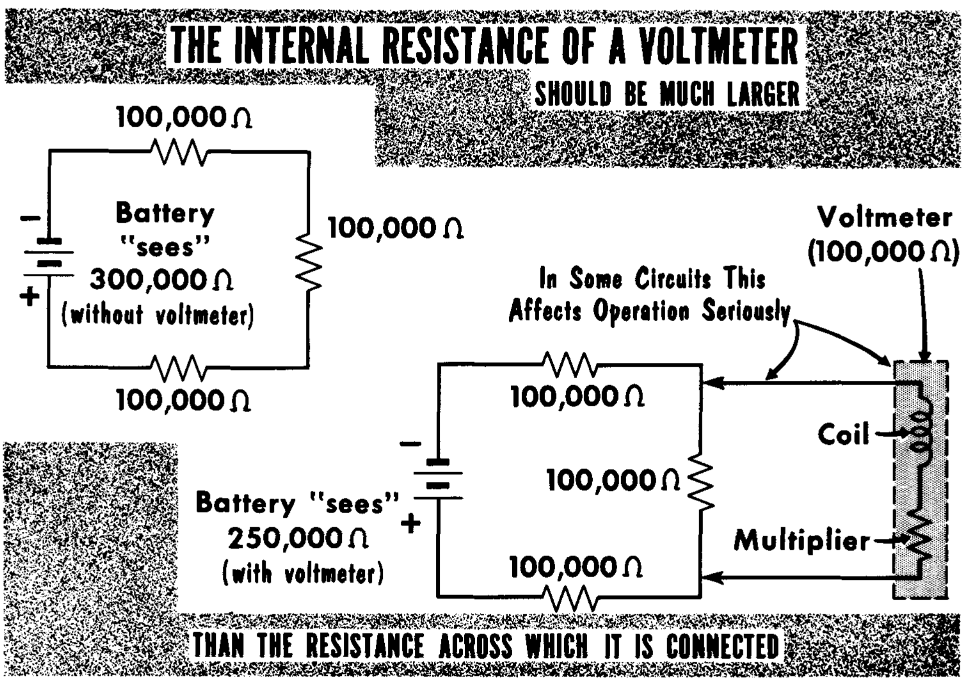
MULTIRANGE VOLTMETER USES SERIES MULTIPLIERS

By using a number of multiplier resistances that can be connected in series with the meter coil by a selector switch, a series of voltage ranges can be provided. Each range is suitable for measuring all d-c voltages between zero and the maximum value. As we will see, any current meter can be converted into a voltmeter by suitable selection of a series multiplier; however, the most desirable current meters are those having high sensitivity (1 ma or less for full-scale deflection.)

D-C Voltmeters—Ohms-Per-Volt Rating

An important electrical characteristic of voltmeters is their ohms-per-volt rating. This term is the basis for the total resistance of the voltmeter on each of its voltage ranges (including the multiplier resistance used on each range). The ohms-per-volt rating is based on the current requirement of the meter coil for full-scale deflection. If the meter coil in a voltmeter requires 50 μa (.00005 ampere) for full-scale deflection, the multiplier resistance required for each volt indication is R = E/I, or 1/.00005 = 20,000 ohms. The instrument therefore has a rating of 20,000 ohms-per-volt on each range.

The total resistance of the instrument on each range, then, is the ohms-per-volt rating multiplied by the full-scale limit on each range. For example, if the range selected is 5 volts full scale, the total resistance presented by the meter is 20,000 × 5, or 100,000 ohms. A meter that requires 1 ma (.001 ampere) for full-scale deflection has an ohms-per-volt rating of 1/.001 or 1000. If the voltmeter movement uses a 10-μa moving coil, the instrument has a rating of 1/.00001 or 100,000 ohms per volt.

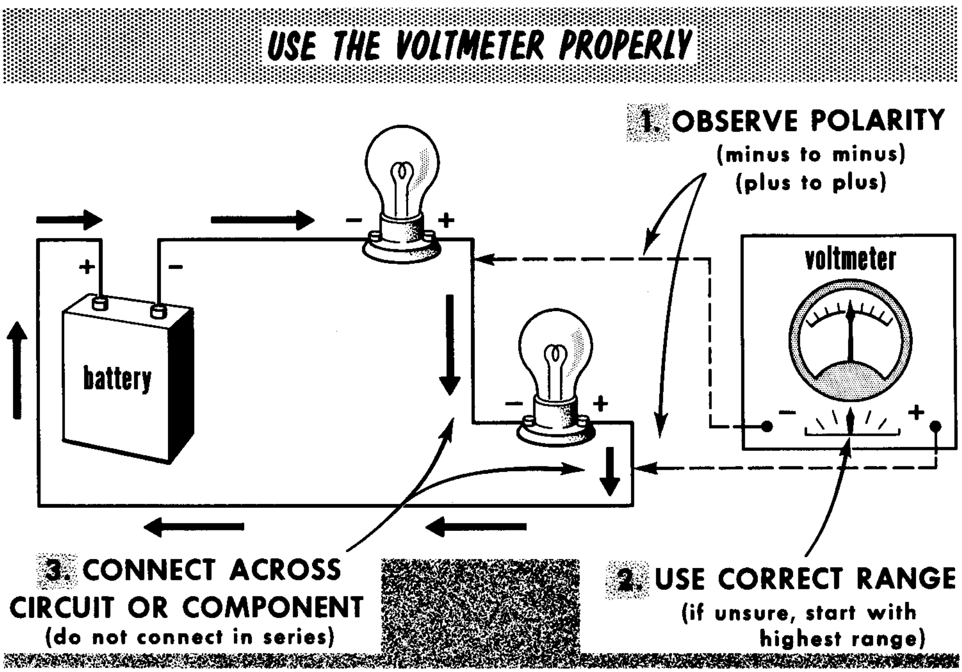


Voltmeters are connected across circuits; hence, the resistance of the voltmeter is in parallel with the resistance of the circuit being measured. To minimize the shunting effect of the voltmeter resistance across the circuit resistance, it is essential that the voltmeter resistance be as high as possible. It is therefore best to use the highest range possible consistent with readability.

Using the Voltmeter Properly

There are three general rules to keep in mind when using the voltmeter. The first rule concerns polarity. The d-c moving coil movement must be connected so that current flows through the coil in the proper direction. If current passes in the reverse direction, the pointer will deflect backwards and the meter possibly will be damaged. All d-c meters have positive and negative terminals, and connection to a circuit should be made in accordance with the polarities of the voltages or voltage drops being measured.

The second rule deals with the selection of the correct voltage range. If the approximate voltage is known, use a range that allows a safe tolerance. If the voltage to be measured is unknown and unpredictable, start with the highest range the voltmeter permits. Then, move to lower ranges until one is reached that permits an accurate reading.

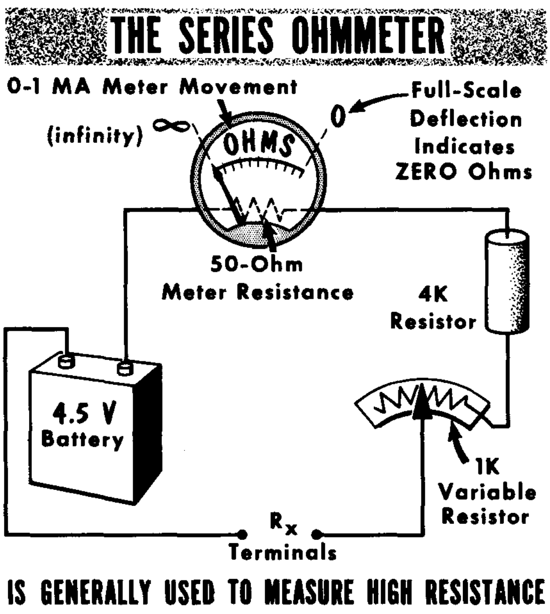


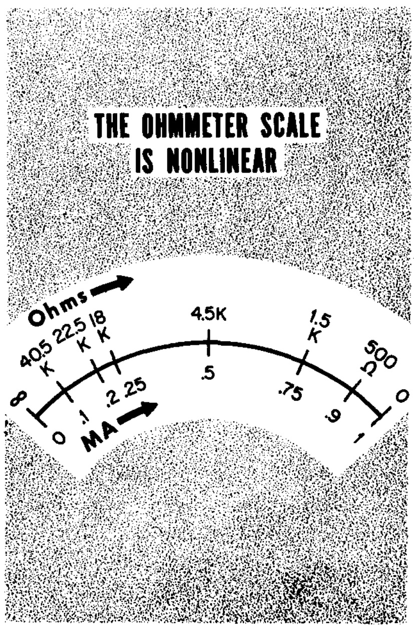
The third rule involves the actual connection of the voltmeter. Since a voltmeter is a comparatively high-resistance instrument, it is always connected across, or in parallel, with the load or voltage source. Since the voltmeter shunts the component which it measures, the voltmeter resistance should be as high as possible so as not to affect circuit operation. The highest range that provides a reliable reading should always be used. On a 20,000-ohms-per-volt meter, a 10-volt range presents a resistance of 200,000 ohms. If the actual reading is about 8 or 9 volts, then the 50-volt range, which presents a shunt resistance of 1 megohm, can be used. This would not affect the operation of a circuit unless its resistance was in excess of 100K ohms.

The Series Ohmmeter

The same moving-coil meter that we used in our ammeter and voltmeter can be used to make an ohmmeter, an instrument for measuring resistance. In building a basic ohmmeter circuit, we start once again with a 0-1 ma meter movement connected in series with a fixed resistor (4K), a variable resistor (1K), and a 4.5-volt battery (see illustration). This entire circuit ends in a pair of terminals across which the unknown resistance (R_X) is to be placed.

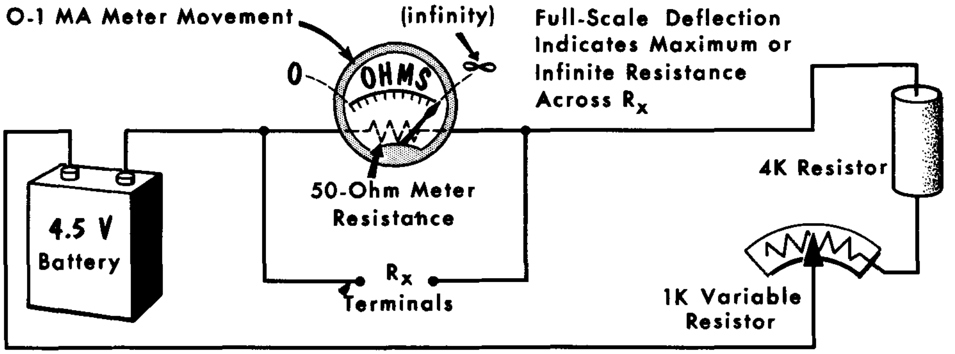
When the unknown or external resistance is zero, or the output terminals are short circuited, we want enough current to flow for full-scale meter deflection. By Ohm's law, 1 ma will flow when the total series circuit resistance is: $R = 4.5/.001$, or 4500 ohms. Since it is possible for the voltage of a new battery to be slightly higher than 4.5 volts, the total series resistance should be about 5000 ohms to be safe. Having a fixed resistance of 4000 ohms and a variable resistance of 1000 ohms permits "zero-adjust" to be made on the ohmmeter. As the battery voltage lowers with age, the potentiometer is varied to give full-scale deflection when R_X is zero.





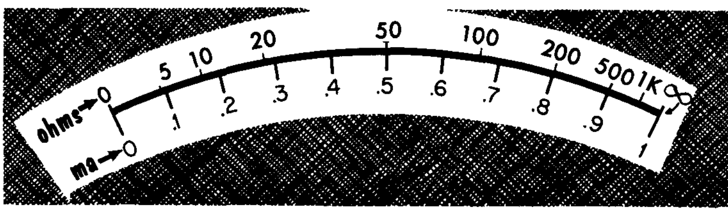
From this, we see that in the series ohmmeter, full-scale deflection is equal to zero ohms at the input terminals. With R_X "open", this is equivalent to an infinitely-high resistance. Thus, the left side of the meter scale reads "infinity", or some very high value of resistance. We see then that the ohmmeter scale reads opposite in direction to the ammeter and voltmeter scales. In addition, it is not a linear scale. If R_X is 1500 ohms, circuit current drops .25 ma, to .75 ma ($I = 4.5/6000$). However, the addition of another 1500 ohms across R_X produces a current of $I = 4.5/7500$, or .60 ma, a drop of an additional .15 ma. This produces the nonlinear scale shown.

The Shunt Ohmmeter



THE SHUNT OHMMETER is used generally to measure low resistance

For reading very low values of resistance, the shunt-type ohmmeter is better suited than the series type. In the shunt unit, the unknown resistance R_X is now shunted or placed in parallel with the meter, instead of in series with it. With the unknown resistance connected in this manner, some of the current in the ohmmeter circuit now takes the path through R_X. The current through the meter movement is reduced accordingly, and the amount of deflection drops in proportion to the reduction in current. The amount of current through the meter depends upon the ratio of the shunt resistance of R_X to the internal resistance of the meter.



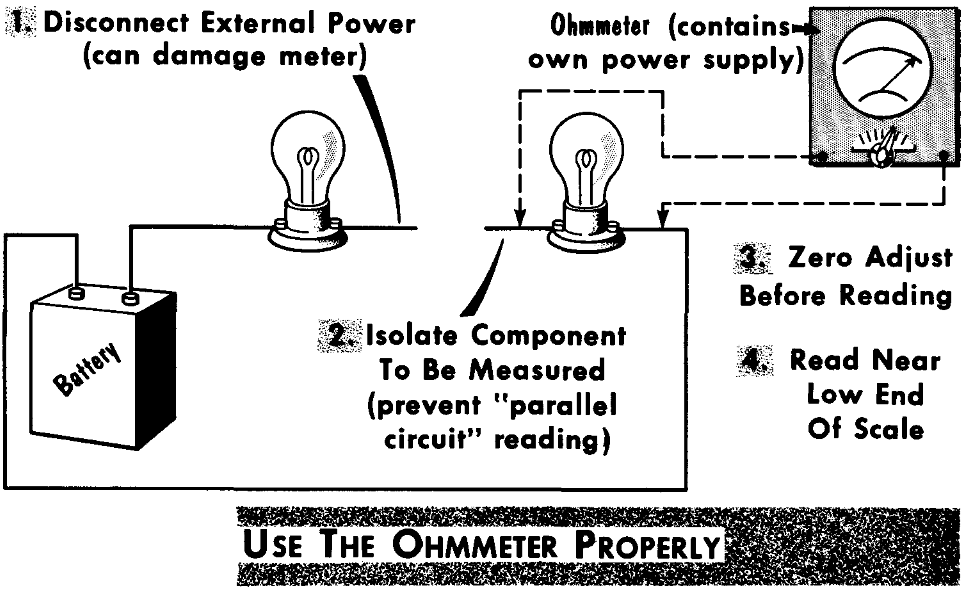
typical shunt ohmmeter scale

The zero-adjust in the shunt ohmmeter operates oppositely from that in the series ohmmeter. In the shunt instrument, the variable resistor is adjusted so that full-scale current flows through the moving coil when there is an open circuit across the R_X terminals. Thus, maximum or infinite resistance is indicated on the right-hand side of the meter scale. A short circuit across terminals R_X would bypass all current around the meter and produce no pointer deflection. Thus, zero ohms would appear at the left-hand side of the meter scale. Any resistance connected across the R_X terminals will provide a path for current, and will cause the current through the meter to be less than full scale. Low resistances will bypass considerable current and produce little deflection; high resistances will bypass little current and produce large deflection. By using a selector switch and various-sized shunts, a multirange ohmmeter can be made, with each range having a different multiplying factor.

Using the Ohmmeter

There are four basic rules to keep in mind when using the ohmmeter for measuring resistance. First, we must remember that an ohmmeter carries its own power supply. That is, the power necessary to deflect the pointer comes from the battery within the ohmmeter, and no external power is necessary. As a matter of fact, any external voltage in the circuit being measured will not only cause an erroneous reading, but may damage the moving coil and pointer. Thus, when measuring the resistance of a component or circuit, remove the voltage source from the external circuit.

Second, when measuring the resistance of a component, the component must be "isolated" from the rest of the circuit. This can be done simply by disconnecting one end of the component and letting it hang free. By measuring its resistance while connected in the circuit, you may actually be measuring the combined resistance of the component and that of some shunt component.



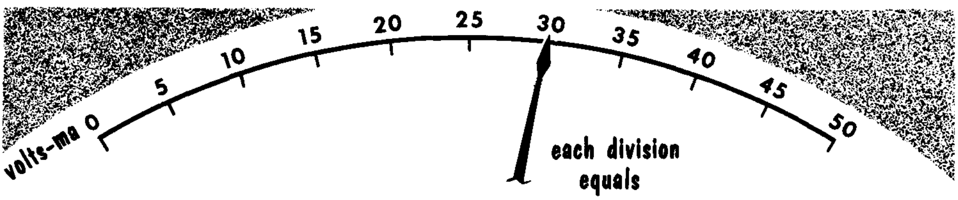
Third, make the zero adjustment before making a reading. If the ohmmeter is not properly adjusted to read zero ohms accurately, all other readings made will be incorrect.

Fourth, always make a resistance reading near the low end of the resistance range. At this point, the resistance values on the meter scale are spread out and a more accurate reading can be made. Avoid reading at the high end; here, the resistance values are crowded and the slightest inaccuracy of the ohmmeter can cause a considerable error in a resistance reading. If a particular resistance reads near the high end of a resistance range, switch to the next higher range; this will cause the pointer to indicate at a much lower place on the meter scale.

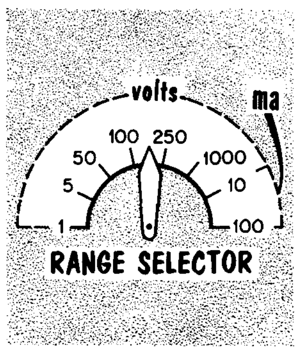
D-C Meter Scales

Meter scales used for d-c current and voltage meters have uniformly-spaced (linear) calibration markings. Single-range meters have a single scale showing only one set of values. This is so either for current or voltage. Multirange meters may have a single scale on which several sets of values are marked, with each set serving a different range, or separate scales for each range covered by the instrument. When a meter is designed to be used for many ranges, it is equipped with a range-selector switch. Each position on the switch indicates the highest value of the range of measurement. All ranges begin with zero.

The d-c meter used for resistance measurement has non-uniformly spaced (nonlinear) calibration markings. Although only one set of numerals is associated with a scale, a resistance range-selector switch is always part of the instrument. Each switch position bears a multiplying factor; i. e., X1, X10, X100, etc. The proper reading on that particular range is equal to the meter scale reading multiplied by the multiplying factor on the range switch.



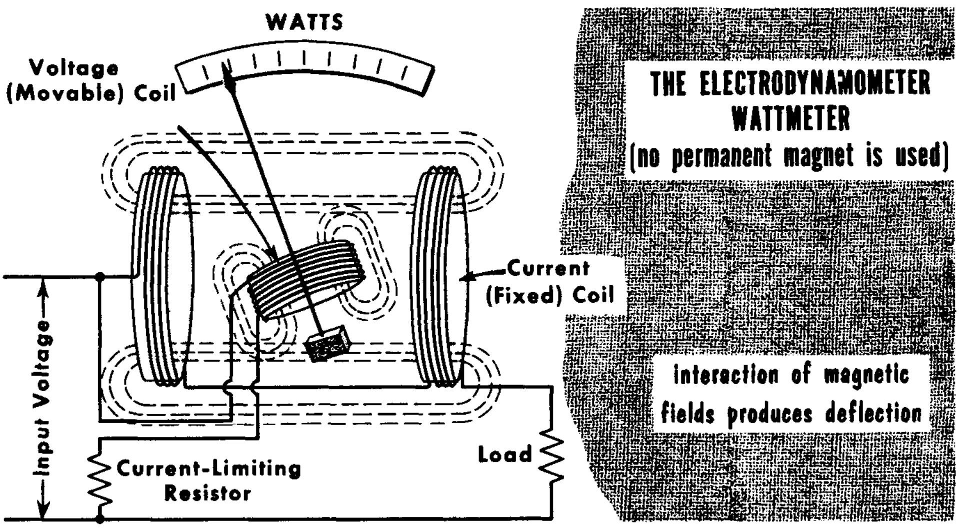
			each division equals
ON	1-volt range meter reads	.6 volt	$\frac{1}{50}$ =.02 V
ON	5-volt range meter reads	3 volts	$\frac{5}{50}$ =.1 V
ON	50-volt range meter reads	30 volts	$\frac{50}{50}$ =1 V
ON	100-volt range meter reads	60 volts	$\frac{100}{50}$ =2 V
ON	250-volt range meter reads	150 volts	$\frac{250}{50}$ =5 V
ON	1000-volt range meter reads	600 volts	$\frac{1000}{50}$ =20 V
ON	10-ma range meter reads	6 ma	$\frac{10}{50}$ =.2 ma
ON	100-ma range meter reads	60 ma	$\frac{100}{50}$ =2 ma



The amount of change between any two adjacent divisions on any one range is equal to the maximum value of the range divided by the number of divisions. If a voltmeter scale consists of 50 divisions and full-scale deflection is 250 volts, each division then corresponds to 250/50 or 5 volts. If this same meter were indicating 25 ma full scale, each division would be equal to 25/50 or 0. 5 ma. If the scale consisted of 100 divisions, each division would represent 2. 5 volts or 0. 25 ma. Occasionally, the pointer comes to rest between scale markings. In this case, the user must determine the value of the scale markings on either side of the pointer, and then estimate the approximate value of the reading.

The Wattmeter

We have discussed thus far the various ways and means of measuring voltage, current, and resistance. Another important measurement is that of power. Since power is equal to the product of voltage and current (E × I), we can obtain this measurement by voltmeter and ammeter readings in a circuit. There is, however, a direct means for measuring power in watts--the electrodynamometer wattmeter. This wattmeter consists of two coils--one stationary and one movable. There are two ways of connecting a wattmeter. In one method, the stationary coil is the voltage coil and consists of many turns of small wire having a high resistance; the movable coil is the current coil and consists of a few turns of large wire having low resistance. The voltage (or high-resistance) coil is connected across the voltage source, or across the device whose power consumption is being measured. The current (or low-resistance) coil is connected in series with the load, and current through the load passes through the current coil.



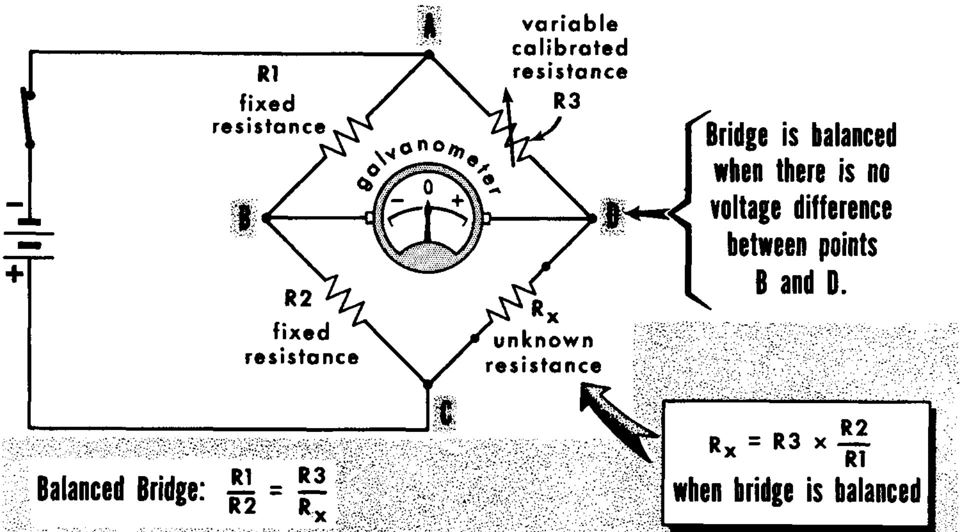
The second method of wattmeter connection has the high-resistance voltage coil as the movable coil, with the low-resistance current coil being stationary. This second method is superior for large currents because it removes the difficulty in conducting large currents into and out of the spring suspension of the moving coil.

Current through the voltage coil is proportional to the voltage across the load. The interaction of the magnetic fields from the fixed and movable coils causes the movable coil to rotate. The effect is almost the same as if the voltage applied across the load and the current through the load were multiplied together. The torque on the movable coil is proportional to the current and also to the voltage; it is therefore proportional to their product. The meter pointer thus registers according to the power consumption E × I. As we will learn later, wattmeters are used more in a-c circuits than in d-c circuits.

The Wheatstone Bridge

It is often necessary to make resistance measurements of greater accuracy than is possible with an ohmmeter. Measuring ammeter shunts, voltmeter multipliers, and other high-precision resistances can be done with an instrument known as the Wheatstone bridge. In the schematic of the bridge, resistances R1 and R2 are fixed and form one leg of the bridge. The other leg of the bridge consists of variable resistance R3 and the unknown resistance R_X. The applied battery voltage appears across both branches of this parallel circuit and produces current flow.

WHEATSTONE BRIDGE PROVIDES HIGH-ACCURACY RESISTANCE MEASUREMENTS



Let us assume that the battery voltage is 10 volts, R1 = 1 ohm, and R2 = 9 ohms. The current through the A-B-C leg would be 1 ampere, with point B being 1 volt positive with respect to point A. In leg A-D-C there will also be current flow. Let us assume the unknown resistance is 90 ohms, and the variable resistance is adjusted to 10 ohms. The current flow in this leg would then be 0.1 ampere or 100 ma ($I = 10/100$). The 0.1 ampere flowing through R3 would cause a voltage drop of 1 volt, with a 9-volt drop across R_X. Thus, point D would also be 1 volt positive with respect to point A. Since points B and D are both 1 volt positive with respect to point A, they are at the same potential, and no current will flow through the galvanometer. However, should the ratio of R1 to R2 differ from that of R3 to R_X, points B and D would not be at the same potential, and current would flow from B to D or D to B, depending on which point was at the higher potential.

R3 is usually a calibrated resistor which is varied until the galvanometer reads zero, indicative of no difference of potential across it. It is important to remember that R1 and R2 need have only a known ratio. Commercial bridges often have plug-in units or means of changing the R1-R2 ratio in terms of 1:1, 10:1, 100:1, 1000:1, etc.

The basic d-c meter is the permanent-magnet, moving-coil type (called the D'Arsonval movement.

The moving coil d-c meter is used to measure voltage, current, and resistance.

Voltmeters are always connected across a circuit being measured; ammeters are always connected in series with a circuit being measured.

D-c current meters include ammeters, milliammeters, and microammeters.

Correct polarity must be observed when measuring current to avoid damaging the meter.

Always begin at the highest range of a current meter whenever in doubt as to how much current is in a circuit.

A current meter can be used as a voltmeter to measure voltages exceeding the internal voltage drop across the moving coil by placing a multiplier resistance in series with the moving coil.

Current meters can be used to measure more current than can safely flow through their moving coils by using either an external or an internal shunt around the coil.

When using a voltmeter, always select the highest range which provides a reliable reading.

When using an ohmmeter, always make a resistance reading near the low end of the resistance range. At this point, a more accurate reading can be made.

Meter scales used for d-c current and voltage meters are linearly calibrated; the d-c meter scale used for resistance measurements is nonlinearly calibrated.

The zero adjust in the shunt ohmmeter operates oppositely from that in series ohmmeter.

The series ohmmeter is generally used to measure high resistance.

The shunt ohmmeter is used to measure very low values of resistance.

An electrodynamometer wattmeter measures power directly in watts.

A Wheatstone bridge is used to make very accurate resistance measurements.

REVIEW QUESTIONS

1. Why must the power be disconnected when an ohmmeter is placed in the circuit?
2. What is the basic meter movement used in d-c ammeters, voltmeters, and ohmmeters?
3. When is a meter shunt resistance used?
4. What formula is used to calculate the resistance of a meter shunt?
5. How should current meters be connected in a circuit where current is to be measured?
6. How is a current meter used as a voltmeter to measure voltage greater than the internal voltage drop across the moving coil?
7. What is meant by the ohms-per-volt rating of a meter?
8. How are voltmeters connected in a circuit?
9. What three rules should be remembered for correct use of a voltmeter?
10. How do series ohmmeters and shunt ohmmeters differ?
11. What type of wattmeter directly measures power in watts?
12. When is a Wheatstone bridge used?

AMERICAN WIRE GAGE TABLE (B & S)

FOR STANDARD ANNEALED BARE COPPER WIRE (at 68°F)

Gage	Diameter (inches)	Area (circ. mils)	Weight (lbs. per 1000 ft.)	Resistance (ft. per ohm)	Resistance (ohms per 1000 ft.)	Current Capacity (amps-rubber insul)
0000	.4600	211600.	640.5	20400.	.04901	225
000	.4096	167800.	507.9	16180.	.06180	175
00	.3648	133100.	402.8	12830.	.07793	150
0	.3249	105500.	319.5	10180.	.09827	125
1	.2893	83690.	253.3	8070.	.1239	100
2	.2576	66370.	200.9	6400.	.1563	90
3	.2294	52640.	159.3	5075.	.1970	80
4	.2043	41740.	126.4	4025.	.2485	70
5	.1819	33100.	100.2	3192.	.3133	55
6	.1620	26250.	79.46	2531.	.3951	50
7	.1443	20820.	63.02	2007.	.4982	
8	.1285	16510.	49.98	1592.	.6282	35
9	.1144	13090.	39.63	1262.	.7921	
10	.1019	10380.	31.43	1001.	.9989	25
11	.09074	8234.	24.92	794.	1.260	
12	.08081	6530.	19.77	629.6	1.588	20
13	.07196	5178.	15.68	499.3	2.003	
14	.06408	4107.	12.43	396.0	2.525	15
15	.05707	3257.	9.858	314.0	3.184	
16	.05082	2583.	7.818	249.0	4.016	6
17	.04526	2048.	6.200	197.5	5.064	
18	.04030	1624.	4.917	156.5	6.385	3
19	.03589	1288.	3.899	124.2	8.051	
20	.03196	1022.	3.092	98.5	10.15	
21	.02846	810.1	2.452	78.11	12.80	
22	.02535	642.4	1.945	61.95	16.14	
23	.02257	509.5	1.542	49.13	20.36	
24	.02010	404.0	1.223	38.96	25.67	
25	.01790	320.4	.9699	30.90	32.37	
26	.01594	254.1	.7692	24.50	40.81	
27	.01420	201.5	.6100	19.43	51.47	
28	.01264	159.8	.4837	15.41	64.90	
29	.01126	126.7	.3836	12.22	81.83	
30	.01003	100.5	.3042	9.691	103.2	
31	.008928	79.7	.2413	7.685	130.1	
32	.007950	63.21	.1913	6.095	164.1	
33	.007080	50.13	.1517	4.833	206.9	
34	.006305	39.75	.1203	3.833	260.9	
35	.005615	31.52	.09542	3.040	329.0	
36	.005000	25.00	.07568	2.411	414.8	
37	.004453	19.83	.06001	1.912	523.1	
38	.003965	15.72	.04759	1.516	659.6	
39	.003531	12.47	.03774	1.202	831.8	
40	.003145	9.888	.02993	0.9534	1049.	
41	.00280	7.8400	.02373	.7559	1323.	
42	.00249	6.2001	.01877	.5977	1673.	
43	.00222	4.9284	.01492	.4753	2104.	
44	.00197	3.8809	.01175	.3743	2672.	
45	.00176	3.0976	.00938	.2987	3348.	
46	.00157	2.4649	.00746	.2377	4207.	

GLOSSARY

Alnico Magnet—Permanent magnet consisting of aluminum, nickel, and cobalt.

Ampere-Hour—Unit of electricity equal to a current of one ampere flowing for a period of one hour.

Ampere-Turns—Product of the number of turns in a coil times the current in amperes flowing through the coil.

Atom—The smallest particle into which matter can be divided.

Ballast Resistor—Resistor whose resistance increases rapidly with increases in current flow through it.

Battery—Device for converting chemical energy into electrical energy.

Charged Bodies—Bodies with an excess or a deficiency of electrons.

Coulomb—The unit of quantity for electric current.

D'Arsonval Meter Movement—Most commonly used movement in precision instruments for making d-c measurements.

Difference of Potential—Voltage between two points.

Direct Current—Unidirectional electric current flowing in one direction through a circuit, and essentially constant in magnitude.

Discharge—In a storage battery, conversion of chemical energy into electrical energy.

Electrolyte—Chemical compound, either liquid or pastelike, whose chemical action causes a current flow, or in which a chemical action is caused by the flow of current.

Electromagnet—Core of magnetic material (such as soft iron) which becomes temporarily magnetized by an electric current passed through a coil of wire wound around the core.

Electromotive Force (emf)—Force which tends to alter the motion of electricity or to maintain its motion against resistance. Measured in volts.

Electron—Elementary negative charge that revolves around the nucleus of an atom; negatively-charged particles of matter.

Flux—Term used to designate all the magnetic lines of force in a region.

Flux Density—Number of magnetic lines of force passing through a given area.

Free Electrons—Electrons which are not bound to a particular atom but wander at random among the many atoms of a substance.

Fuse—Protective device used in an electric circuit containing a wire, bar, or strip of fusible metal. Wire melts and breaks when current through it exceeds the rated value of the fuse.

Induced Voltage—Voltage produced by a change in the number of magnetic lines of force passing through a coil in a circuit.

Insulator—Device having high electric resistance used for separating conductors to prevent undesired flow of current from the conductors to other objects.

Ion—Electrified particle which is formed when an atom or group of atoms loses or gains one or more electrons.

IR drop—Voltage drop produced across a resistance R by the flow of current I through the resistance.

I^2R Loss—Power loss in connecting wires and other parts of a circuit caused by current flow I through resistance R of the conductors.

Junction—Connection point between two or more conductors.

Kilowatt—Unit of electrical power equal to 1000 watts.

Law of Magnetism—Like poles repel; unlike poles attract.

Load—Device used to absorb power and convert it into a desired useful form.

Lodestone—A kind of iron ore (magnetite) which is a natural magnet.

Magnetic Flux—Lines of force generated by a magnet.

Magnetic Lines of Force—Imaginary lines used to designate the direction in which magnetic forces are acting as a result of a magnetomotive force.

Molecule—Smallest subdivision of a compound which still retains the chemical properties of that compound.

Ohm—Unit of electrical resistance.

Ohm's Law—Expresses the relationship existing in an electrical circuit between the voltage E, the current I, and the resistance R. $E = IR$; $I = E/R$; $R = E/I$.

Parallel Circuit—Two or more electrical devices connected so the line current divides between them.

Permeability—Measure of how much better than air a given material is as a path for magnetic lines of force.

Polarity—Condition in an electrical circuit whereby the direction of current flow can be determined.

Potential—Difference in voltage between two points of a circuit. Expressed in volts.

Potentiometer—A variable voltage divider.

Power—Rate of doing work or expending energy. In d-c circuits, multiplying volts by amperes gives power in watts. $P = EI$; $P = I^2R$; $P = E^2/R$.

Primary Cell—Cell designed to produce an electric current through electrochemical reaction. Cannot be recharged by electric current when completely discharged.

Proton—Positive particle in an atom. Smallest quantity of positive electricity that can exist in a free state.

Reluctance—Property of a magnetic circuit which determines the amount of magnetic flux that will be produced as a result of applying a given magnetomotive force.

Resistance—Opposition which a device or material offers to the flow of current.

Rheostat—Resistor whose value can be varied. Used to handle large currents.

Secondary Cell—Source of electrical energy where the cell may be recharged after being discharged by sending an electric current through it opposite in direction to the discharge current.

Series Circuit—Two or more electrical devices connected so that the total current must flow through each of them in turn.

Shunt—Any part, or component, connected in parallel with some other part, or component.

Solenoid—Electromagnet having an energizing coil, cylindrical or tubular in form, acting on an armature positioned in the center of a coil. A coil used to produce a magnetic field.

Static Charge—Electric charge accumulated on an object.

Thermistor—Temperature-sensitive resistor whose resistance changes negatively as its temperature increases.

Variable Resistor—Resistor whose value can be changed mechanically.

Varistor—Voltage-sensitive resistor whose resistance changes as the voltage applied to it changes. Its resistance lowers when a high voltage is present.

Volt—Unit of electromotive force or electrical pressure.

Voltage Drop—A difference, or loss of voltage, between two points.

Watt—Practical unit of electrical power.

Watthour Meter—Meter that registers and measures electrical energy in watthours or kilowatthours.

Wheatstone Bridge—Null-type resistance-measuring circuit in which resistance is measured by direct comparision with a standard resistance.

INDEX TO VOL. I

(Note: A cumulative index covering all six volumes in this series is included at the end of Volume VI.)

basic radio

by **MARVIN TEPPER**

Electronic Services Division
Raytheon Company

Author of
FUNDAMENTALS OF RADIO TELEMETRY

VOL. 2

JOHN F. RIDER PUBLISHER, INC., NEW YORK

Copyright September 1961 by John F. Rider Publisher, Inc.

All rights reserved. This book or any parts thereof may not be reproduced in any form or in any language without permission of the publisher.

Library of Congress Catalog Card Number 61-11229

Printed in the United States of America

Fifth Printing, 1968

PREFACE

The purpose of this book is to fill a need for a text stating in plain, everyday language, what many people consider a complex technical subject. A technical subject need not be complex. Careful filling in of the background with essential information, and then leading step by step to the final explanation, provides a logical method of explaining the most difficult subject.

It would be impossible to cover in a single book or series of books, the immense scope implied in the word *electronics*. However, an understanding of radio circuits serves as a foundation for advanced study in all fields of electronics, such as television, radar, computers, etc. For teaching radio, the all-important basic tool of electronics, most available textbooks are woefully inadequate. One type contains information so brief as to acquaint rather than instruct. Another type is based on the premise that teaching a student to design a circuit is the best method of having him understand that circuit's operation.

Basic Radio represents the neglected middle ground. It is a course in radio communications, as distinct from a general course in electronics. The text deals with the circuitry and techniques used for the transmission and reception of intelligence via radio energy. Assuming no prior knowledge of electricity or electronics, the six volumes of this course "begin at the beginning" and carry the reader in logical steps through a study of electricity and electronics as required for a clear understanding of radio receivers and transmitters. Illustrations are used on every page to reinforce the highlights of that page. All examples given are based on actual or typical circuitry to make the course as practical and realistic as possible. Most important, the text provides a solid foundation upon which the reader can build his further, more advanced knowledge of electronics.

The sequence of *Basic Radio* first establishes a knowledge of d-c electricity. Upon this is built an understanding of the slightly more involved a-c electricity. Equipped with this information the reader is ready to study the operation of electron tubes and electron tube circuits, including power supplies, amplifiers, oscillators, etc. Having covered the components of electronic circuitry in Volumes 1 through 3, we assemble these components

in Volume 4, and develop the complete radio receiver, AM and FM. In Volume 5 we recognize the development of the transistor, and devote the entire volume to the theory and circuitry of transistor receivers and semiconductors. The last volume of the course, Volume 6, covers the long-neglected subject of transmitters, antennas, and transmission lines.

No prior knowledge of algebra, electricity, or any associated subject is required for the understanding of this series; it is self-contained. Embracing a vast amount of information, it cannot be read like a novel, skimming through for the high points. Each page contains a carefully selected thought or group of thoughts. Readers should take advantage of this, and study each individual page as a separate subject.

Whenever someone is presented with an award he gives thanks and acknowledgement to those "without whose help . . . " etc. It is no different here. The most patient, and long-suffering was my wife Celia, who typed, and typed, and typed. To her, the editorial staff of John F. Rider, and others in the "background", my heartfelt thanks and gratitude for their assistance and understanding patience.

MARVIN TEPPER

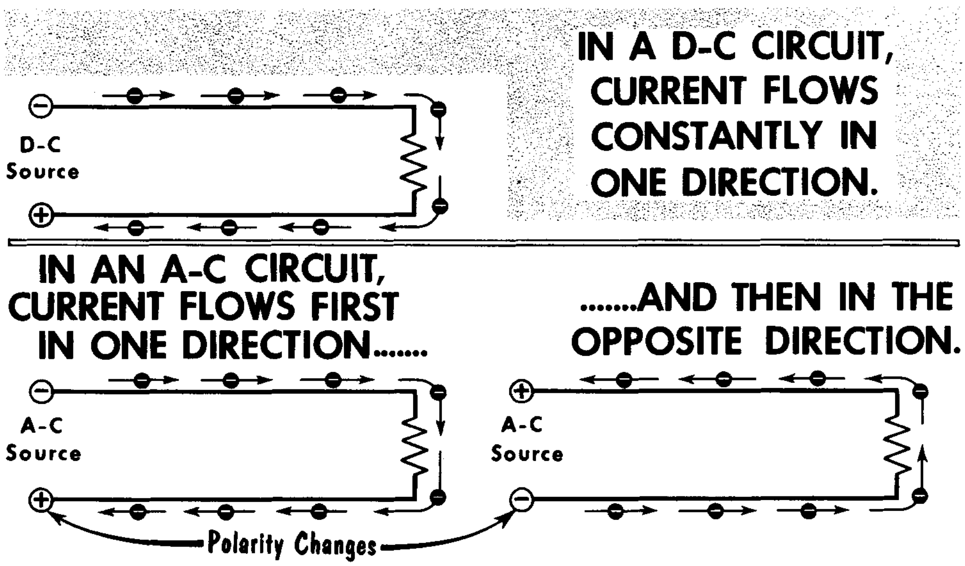
Malden, Mass.
September 1961

CONTENTS

VOL. II — BASIC RADIO

Introduction to AC

In Volume 1, we covered the subject of d-c electricity. In d-c circuits, the polarity of the voltage source remained constant, as did the difference of potential, or voltage. Under these conditions, electron flow was always in one direction, from minus to plus, and of constant quantity. In alternating current (a-c) electricity, we have a condition where the polarity of the voltage source is constantly changing. What was the positive terminal at one instant becomes the negative terminal some time later; what was the negative terminal at one instant becomes the positive terminal some time later. As a result of the constantly changing polarity of the voltage source, the direction of electron flow in the circuit also keeps reversing. In addition to reversing direction, current in an a-c circuit will also keep varying in quantity--from zero to maximum in one direction and back to zero, and from zero to maximum in the opposite direction and then back to zero. Thus, the alternating voltage will cause an alternating current.

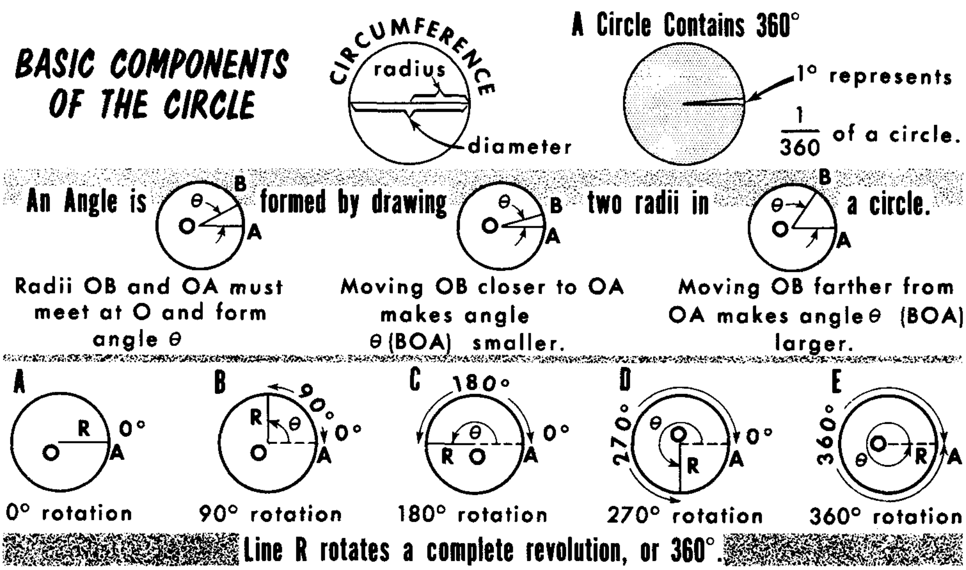


A-c electricity is not "better" than d-c; it is another type of electricity that has certain advantages. With a-c we can use transformers which enable us to transform a-c voltage to as high or as low a voltage as we wish. This permits efficient distribution of electrical power. In addition, there are many kinds of electrical components and devices that can do certain "jobs" in a-c circuits that cannot be done in a d-c circuit. An important point to remember is that a-c does not replace d-c.

A-c makes possible radio communication. However, most of the circuits in a-c communications equipment are controlled by d-c voltages. Because a-c involves constantly changing voltages and currents, we must give a little more thought to them. Thus, we begin our study of a-c with some fundamental mathematics which will help us in this study.

The Circle--Angular Rotation

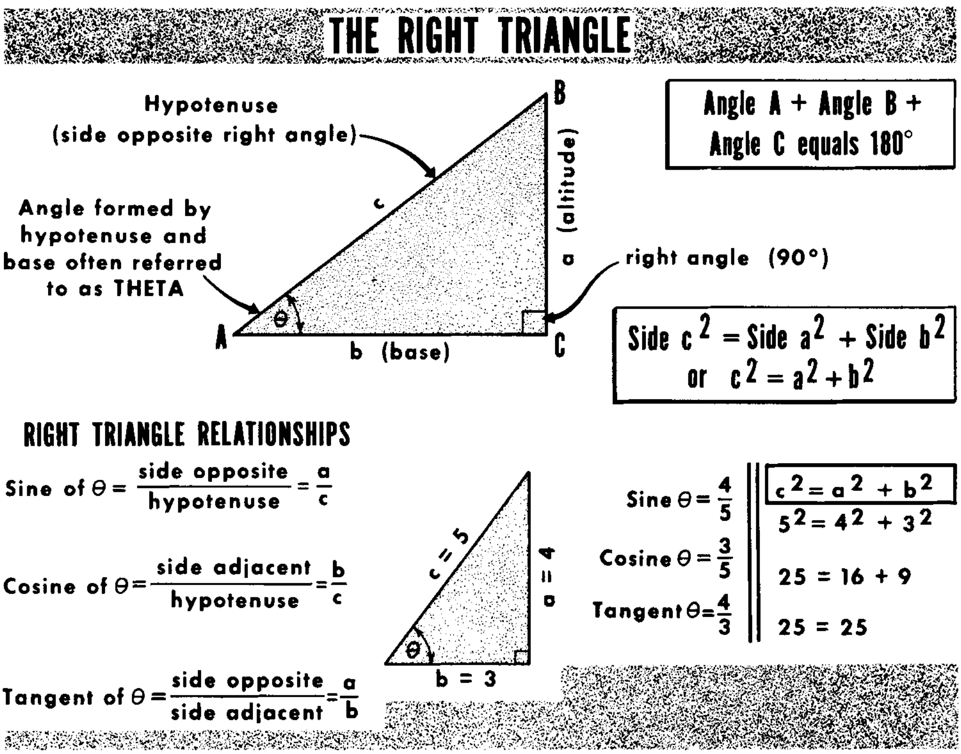
The circle is a simple figure, and yet it represents an important consideration in our study of a-c electricity. Let us review it briefly. The constantly curving line that forms the circle is called the circumference. If we draw a straight line from the center of a circle to any point on the circumference, that line is called a radius. Any line drawn through the center of the circle and dividing the circle in half is called the diameter. Looking further into the circle, we find that the circle is divided into degrees; let us see how they are formed.



From the center of the circle, we draw a radius to the circumference and call that line OA. We will keep this line in this position and use it as a reference line. We now draw a second radius, OB. The position of OB to OA forms an angle. We refer to this angle as angle BOA, with point O being the vertex or origin of lines OA and OB. If we move line OB closer to OA, the angle thus formed becomes smaller; if OB is moved farther away from OA, the angle becomes larger. If line OB is rotated farther and farther away from our reference line OA, it will spin past the entire circumference and end up overlapping line OA. The entire rotation of a radius from one point on the circumference, all around the circumference, and back to the starting point, covers 360°. Thus, if line R started at position OA and rotated until it pointed straight up on the page, it would have rotated 90°, and we would say that it formed a 90° angle with line OA. If we keep rotating the radius 90° more, line R would now form a diameter together with OA, and we call this a straight angle, or 180°. As line R now moves downward to the bottom of the page, it has gone through 90° more, or a total of 270°. Finally, we rotate it 90° more until it reaches its starting point. In one complete rotation, line OB has moved through 360°. The angle formed by the rotating line R with respect to line OA is given the name "theta," after the Greek letter (θ).

The Right Triangle

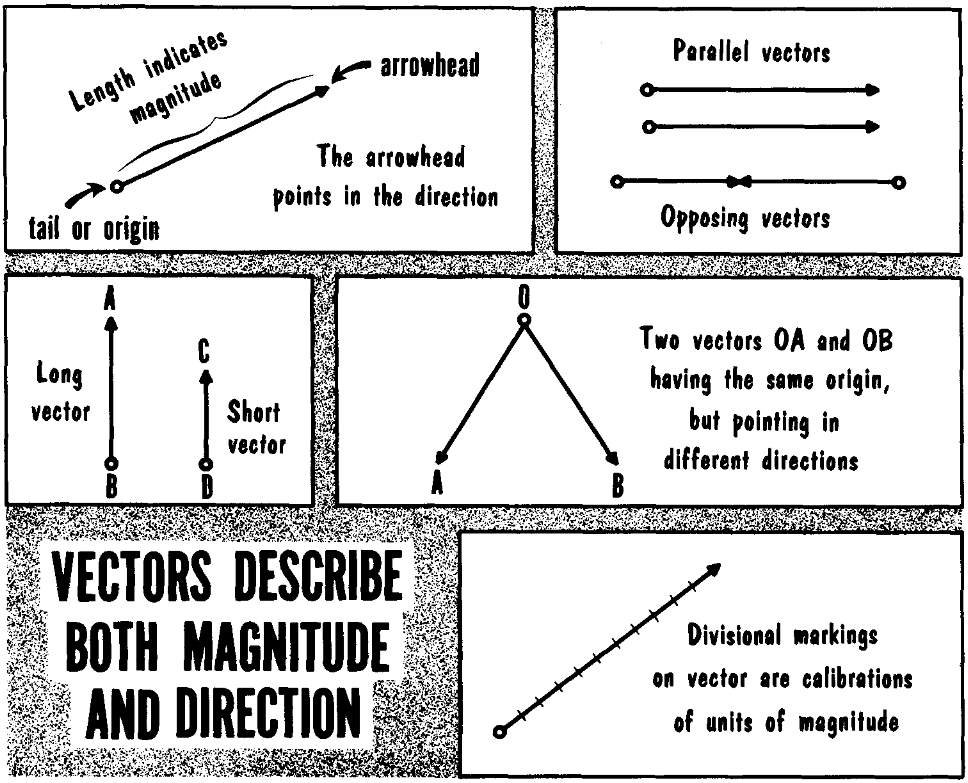
The right triangle is a special triangle in that one of its three angles is a right (90°) angle. The number of degrees included in the three angles of any triangle is 180. Thus, since one of the angles of a right triangle is equal to 90°, the sum of the remaining two angles must be 90°. In studying the right triangle, we assign a particular group of names to the various sides and angles. The side that lies horizontal to the page is called the base (b), the vertical side is called the altitude (a), and the side opposite the right angle is called the hypotenuse. The length of these sides has a particular relationship--the square of the length of the hypotenuse is equal to the sum of the squares of the lengths of the other two sides ($c^2 = a^2 + b^2$). Also, the length of the hypotenuse is greater than the length of either of the other sides but is less than their sum.



The angle formed by the base and hypotenuse is often referred to as the angle theta (θ). With regard to this angle, we will often refer to the side opposite angle θ, the side adjacent to angle θ, and the hypotenuse. These relationships are referred to as the sine, cosine, and tangent of angle θ. The sine of this angle is equal to the side opposite divided by the hypotenuse; the tangent of θ is equal to the side opposite divided by the side adjacent. In rotating from 0° to 90°, the sine of θ will vary from a value of 0 to 1; the cosine of θ will vary from 1 down to 0; and the tangent of θ will vary from 0 to infinity.

Vectors

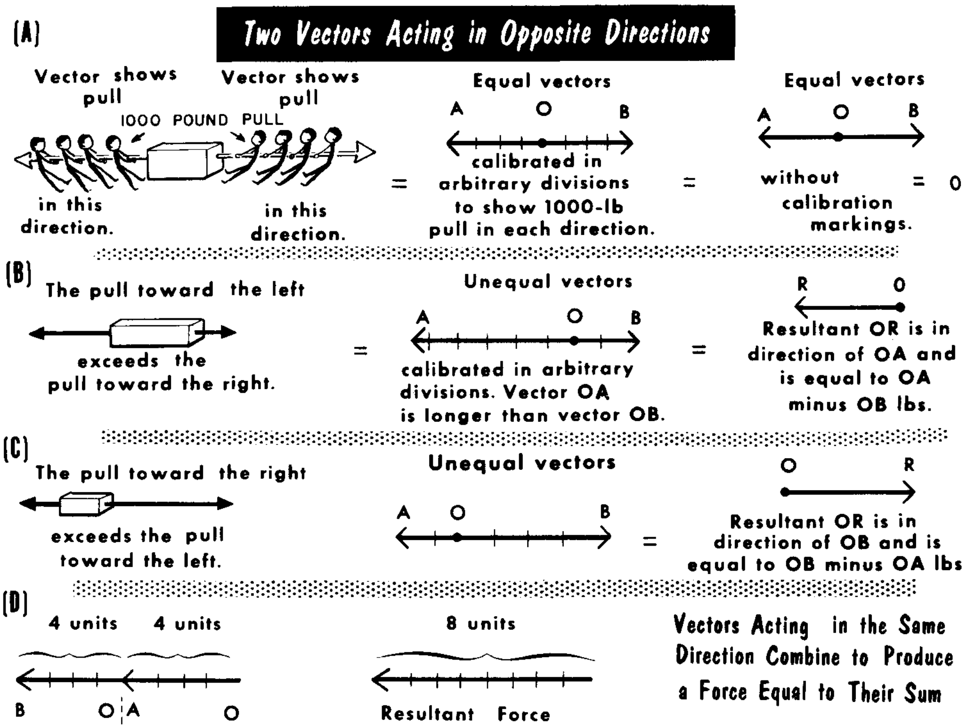
Some things or situations can be expressed by a single number, (e.g., the population of a town, the number of feet in a mile, or the number of chairs in a room). Anything which can be described fully by a single number is called a scalar quantity. There are, however, many situations or actions that cannot be described in this manner. For example, the movement of a jet plane. To say that it is flying at 600 miles an hour is not sufficient--the direction in which it is going also must be stated. If we desire to identify the force being applied to an object, it is not sufficient to say it has a force of 100 pounds. Is the force being applied upward or downward, to the right or to the left? Both magnitude (amount) and direction must be stated. Any situation or action that requires mention of both magnitude and direction to describe it is called a vector quantity. Alternating voltage and current and related phenomena are vector quantities.



A vector is a straight, or directed line with an arrowhead at one end. This is the head end; the other end is the origin or the tail of the vector. The length of the vector identified in any suitable units indicates the magnitude of the quantity, whereas the direction in which the arrowhead points is the direction of the quantity. Letters assigned to the head and to the tail of the vector readily identify the vectors, such as OA, OB, AB, CD, etc.

Multiple Vectors

Assume two like teams of men pulling on an object as in (A). One team exerts a pull of 1000 pounds in one direction while the other team is exerting a similar pull of 1000 pounds in the opposite direction. The situation can be illustrated by two vectors OA and OB of equal length (equal magnitude of force) directed in opposite directions. One pull cancels the other, hence, the net force acting on the object is zero.



If vector OB is shorter than vector OA as in (B), it means that the amount of force applied in one direction (vector OA) exceeds the amount of force applied in the other direction (vector OB). The net force is established by subtracting vector OB from OA. The resultant force (OR) acting on the object is in the direction of the greater force--vector OA. The same method is used to establish the resultant when the greater force is in the direction of OB as in (C). The resultant is vector OR in the direction of vector OB. The vector subtraction obviously is simple arithmetical subtraction.

Vectors representing forces acting in the same direction can be added to each other, as shown in (D). They are joined head to tail, as shown by the tail of OB being joined to the head of OA. The resultant is indicated by the sum of the lengths of OA and OB interpreted in whatever units express the magnitudes of OA and OB. Again, the vector addition is simple arithmetical addition.

The Parallelogram

Forces do not always counteract or aid each other completely. Sometimes, they act on the same object in directions which are at right angles to each other. This condition can be shown graphically by two vectors, OA and OB, having a common origin and forming a right angle. Four positions of the vectors are shown in (A) on page 2-7. Each of the vectors in the presentation is referred to as a component vector. Also, one of them is selected as the reference vector, usually the one which is positioned horizontally. Vector OB typifies this.

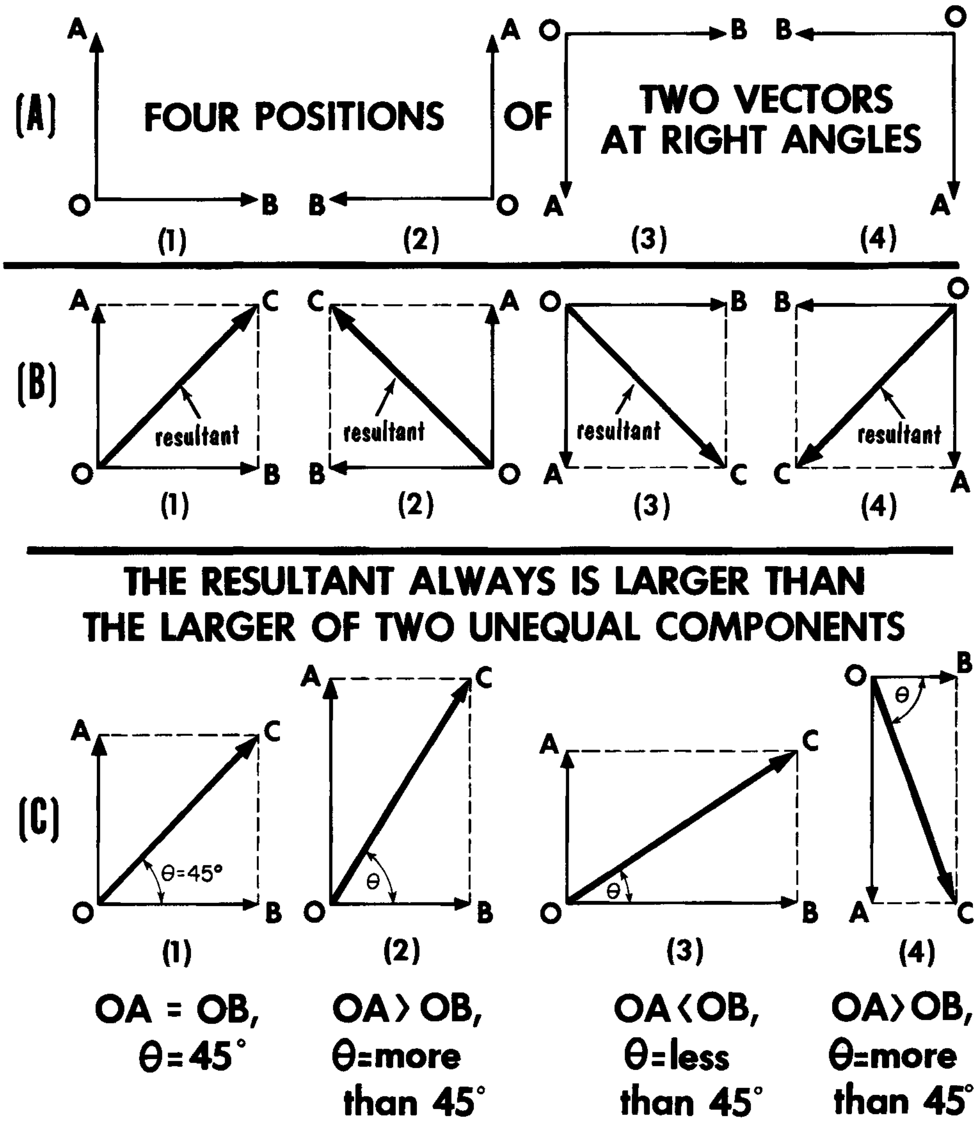
Two forces acting at right angles to each other produce a resultant force which also has magnitude and direction. It is established in a particular way known as the parallelogram method. The parallelogram is formed by using the vectors OA and OB as adjacent sides and adding two new sides, BC and AC, shown by the dotted lines in the figures shown in (B). Side BC is parallel and equal to side OA, and side AC is parallel and equal to side OB. The diagonal drawn between the origin (point O) of the component vectors and the opposite corner of the parallelogram is the resultant. If the resultant is calibrated in the same units that are used for the two component vectors OA and OB, the magnitude of the resultant can be interpreted directly from the length of the resultant OC.

As can be seen, the resultant has a direction of action which differs from that of the two component vectors OA and OB. The original right angle (90°) relationship between OA and the reference vector OB is modified, and the resultant now has an angular relationship COB relative to the direction of vector OB.

The magnitude of the resultant is a function of the relative magnitudes of the two components. When the two components have equal magnitudes (equal length), the resultant is the smallest possible, but even then it is greater than either of the components but not equal to their sum. When one component vector exceeds the other, the resultant is always greater than the larger of the two components. Examples are given in (C) on page 2-7.

As to the direction of the resultant OC, it too is a function of the relative magnitudes of the two component vectors. When the component vectors are of equal length, the direction of the resultant is mid-way between the directions of the individual components. These are 90° apart; the resultant is always at 45° see the first example in (C) . When the vertically directed component (in these examples it is OA) is greater in magnitude than the horizontally-directed component OB as shown in (C), examples 2 and 4 , the resultant has a direction which is closer to the vertically directed component, and the angle COB exceeds 45°. When the situation is reversed and the horizontally directed component vector OB is the larger of the two components (C), example 3 the resultant is directed more in the direction of the larger component. As can be seen, the angle COB is then less than 45°. As long as the two component vectors are present, the angle COB will never be 0°, nor will it ever be 90° -- its angle will always have some value in between these limits.

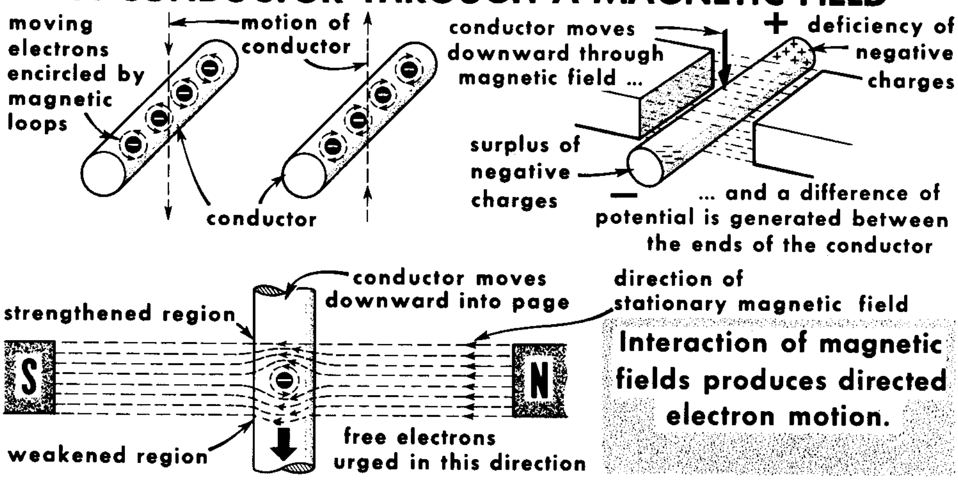
VECTORS IN DIRECTIONS FORMING A 90° ANGLE



Generating a Voltage in a Moving Conductor

There are several ways of generating an a-c voltage. The basic method is to induce an emf in a conductor by moving it across the lines of force of a stationary magnetic field, as we discussed in Volume 1. We will now consider this a little more closely. When a conductor is moved, the free electrons it contains move with it regardless of which direction the conductor moves. Every moving electron is encircled by magnetic loops of force, and these loops always position themselves at right angles, or perpendicular to the direction of the moving negative charge. When a conductor moves downward, the electrons it contains move downward with it. Thus, the magnetic loops encircling the electrons are perpendicular to the downward motion, or in a horizontal plane. Applying the left-hand rule to the motion of the electrons, the magnetic loops will rotate counterclockwise around the electrons (viewing the electrons from the top down). When a conductor moves upward, the reverse occurs.

GENERATING A VOLTAGE BY PASSING A CONDUCTOR THROUGH A MAGNETIC FIELD

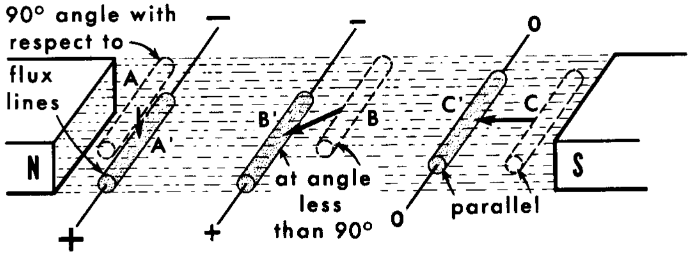


When a conductor cuts through a stationary magnetic field, there is interaction between this field and the magnetic loops encircling the moving electrons. This interaction produces a resultant magnetic field around the free electrons, which in turn produces a strengthened and a weakened region on opposite sides of the electrons. The strengthened region results from the two magnetic fields aiding each other; the weakened region from the two magnetic fields opposing each other. The free electrons present are thus "urged" from the strengthened region in the direction of the weakened region, creating an accumulation of free electrons at one end and a corresponding shortage at the other. The area of electron accumulation is called the negative end of the conductor; the area of electron shortage is called the positive end. Thus, a potential difference is produced between the ends of the moving conductor, and the conductor becomes a voltage source.

Generating a Voltage in a Moving Conductor (Contd.)

The polarity of the electromotive force induced in a conductor cutting through magnetic lines of force is a function of the relative directions of the lines of force of the stationary magnetic field and the direction of motion of the conductor. The direction of the stationary field is fixed – from the north pole to the south pole. Therefore, the direction of motion of the moving conductor is the controlling factor in determining polarity.

Maximum Voltage is Induced in a Conductor When It Cuts a Maximum Number of Lines of Force Per Unit Time.

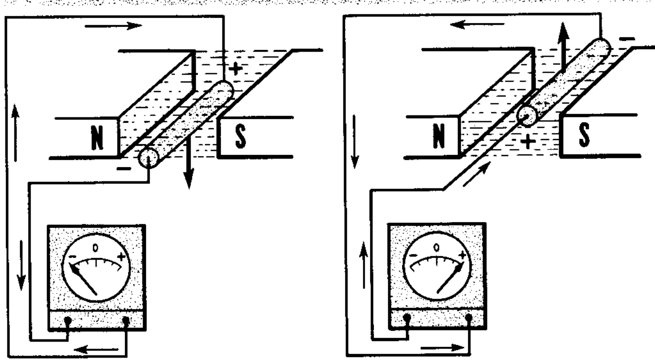


It varies with the cutting speed and the strength of the magnetic field.

A—A'
Maximum number of flux lines cut; maximum voltage induced.

B—B'
Fewer flux lines cut; less induced voltage.

C—C'
Minimum flux lines cut; minimum or zero voltage induced.



POLARITY OF INDUCED VOLTAGE VARIES WITH THE DIRECTION OF CUTTING.

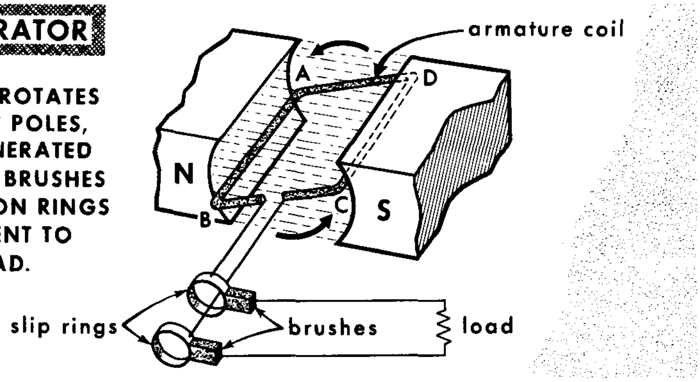
The amount of voltage that is induced or generated depends on the velocity at which the lines of force of the stationary field are cut by the conductor, and the strength or flux density of the magnetic field. Assuming the conductor velocity remains constant, the rate at which the stationary flux lines are cut then will depend on the <u>angle</u> at which the conductor cuts through the flux lines. Minimum or zero <u>voltage</u> is induced when the conductor moves parallel to the lines of force. Maximum voltage is induced when the conductor cuts the lines of force at right angles, or 90°. This is the greatest possible <u>cutting angle</u>, and the angle at which maximum flux lines are cut per unit time. Between these two points, an intermediate amount of voltage is induced.

The Basic A-C Generator

The basic generator of alternating voltage is a pivoted-loop armature having two coil sides, A-B and C-D, which rotate between the two pole pieces of a horseshoe magnet with uniform velocity and through a uniform stationary magnetic field. The rotating motion causes the coil sides to cut the flux lines of the stationary field. Because the voltage generated in the two sides of the rotating loop is equal (though opposite in polarity), we can examine the process of voltage generation by considering one coil side only. For this purpose, we select the side C-D and use its slip-ring as the voltage reference.

BASIC A-C GENERATOR

AS ARMATURE COIL ROTATES BETWEEN MAGNETIC POLES, A-C VOLTAGE IS GENERATED BETWEEN SLIP RINGS. BRUSHES MAINTAIN CONTACT ON RINGS AND CARRY CURRENT TO AND FROM LOAD.

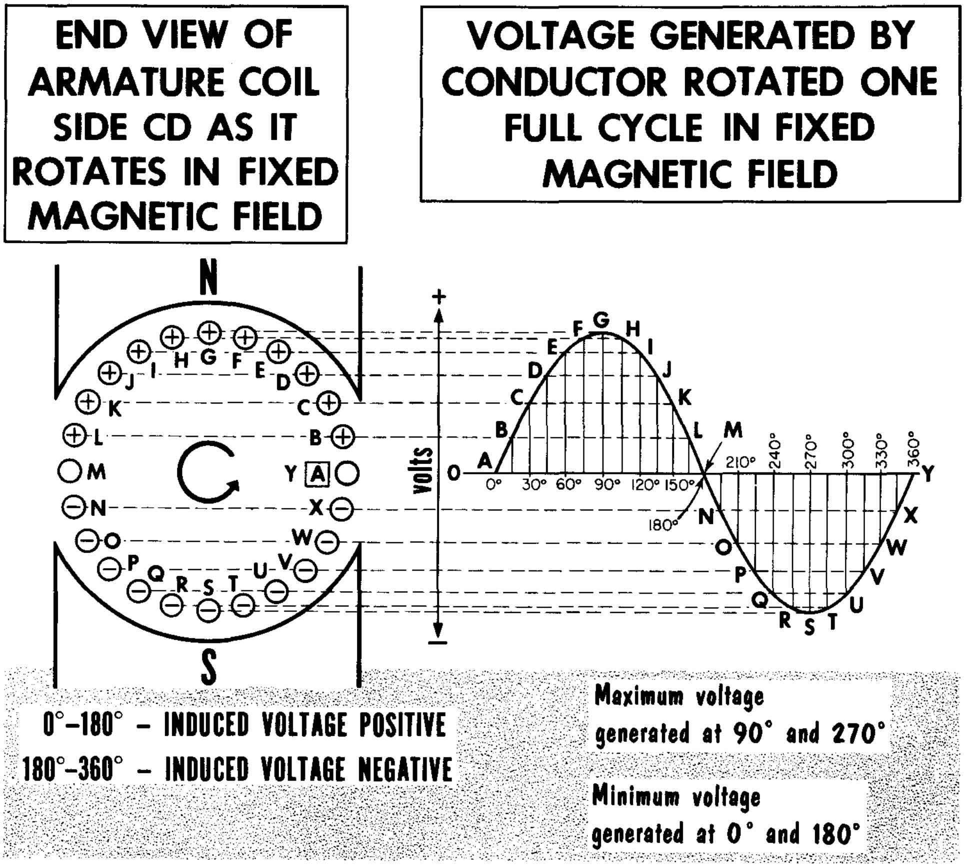


Assume that the action begins with side C-D momentarily positioned at A. At this moment, the coil side is moving parallel to the flux lines of the stationary field. Thus, the angle of cutting of the flux lines is zero (or the rate of cutting of the flux lines is zero); hence, the voltage induced in the coil side is zero. As rotation continues, the coil side moves upward and passes through progressively increasing angles of rotation, as shown by points B, C, D, E, F, and G, which correspond to 15°, 30°, 45°, 60°, 75°, and 90°. In doing so, the angle of cutting of the flux lines by the rotating coil side increases from 0° at A to a maximum of 90° at G; therefore, the rate of cutting of flux lines increases and the output voltage increases. Maximum output voltage is developed at G, or when the angle of cutting is 90°, thus completing one-quarter turn. A plot of the output voltage in steps of 15° of angular time between 0° (A) and 90° (G) is shown.

As the coil side continues rotating, it inscribes increasing angles of rotation; 105° (H), 120° (I), 135° (J), 150° (K), 165° (L), and 180° (M), but the angle and the rate of cutting of the flux lines decreases progressively from G to M. And so does the voltage output, reaching zero at M. Here, the coil side again is moving parallel to the flux lines. Note that the amount of decrease in voltage for each 15° change in angular rotation between 90° and 180° (G to M) is exactly the same as the amount of increase in voltage between 0° and 90° (A to G). Note also that the output voltage remains positive while the coil side is completing the half turn from 0° to 180° of angular time, the reason being that the motion of the coil side through the flux lines (past the Npole) continues throughout, except at the angles of 0°(A) and 180° (M).

The Basic A-C Generator (Cont'd)

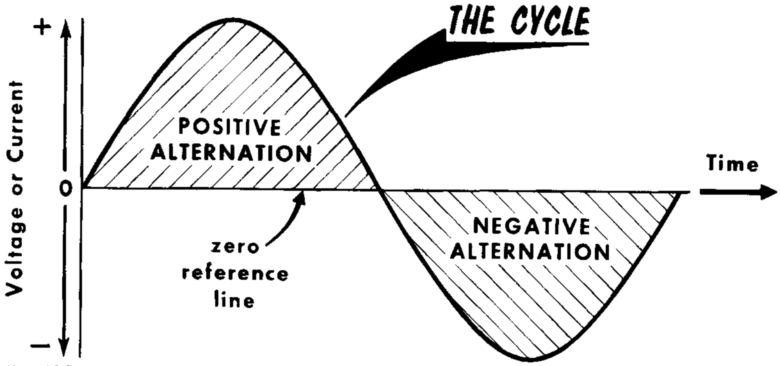
As the coil side moves past the 180° point, the angle of rotation continues to increase, as does the angle at which the conductor cuts the field. The conductor now is moving downward past the S pole. The direction of electron flow in the conductor is the opposite of before and the induced voltage is negative.

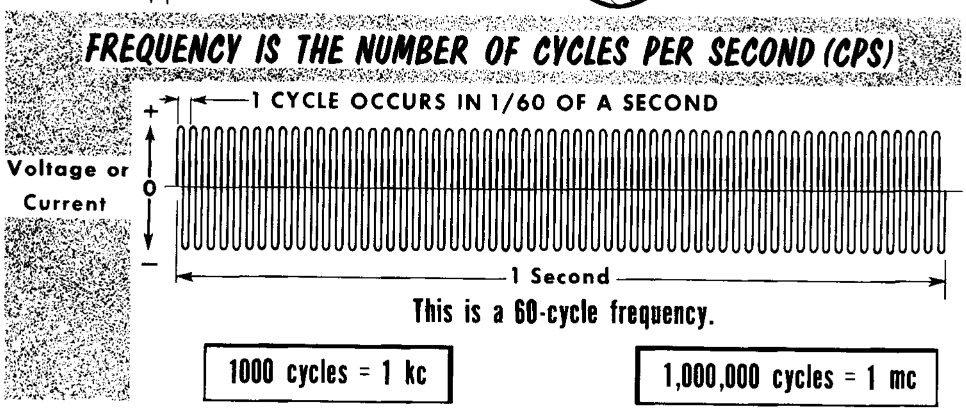


Now we show the voltage below the zero voltage reference line. The output voltage has been reversed because of the change in the relative direction of the moving conductor and the stationary flux lines. The change in output voltage of negative polarity for angular steps of 15° between 180° (M) and 270° (S) is exactly the same as the voltage between 0° (A) and 90° (G) of positive polarity. That is, the 90° (G) and 270° (S) points are maximum voltage points of opposite polarity. Further movement of the coil side from the 270° (S) position to the 360° (Y) position results in a fall in voltage from maximum negative to zero. The angular rotation increases but the rate of cutting of the flux lines decreases from the maximum at the 270° (S) point to 0 at the moment of 360° (Y) of rotation. The coil has completed one full turn. It corresponds to 360° of rotation and is the equivalent of 1 cycle of the output voltage.

The Cycle--Frequency

When the armature in the basic a-c generator completes one full turn (360°) of rotation, it has generated one cycle of voltage. The voltage has gone from a starting value of zero, risen to a maximum positive value, fallen back to zero, risen to a maximum negative value, and fallen back to zero. A cycle refers to a complete chain or sequence of events. A cycle of voltage applied to a resistance load will cause a similar cycle of current to flow in a circuit. The cycle of current will go through the same fluctuation as the cycle of voltage. When all the voltage or current values are joined together, they form a "picture" or pattern called a waveform.

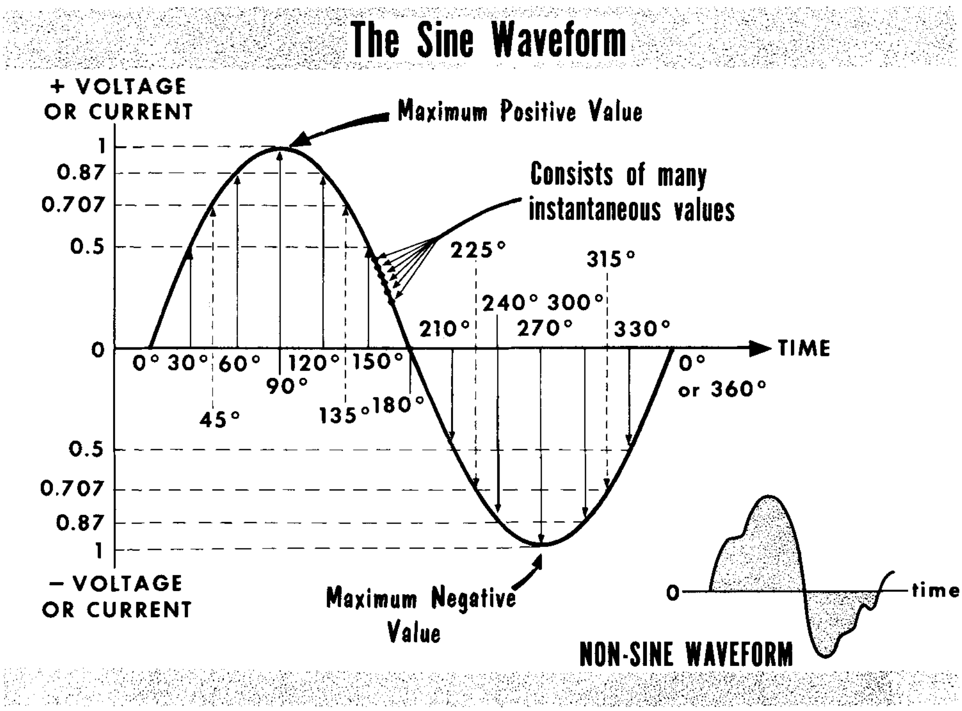




The number of complete cycles that occur in one second is called the frequency of the waveform. When a voltage or current waveform passes through 60 cps, it is called a 60-cycle frequency. Each half-cycle is called an alternation; each complete cycle thus contains two alternations--a positive and a negative. Since a 60-cycle frequency represents 60 complete cycles per second, the time duration of each cycle is 1/60 of a second. In high frequencies such as 1 megacycle (1,000,000 cycles per second), the time duration of each cycle is 1/1,000,000 of a second. The faster our basic generator rotates, the more cycles per second will be generated, and the higher will be the output frequency. The strength of the magnetic field will determine the strength or amplitude of the output waveform, but not its frequency.

The Sine Waveform--Voltage and Current

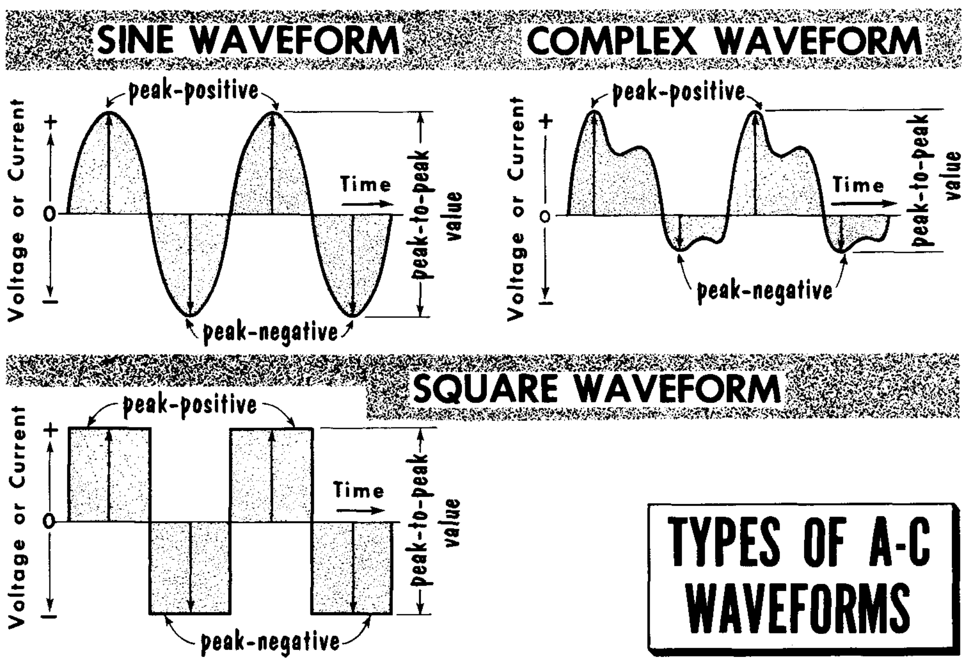
The sine or sinusoidal waveform is a pattern of instantaneous changes in the value of an alternating voltage or current. The word "sine" is taken from the sine table (see Appendix) used in mathematics because the amplitude of the sine wave varies from zero to maximum in the same manner as the values in the sine table. When we refer to a waveform as a sine wave, it indicates that we are considering only a single frequency. When various sine waves of different frequencies are combined, they form a complex waveform which is not a sine wave.



An important characteristic of the sine waveform is that the positive and negative half-cycle are mirror images of each other. The rate of rise and fall of both alternations is identical. At 0°, we see that the value of the sine wave (voltage or current) is zero. At 30° along the zero time axis, the sine waveform value has climbed to 0. 5, or half its maximum value. At 45°, the sine wave is at 0. 707 of its maximum value, and at 60° a value of 0. 87 of maximum is reached. Finally, at 90° or one quarter of the entire cycle, the maximum value of the sine wave is reached. In going from 90° to the half-way point at 180°, the sine wave decreases in value in a manner opposite from the way it increased going from 0° to 90°. The second half-cycle, from 180° to 360°, has identical rise and fall values to those of the first half-cycle except that they are in the opposite direction. Actually, a sine wave consists of many more values than are shown. There are an infinite number of values, and the sine wave is a picture of all their instantaneous values joined together in time.

Instantaneous and Peak Values of A-C Voltage and Current

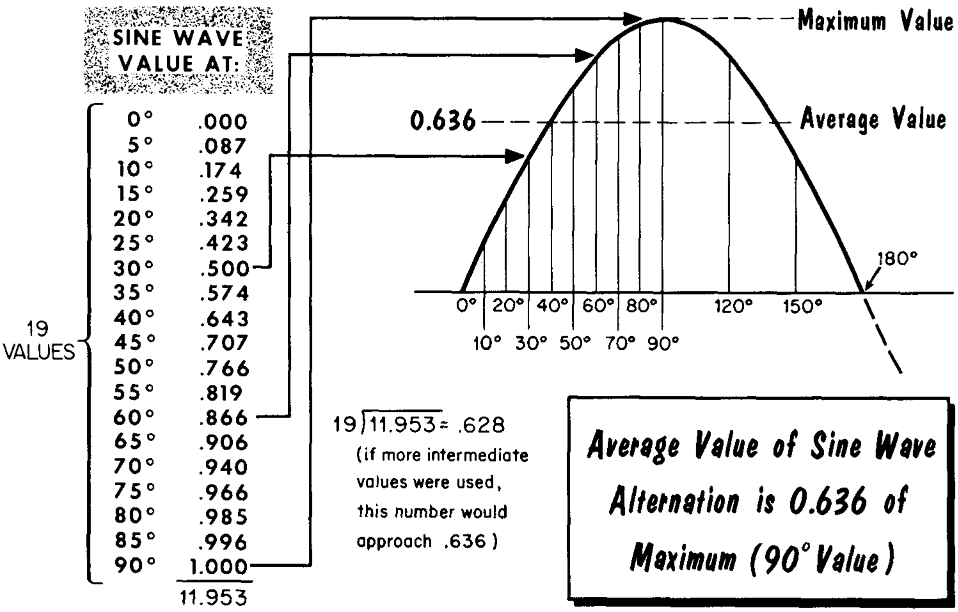
The continually changing values of alternating voltage and current necessitate the use of special terminology to express the amount of the voltage and the current. The instantaneous value of an a-c voltage or current is that value which exists at any specific instant of time. It can have any amplitude between 0 and maximum, and can be of a positive or a negative polarity. When an instantaneous value is indicated, small letter "e" is used to express voltage and small letter "i" to express current. For example, the instantaneous value of a sine waveform voltage at 45° is stated as $e = 28.3$ volts. Similarly, the instantaneous value of a sine waveform current at 345° might be stated as $i = 0.259$ ampere. Reference to an instantaneous value must be accompanied by identification of its time in the cycle, because the value of a sine wave changes constantly. An a-c voltage at 33° is different from an a-c voltage at 34°, and is still different at 34.5°.



The peak value of an alternating voltage or current is the highest value reached by the quantity during a cycle. Maximum and peak value have the same meaning. A peak amplitude of 10 volts means the same as a maximum amplitude of 10 volts. This holds true for any type of waveform. Of course, the polarity may also be referred to, such as peak-positive or peak-negative value. In the sine waveform, the positive-and negative-peak values are always alike; this is not so for voltage or current waveforms which are not of sine waveform variation. The peak-to-peak value (sometime abbreviated p-p), is simply the sum of the positive-peak and negative-peak values, regardless of the waveform. A voltage that has a 10-volt peak-positive value and a 10-volt peak-negative value has a 20 volt (10 + 10) peak-to-peak value.

Average Value of a Sine Waveform Voltage or Current

If we add all the instantaneous values of the positive half-cycle and the following negative half-cycle of a sine waveform, and then find the average of these values, we find it to be zero. The reason for this is that the two adjacent half-cycle are of opposite polarity (one being plus and the other minus) and, when we add a plus quantity to a minus quantity of equal value, the result is zero. So a general statement can be made--the average value of a cycle of a sine waveform is zero.



Average Value of Sine Wave Alternation is 0.636 of Maximum (90° Value)

A different situation prevails if we think in terms of a half-cycle, either positive or negative. Imagine that the peak value of a sine waveform voltage or current is 1 volt or 1 ampere. If we add up the instantaneous values of voltage or current prevailing at each 5° of the cycle, the total is 22.90. Since the half-cycle waveform is made up of 36 instantaneous values, then 22.9/36 gives us an answer of 0.636 volt or ampere. Therefore, 0.636 is the average value of a half-cycle of a sine waveform voltage or current when the peak value is 1 volt or 1 ampere. We say then, that the average value of a sine wave is 0.636 of its maximum value.

$$E_{av} = 0.636 \times E_{max}, \text{ or } I_{av} = 0.636 \times I_{max}$$

Knowing the average value of a sine wave, we can calculate the peak or maximum value. It is the average value multiplied by 1.57 or

$$E_{max} = 1.57 \times E_{av}, \text{ or } I_{max} = 1.57 \times I_{av}$$

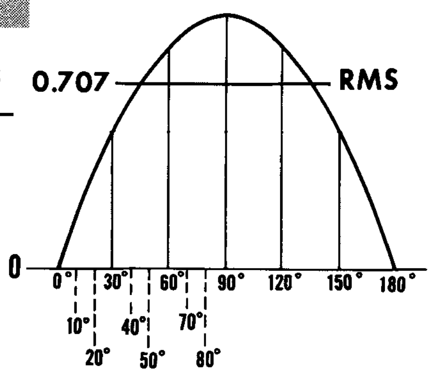
For instance, if the average value of a sine waveform voltage is 140 volts, the maximum (or peak) value is $E_{max} = 1.57 \times 140 = 219.8$ volts.

The Effective or RMS Value of a Sine Wave

You have learned that when direct current flows through a resistance, the amount of heat developed is proportional to I^2, or the square of the current. When alternating current flows through a resistor, the heat developed is proportional to the square of the instantaneous values of current. We can see why this is so. Alternating current changes constantly in value--it changes every instant. First, it rises from zero to a maximum value and then, it falls back to zero. Following this, it rises to a maximum in the opposite direction in a circuit and then, it again falls back to zero. Because of this constant variation, we must find a value that will be equivalent to some value in direct current. This equivalent value is called the effective value, because the effective value of an a-c sine wave tells us that that value of alternating current will do as much work as the same value of direct current. Unless we have an effective value, it would be difficult to discuss a-c voltages and currents in comparison to d-c voltages and currents.

ROOT MEAN SQUARE (EFFECTIVE VALUE) OF SINE WAVE EQUALS 0.707 OF MAXIMUM VALUE

SINE WAVE VALUE AT		INSTANTANEOUS VALUE		INSTANTANEOUS VALUE SQUARED
0°	=	0.00	→	0.0000
10°	=	0.17	→	0.0289
20°	=	0.34	→	0.1156
30°	=	0.50	→	0.2500
40°	=	0.64	→	0.4096
50°	=	0.77	→	0.5929
60°	=	0.87	→	0.7569
70°	=	0.94	→	0.8836
80°	=	0.98	→	0.9604
90°	=	1.00	→	1.0000
				4.9979

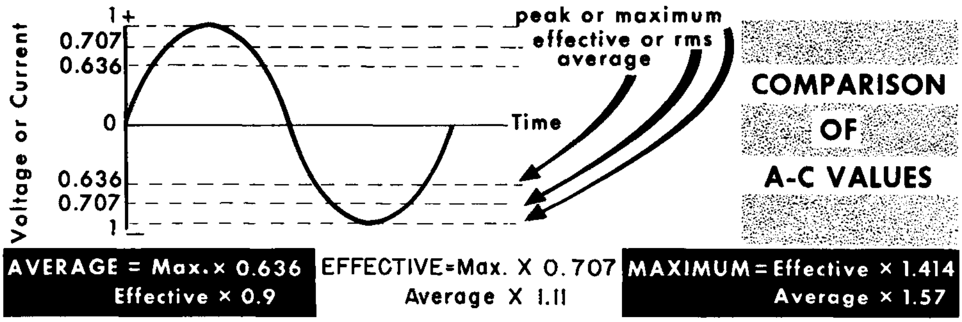


$10\overline{)4.9979} = 0.49979$ (mean square)

$\sqrt{0.49979} = 0.7069$ — (root mean square)
or 0.707

The effective value of a sine wave is obtained as follows: First, we take a large number of instantaneous values of a sine wave and square each one. Then, we add up all the squared instantaneous values and divide this total by the number of values used. This gives us the average or mean square. Finally, we take the square root of the mean square, which gives us the root mean square, or rms value. This rms value is very important, because the rms value of an a-c sine wave indicates that a specific voltage or current will do as much work as the same value of d-c.

The Meaning of Effective Value



When we plug a soldering iron into a 120-volt a-c source, it will reach a certain temperature, depending upon its wattage rating. If we plug this same soldering iron into a 120-volt d-c source, it will arrive at the same temperature. This is because the 120 volts a-c is the effective value of the a-c waveform. Its peak value is much higher than 120 volts, and for much of each alternation its instantaneous values are less than 120 volts. Actually, the effective value of a sine wave is 0.707 of its maximum value. An alternating voltage with a peak or maximum value of 1 volt will have an effective value of 1 × .707 volt, and it will produce the same heat in a given resistor as .707 volt d-c.

There are two simple formulas that can be used: One, to find the effective value of a sine wave knowing its maximum value; the other, to find the peak value knowing the effective value.

$$\text{Effective Value} = \text{Maximum Value} \times 0.707$$
$$\text{Maximum Value} = \text{Effective Value} \times 1.414$$

At the ordinary 120-volt 60-cycle house outlet, the peak value is:

$$E_{max} = 120 \times 1.414 = 169.68 \text{ volts.}$$

If an a-c current has a peak value of 5 amperes, the rms value would be:

$$I_{eff} = 5 \times .707 = 3.535 \text{ amperes}$$

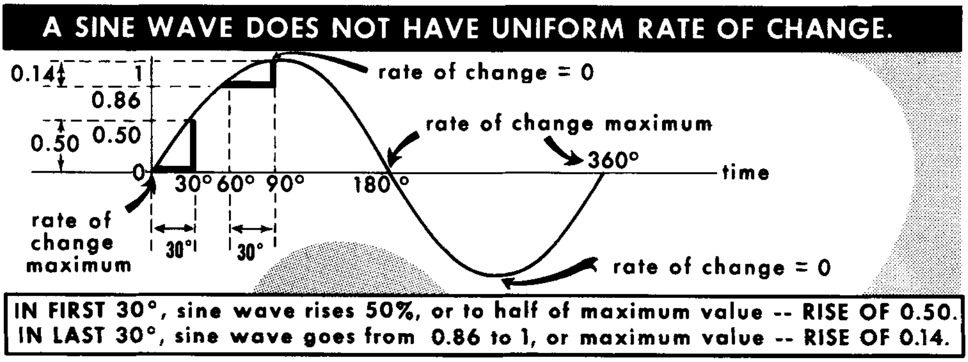
As another example, if the effective value of a current is 180 milliamperes, and the effective value of a voltage is 690 millivolts, their peak values are:

$$I_{max} = 180 \times 1.414 = 254.5 \text{ milliamperes}$$
or
$$.18 \times 1.414 = .2545 \text{ amperes}$$

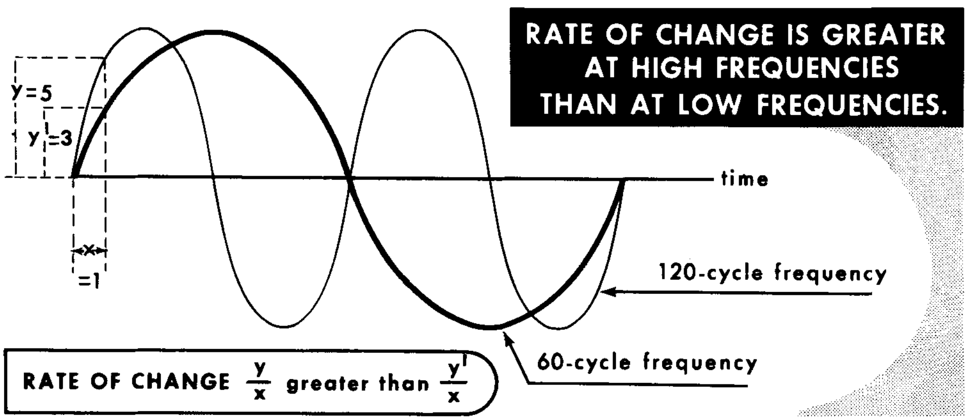
$$E_{max} = 690 \times 1.414 = 975.7 \text{ millivolts}$$
or
$$.69 \times 1.414 = .9757 \text{ volts}$$

Whenever an a-c voltage or current is stated, it always is taken to mean the effective or rms value, unless otherwise indicated. The usual a-c voltmeters and ammeters are calibrated to read rms values.

Rate of Change



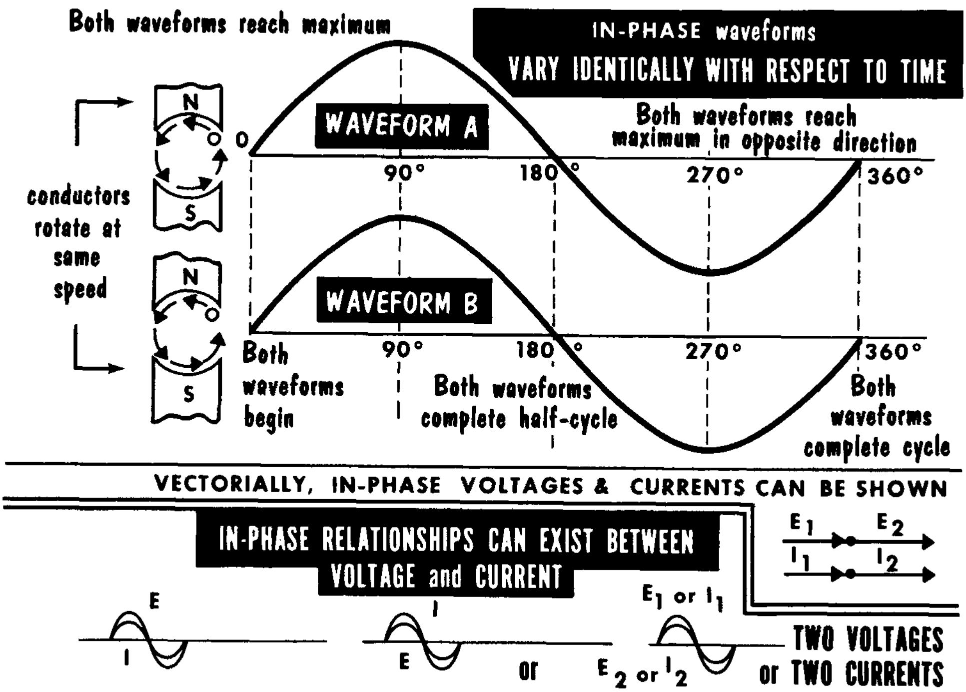
Related to the behavior of alternating current and voltage is a mathematical term known as rate of change. It refers to the relative change in value of an a-c voltage or current in a unit period of time. For example, if a current (or voltage) changes a great deal in value in a given small interval of time, its rate of change is high; if its value changes only a little in the same period of time, its rate of change is low. If an alternating current (or voltage) waveform is plotted on a graph, the slope or steepness of the waveform as it increases or decreases at any point on the graph, relative to the horizontal axis or time scale, is an indication of the relative rate of change. From this, it is evident that the higher the frequency of a voltage or current, the faster the rate of change.



The rate of change of a sine waveform quantity is maximum at the instant that the current (or voltage) is passing through zero in both the positive and negative-polarity directions. It is minimum (zero) at the moment when the waveform is passing through its maximum amplitude. At this instant, the quantity is neither decreasing nor increasing. Thus, the maximum rate of change occurs at 0°, 180°, and 360°; the minimum rate of change occurs at 90° and 270°.

The Concept of Phase (In Phase)

"Phase," sometimes referred to as "phase displacement," "phase difference", or "phase relation," is a concept of a time relationship between two alternating quantities--voltage, currents, or a current and a voltage. By time relationship in a-c, we mean the extent to which the two quantities remain in step or go out of step as their amplitudes change in value. In a d-c circuit, a change in current keeps in step (phase) with a change in voltage; this is not necessarily the case in a-c circuits. Circuit components other than resistance cause changes in phase. When the voltage and current changes keep in step with each other, they are said to be in phase. This always takes place in a resistive circuit, since voltage and current in a resistance are in phase.



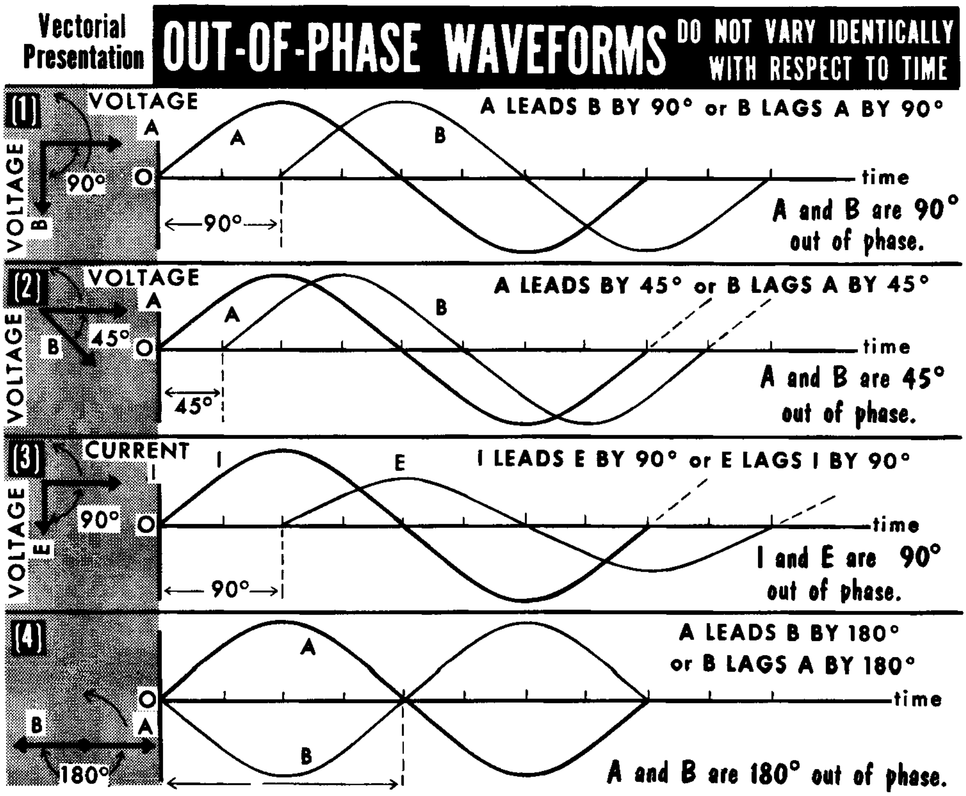
Imagine two identical generators that start functioning at the same instant, with their armatures revolving at the same speed. Each will generate a sine waveform voltage in which zero and maximum amplitudes occur together, and where their relative intermediate values will occur at the same time. We describe such behavior of two generators as being in step, or in phase, and producing two voltages which are in phase. Another way of stating this is to say that the two voltages have 0° phase difference, or that the phase angle of voltage A relative to voltage B (or vice versa) is 0°. When considering the phase relation between two quantities, a suitable point of reference is the instant when the two quantities pass through zero amplitude in the same direction. When shown as vectors, the in-phase quantities have a common origin and lie along the same plane, each vector head having its own identity.

Concept of Phase (Lead and Lag)

Out of phase is a broad expression which indicates that the identical amplitude variations of the waveforms do not occur at the same time. Like two runners (A and B) who are racing, A can arrive at a selected point ahead of B, or lead B, which automatically places B behind A, or lagging A. In electrical considerations, the point of reference is the instant when the two waveforms being compared pass through zero amplitude in the same direction. Whichever quantity passes through this point first is leading the other.

Out-of-phase conditions are expressed in electrical degrees, because this manner of expression is much more convenient than referring to a fractional part of a cycle. It is preferable to say a 90° phase difference than 1/4 cycle; using the term 45° phase difference is preferable to 1/8 cycle.

If you examine (1) below, you will note that voltage A passes through zero going in the positive direction one-quarter of a cycle, or 90° before voltage B. In (2), the two voltages (A and B) have a phase difference of only 45° with A leading B, which is the same as B lagging A by this amount. In (3), current I is leading voltage E by one-quarter of a cycle, or 90°, or E is lagging I by 90°. In (4), voltages A and B are 180° out of phase. They pass through maximum and zero points at the same time but in opposite directions.



Any line drawn through the center of a circle which divides the circle in half is called the diameter.

The constantly curving line that forms a circle is called the circumference.

The three sides of a right triangle are: the base (horizontal side); the altitude (vertical side); and the hypotenuse (side opposite the right angle).

In a right triangle, the angle formed by the base and the hypotenuse is referred to as the angle theta (θ).

In a right triangle, the square of the hypotenuse is equal to the sum of the squares of the other two sides ($C^2 = A^2 + B^2$).

A vector relationship is used to describe any situation or action that involves both magnitude and direction.

The basic method of inducing an emf in a conductor is to move it across the lines of force of a stationary magnetic field.

Maximum voltage is induced in a conductor when the conductor cuts the magnetic lines of force at right angles, or 90°.

The basic a-c generator is a pivoted-loop armature having two coil sides which rotate between magnetic poles with uniform velocity and through a uniform stationary magnetic field.

A cycle is one complete series of changes in an a-c current or voltage. Frequency is the number of cycles which occur in one second.

The sine waveform is a pattern of instantaneous changes in the value of an alternating voltage or current.

The instantaneous value of an a-c voltage or current is that value which exists at any specific instant of time. The peak (maximum) value is the highest value reached by a quantity during a cycle.

The average value of a sine wave equals .636 of its maximum value.

The root mean square (RMS) or effective value of a sine wave equals .707 of its maximum value.

The higher the frequency of a voltage or current, the faster the rate of change.

Phase refers to a time relationship between two alternating quantities--voltages, currents, or voltages and currents.

Out of phase broadly indicates that the identical amplitude variations of two waveforms do not occur at the same time.

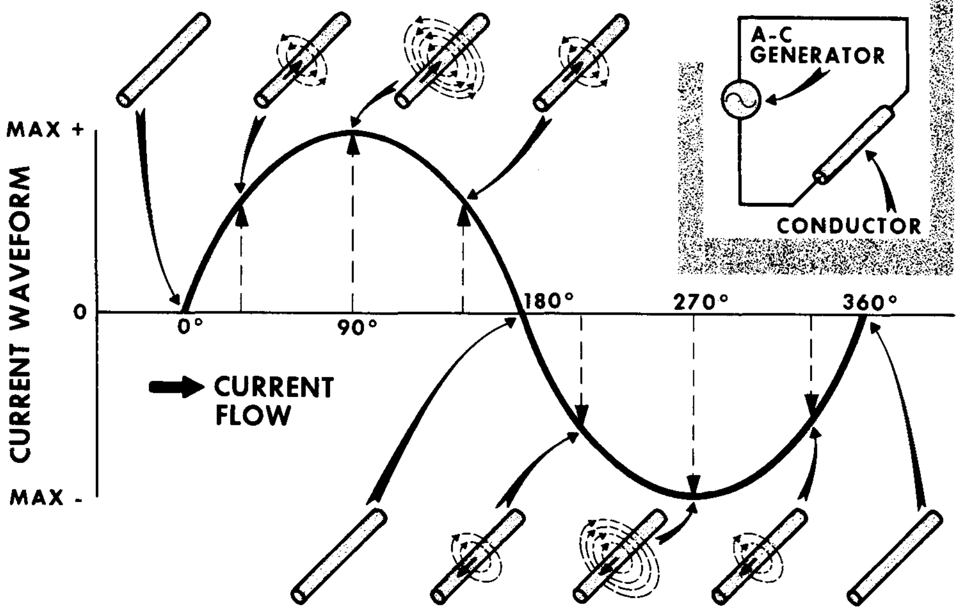
REVIEW QUESTIONS

1. What are the three sides of a right triangle?
2. Explain the parallelogram method and its application to vector analysis.
3. In a right triangle, what is the angle formed by the base and the hypotenuse referred to as?
4. What is the relationship of the three sides of a right triangle to each other?
5. What is a vector relationship used to describe?
6. Give three factors which determine the magnitude of an induced emf.
7. What is the basic method used to induce an emf in a conductor?
8. Explain the principle of operation of the basic a-c generator.
9. What is a sine waveform?
10. Define a cycle. Define frequency.
11. What is meant by the rms or effective value of a sine wave and what is it equal to? What is the average value of a sine wave equal to?
12. Explain what is meant by rate of change.

Magnetic Field around Alternating Current

Alternating current is encircled by loops of magnetic force that change in number instant by instant and periodically change direction.

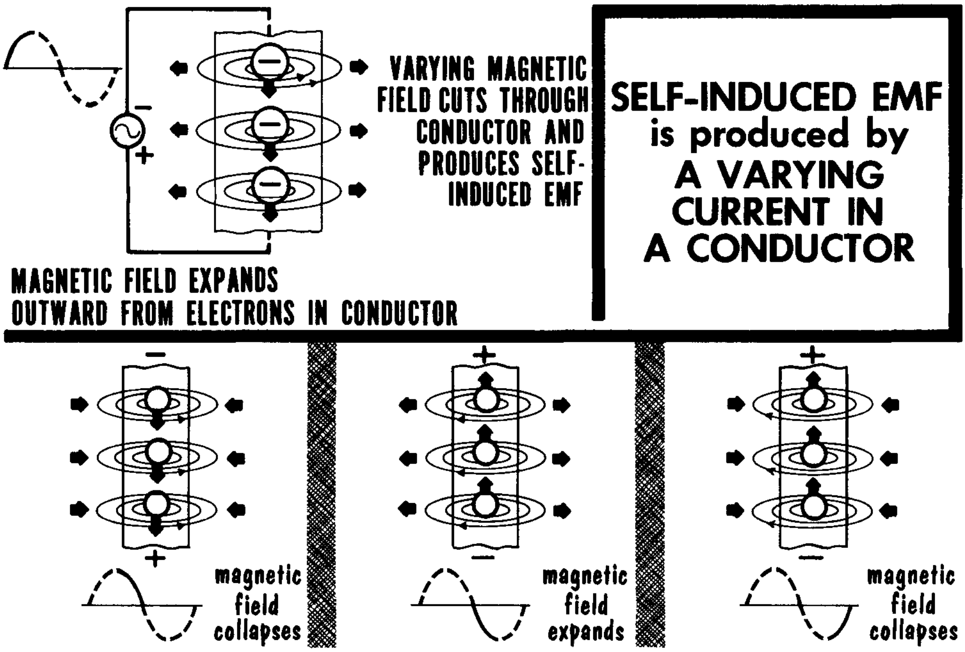
Magnetic Field around a Current-Carrying Conductor during 1 Complete Cycle



Assume a sine waveform current. (Whatever happens during one cycle occurs during the others.) When the current is zero, there is no magnetic field around the conductor. As current begins to increase, the magnetic field builds up in density, reaching maximum value coincident with the maximum current point in the positive half-cycle. The direction of the field is determined by the left-hand rule. We show the field counterclockwise; during the positive alternation, as the current starts to decrease in value but is still in the same direction, the intensity of the magnetic field also decreases. To all intents and purposes, the lines of force of the previously higher value of current in the vicinity of the conductor fall back into the conductor, i.e., the field collapses, reaching zero intensity when the current reaches zero. At this instant, the direction of flow reverses. As the current begins to flow in the opposite direction, increasing in value moment by moment, the magnetic field starts increasing in intensity--but now it has a direction that is the opposite of what existed before. We show it as having a clockwise direction. Maximum field strength is again reached at the peak point of the negative half-cycle; then, the magnetic field begins to decrease, again collapsing into the conductor, reaching zero at the instant the cycle has been completed. An alternating current produces a constantly changing magnetic field around the conductor in which it is flowing.

Self-Induction of EMF

You have learned that relative motion between a magnetic field and a conductor will induce an emf in a conductor. In the examples shown so far, the moving or changing magnetic field was produced by one component (a magnet), and the conductor in which the emf was induced represented another component.

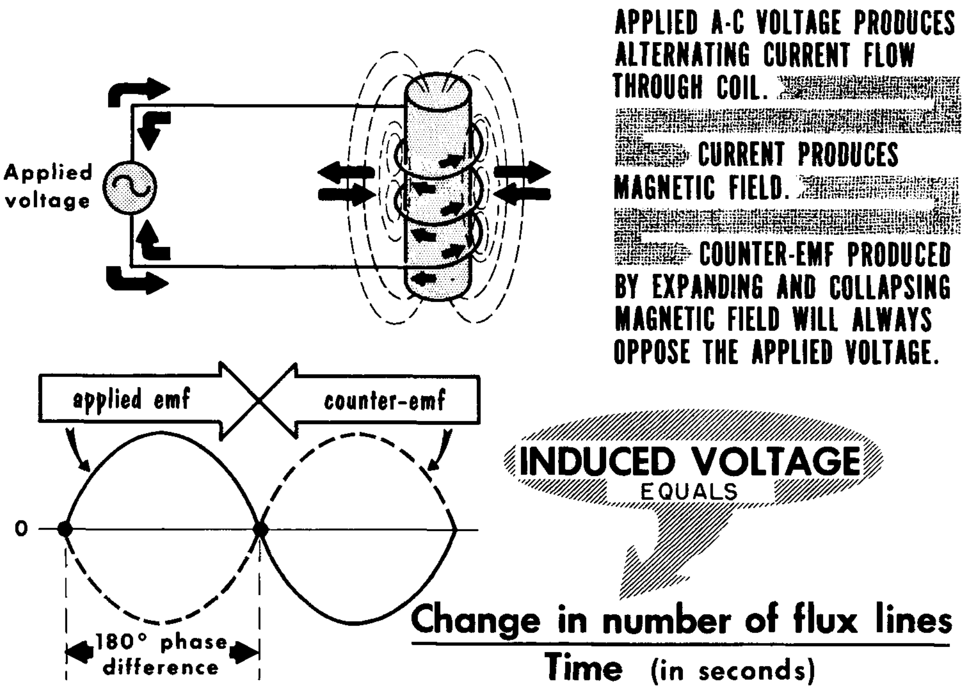


Self-induction of emf involves a changing current, a changing magnetic field, and a conductor; but now, the path of the current, the place of origin of the magnetic field, and the conductor in which it induces a voltage are one and the same. This is why we call the emf that is produced a self-induced emf. To visualize the action, think of it this way. An increasing current creates an increasingly intense magnetic field. The magnetic field originates at the free electrons inside the wire. As the field increases and expands from inside the wire to outside the wire, it must first move through the wire. It is during the time that the expanding magnetic field is cutting the wire that the emf of self-induction is generated.

Now imagine the current to be decreasing. The surrounding field collapses into the wire, i.e., returns to its place of origin, the free electrons. While moving back through the conductor to the electrons, the shrinking loops of flux cut the conductor and induce an emf--a self-induced emf. In one case, the emf of self-induction is generated by lines of force that move outward; in the other case, the emf of self-induction is generated by lines of force that move inward. If one direction of cutting due to a rising current generates a voltage of one polarity, the opposite direction of cutting due to a falling current generates an emf of opposite polarity.

The Action of Self-Induced EMF (Lenz's Law)

Assume an a-c circuit in which the current is increasing. The self-induced emf produced would have a polarity opposite to the applied voltage and, therefore, acts in opposition to this voltage and tends to retard the build-up of the circuit current. When the circuit current decreases, the self-induced emf has a polarity which aids the applied voltage and so tends to maintain the current; that is, prevent it from falling together with the decrease in voltage. Because the action of the self-induced emf is opposite to that of the applied voltage, it is often referred to as counter-emf or back-emf.



The behavior of self-induced emf was first explained by H. F. Emil Lenz and has since become known as Lenz's law. Although stated in different ways, Lenz's law states: "A changing current induces an emf whose polarity is such as to oppose the change in current." Counter-emf is not readily measurable, but its effects can be observed. When a circuit in which a substantial amount of current is flowing through coils is suddenly opened, the sudden collapse of the magnetic field induces a counter-emf which can be greater than the originally applied voltage. In fact, the counter-emf may even cause a momentary arc to bridge the gap where the circuit was opened.

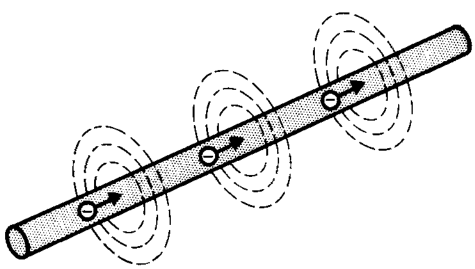
The amount of counter-emf produced will depend upon the rate of change at which the expanding and contracting magnetic lines of force cut the conductor. The greater the current, the more lines of force cutting per unit time; the higher the frequency, the more rapidly the magnetic field moves and, again, the more lines of force cutting per unit time.

Inductance

The physical shape of the conductor in which current flows, sometimes referred to as the "geometry of the inductor," also has a bearing on the control of the current. The loops of flux lines associated with current in a straight wire cut only that conductor during expansion and contraction of the surrounding magnetic field. The number of flux linkages between the lines of force and the conductor is the same as the number of loops of force produced by the current in the conductor. If, however, the conductor is coiled to form a solenoid, each turn links not only with the flux lines from that turn, but also with flux lines from adjacent and nearby turns.

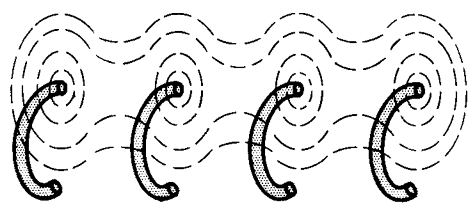
Inductance is that Property of a Circuit or Component which Opposes a Change in Current.

In expanding and contracting, flux linkages induce a counter-emf in a conductor. Induced voltage opposes both a rise and fall of circuit current.



A STRAIGHT CONDUCTOR CONTAINS INDUCTANCE.

Cross sectional view of a coil shows flux linkages between turns of the coil.



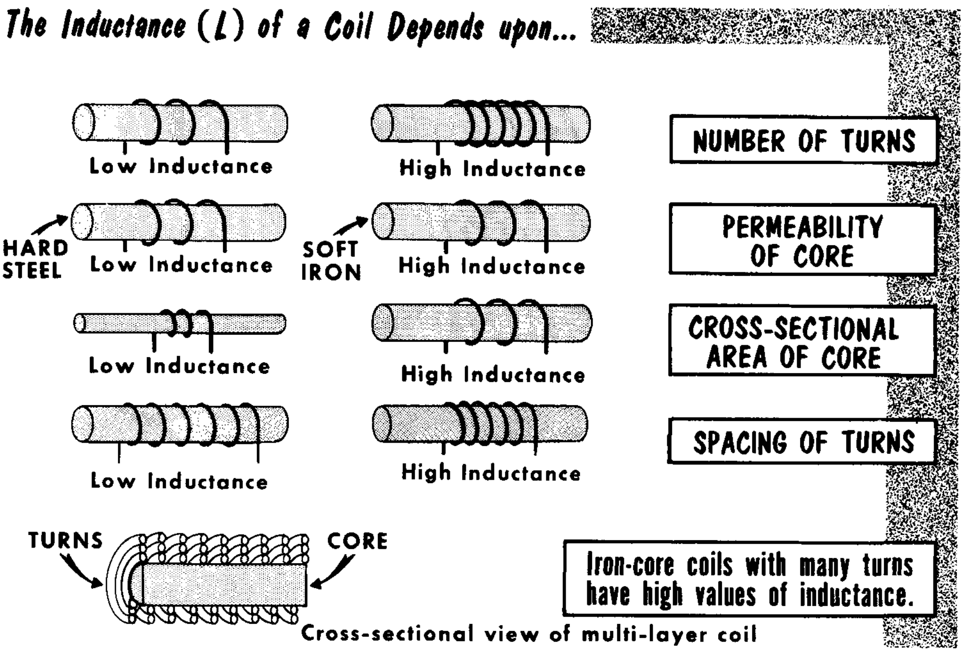
A COIL CONTAINS MORE INDUCTANCE BECAUSE INCREASED NUMBER OF FLUX LINKAGES PROVIDES GREATER COUNTER-EMF.

The total number of flux linkages with each turn is, therefore, several times the number of flux lines from a given current in the turn. If 10 flux lines originate from each of three adjacent turns and the lines link with each turn, the total number of flux linkages is 90, whereas the total number of lines that originate from the three turns is only 30. The greater the number of flux linkages per unit time, the greater the emf induced in each turn; hence, in the solenoid as a whole. For any given current in a coil, the counter-emf induced in the coil is a function of the number of flux lines multiplied by a constant that arises from the shape of the coil. The constant is symbolized by the capital letter L and is called self-inductance, or simply inductance.

Inductance (Cont'd)

The greater the inductance of a coil, the higher is the induced emf and the greater is the opposition to the increase and decrease of current in the solenoid.

Every conductor--short or long--has inductance. When the frequency is low (say up to several hundred cycles), the effect of the inductance of any reasonably long length of straight conductor is negligible. When the straight conductor is coiled, its inductance increases substantially. Even if the conductor is not coiled but its length is great, the amount of inductance possessed by the straight wire can be sufficient to influence current flow. This situation can be a problem for even a reasonably short length of wire when the operating frequency is high.



The greater the number of turns in a coil, the higher is its inductance. The closer the coil turns are to each other, the higher the coil inductance, because the flux linkages increase in number. If the core of a solenoid is made of a high-permeability material, such as soft iron, the inductance increases still more. On the other hand, if a coil is wound by doubling the wire back on itself, the inductance is held to a minimum. The self-induced emf generated in one half of the total length of the conductor offsets the self-induced emf generated in the other half of the conductor; hence, the coil as a whole displays minimum or even negligible inductance. Such a winding is known as a "non-inductive" winding. This method of winding is used to form wire-wound resistors wherein d-c resistance is desired but where inductance is an undesirable effect.

Unit of Inductance--The Henry

The unit of inductance is the henry, named after the American physicist, Joseph Henry. By definition, a conductor, or coil, has an inductance of 1 henry when a current which changes at the rate of 1 ampere per second induces an emf of 1 volt. The number of flux lines corresponding to this rate of change in the current is 100,000,000, or 10^8. In defining inductance as flux linkages per ampere of current producing the flux, we can say:

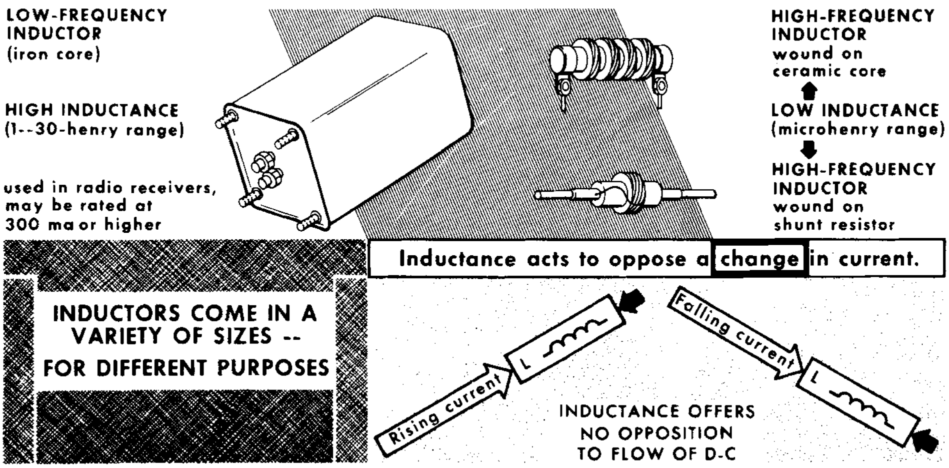
$$\text{Inductance (in henries)} = \frac{\text{flux linkages}}{\text{current producing flux}} \times 10^{-8}$$

Having established the above, we can now study the formula for determining the magnitude of a counter-emf:

$$\text{Counter-emf (induced voltage caused by changing current)} =$$

$$-L \times \frac{\text{change in current}}{\text{change in time}}$$

The minus sign means that the voltage developed is a counter voltage and opposes the force producing it. From this, we can see that the greater the inductance or the faster the rate of current change, the greater the counter-emf induced in the circuit.

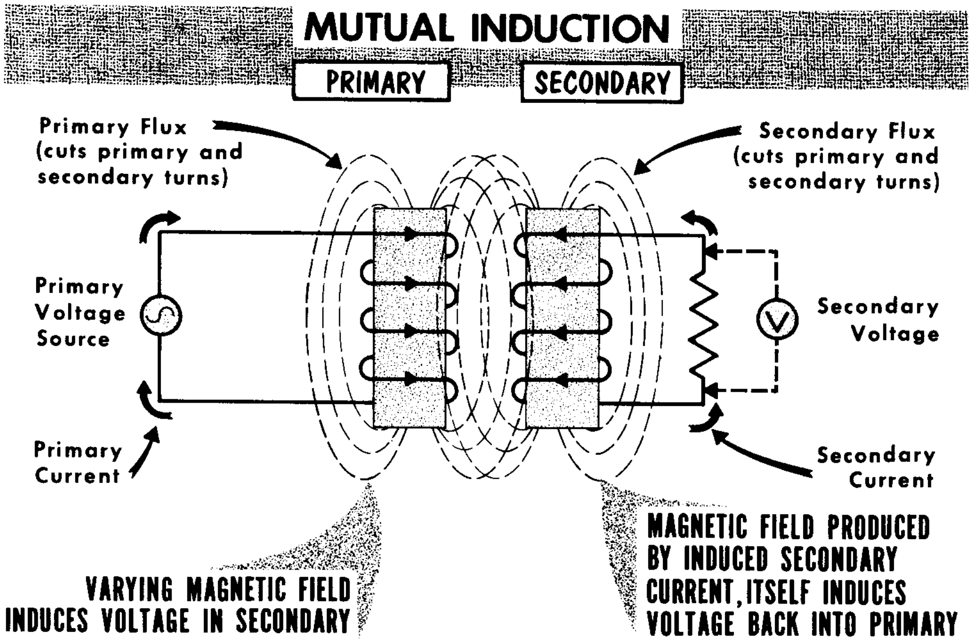


The inductance of coils used in radio communications equipment ranges from extremely small air-core units of 1 microhenry and less to large iron-core multilayer units of 30 henries and more. Inductors are also usually rated at some specific current. When excessive current flows through an iron-core inductor, the core may become "saturated" and the inductance decreases.

From what we have learned, we see that in addition to resistance, another circuit property, inductance, also is involved in the control of current. Of course, we must remember that while resistance opposes the flow of both a-c and d-c, the effect of inductance comes into play only under conditions of alternating or changing current.

Mutual Induction

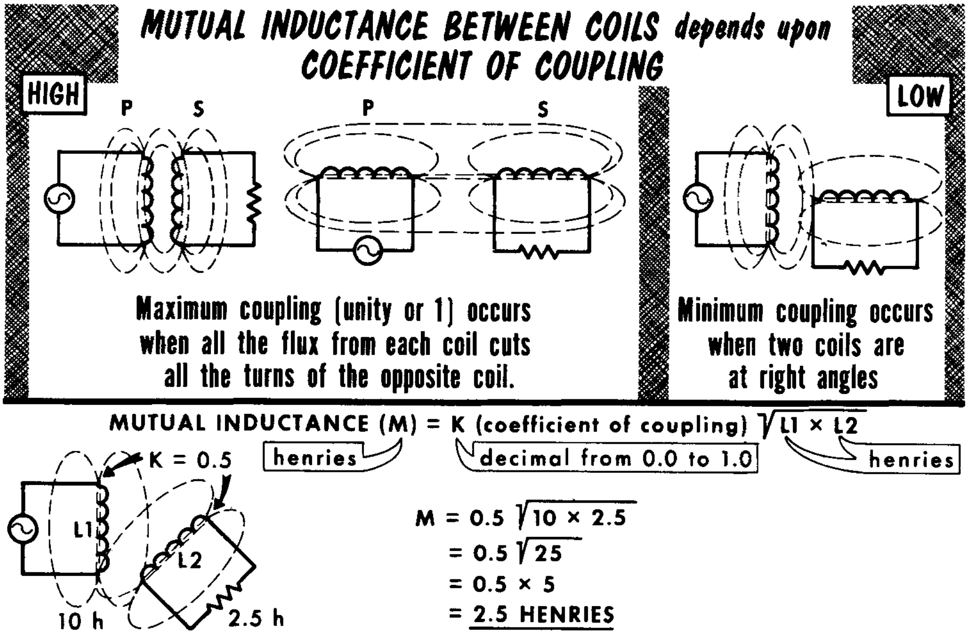
When a changing magnetic field produced by one coil cuts the turns of a second coil and induces an emf in the second winding, the action is known as mutual induction. The winding from which the flux originates is called the primary, usually indicated by the letter P. The voltage that is applied to the primary winding and causes current to flow is called the primary voltage. The changing current that flows in the primary winding and produces the changing flux is the primary current, sometimes referred to as the inducing current. The winding in which the emf is induced by the changing magnetic field is known as the secondary, usually indicated by the letter S. The emf induced in the secondary winding is known as the secondary voltage. If the secondary is part of a closed circuit wherein current flows, this current is called the secondary current. Mutual induction is a basis for transferring electrical energy from one circuit to another by means of a changing magnetic field. This is the basis of transformer operation.



The amount of emf induced in the secondary is, among other conditions, a function of the physical positioning between the primary and secondary windings. The positioning determines the flux linkages between the windings and, therefore, the rate of cutting by the lines of force. This physical relationship is known as "coupling." Coupling, flux linkage, and induced emf are maximum when the primary and secondary turns are interwound, or when the primary and secondary windings are one above the other and very close together. The more the number of turns of the secondary winding that are cut by the changing flux from the primary, the higher the emf induced in the secondary. The emf induced in each turn of the secondary winding is additive to the others.

Mutual Inductance

If a sine waveform voltage is applied to the primary winding of a primary-secondary assembly, a similarly varying current will flow in the primary winding. During the time that the primary current increases, its magnetic field expands, cutting the turns of the secondary and inducing a voltage in it. The secondary voltage, in turn, causes the flow of secondary current which has such direction (opposite to the primary current) as to create a magnetic field that opposes the field produced by the primary current. This action conforms with Lenz's law.



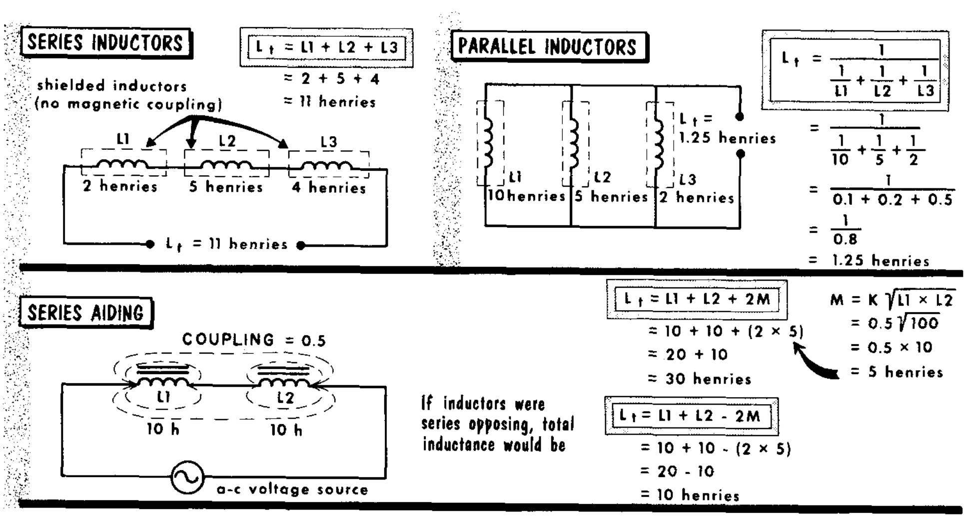
The magnetic field produced by the secondary current expands and, in so doing, cuts the turns of the primary winding. Here, it induces an emf which acts in opposition to the emf that is self-induced by the primary current. The resultant voltage of these two oppositely acting voltages is lower in value than the original self-induced voltage. Therefore, the primary current rises higher than it would were the field from the secondary current absent. During the period of decreasing primary current, the collapsing magnetic field cuts the secondary winding and induces an emf. The secondary current now has a direction that produces a magnetic field which tends to offset the collapsing field around the primary; i.e., it aids the self-induced emf in the primary and thus tends to prevent the primary current from falling.

The control of the primary current is presumed to be the result of magnetic lines of force common to both the primary and secondary windings. These common flux linkages are given the name mutual inductance, which is designated by the letter M and uses the henry as its unit. Any two coils positioned so that flux from one links with the other have mutual inductance.

Inductance in Series and Parallel

In order to achieve certain desired amounts of inductance, it is sometimes necessary to conduct inductors in series or parallel. When connecting inductors in series, the total inductance will be the sum of all the individual inductances: $L_t = L1 + L2 + L3 + \text{etc.}$ This formula holds true, however, only when the inductors are shielded from each other, or so positioned physically that there is no mutual inductance between them. If however, two inductors are located so that the flux lines from each cut the turns of the other, then the total inductance must take into consideration the mutual inductance between them.

We use the formula: $L_t = L1 + L2 \pm 2M$. The plus-or-minus 2M is used to take into consideration that the two inductors can be connected either series aiding or series opposing. In series aiding, the two inductors are arranged so that their flux lines move in the same direction and thus aid each other. This additional mutual inductance adds to the basic inductances of L1 and L2. When two inductances are arranged so that their flux lines oppose each other, we say that they are connected series opposing. Thus, the coefficient of coupling and the direction of magnetic fields help determine the total inductance. We can note an interesting point here. If we double the number of turns of a coil and assume perfect coupling (K - 1), we can achieve four times the inductance.



The total inductance of a circuit containing inductances in parallel is calculated in the same manner as resistances in parallel:

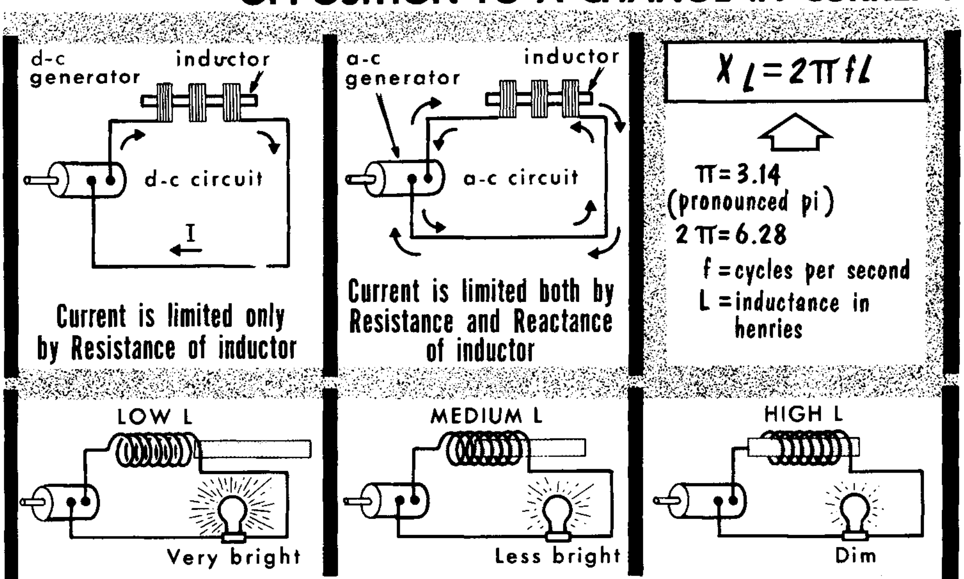
$$L_t = \frac{1}{\frac{1}{L1} + \frac{1}{L2} + \frac{1}{L3} + \text{etc.}}$$

The above formula is valid, as in series opposing, only when each inductor is shielded from the other. Any mutual inductance existing between inductors in parallel tends to reduce the total inductance.

Inductive Reactance

The opposition offered by an inductance to a change in current is measured at any given instant in terms of counter-emf (the voltage opposing the applied voltage). We saw that the characteristic of an inductance was to oppose any change in current, be it an increase or a decrease. This opposition presented by an inductance to an a-c or changing current is called inductive reactance (indicated as X_L). It can be compared somewhat to resistance (R). In d-c circuits and in a-c circuits containing only resistance, the total opposition to the flow of current is the resistance, in ohms ($R = E/I$).

INDUCTIVE REACTANCE -- OPPOSITION TO A CHANGE IN CURRENT

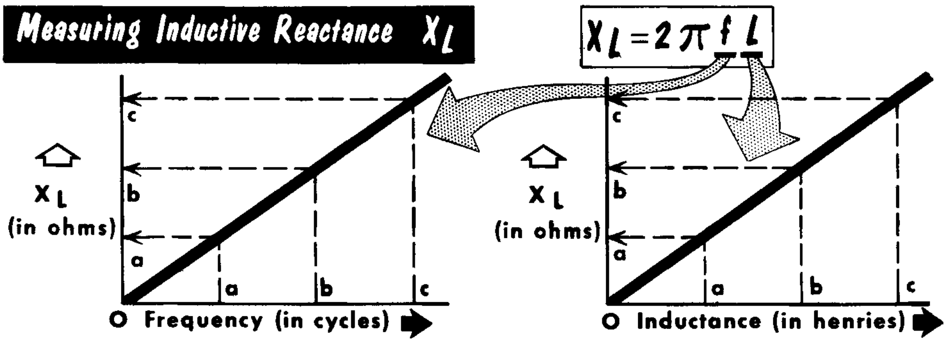


As inductance of circuit increases, reactance increases.

Reactance comes into play only under varying conditions. It represents an opposition to the flow of a varying current. Thus, the opposition offered by an inductor is called inductive reactance and, like resistance, is also measured in ohms. (Later in this course, we will study the reactance presented by a capacitor, called capacitive reactance.) Since the magnitude of induced emf depends on the amount of inductance and the frequency (rate of change) of the current, the formula for inductive reactance takes both into account. Inductive reactance is calculated: $X_L = 2\pi fL$. The $2\pi f$ represents the rate of change of the current. There are $2\pi(6.28)$ radians in a cycle, so $2\pi f$ represents the rate of change in current per second (angular velocity). Frequency (f) is in cycles per second, and L is equal to the inductance in henries. From this formula, we can see that the higher the frequency or the greater the inductance, the greater will be the inductive reactance. This is logical, since an increase in either will cause flux lines to be cut at a greater rate, and produce a greater counter-emf.

Inductive Reactance--Solving Problems

Having established the formula for inductive reactance: $X_L = 2\pi fL$, let us solve some problems concerning inductive reactance so as to gain a greater familiarity with this "new" type of opposition to current flow. In each of the following problems, we will assume that the resistance of the inductor is zero ohms. Actually, this is never the case. Since an inductor is wound with turns of wire, there must be some d-c resistance. Later, we will discuss the practical inductor which contains both resistance and reactance.



As FREQUENCY increases or decreases, inductive reactance increases or decreases. **As inductance increases or decreases, inductive reactance increases or decreases.**

In a simple circuit containing a 60-cycle voltage source and a 10-henry coil we will find the inductive reactance of the coil using the basic formula: $X_L = 2\pi fl = 6.28 \times 60 \times 10 = 3768$ ohms.

Leaving L fixed at 10 henries, and doubling the frequency to 120 cycles, $X_L = 2\pi fL = 6.28 \times 120 \times 10 = 7536$ ohms.

We see that doubling the circuit frequency doubled the inductive reactance. Leaving L at 10 henries but changing the frequency to 30 cycles,

$$X_L = 2\pi fL = 6.28 \times 30 \times 10 = 1884 \text{ ohms.}$$

We see that halving the circuit frequency halved the inductive reactance. Just as doubling and halving the circuit frequency doubled and halved the inductive reactance, the same would be true with doubling and halving the inductance of the coil. We say, then, that the inductive reactance of a coil varies directly with the frequency and with the inductance.

Working with a small 50-microhenry coil at a frequency of 4 megacycles, we get an inductive reactance through the coil of:

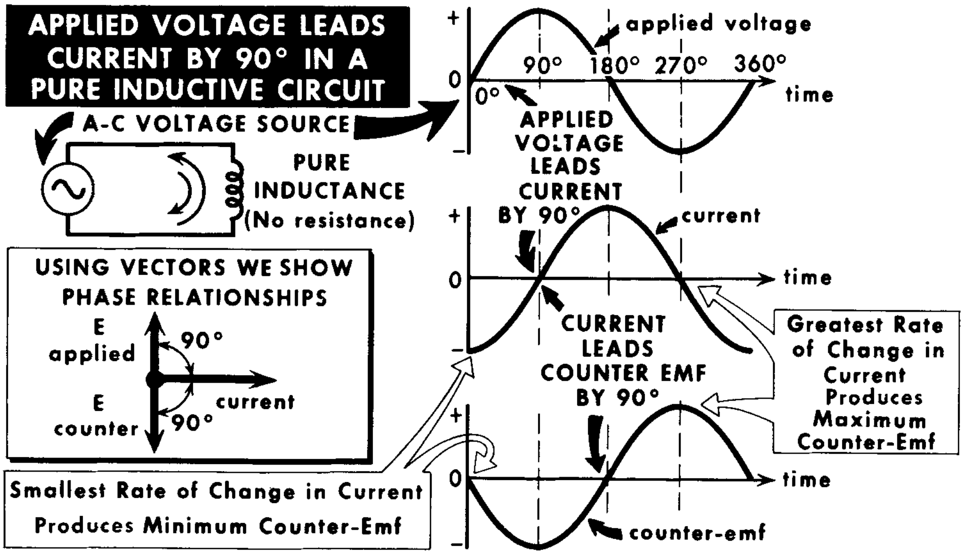
$$X_L = 2\pi fL = 6.28 \times 4{,}000{,}000 \times 0.00005 = 1256 \text{ ohms.}$$

Let us find the inductive reactance of a 1-millihenry coil at 10 kilocycles.

$$X_L = 2\pi fL = 6.28 \times 10{,}000 \times 0.001 = 62.8 \text{ ohms.}$$

Alternating Voltage and Current in an Inductive Circuit

In discussing voltage and current in an inductive circuit, we shall first assume an ideal inductance--one without any resistance. This will establish basic theory about a "pure" inductive circuit which, while it never occurs, enables us to understand practical inductive circuits. We learned that inductive reactance not only limits the amount of current flowing in an inductive a-c circuit, but also delays the increase or decrease of current in the circuit. The current in an inductive circuit takes the form of a sine wave if the applied voltage is of a sine waveform, except that the current is delayed or lags behind the voltage variations.



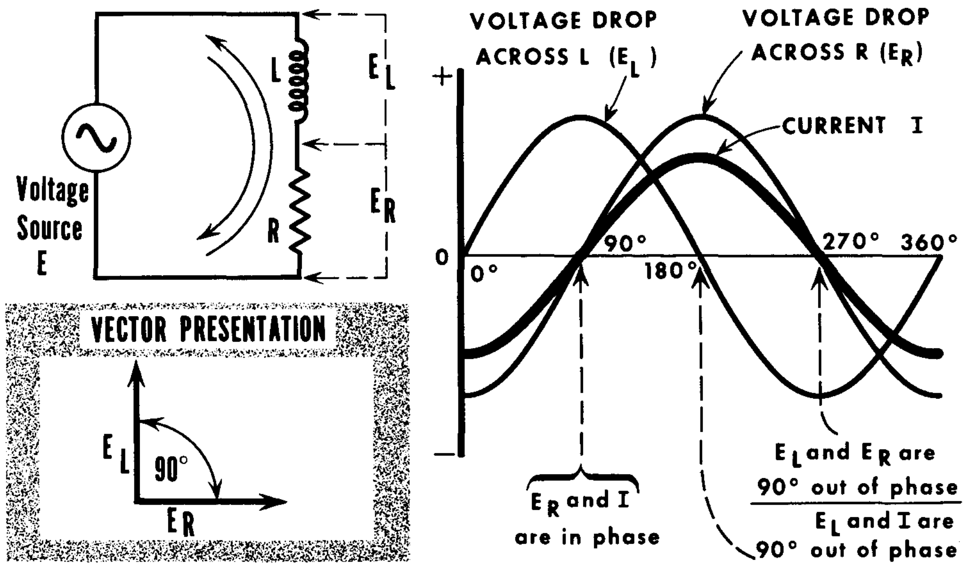
We learned that in an inductive circuit, a change in current produces a counter-emf which acts to oppose that change. It was established that the counter-emf was 180° out of phase with the applied voltage. We can now consider the relationship of the current in an inductive circuit to the applied voltage and the counter-emf. Since the counter-emf is induced by the changing current, it follows that the maximum counter-emf is induced when the current is changing at its greatest rate. We learned that the greatest rate of change in a sine waveform occurs when the waveform passes through its 0°, 180°, and 360° points; the least rate of change (but maximum value) occurs at its 90° and 270° points. Thus, when the current waveform is at maximum, for example, the counter-emf waveform will be zero; when the current waveform is at zero, the counter-emf is at maximum.

From this, we can see that there is a 90° phase difference between the current in an inductive circuit and the counter-emf it produces. Since the applied voltage is 180° out of phase with the counter-emf, there is a 90° phase difference between the applied voltage and the current. The applied voltage causes the current to flow, so we say that the applied voltage leads the current by 90°, or the current lags the applied voltage by 90°.

Alternating Voltage and Current in R-L Circuits

The practical coil consists of both inductance and resistance acting in series. We have just discussed the effect of inductance on voltage and current. Let us now review the effect of resistance on voltage and current in an a-c circuit. Since the property of resistance has no association with magnetic effects (actually, resistors contain some inductance), current flow through a resistance is assumed to be free of a magnetic field. The absence of a magnetic field prevents the self-induction of an emf; hence, a varying voltage applied to a resistance causes a simultaneously varying current. In other words, voltage and current are in phase in a resistance. We can use Ohm's law to find the current in an a-c resistive circuit just as in a d-c circuit, except that in an a-c circuit, we must think of I and E in the same terms--average, effective, or peak values.

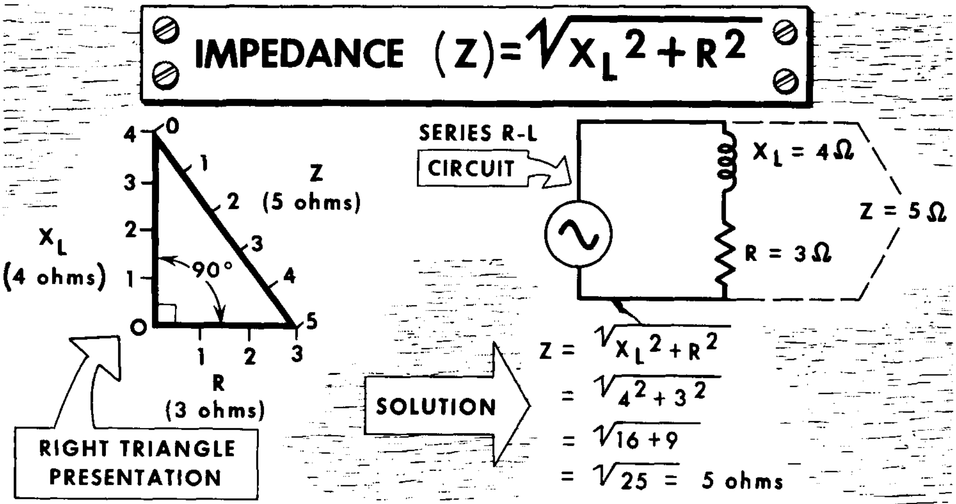
In a Series A-C Circuit Containing L and R, Voltage Drops across L and R are 90° Out of Phase



When an alternating voltage is applied to a practical coil, the same current (I) flows in the inductive and resistive parts of the coil. In flowing through the R-L circuit, the current produces two voltage drops--one across the inductance (E_L) and one across the resistance (E_R). The inductance voltage drop is equal to IX_L; the resistance voltage drop, to IR. With the same current flowing through the coil, E_L leads I by 90°, and E_R is in phase with I. Thus, voltage drops E_L and E_R are 90° apart, with E_L leading E_R. Since the effect of the inductance is to produce a voltage drop 90° out of phase with the current, and resistance produces a voltage drop in phase with the current, the net effect is that the resultant or applied voltage will lead the current in an R-L circuit by a phase angle 90° or less. Later, we will learn how to find the exact phase angle.

Impedance (Z)

Two sources of opposition to current flow exist in the practical inductor--one is inductive reactance (X_L), arising from the action of inductance (L); the other is resistance (R), arising from the nature of the conductor material. The combined actions of X_L and R constitute the total opposition to current flow known as impedance. Impedance is expressed in terms of ohms and is indicated by the letter Z.



THE IMPEDANCE OF A SERIES R-L CIRCUIT CAN NEVER BE EQUAL TO OR AS GREAT AS THE SUM OF X_L AND R, NOR CAN IT BE EQUAL TO OR LESS THAN EITHER X_L OR R.

The two current-opposition components--inductive reactance and resistance--are considered as being in series in an inductor. However, to find their impedance, we cannot add them arithmetically for our answer. To determine the impedance of a series R-L circuit, it is necessary to take into account the 90° phase difference between the voltage drops across the inductance and the resistance. This can be done in two ways: by using the right triangle equation; or by vectors (graphically). The equation method uses the right triangle as its basis. Early in this volume, we learned that if we squared the hypotenuse of a right triangle, that sum would be equal to the sum of the other two sides squared. When we consider the right triangle for calculating impedance, we make the vertical side, or altitude, represent the inductive reactance. The horizontal side, or base, represents the resistance. The hypotenuse which joins the ends of these two sides represents the impedance of the circuit. From this, we can see a basic formula for finding the impedance of a series R-L circuit: $Z^2 = X_L^2 + R^2$. To simplify this and find Z directly, we take the square root of both sides of the equation and get the highly usable formula: $Z = \sqrt{X_L^2 + R^2}$.

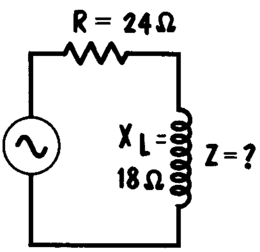
In a series R-L circuit, X_L and R must be considered 90° apart because the same current flows through R and L, but the voltage drops are 90° displaced.

Solving Impedance Problems

Let us now solve some impedance problems in order to familiarize ourselves with this procedure. In the first two, we will assume we already know the inductive reactance; in the third, we will work out the problem.

We have a series R-L circuit in which the inductive reactance is 18 ohms and the resistance is 24 ohms. Find the impedance of the circuit. Using the impedance formula $Z = \sqrt{X_L^2 + R^2}$, we first find the square of X_L. Since X_L is equal to 18 ohms, X_L^2 is equal to 18×18, or 324 ohms. R is equal to 24 ohms, so R^2 equals 24×24, or 576 ohms. We now add X_L^2 (324 ohms) and R^2 (576 ohms) and get a total of 900 ohms. Taking the square root of 900 gives us 30 (30×30 equals 900). Thus, this circuit's impedance is 30 ohms.

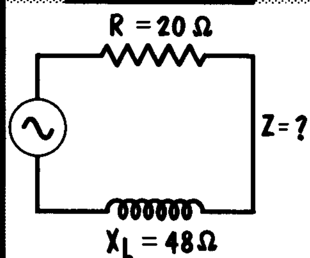
PROBLEM



SOLUTION

$$\begin{aligned} Z &= \sqrt{X_L^2 + R^2} \\ &= \sqrt{18^2 + 24^2} \\ &= \sqrt{324 + 576} \\ &= \sqrt{900} \\ &= \underline{30 \text{ ohms}} \end{aligned}$$

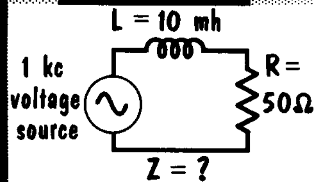
PROBLEM



SOLUTION

$$\begin{aligned} Z &= \sqrt{X_L^2 + R^2} \\ &= \sqrt{48^2 + 20^2} \\ &= \sqrt{2304 + 400} \\ &= \sqrt{2704} \\ &= \underline{52 \text{ ohms}} \end{aligned}$$

PROBLEM



SOLUTION

1st: FIND X_L

$$\begin{aligned} X_L &= 2\pi fL \\ &= 6.28 \times 1000 \times 0.01 \\ &= 62.8 \text{ ohms} \end{aligned}$$

2nd: FIND Z

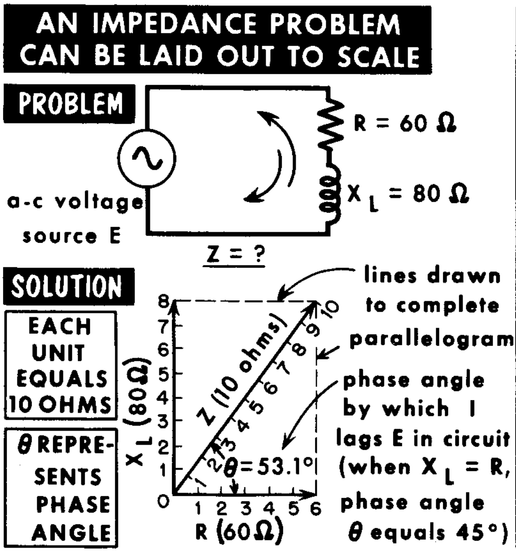
$$\begin{aligned} Z &= \sqrt{X_L^2 + R^2} \\ &= \sqrt{62.8^2 + 50^2} \\ &= \sqrt{3943.84 + 2500} \\ &= \sqrt{6443.84} \\ &= \underline{80.3 \text{ ohms}} \end{aligned}$$

We now have a series R-L circuit having an inductive reactance of 48 ohms and a resistance of 20 ohms. Find the impedance of the circuit. Once again, we first square X_L, and get 48×48, or 2304 ohms. We then square R, and get 20×20, or 400 ohms. Adding these two squared numbers, we get 2304 plus 400, or 2704. Taking the square root of 2704, we get 52 (52×52 equals 2704). The impedance is 52 ohms.

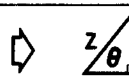
In our third problem, we have a series R-L circuit in which L is a 10 millihenry coil and R is a 50-ohm resistor. The frequency of the applied voltage is 1 kc (or 1000 cycles). Find the impedance of the circuit. We know the resistance is 50 ohms, so we must first find the inductive reactance. This is equal to: $X_L = 2\pi fL$. Filling in the formula, we get $X_L = 6.28 \times 1000 \times 0.01$, or 62.8 ohms. Knowing the inductive reactance is 62.8 ohms, we can now find the impedance of the circuit. We square 62.8 (62.8×62.8) and get 3943.84. We then square the resistance (50 ohms) and get 2500. Adding 3943.84 and 2500, we get 6443.84. Finally, taking the square root of 6443.84 we get 80.3 ohms, the impedance of this circuit.

Graphical Determination of Impedance (R and L in Series)

There is a simple graphical method that can be used to determine the impedance of an R-L circuit. It makes use of a parallelogram method that has certain advantages and disadvantages. To apply this method, let us first state a problem with which we can work. We will assume a series R-L circuit in which the inductive reactance is 80 ohms and the resistance is 60 ohms at the frequency of the voltage source. The problem--to find the impedance of the circuit. Instead of using the impedance formula, the problem is laid out to scale. We draw the vertical line which represents X_L to some exact length to represent the number "80". It could be 8-inches long, with each inch representing 10 ohms, or any other unit of measurement can be used. Then, with X_L being 8 units long, we draw the horizontal R axis exactly 6 units long to represent 60 ohms.



PRECISE METHOD FOR FINDING PHASE ANGLE (θ)

using triangle  $X_L = 8$, $R = 6$

$$\text{Tangent of angle } \theta = \frac{\text{opposite side}}{\text{adjacent side}} = \frac{X_L}{R} = \frac{8}{6} = 1.333$$

Looking in tangent column of trigonometry table, we find: 53.1° = 1.3319
53.2° = 1.3367

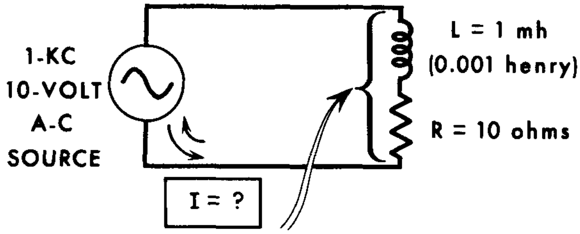
TO NEAREST TENTH OF DEGREE,
$\theta = 53.1°$

A parallelogram is now drawn, with one side parallel to the X_L axis and one side parallel to the R axis. Each new side is drawn from the end of the X_L and R axis. We now draw the resultant from the point where X_L and R meet to the diagonal corner of the parallelogram. This diagonal represents the resultant of the X_L and R vectors and indicates the impedance of the circuit. If the parallelogram were scaled and drawn properly, the resultant would be exactly 10 units long, representing 100 ohms impedance. Naturally, the same units of length would have to be applied to X_L, R, and Z in any given problem.

An advantage of this method is that it gives a quick, rough approximation. Its disadvantage is that it is somewhat clumsy and impractical where high precision is required. Another thing can be seen. The angle formed by the resultant and the R axis is the angle by which the current lags the voltage in an R-L circuit, and can be measured with a protractor. In our circuit, it is about 53.1°.

Alternating Current in an Inductor

When the impedance of a coil and the applied voltage are known, the alternating current flowing through the inductor can be readily calculated. Previously explained versions of Ohm's law for current are used except that R (resistance) in the equation is replaced by Z (impedance). Ohm's law as applied to a-c then reads: $I = E/Z$; $Z = E/I$; and $E = I \times Z$.



TO FIND CURRENT FLOW IN AN A-C CIRCUIT, WE USE OHM'S LAW FOR A-C CIRCUITS
$I = E/Z$

L and R are shown as separate components for the purpose of this problem only. Actually, in an inductor the two are inseparable since the same winding produces the resistance and the inductance.

1st FIND X_L:	2nd FIND Z:	3rd FIND I:
$X_L = 2\pi fL$	$Z = \sqrt{X_L^2 + R^2}$	$I = \frac{E}{Z}$
$= 6.28 \times 1000 \times 0.001$	$= \sqrt{6.28^2 + 10^2}$	$= \frac{10}{11.8}$
$= 6.28$ ohms	$= \sqrt{39.4 + 100}$	$= 0.847$ ampere
	$= \sqrt{139.4}$	
	$= 11.8$ ohms	

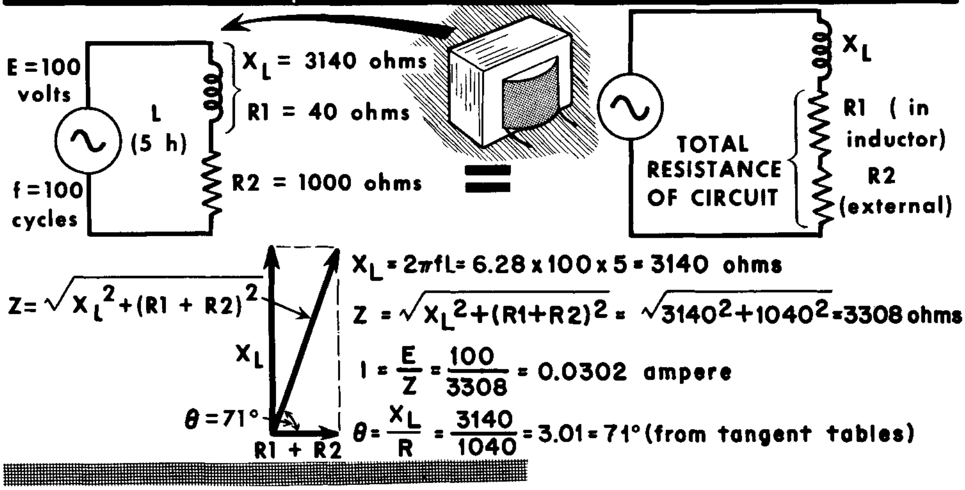
In each of these ratios, current and voltage must be expressed in the same terms. If we are considering the effective value of I, then we must consider the effective value of E. If we use peak values or average values of I, we must use like values of E. By so doing, it does not matter what values of E and I are being used. In virtually all instances, except where specifically noted, it is assumed that the effective value of E and I are being used. Note that nothing really new is being introduced--Ohm's law is still perfectly usable. The only thing new is that we must now consider other things such as X_L and Z since we are dealing with an inductive circuit.

The following problem illustrates a typical situation of an inductor in an a-c circuit. The resistance in the inductor is the actual d-c resistance of the copper wire that makes up the coil. Assume we have a 1-millihenry coil to which is applied a 1-kilocycle 10-volt a-c source. The d-c resistance of the inductor is 10 ohms. How much current flows through the coil? To solve the problem, we must first determine the inductive reactance and then the impedance. We find the inductive reactance using the formula $X_L = 2\pi fL = 6.28 \times 1000 \times .001 = 6.28$ ohms. We can now find the impedance: $Z = \sqrt{X_L^2 + R^2} = \sqrt{39.4 + 100} = \sqrt{139.4} = 11.8$ ohms. Now using Ohm's law for a-c circuits, we find the current: $I = E/Z = 10/11.8 = .847$ ampere. We can go one step further and find the phase angle Θ by which current lags the voltage: Tangent $\Theta = X_L/R = 6.28/10 = .628$. From the tangent table, $.6273 = 32.1^{\circ}$.

Determining the Current in an R-L Series Circuit

The theory you have learned about the series-connected R and L components of an inductor applies equally when an external resistance is series connected with the inductor. Now, two values of resistance are involved--the d-c resistance of the coil winding, and the external resistance. Assume a series-connected circuit in which inductance L = 5 henries, coil resistance R1 = 40 ohms, external resistor R2 = 1000 ohms, and the applied voltage is 100 volts at 100 cycles. Let us find the current in this circuit.

PRACTICAL INDUCTOR presents both INDUCTANCE and RESISTANCE to the circuit.



Before any calculating is done, let us examine the circuit. The coil resistance of R1 (40 ohms) is negligible (less than 1/10) relative to the external resistance of R2 (1000 ohms), but we shall take R1 into account just the same. Being series connected, we can visualize the resistance elements as a single sum R1 + R2. Then, the equation for impedance Z reads:

$$Z = \sqrt{X_L^2 + (R1 + R2)^2}$$

To calculate the impedance, we must first solve for the inductive reactance (X_L).

$$X_L = 2\pi fL = 6.28 \times 100 \times 5 = 3140 \text{ ohms}$$

Then:

$$\begin{aligned} Z &= \sqrt{3140^2 + (40 + 1000)^2} \\ &= \sqrt{3140^2 + 1040^2} \\ &= \sqrt{10,941,200} \\ &= 3308 \text{ ohms} \end{aligned}$$

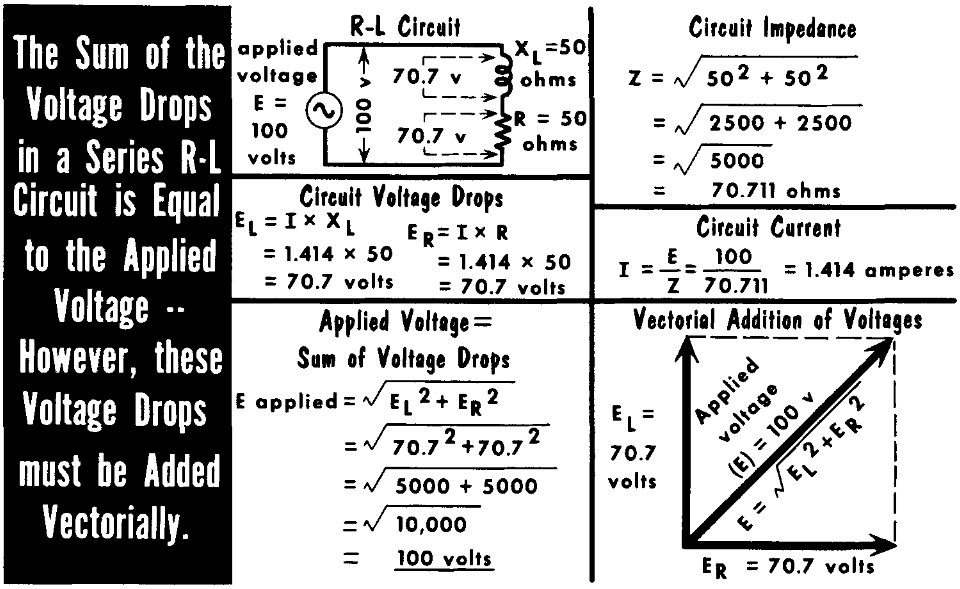
(If the coil resistance of R1 (40 ohms) is neglected, Z = 3296 ohms.)

$$\text{Current } I = \frac{E}{Z} = \frac{100}{3308} = .0302 \text{ ampere, or } 30.2 \text{ milliamperes.}$$

The phase angle of the current is: tangent $\theta = \frac{X_L}{R} = \frac{3140}{1040} = 3.01 = 71°$

Voltage Distribution in a Series R-L Circuit

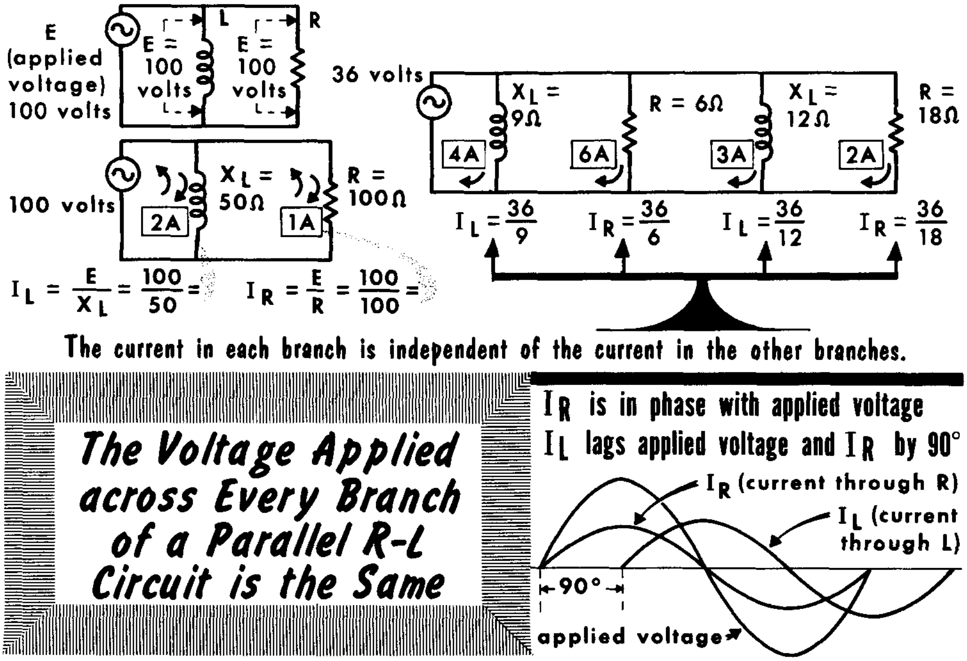
When we studied d-c electricity, it was established that the sum of all the voltage drops in a series circuit was equal to the battery or applied voltage. We discussed this further in Kirchhoff's laws. The same is true of the voltage drops in an R-L series a-c circuit, with one single exception. The voltage drops across R and L are not simply added together. The reason for this is that there is a 90° phase difference between the inductive voltage drop (E_L) and the resistive voltage drop (E_R). The 90° phase difference between these voltage drops is caused by the fact that while the same current flows through R and L in a series circuit, the current through R is in phase with the voltage, but through L, the current lags behind the voltage by 90°.



This does not present any new problem. Our fundamental rule that the sum of all the voltage drops in a series circuit is equal to the applied voltage still holds true. The only difference is that in order to get the sum of an inductive and a resistive voltage drop, it is necessary to add them vectorially. Since the same current flows through R and L, we can find the sum of their voltage drops in the same manner that we found the sum of their resistance and reactance, because the voltage drops are equal to IR and IX_L. We found the impedance of an R-L series circuit using the formula $Z = \sqrt{X_L^2 + R^2}$. By a very simple substitution, we can find the sum of two voltage drops 90° out of phase: E (applied) = $\sqrt{E_L^2 + E_R^2}$. Thus, while measuring the voltage drop across R and L separately, it would seem to give a ridiculous answer (greater than the applied voltage); vectorial addition of these voltages would give an answer equal to the exact applied voltage. Thus, Kirchhoff's laws for d-c circuits holds up equally well with a-c circuits. In an inductor, it is impossible to measure separately the voltage drop across R and L; we get one voltage drop across the R-L impedance.

The Parallel R-L Circuit--Voltage Distribution

The parallel R-L circuit consists of a voltage source across which an inductive element and a resistive element are connected. By definition, there must be one or more of each of these elements in this type of circuit. Once again, we assume that the inductive element has zero resistance. Of course, in practice this is impossible. Later, we will discuss parallel circuits in which a particular branch contains both R and L. First, let us analyze the voltage distribution of a parallel R-L circuit.



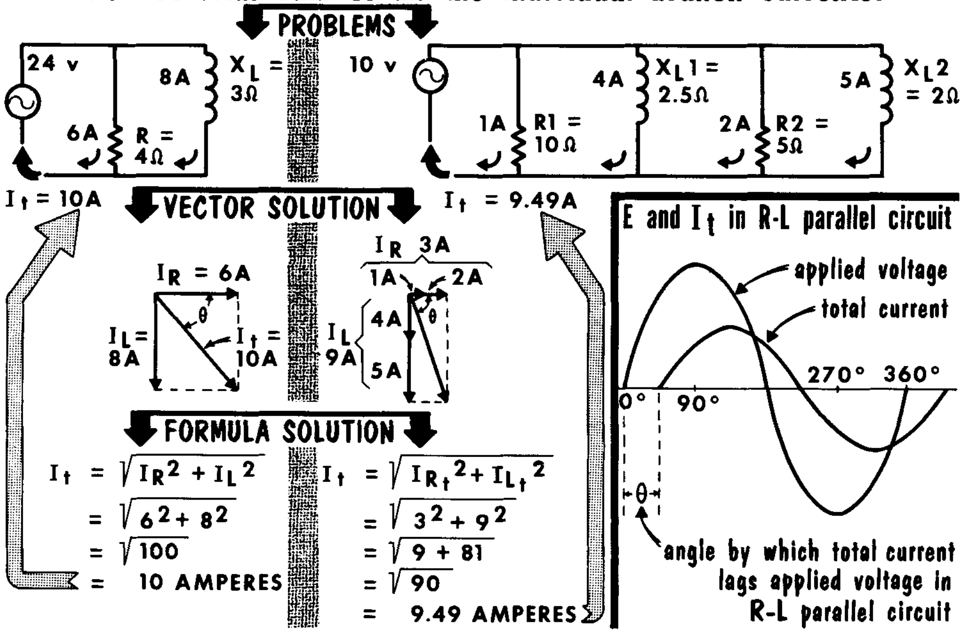
As in the case of the parallel circuit in our study of d-c circuits, a parallel circuit contains two or more branches. The applied voltage is across each and every branch of this circuit. Thus, a voltage or difference of potential equal to the full applied voltage is across each branch. In this respect, the current flow through each branch acts independently of the current flow in every other branch. Should one branch of the parallel circuit be opened, the stoppage of current flow in that branch would not affect the operation of any of the other branches; only the total current (I_t) would be affected.

The amount of current in each branch of a parallel circuit is determined by the voltage applied to that branch and the R or X_L of that branch. In short, the current in each branch would be equal to $I_R = E/R$ or $I_L = E/X_L$, as the case may be. The current flow in each branch must be treated separately. However, there is one important new consideration. The current through a resistive branch is in phase with the applied voltage; the current through an inductive branch lags the applied voltage by 90^o. This is an important consideration when computing the total current.

The Parallel R-L Circuit--Current Distribution

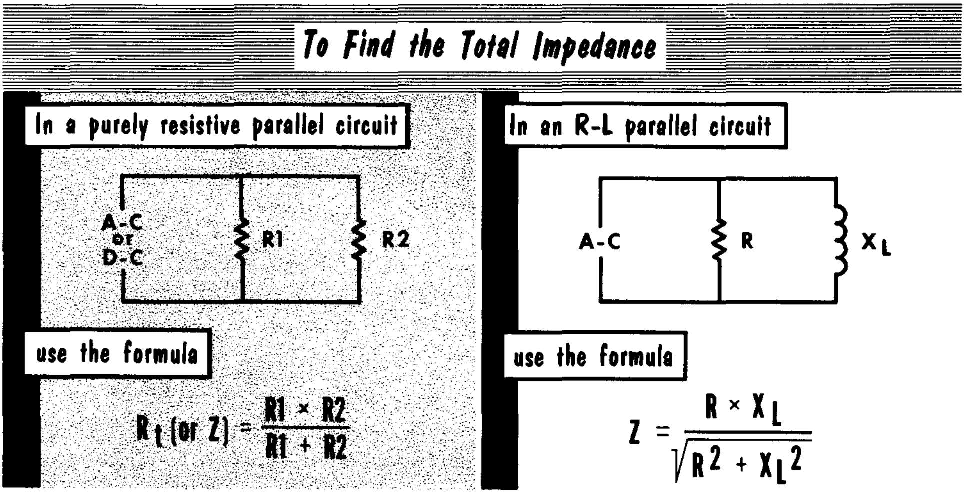
We have stated that the current flow in each branch of a parallel circuit is completely independent of the current flow in every other branch. The important difference between a purely resistive parallel circuit and an R-L parallel circuit is in finding the total current. In a purely resistive parallel circuit, we simply find the total of all the individual branch currents, and this sum equals the total current. But in the R-L parallel circuit, the currents in the inductive branches are 90° out of phase with the current in the resistive branches. Thus, once again we are faced with vectorial addition. We must first add up all the inductive currents and all the resistive currents, and then add them vectorially.

The total current in a parallel R-L circuit is equal to the vectorial sum of all the individual branch currents.

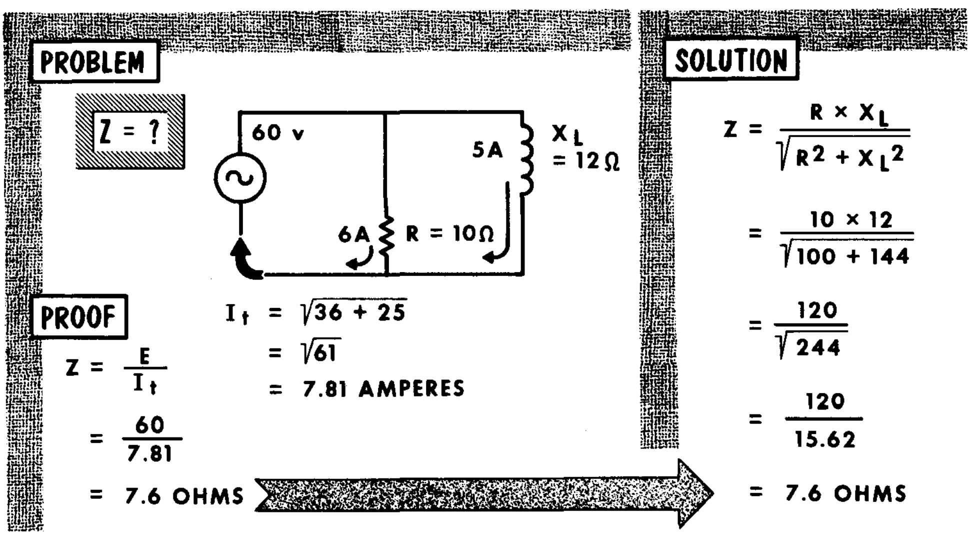


The total current in an R-L circuit can be found in two ways: by a graphical layout of the current vectors; and by direct formula. In the graphical representation, the current vectors for the inductive current (I_L) and the resistive current (I_R) are placed at right angles. Since the resistive current is in phase with the applied voltage, I_R is the reference point and is located on the horizontal axis. With the inductive current lagging 90° behind it, we place the inductive current vector straight down, representing a 90° lag behind I_R. Using the parallelogram, the resultant represents the total current and the angle of lag between the applied voltage and the total current. Using the formula for impedance, $Z = \sqrt{X_L^2 + R^2}$, which we developed for right triangle problems, we simply substitute and get: total current (I_t) $= \sqrt{I_L^2 + I_R^2}$. The angle of current lag behind the applied voltage is equal to: tangent $\Theta = I_L/I_R$.

Impedance of the Parallel R-L A-C Circuit



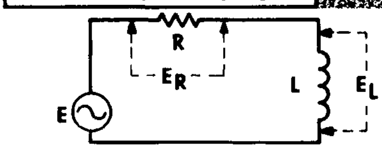
The impedance of the parallel R-L a-c circuit is computed by a method very much like that used for calculating the total resistance of resistors connected in parallel. We learned that to find the total resistance of two resistances in parallel, we used the formula $R_t = (R1 \times R2)/(R1 + R2)$. We can now substitute in this formula to bring our R-L circuit into play. To find impedance, we say $Z = (R \times X_L)/(R + X_L)$. However, the addition of two vector quantities, as we have seen, cannot be made by simple addition. Therefore, to take into consideration the fact that R and X_L must be added vectorially, we change the formula to: $Z = R \times X_L/\sqrt{R^2 + X_L^2}$. Using this formula, we would be accounting for the 90° phase difference between the currents in the resistive and inductive branches.



2-44 COMPARISON BETWEEN SERIES AND PARALLEL R-L CIRCUITS

SERIES R-L CIRCUIT	PARALLEL R-L CIRCUIT
The current is the same everywhere.	The current divides between the branches; each branch current is a function of the branch resistance or reactance.
The current is in phase throughout the circuit.	The current in the inductive branch lags the current in the resistive branch by 90°.
The voltage across the inductance leads the voltage across the resistance by 90°.	The voltage across the inductance is in phase with the voltage across the resistance.
The angle of lag between the total circuit current and the applied circuit voltage is determined by the amount of reactance and resistance.	The angle of lag between the line current and the applied voltage is determined by which circuit component is smaller--the reactance or the resistance.
Increasing the frequency makes the circuit more inductive because the inductive reactance exerts greater control on the circuit current. The angle of lag increases.	Increasing the frequency makes the circuit more resistive because the line current is predominantly the resistive branch current.
The applied voltage divides vectorially between the series reactance and resistance.	The applied voltage is the same across all parallel-connected elements.
Increasing resistances makes the circuit more resistive. The angle of lag of the circuit current approaches 0° more closely.	Increasing resistances makes the circuit more inductive. The angle of lag of the line current approaches 90° more closely.
Increasing the inductance makes the circuit more inductive. The angle of lag of the circuit current approaches 90° more closely.	Increasing the inductance makes the circuit more resistive. The angle of lag of the line current approaches 0° more closely.

THE SERIES R-L CIRCUIT

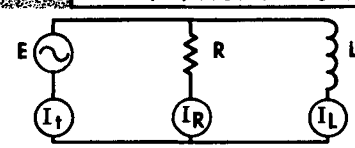


$E = E_R + E_L$ (added vectorially)

SAME CURRENT THROUGH R AND L

$$Z = \sqrt{R^2 + X_L^2}$$

THE PARALLEL R-L CIRCUIT



$I_t = I_R + I_L$ (added vectorially)

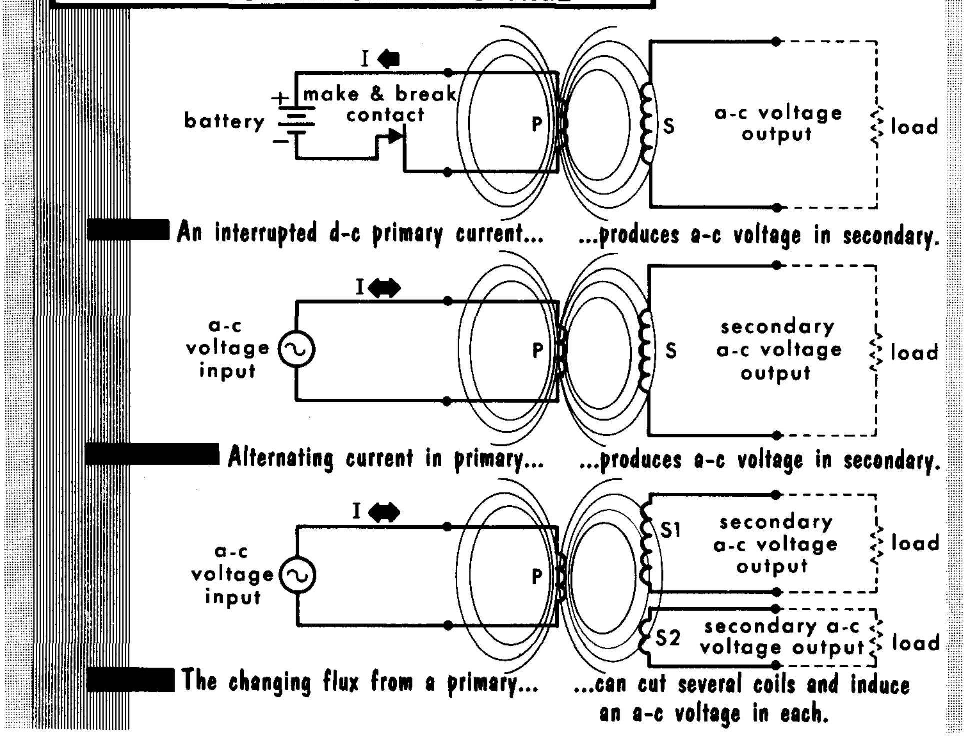
SAME VOLTAGES ACROSS R AND L

$$Z = \frac{R \times X_L}{\sqrt{R^2 + X_L^2}}$$

The Transformer

You have learned that an alternating current or a varying d-c current flowing in one coil can induce a varying voltage in a neighboring coil. The changing magnetic lines of force from the varying current in one coil (which we call the primary) cuts the turns of the other coil (called the secondary) and induces a changing voltage in each of the turns of the secondary. When two coil windings are arranged so that a changing current in one induces a voltage in the other, the combination of windings constitutes a transformer. Every transformer has a primary winding and one or more secondary windings. The primary winding (usually labeled P) receives the input electrical energy from a voltage source, whereas the secondary winding or windings (usually labeled S) delivers the induced output voltage to a load.

CHANGING FLUX LINES WHICH CUT TURNS OF A COIL INDUCE A VOLTAGE



The transformer serves many functions. It enables the transfer of electrical energy from one electrical circuit to another by using changing magnetic lines of force as the link between the two. In this way, it behaves as a coupling device. Also, it provides a means whereby an alternating voltage of a given amount can be changed (transformed) to higher or lower amounts, making electrical power distribution practical. Such transformation also can be applied to current and impedance. These functions are explained later.

Transformer Action (Unloaded Secondary)

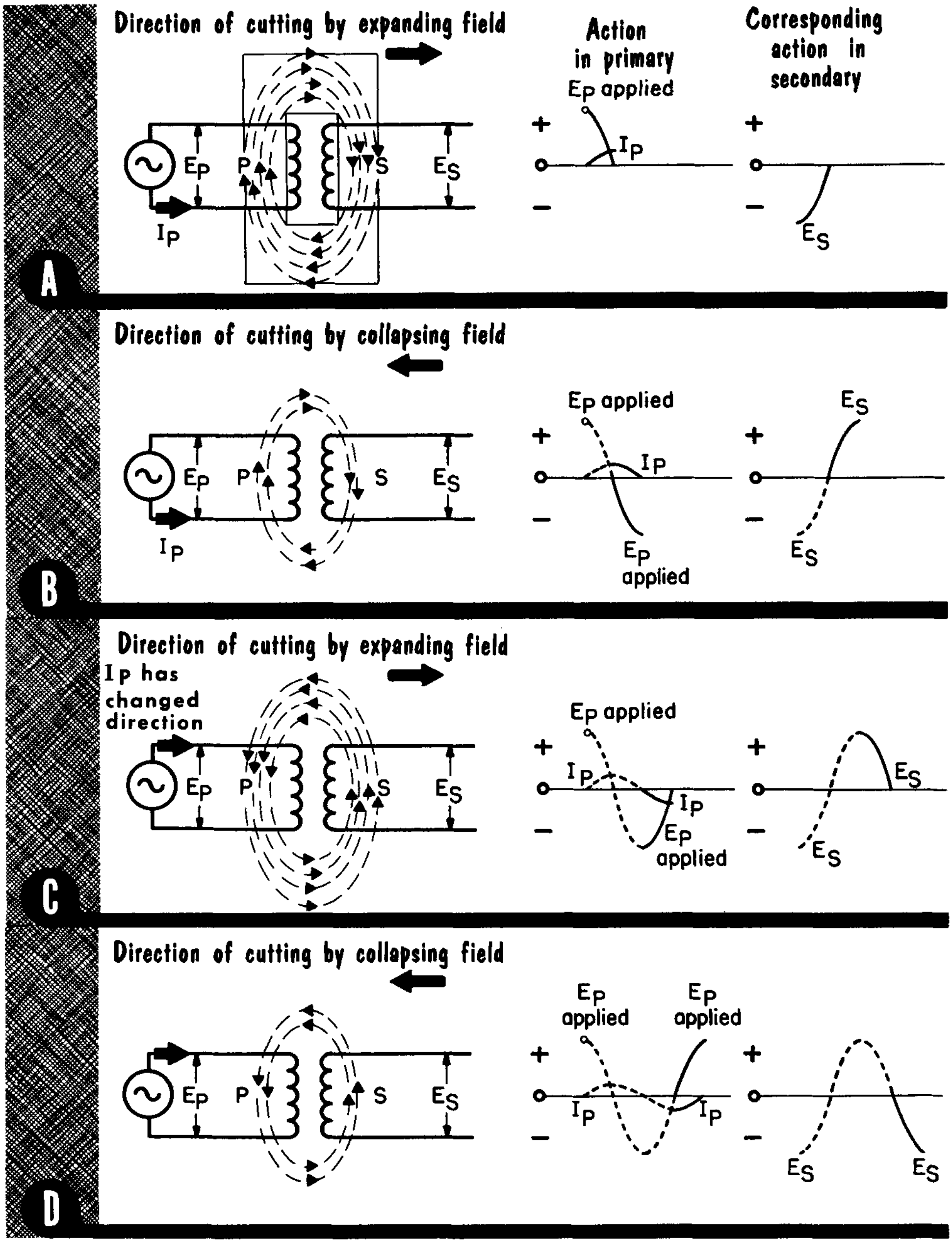
Assume a two-winding iron-core transformer with a primary (P) and a secondary (S), both of equal number of turns and of very low resistance. There is no load connected to the secondary winding. The a-c voltage applied to the primary is shown as a single cycle starting at maximum positive.

As the primary voltage E_p starts decreasing from maximum positive value, the primary current I_p starts increasing from zero in a positive direction. (See A on facing page.) The 90^o lag of primary current relative to primary voltage is due to the inductance of the primary winding. As the primary current starts increasing from zero, its associated magnetic field starts expanding. At this instant, the rate of change of the current, and of the flux, is maximum. The flux lines cut the turns of the primary winding and generate a self-induced emf of maximum value in the primary winding. This emf acts in opposition to the applied primary voltage E_p. Since there is nothing to prevent the generation of a maximum number of flux linkages, the self-induced emf is high, thereby causing the primary current I_p to be very low in value. Stated another way, the primary current is held low by the high inductive reactance of the iron-core primary.

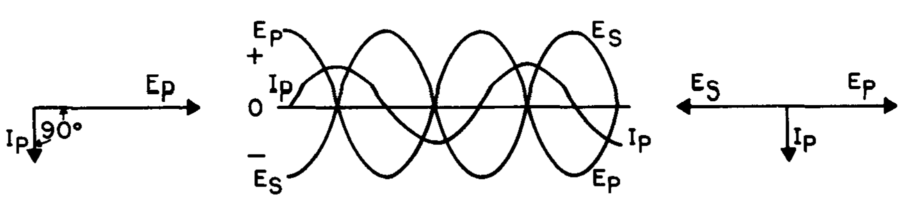
At the same instant in time (still A on the facing page), the expanding field produced by the primary current cuts the turns of the secondary (S), where it induces the secondary voltage, E_s. Inasmuch as the rate of change of the magnetic field is maximum, the voltage E_s is maximum. This secondary voltage has a polarity that is opposite to that of the primary voltage. It appears across the secondary, but since the secondary is unloaded (open) there is no secondary current. Hence, the action in the secondary has no effect on the action in the primary circuit. When the primary current reaches its maximum positive value, the rate of change of its field is theoretically zero; hence, the voltage induced in the secondary is zero. This coincides with the instant in time when the primary voltage E_p is zero.

As the primary voltage E_p passes through zero, changes polarity, and starts increasing towards its negative peak, the primary current I_p (still of the same direction) starts decreasing from maximum to zero, accompanied by the collapse of the magnetic field back into the primary. (See B.) The flux lines again cut the turns of the secondary winding, but now, in a direction opposite to that when the field was expanding. The result is a secondary voltage opposite in polarity to the previous voltage, and opposite to that of the applied primary voltage. As E_s increases towards its maximum positive value, the applied primary voltage increases towards its maximum negative value, both peaks being reached at the same instant. Also, at the same moment, the primary current I_p passes through zero.

The action of the transformer during the remainder of the primary voltage and current cycle is shown in C and D. It is the same as previously described except for the reversal in direction of the primary current. The expansion and collapse of the field is as before, during which time the remainder of the secondary voltage cycle is generated. At each instant of time, the secondary voltage is 180^o out of phase with the primary voltage. The primary current is 90^o behind the primary voltage, but 90^o ahead of the secondary voltage E_s.

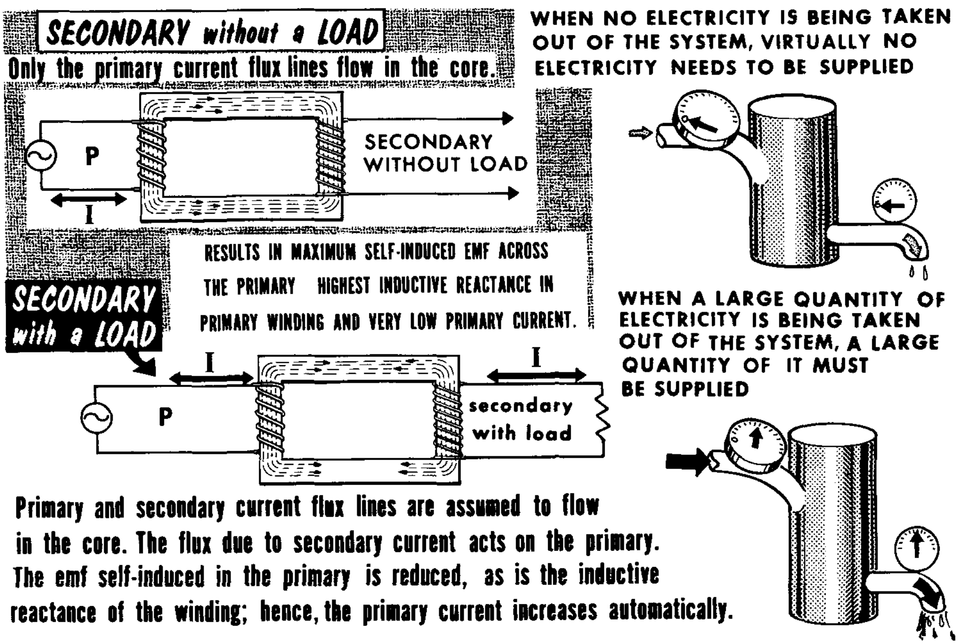


SUMMARY OF VOLTAGE-CURRENT CONDITIONS IN UNLOADED SECONDARY TRANSFORMER



Transformer Action (Loaded Secondary)

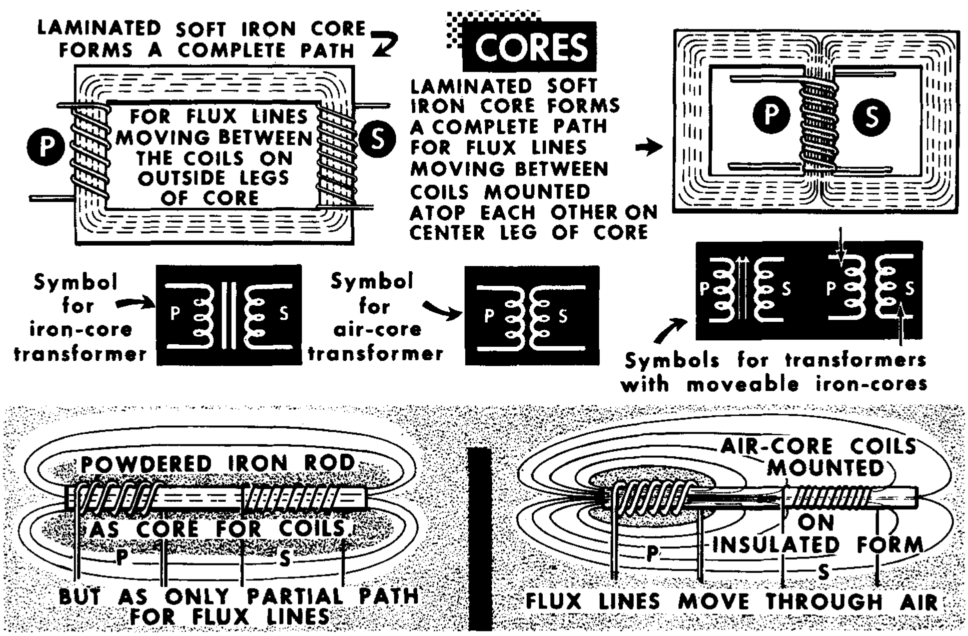
When the transformer is delivering voltage to a load, current flows in the load and in the secondary winding. This current affects the primary current. We will assume that the load connected to the transformer is a resistance R. When a voltage is induced in the secondary winding, a current flows through the load. This current also flows through the secondary winding. The load current is a current drain on the transformer. Like any other alternating current, the current in the secondary is accompanied by changing flux lines. The path for these flux lines is the transformer core, but the direction of these lines is opposite to that of the flux lines associated with the primary current. So, in effect, two sets of flux lines flow in the core--one due to the primary current, and the other due to the secondary current.



By virtue of its direction, the secondary current flux opposes the primary current flux. In doing so, the number of linkages which occur between the primary current flux and the primary turns, and which accounts for the self-induced emf in the primary, is reduced. This action is like a self-regulating valve that permits the primary current to increase above the small amount which flows when there is no load on the secondary. The amount of increase in primary current is determined by the amount of current drawn from the secondary winding. In other words, when the secondary winding delivers power (E × I) to a load, the primary winding draws more power from the voltage source than when the secondary winding is not delivering power to a load. With the primary voltage being fixed in value, the increase in power required by the primary appears as an increase in primary current. Of course, the converse is true--if the secondary current drain decreases, the primary current automatically decreases to adjust to the new situation.

Iron-Core and Air-Core Transformers

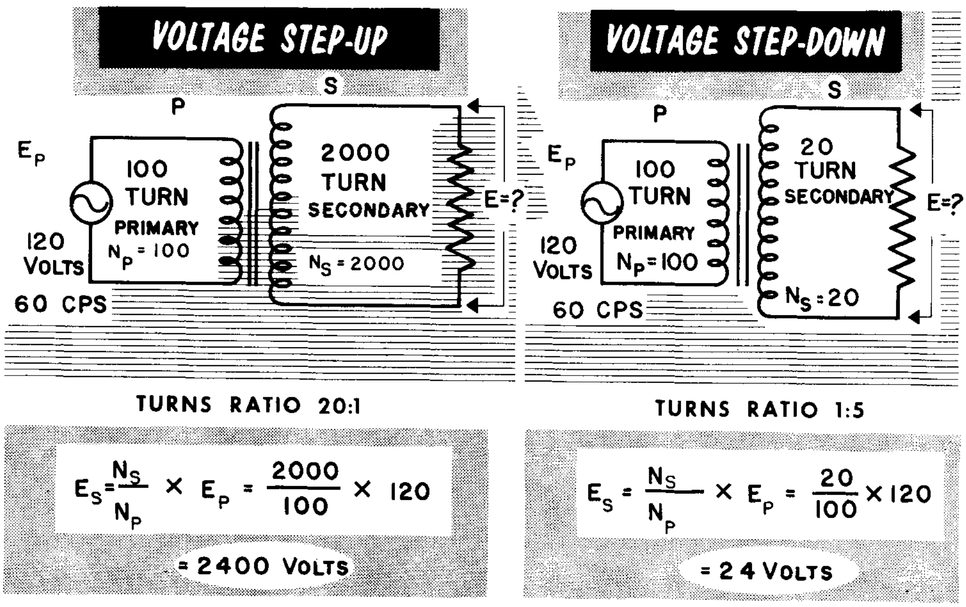
There are many kinds of transformers. In a broad sense, they fall into two categories--iron-core and air-core. Each category has numerous subdivisions relating to its particular uses. The two types mentioned state the kind of material that serves as the path over which the magnetic lines of force travel from the primary to the secondary and in the reverse direction. Since soft iron is a much better path for magnetic lines of force than air, it is used as the core for the transformer windings, except at very high frequencies. The iron core conducts most of the flux lines originating from the primary current in the primary winding to the turns of the secondary winding, thereby allowing the maximum number of flux linkages (tight coupling between the windings), or the transfer of the greatest amount of electrical energy from the primary to the secondary. To improve the action in many iron-core transformers, the primary and secondary windings are wound on top of each other. Some iron-core transformers contain a powdered-iron core in the form of a rod on which the primary and secondary coils are located side by side. The flux linkage is reduced; hence, the coupling is not as tight. Transformers of this type may have iron cores that are movable for "tuning" (varying inductance) purposes. The presence of a ferrite (iron) material as the core in a transformer is symbolized by two or more straight lines located in the space between the symbols for the coils, or above or below them.



The air-core transformer has its coils wound on insulated forms that use air as the path for the flux lines moving between the windings. The absence of the iron core provides low values of inductance and limited flux linkages; hence, very loose coupling between the coils. Such transformers are used at very high frequencies. They are discussed at greater length later in this course.

Voltage Step-Up and Step-Down in Transformers (Turns Ratio)

One of the fundamental considerations in transformers is the amount of voltage derived from the secondary winding relative to the amount of voltage that is applied to the primary winding. If the voltage output from the secondary winding is higher than the voltage applied to the primary winding, a voltage step-up has taken place; if the secondary voltage is less than the primary voltage, a voltage step-down has taken place. Some transformers are designed to furnish both voltage step-up as well as voltage step-down. When a transformer is desired with a single secondary winding intended to furnish a voltage equal to the primary voltage, the voltage transformation is 1-to-1 (1:1) and the device is known as an "isolation" transformer. Its only function is to isolate one circuit from the other.



Whether the secondary voltage exceeds the primary voltage or is less than the primary voltage is determined by the turns ratio between the secondary winding and the primary winding. This is expressed as an equation as follows:

$$\frac{\text{secondary voltage}}{\text{primary voltage}} = \frac{\text{number of turns in secondary winding}}{\text{number of turns in primary winding}}\ ;\ \frac{E_S}{E_P} = \frac{N_S}{N_P}$$

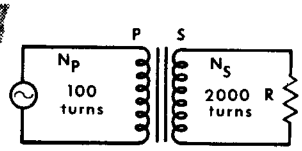
As you can see, the secondary-primary turns ratio equals the secondary-primary voltage ratio. When the number of turns in the secondary (N_S) exceeds the number of turns in the primary (N_P), or N_S is greater than N_P, a voltage step-up occurs. When the reverse is true, a voltage step-down occurs. The actual voltage derived from the secondary winding is equal to the product of the secondary-primary turns ratio and the voltage applied to the primary. The above assumes perfect (100%) coupling between primary and secondary. This is seldom the case. However, in some power transformers, coupling is almost perfect.

Current Turns Ratio

The secondary-primary turns ratio determines the amount of primary current that will flow for a given secondary current. You have learned that the more the number of turns (N) in a coil through which a current (I) in amperes is flowing, the greater the number of flux lines that are established by the current. The product of the number of turns and the current (or N × I) was identified as the ampere-turns. In the ideal transformer, the number of ampere-turns in the primary equals the number of ampere-turns in the secondary. Imagine that you are working with a loaded transformer in which the primary winding has 100 turns and the secondary winding has 2000 turns. The secondary-primary turns ratio then is 2000/100 = 20. If the secondary load current is 0.2 ampere, the secondary ampere-turns are 2000 × .2 = 400. For the same number of ampere-turns to exist in the primary of 100 turns, the primary current must be increased in the same proportion as the secondary-primary turns ratio. This ratio is 20; hence, the primary current must be 20 times greater than the secondary current. The self-regulating action of the primary winding establishes the primary current at the required value.

The Primary-Secondary Current Relationship

IN A VOLTAGE STEP-UP TRANSFORMER



Assume

I_S = 0.2 amp load current

TURNS RATIO = N_S/N_P = 2000/100 = 20

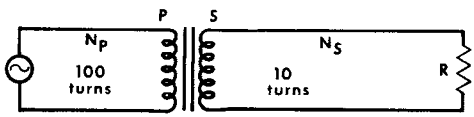
then

$$I_P = \frac{N_S}{N_P} \times I_S = \frac{2000}{100} \times 0.2 = 4 \text{ amperes}$$

PRIMARY AMPERE-TURNS = $N_P \times I_P = 100 \times 4 = 400$

SECONDARY AMPERE-TURNS = $N_S \times I_S = 2000 \times 0.2 = 400$

IN A VOLTAGE STEP-DOWN TRANSFORMER



Assume

I_S = 2 amps load current

TURNS RATIO = N_S/N_P = 10/100 = 0.1

then

$$I_P = \frac{N_S}{N_P} \times I_S = \frac{10}{100} \times 2 = 0.1 \times 2 = 0.2 \text{ ampere}$$

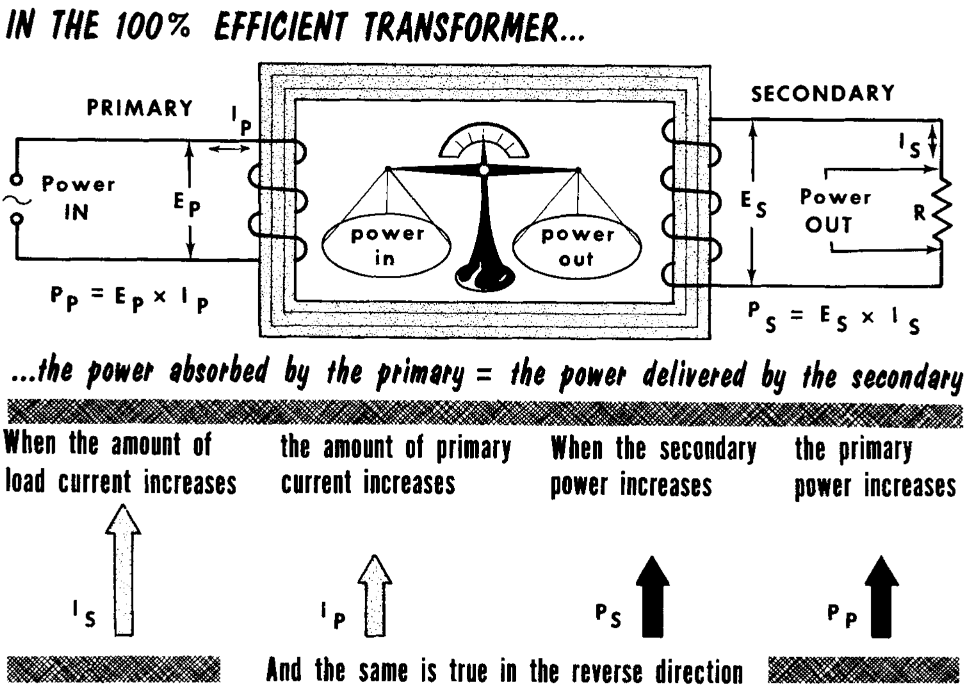
PRIMARY AMPERE-TURNS = $N_P \times I_P = 100 \times 0.2 = 20$

SECONDARY AMPERE-TURNS = $N_S \times I_S = 10 \times 2 = 20$

The same conditions hold true when the secondary winding has fewer turns than the primary winding. In this case, the equality of ampere-turns is gained by the higher current in the secondary of fewer turns, and the lower current in the primary of more turns. Thus, the primary-secondary current ratio is opposite to that of the primary-secondary voltage ratio. A 1:10 voltage step-up transformer will exhibit a 10:1 current step-down characteristic. We see that E × I in the primary will equal E × I in the secondary. In short, primary power equals secondary power.

Current Transformation (Power)

A transformer is not a generator of electrical power; its primary circuit absorbs power from a voltage source and its secondary circuit delivers power to a load. Assuming an ideal transformer with a resistive load in the secondary, the power absorbed by the primary equals the power consumed by the secondary. (In practice, the power consumed by the secondary is slightly less than that absorbed by the primary, the difference being due to electrical losses in the transformer, as will be explained later.)

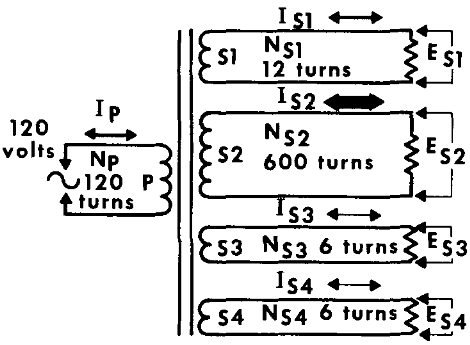


The ideal input-output power relationship is stated as: power (in watts) in primary = power (in watts) in secondary, or P_P watts = P_S watts. Since power is equal to voltage times current ($E \times I$), the power relationship can be restated as: $E_P \times I_P = E_S \times I_S$. Assume a loaded transformer with a 1:1 turns ratio. Then, the primary voltage E_P and the secondary voltage E_S will be equal. Whatever the secondary load current I_S may be, the primary current I_P will adjust itself to the same value so as to satisfy the condition $E_PI_P = E_SI_S$, and the power is the same in both circuits. If the load on the secondary is changed, thereby changing secondary current I_S, the primary current I_P will readjust itself to be equal to I_S. For any amount of power delivered by the secondary within the capabilities of the transformer, the primary circuit behaves as a self-regulating system in which the current changes in value so that the primary circuit power equals the secondary circuit power. The primary current change is the result of the increased or decreased action of the secondary current flux lines on the primary current flux lines; hence, on the emf self-induced in the primary.

Multi-Secondary Winding Transformers

A broad category of iron-core transformers is known as "power transformers." They have several secondary windings and a single primary winding which is common to all the secondaries. The purpose of these transformers is to supply a number of operating voltages required by radio communications equipment. Sometimes, as many as two voltage step-up secondaries and three voltage step-down secondaries are part of the same transformer. The voltage derived from each secondary winding is independent of the others, its amount being determined by the individual secondary-common primary turns ratio.

VOLTAGE RATIOS in MULTI-SECONDARY WINDINGS



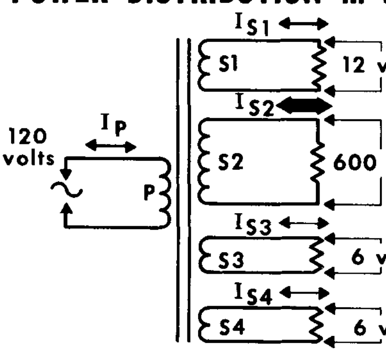
$E_{S1} = \frac{N_{S1}}{N_P} \times E_P = \frac{12}{120} \times 120 = 12$ volts

$E_{S2} = \frac{N_{S2}}{N_P} \times E_P = \frac{600}{120} \times 120 = 600$ volts

$E_{S3} = \frac{N_{S3}}{N_P} \times E_P = \frac{6}{120} \times 120 = 6$ volts

$E_{S4} = \frac{N_{S4}}{N_P} \times E_P = \frac{6}{120} \times 120 = 6$ volts

POWER DISTRIBUTION in MULTI-SECONDARY WINDINGS



12 volts @ 2 amps = $P_{watts} = 12 \times 2 = 24$ watts

600 volts @ 0.1 amp = $P_{watts} = 600 \times 0.1 = 60$ watts

6 volts @ 4 amps = $P_{watts} = 6 \times 4 = 24$ watts

6 volts @ 1 amp = $P_{watts} = 6 \times 1 = 6$ watts

Secondary winding powers are additive and equal to primary power.

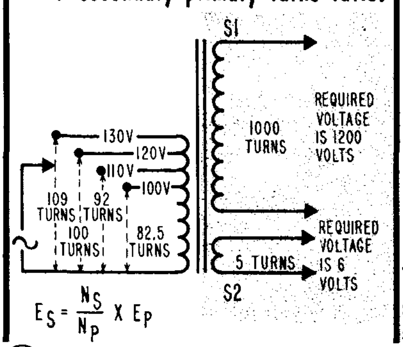
$P_{primary} = P_{S1} + P_{S2} + P_{S3} + P_{S4} = 24 + 60 + 24 + 6 = 114$ watts

Concerning the power consumed by the secondary windings relative to the power absorbed by the primary winding, if we assume the ideal case, the power in the primary equals the arithmetical sum of the power delivered by each of the secondary windings, or $P_P = P_{S1} + P_{S2} + P_{S3}$. In the practical case, the primary power may be from 5 to 10% higher than the sum of the power delivered by the secondary windings, this much being wasted as electrical losses in the transformer.

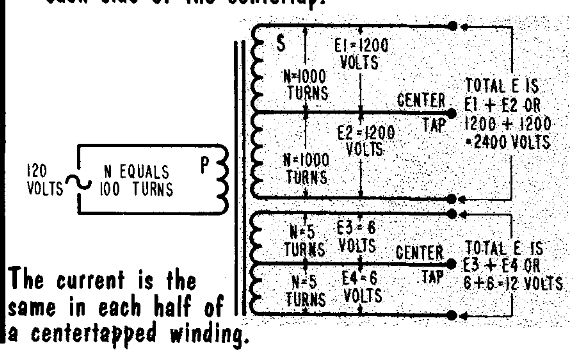
Tapped Primary and Secondary Windings

Some types of iron-core transformers have a multi-tapped primary winding and a centertapped secondary winding. A tap is simply a wire connection to the winding. Usually, for convenience, it is joined to a terminal. The tapped primary is a continuously wound coil that affords a selection of the number of primary turns which are active during the operation of the transformer. It permits the use of the transformer over a range of primary voltages for a given secondary voltage. Assume a transformer that is rated 1200 volts from one secondary winding (S1) and 6 volts from another secondary winding (S2) with 120 volts a-c applied to the primary. Winding S1 has 1000 turns and S2 has 5 turns. To satisfy the above voltage conditions, the primary must have 100 active turns. Now suppose that the available primary voltage is only 110 volts. If this voltage were applied to the transformer, the two secondaries would deliver less than the required 1200 and 6 volts. But if we increased the secondary-primary turns ratio by the correct amount, it would compensate for the reduced primary voltage. The taps on the primary permit such a change in turns ratio to allow for the condition when the primary voltage is either lower or higher than the rated optimum voltage. (See A).

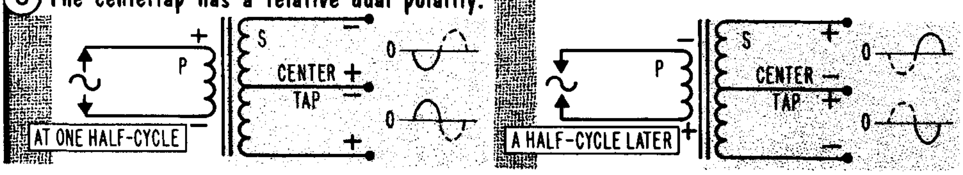
(A) Tapped primary affords changes in secondary-primary turns ratio.



(B) Centertapped secondary affords equal voltages on each side of the centertap.



(C) The centertap has a relative dual polarity.

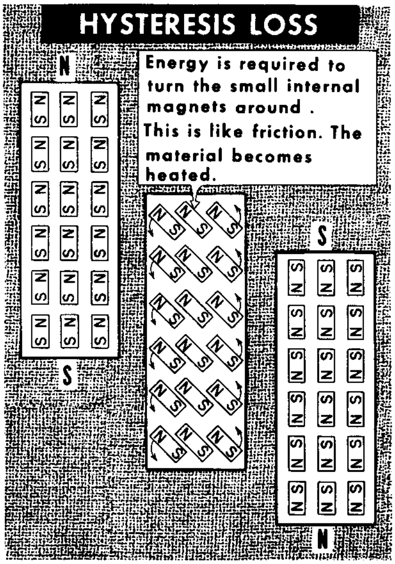


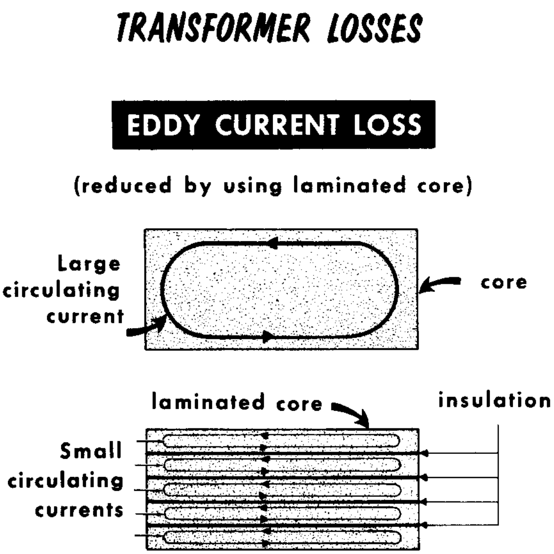
The centertapped secondary (B) is simply a means of achieving equal voltages on both sides of a common reference point--the centertap--to suit certain operating conditions. The voltage available from each half of the winding relative to the primary voltage is determined by the turns ratio between each half of the winding and the primary. At any one instant, the polarity of the voltage available from the whole centertapped winding is 180° out of phase with the primary polarity, but the centertap behaves as if it has a dual polarity (C). With respect to the centertap, the secondary provides two equal voltages 180° out of phase with each other.

Transformer Losses

We have stated that the transformer can be made almost a 100% efficient device. There are, however, certain inherent losses in a transformer that can be minimized but never completely eliminated. The most apparent losses are called copper losses. Since the primary and secondary are wound with many turns of copper wire, there will be wasted I^2R losses. This accounts for the secondary voltage being slightly lower under load than when unloaded. These losses are reduced by using the largest practical cross-sectional area wire.

Hysteresis losses are due to the lagging of the magnetization and demagnetization of the soft steel core behind the alternating current in the circuit. The atoms of the core material must keep changing polarity, and a sort of frictional loss is developed. The use of material such as soft silicon steel for the core greatly reduces hysteresis losses.



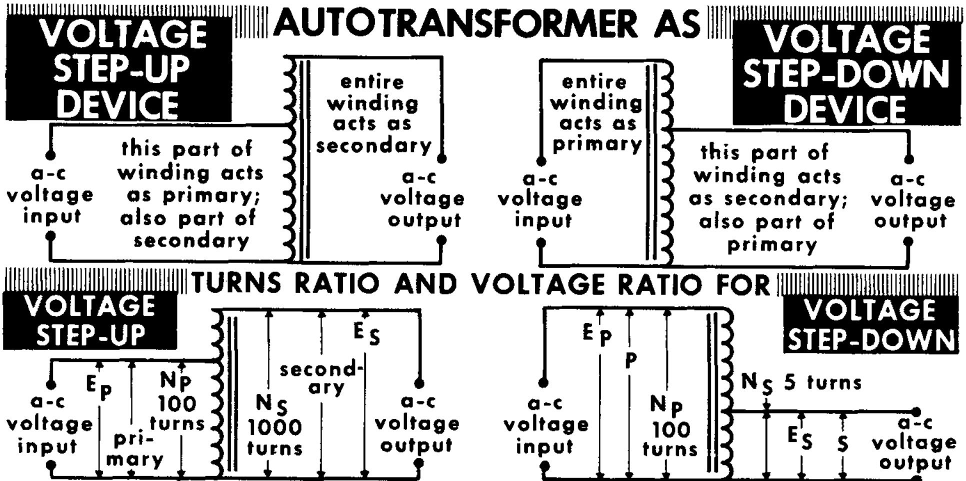


Magnetic core materials cause additional transformer losses because, as conductors, small short-circuited currents called eddy currents are induced in them. To reduce these currents, transformer cores are laminated. Each lamination (strip) is sprayed with an insulated coating so that the d-c resistance between them is very high. The strips are then pressed together to form the core.

A source of inefficiency stems from the fact that all the lines of flux produced by the primary and secondary windings do not move through the iron core--some leak directly out of the windings into space and do not link the windings. This is known as flux leakage. Another core inefficiency occurs during core saturation. Above a certain point, an increase in magnetizing force causes no additional magnetization. Thus, more magnetizing current is being used than required, resulting in a loss in efficiency.

The Autotransformer

The iron-core autotransformer differs from the multi-coil device we have been studying. Its operation, as before, depends on flux lines being produced by the primary current, cutting the turns of the secondary and inducing a voltage. However, its advantage over the conventional transformer is that the secondary voltage remains substantially constant when the load on the secondary is changed. However, the disadvantage of the autotransformer is that its primary and secondary are part of the source coil; hence, there is no isolation of the external circuits that are connected to the primary and the secondary. This does not prevent the use of the autotransformer in many circuits.



Turns ratio $= \frac{N_S}{N_P} = \frac{1000}{100} = 10$

Secondary voltage $E_S = \frac{N_S}{N_P} \times E_P$

If $E_P = 120$ volts, $E_S = \frac{1000}{100} \times 120$
$= 1200$ volts

Turns ratio $= \frac{N_S}{N_P} = \frac{5}{100} = 0.05$

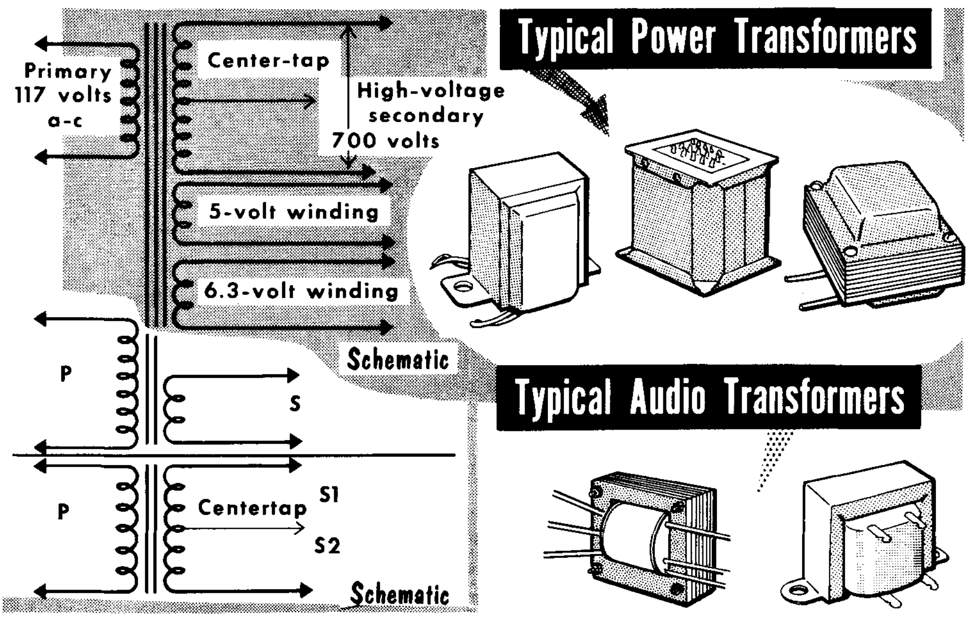
Secondary voltage $E_S = \frac{N_S}{N_P} \times E_P$

If $E_P = 120$ volts, $E_S = \frac{5}{100} \times 120$
$= 6$ volts

When used as a voltage step-up transformer, the entire winding is the secondary and a part of the winding is the primary. When used as a voltage step-down transformer, the entire winding is the primary and part of the winding is the secondary. Obviously, one part of the winding is always common to both the primary and secondary functions. The turns ratio between the secondary and primary portions of the winding determines the output (secondary) voltage relative to a given input (primary) voltage, just as in the case of the conventional transformer. In this regard, note that a certain number of turns is common to both the so-called secondary and primary windings. This does not change the usual manner of determining the turns ratio--the secondary and primary are considered as if each were separate and individual. Usually, taps on the winding permit changing the secondary-primary turns ratio; hence, the output voltage.

Transformer Applications

The transformer truly has an unusual variety of applications in radio communications. We will discuss a few here, and many others will be covered during this course when applicable. The most commonly used is the power-transformer type. This transformer is used in the power supply of electronic equipment to furnish the various a-c voltages necessary for the production of d-c voltages, and for the operation of particular circuits. The basic power transformer has a single primary winding, with two or more secondary windings. One secondary winding usually provides high-voltage a-c for the rectifier, and one or more others provide various filament voltages for tubes (these terms will be discussed later in the course). The high-voltage winding commonly provides from 400 to 800 volts a-c at currents from 25 to 400 milliamperes, and is usually center-tapped. The filament windings usually provide 6.3 or 12.6-volts at 1 to 5 amperes, and 5 volts at 2 or 3 amperes.

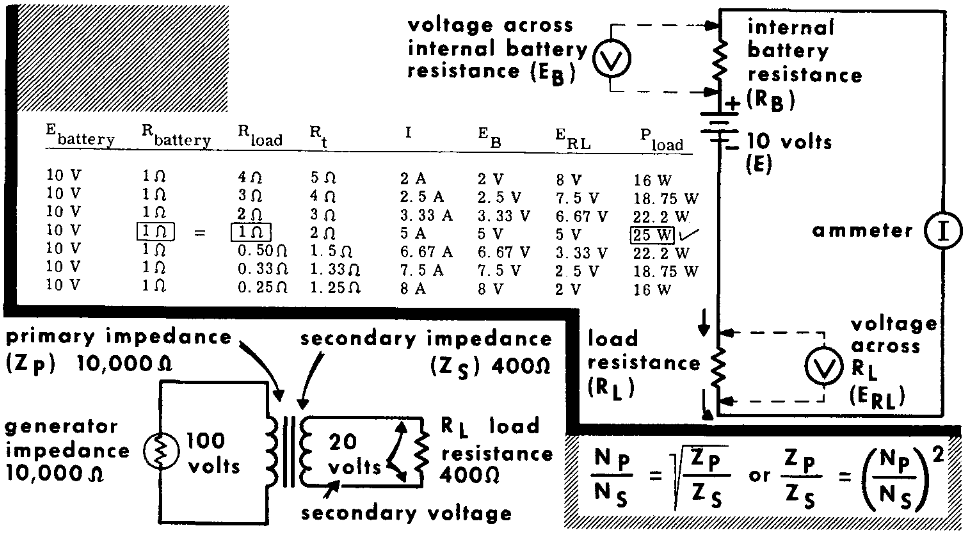


Another commonly used transformer is the audio type. Designed to operate at audio frequencies (20-16,000 cycles), it is usually smaller than the power transformer and has a wide variety of functions. These transformers are used primarily for impedance matching and, in some instances, for voltage amplification. Audio transformers are usually designated by their particular application--input transformer, output transformer, microphone transformer, modulation transformer, interstage transformer, driver transformer, etc. Usually, they are rated by their primary and secondary impedances and current-handling capabilities.

Transformers designed to operate at high frequencies (above audio range) are referred to as intermediate-frequency and radio-frequency types, and will be discussed where applicable in this course.

Impedance Matching

In the transfer of power from any electrical source of its load, the impedance of the load must be equal to or match the internal impedance of the source for maximum transfer of power. From our table, we see how this is so. Assuming a 10-volt battery having an internal resistance of 1 ohm, we connect various loads ranging from 0.25 ohm to 4 ohms. From the calculations, it is seen that the greatest amount of power is delivered to the load when the load has a resistance or impedance of 1 ohm--the same as the internal impedance of the battery or voltage source.



$E_{battery}$	$R_{battery}$		R_{load}	R_t	I	E_B	E_{RL}	P_{load}
10 V	1 Ω		4 Ω	5 Ω	2 A	2 V	8 V	16 W
10 V	1 Ω		3 Ω	4 Ω	2.5 A	2.5 V	7.5 V	18.75 W
10 V	1 Ω		2 Ω	3 Ω	3.33 A	3.33 V	6.67 V	22.2 W
10 V	$\boxed{1\ \Omega}$	=	$\boxed{1\ \Omega}$	2 Ω	5 A	5 V	5 V	$\boxed{25\ W}$ ✓
10 V	1 Ω		0.50 Ω	1.5 Ω	6.67 A	6.67 V	3.33 V	22.2 W
10 V	1 Ω		0.33 Ω	1.33 Ω	7.5 A	7.5 V	2.5 V	18.75 W
10 V	1 Ω		0.25 Ω	1.25 Ω	8 A	8 V	2 V	16 W

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}} \quad \text{or} \quad \frac{Z_P}{Z_S} = \left(\frac{N_P}{N_S}\right)^2$$

The transformer is a useful device for matching the impedance of a generator to that of its load. This is important because in radio work, it is often necessary to connect a low-impedance load to a high-impedance generator, and vice versa. Unless there is an impedance match, there will not be maximum transfer of power. Assuming a source or generator impedance of 10,000 ohms, we will match it to a load of 400 ohms. Using a transformer, the primary impedance must match the generator impedance, and the secondary impedance must match the load impedance. The turns ratio of the transformer must be:

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}} = \sqrt{\frac{10,000}{400}} = \sqrt{\frac{25}{1}} \qquad \text{Then, } \frac{N_P}{N_S} = 5:1$$

If 100 volts are applied to the primary, the secondary voltage is 20 volts. Secondary current is 20/400, or .05 ampere. The primary current is 100/10,000, or .01 ampere. Since the primary power (1 watt) is equal to the secondary power, the transformer has matched a 400-ohm load to a 10,000-ohm source with maximum transfer of power. We can say that the source "sees" the primary impedance as a matching impedance, and the secondary, which by transformer action receives the primary power, "sees" the load impedance as a matching impedance.

Any change of current flowing in a circuit containing inductance produces a counter emf which opposes the change taking place. This self-induced emf tends to prevent an increasing current from increasing and a decreasing current from decreasing.

Basically, Lenz's law states, "A changing current induces an emf whose polarity is such as to oppose the change in current."

The greater the coil inductance, the higher the induced emf, and the greater the opposition to the increase and decrease of current in the coil.

Mutual induction occurs when a changing magnetic field produced by one coil cuts the windings of a second coil and induces an emf in the second coil.

Inductive reactance (X_L) is the opposition presented by an inductance to an alternating current. $X_L = 2\pi fL$.

In a series R-L circuit, the voltage drops across R and L are 90° out of phase.

In a circuit containing both inductance and resistance, impedance (Z) is the total opposition to the flow of alternating current, and is a combination of X_L and R. It is expressed in ohms. $Z = \sqrt{X_L^2 + R^2}$.

Ohm's law for a-c circuits is: $E = IZ$; $I = E/Z$; and $Z = E/I$.

The vectorial sum of all the voltage drops in a series R-L circuit is equal to the applied voltage.

The primary winding of a transformer absorbs the input electrical energy from a voltage source; the secondary winding delivers the induced output voltage to a load.

The turns ratio between the secondary and primary windings of a transformer determines whether the secondary voltage is greater or less than the primary voltage.

The primary-secondary current ratio is opposite to that of the primary-secondary voltage ratio.

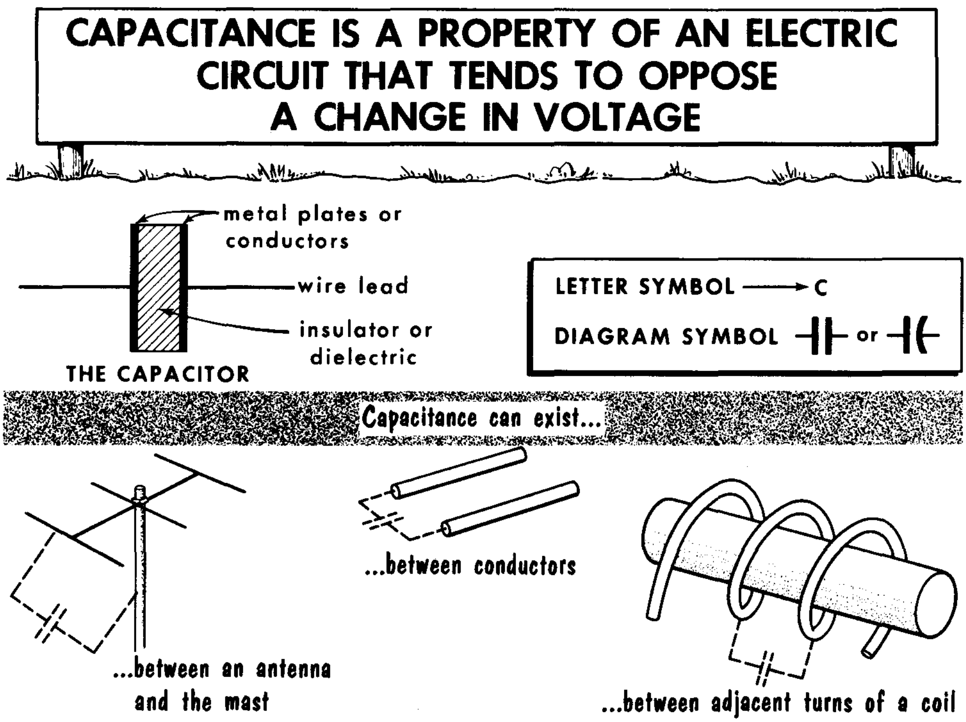
In the transfer of power from an electrical source to a load, the load impedance must be equal to, or match, the internal impedance of the source for maximum transfer of power.

REVIEW QUESTIONS

1. What is self-induction of emf? Describe its action.
2. Define inductive reactance and give the formula for calculating the X_L of a circuit.
3. Describe the action of mutual induction and tell what its purpose is.
4. In an inductive circuit, what are the phase relationships between: (1) the current and the counter emf it produces; (2) the applied voltage and the counter emf; (3) the applied voltage and the current?
5. What is impedance? How is it calculated in series R-L circuits?
6. What two methods can be used to determine the impedance of a series R-L circuit?
7. State Ohm's law for a-c circuits.
8. What is the total current in a parallel R-L circuit equal to?
9. What formula is used to find the impedance of a parallel R-L circuit?
10. Give two main functions of a transformer.
11. What determines whether a voltage step-up or step-down takes place in a transformer?
12. Name three types of losses inherent in a transformer.

Definition, Function, Construction

An electrical capacitor (also known as a condenser) stores electricity by accumulating free electrons on a metal surface, and then releases them as a current into the circuit of which the capacitor is a part. It can be said that capacitance is a property of a circuit in which energy may be stored in the form of an electric field. This capability that we call capacitance is generally designated by the letter C. Although electrical capacitors are used in many different ways (as you shall see), and seemingly for different purposes, every use entails the storage and release of electrical energy. This action underlies the definition of capacitance — that property of a circuit which opposes a change in voltage. This property differs from that of an inductance, which opposes any change in current.

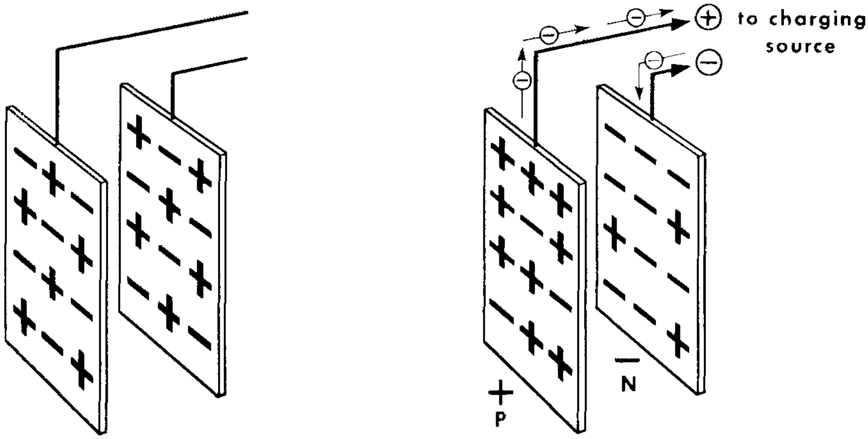


Any two conductors separated by an insulator (called a dielectric) can behave as a capacitor. The conductors may be long or short lengths of wire, large or small pieces of metal or metal foil, or any other conducting material. The dielectric that separates the conducting materials may be the insulation around the wires, a thin insulating film chemically deposited on the metal, ceramics, mica, oil, or wax-impregnated paper to mention just a few. In some cases, the dielectric is air, and in still others, a vacuum. As a rule, capacitors used in electrical and radio circuits are specially manufactured items but, as will be explained later, the metal parts of electrical and radio systems often behave as capacitors.

Charging a Capacitor

The action of storing electricity in a capacitor is called charging. The electricity stored in the capacitor is called the charge. As used here, "charge" refers to an amount of electricity being stored, rather than to the fundamental particles: the electron and the proton. The action of releasing the stored electricity is called discharging. For purposes of explanation, let us assemble a basic capacitor consisting of two thin sheets of aluminum 10 inches square separated by air and positioned 1/2 inch apart. Because air separates the two active plates (conductors) of the capacitor, the unit is called an air-dielectric capacitor.

IN THE UNCHARGED CAPACITOR | **IN THE CHARGED CAPACITOR**



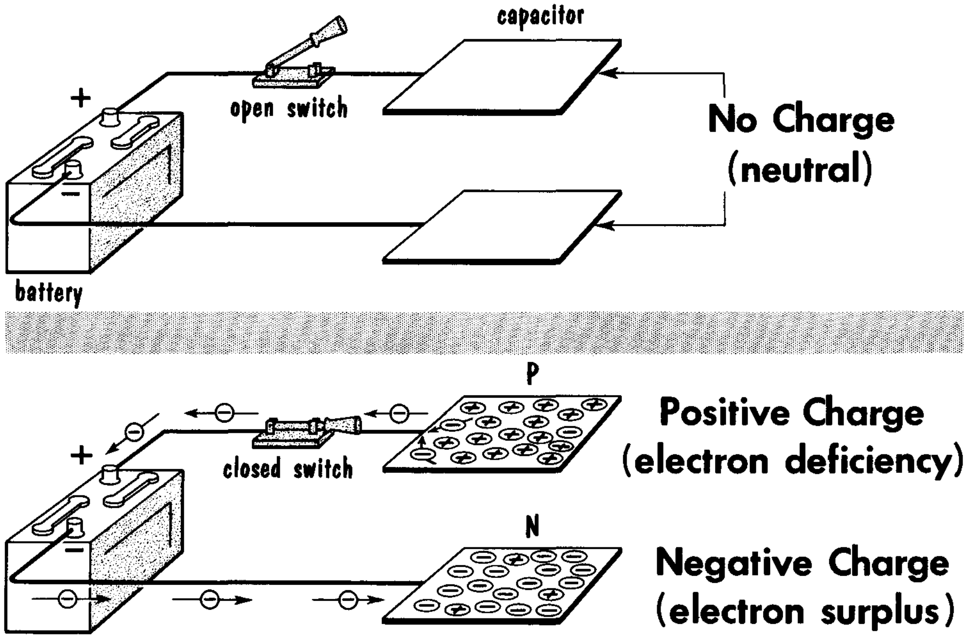
there are equal amounts of positive and negative electricity on each plate

one plate has an excess of positive electricity; the other plate has an excess of negative electricity

PRODUCED BY FORCING FREE ELECTRONS FROM ON PLATE ONTO ANOTHER

We begin by assuming that the two metal plates are electrically neutral — each plate contains equal amounts of positive and negative electricity. The capacitor is therefore in an uncharged state. In the process of charging, one plate (plate P) of the capacitor is made to give up free electrons and be left with a preponderance of positive electricity. The other plate (plate N), is made to accept as many free electrons as were released by plate P, and now has a surplus of negative electricity. Both electrical conditions are created simultaneously. When this electrical condition prevails, the capacitor is said to contain a charge, or be charged.

Charging a Capacitor (Cont'd)



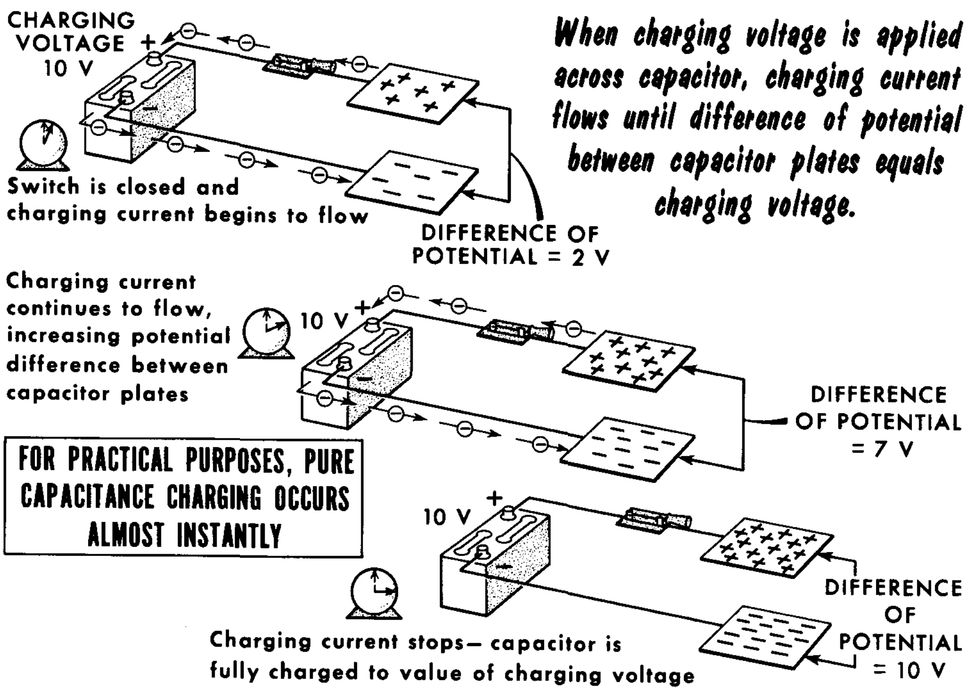
A CAPACITOR IS CHARGED BY PLACING A DIFFERENCE OF POTENTIAL ACROSS PLATES

To charge a capacitor, it is necessary to apply a voltage (or difference of potential) to its plates (across its terminals). In other words, the electrical energy stored in the capacitor must come from a source of voltage (we show a battery). Assume the circuit elements shown. The open switch isolates the battery from the capacitor. Then, the switch is closed and the capacitor becomes charged from the voltage applied to its plates by the battery.

The positive terminal of a voltage source is always deficient in free electrons. It draws these electrons from one plate of the capacitor and leaves that plate (P) with a deficiency of free electrons; hence, with a positive charge. At the same time, the negative terminal of the voltage source releases an equal number of free electrons into the wire connected to it, thereby forcing free electrons onto the other plate (N). These added free electrons create a surplus of negative charge on this plate, thus giving the N plate a negative charge. The creation of such an electrical condition on the plates of the capacitor is known as charging. The free electrons that are pulled from the P plate of the capacitor to the positive terminal of the battery and the free electrons that move from the negative terminal of the battery to the N plate constitute a momentary current usually referred to as a charging current.

Building Up Voltage in the Capacitor – Charging Current

From the instant that free electrons leave one plate and begin accumulating on the other, charging current flows, and a difference of potential (voltage) appears between the plates of the capacitor. (Note that the charging current flows first; then, the voltage buildup occurs.) Unless deliberately prevented from doing so, the charging (applied) voltage will continue attracting free electrons from one plate and forcing them onto the other. This process builds up the capacitor voltage until it becomes equal to the maximum value of the charging voltage. The voltage built up across the capacitor has the same polarity as the charging voltage; consequently, it acts in opposition to the charging voltage. When the capacitor voltage equals the charging voltage, the two voltages offset each other, and there is no further movement of free electrons (no charging current).



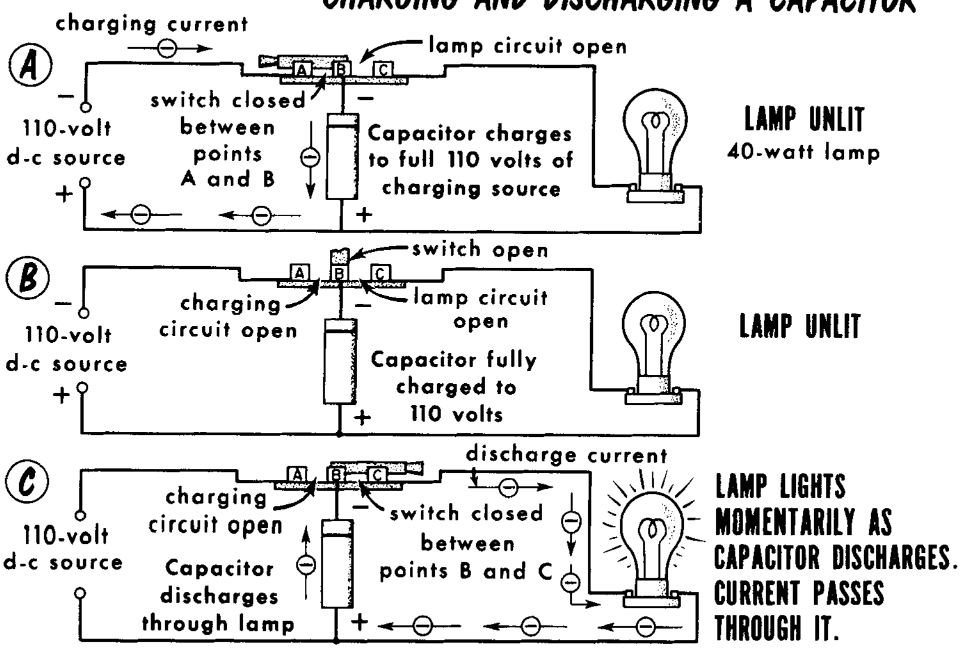
There is a limit to how much voltage can be built up across a given capacitor by a given charging voltage. The capacitor voltage cannot exceed the maximum value of the charging voltage at any time. But it is possible to subject a capacitor which has been charged by a lower value of charging voltage to an increased value of charging voltage. The capacitor voltage then rises to the higher value. The ability to charge a capacitor to higher and higher values of voltage is not without limitations. This limitation arises from a constructional characteristic of the capacitor. Every capacitor has a maximum d-c working voltage rating (which will be explained later). For the present, let us say that the capacitor voltage rating sets the limit on the highest value of charging voltage that may be applied to the capacitor.

Demonstrating the Voltage across a Charged Capacitor

When a charged capacitor is disconnected from the charging voltage source, the charge remains in the capacitor. If the capacitor is very "good," it retains its charge for a long period of time, during which time a voltage will be present across its terminals. This is not the usual way in which capacitors are used in electrical and communication systems, but this capability does have its uses.

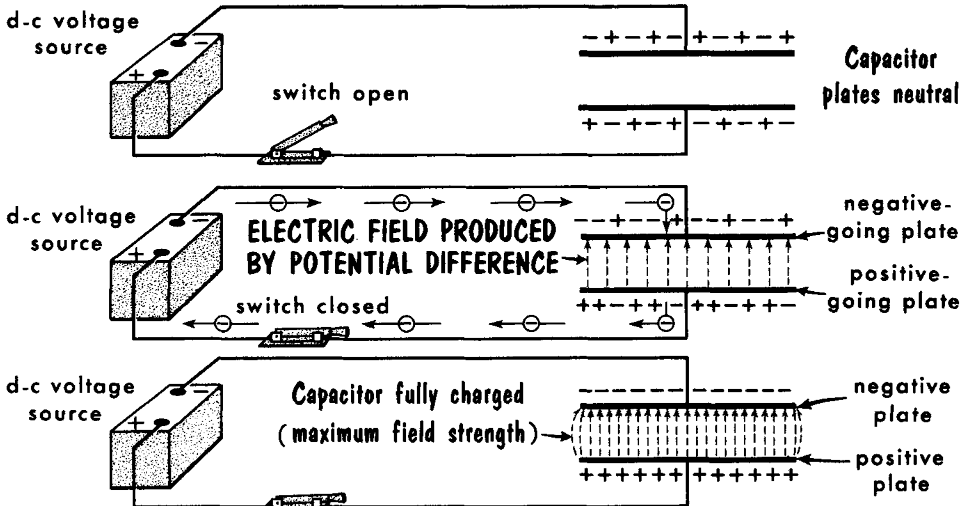
Assume that a capacitor is connected to a 110-volt d-c charging source by means of a single-pole double throw (spdt) switch. After the lapse of sufficient time (a short interval) to charge the capacitor fully, the switch is opened, thus removing the capacitor from contact with the charging circuit. To prove that the capacitor is charged and that a voltage exists across its terminals, we connect an ordinary 40-watt, 110-volt household electric light across the capacitor by closing the switch. The instant the switch is closed, the lamp glows brightly for a moment. The electrical energy required to light the lamp was obtained from the charged capacitor. The voltage across the terminals of the charged capacitor overcomes the resistance of the lamp filament (and the connecting wires), and allows the surplus free electrons on the negative plate of the capacitor to behave as current and move to the positively charged plate of the capacitor through the filament. This action discharges the capacitor. The movement of the free electrons from the negatively charged plate of the charged capacitor constitutes a discharge current; this is accompanied by a fall of voltage across the capacitor to zero. After this discharge, the filament no longer glows, because there is no flow of electrons (current) through the lamp filament.

CHARGING AND DISCHARGING A CAPACITOR



The Electric Field between the Capacitor Plates

You have learned that every fundamental particle of electricity – each electron and each proton – is inseparably associated with an invisible region of energy that exists all around the charge. The zone of energy is referred to as an electric field made up of electric lines of force.

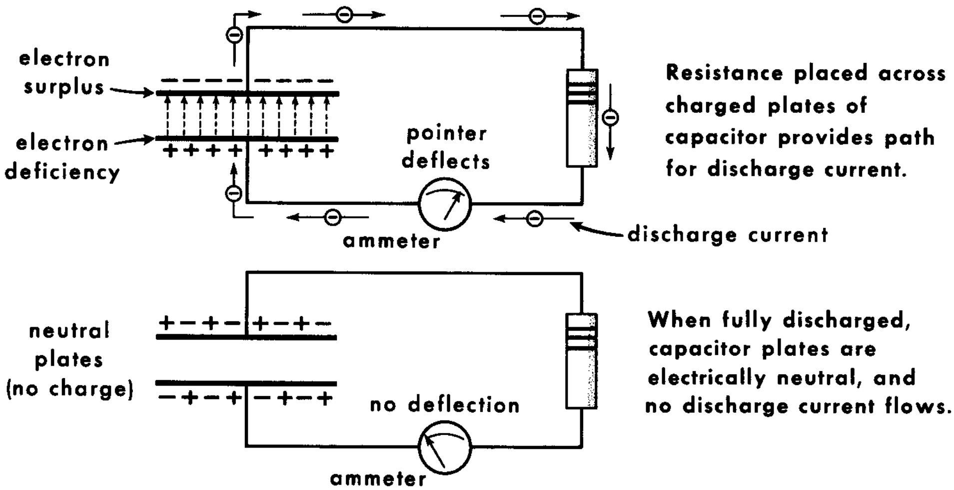


Arrows between plates represent ELECTROSTATIC LINES OF FORCE which make up ELECTROSTATIC FIELD between plates. FIELD STRENGTH is proportional to POTENTIAL DIFFERENCE between plates.

Beginning with the first instant after the flow of charging current (the appearance of a positive charge on one plate of the capacitor and a negative charge on the other), an electric field is developed between the two charged plates within the space between the plates. This field is between the positive charges on the positive plate and the negative charges on the negative plate. As more and more electrons are removed from the plate and more and more electrons are added to the negative plate, the lines of force increase in number, indicating increased field intensity. This action accompanies the rise in voltage. When the capacitor voltage has reached its maximum value, the field intensity has reached a maximum, and remains as long as the charge given the capacitor remains unaltered. The energy stored in the capacitor is in the electric field.

Discharging a Capacitor

All of the electricity stored in a theoretically perfect capacitor can be recovered from it by providing a suitable electrical conducting path between the terminals of the capacitor. Such a current path is called a discharge path. The electric light filament referred to previously formed such a path. Whether the energy taken out of the charged capacitor is used or wasted is a function of the electrical makeup of the discharge path. The usual discharge path is the circuit connected between the terminals of the capacitor. It is never the battery which serves as the charging voltage source.

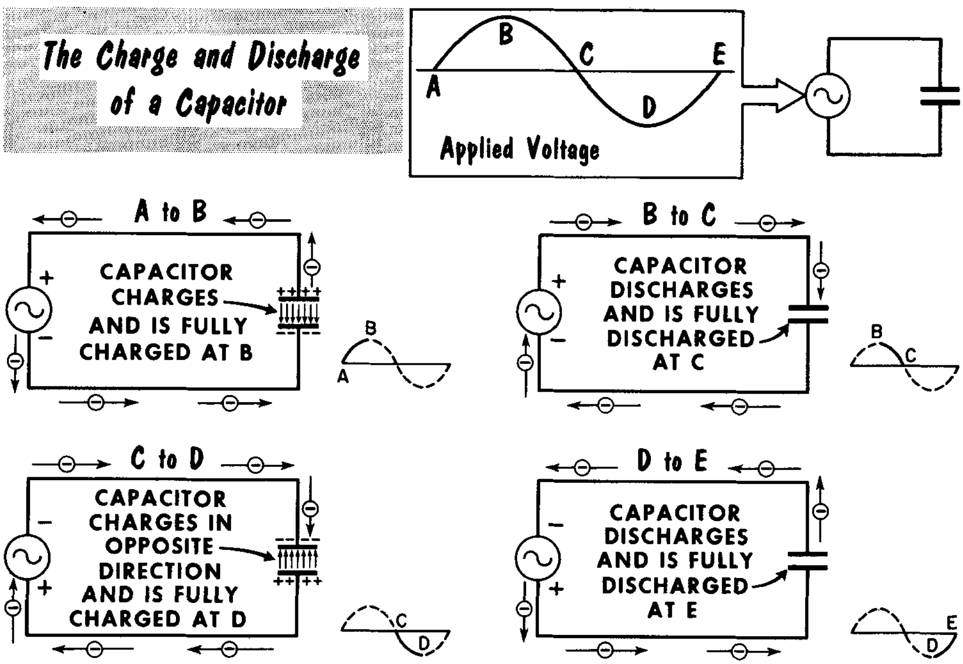


CAPACITOR DISCHARGES WHEN THERE IS A DIFFERENCE OF POTENTIAL BETWEEN PLATES AND A COMPLETE EXTERNAL PATH BETWEEN CAPACITOR TERMINALS

The action of recovering the energy stored in a charged capacitor is referred to as discharging the capacitor. During discharge, the surplus free electrons on the negatively charged plate move toward the positive charges on the positively charged plate via the discharge path. This movement of surplus electrons reduces the negative charge on the negative plate and the positive charge on the positive plate. Since the movement of free electrons during discharge is a directed motion, it is actually a current, and is referred to as the discharge current. The loss of charge on the negatively and positively charged plates by the flow of the discharge current causes the voltage originally built up across the capacitor plates to decrease. When all the surplus electrons have moved from the negatively charged plate to the positively charged plate, thus making both plates electrically neutral, there is no further charge in the capacitor, and the voltage across its terminals is zero. The capacitor is then fully discharged.

Charging a Capacitor from an A-C Voltage Source

A capacitor can be charged as readily by an a-c voltage as by a d-c voltage. However, the constantly changing amplitude and periodic polarity reversal of the a-c voltage point up many interesting capacitor characteristics. Assume a sine waveform voltage from an a-c source. The charging voltage starts at zero amplitude and increases in a positive direction. The first increase in charging voltage results in the flow of maximum charging current. The instant charge is applied to the capacitor, a voltage, or potential difference, starts building up across the capacitor. As the charging voltage increases in amplitude, more and more charge is added to the capacitor by progressively decreasing amounts of current. The decrease in current is due to the increased bucking action by the voltage (sometimes called counter-voltage) building up in the capacitor. When the charging voltage reaches its peak value, there is zero charging current and maximum voltage built up in the capacitor.

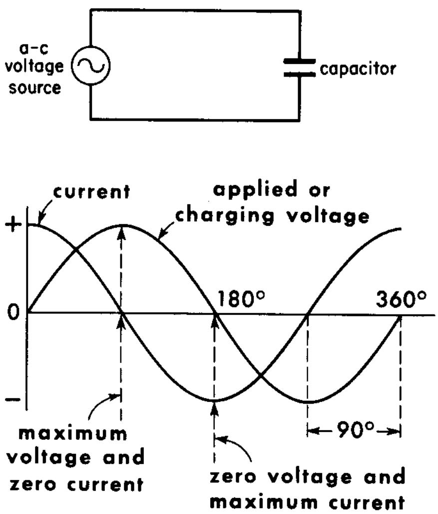


Having reached its peak positive value at the end of the first quarter-cycle (90°), the charging voltage begins to decrease in amplitude. At the first instant of decrease, the voltage built up across the capacitor exceeds the value of the charging voltage source. The voltage in the capacitor begins to fall as charge decreases. Note that the flow of discharge current began first, followed by a fall in capacitor voltage. Similarly, the charging current flow was ahead of the rise in capacitor voltage. The time sequence between the charging current and the rise in capacitor voltage, and between the discharge current and the fall in capacitor voltage, is described by saying that the capacitor current leads the capacitor voltage by a quarter-cycle, or 90°.

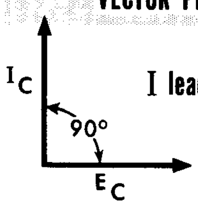
Voltage and Current Phase Shift in a Capacitor

Current and voltage variations in the capacitor during the positive half-cycle of the charging voltage repeat themselves during the negative half-cycle. Of course, the polarity change in charging voltage produces changes in the direction of current flow and in the polarity of the voltage across the capacitor, but the 90° phase difference between I and E is constant throughout the cycle. (This is an important point to remember.) It is also important to remember that the capacitor charge is zero when the current in the circuit is maximum, and maximum when the current is zero. Therefore, the charge on the plate of a capacitor is said to lag the current through it by 90°. However, since the building up and falling off of charge is the building up and falling off of voltage, the voltage across the capacitor is said to lag the current through it by 90°, or the current is said to lead the voltage by 90°.

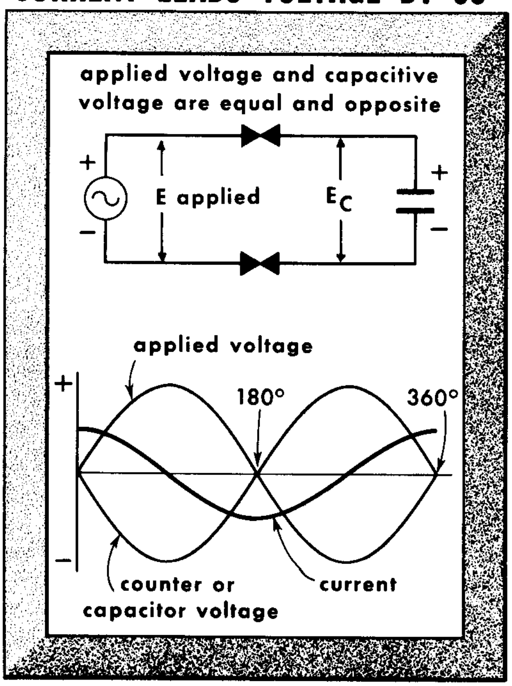
From Kirchhoff's law, we know that the sum of the voltage drops in a series circuit is equal to the applied voltage. Therefore, the voltage across the capacitor is, by definition, a voltage opposite to the applied voltage, or 180° out of phase with the applied voltage. Thus, when the applied or charging voltage is zero, there is no opposition, and when the applied voltage is maximum, there is a maximum opposition – a voltage produced by the charge on the capacitor. The variation in charge on the plates of the capacitor also follows the form of the sine wave and is in phase with the voltage, since for any given capacitor, the voltage across it depends directly on the charge.



VECTOR PRESENTATION



IN A CAPACITIVE CIRCUIT CURRENT LEADS VOLTAGE BY 90°



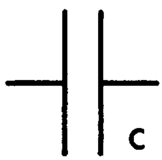
Unit of Capacitance – the Farad

The unit of capacitance is the farad, named in honor of Michael Faraday, the scientist who advanced the concept of electromagnetic induction. The number of electrons entering and leaving the capacitor plates depends upon the free electrons available and on the applied voltage. If the applied voltage is high, the forces of attraction and repulsion are great, and the charge deposited on the plates is also great. It was discovered that for a given capacitor, the ratio between the amount of this charge and the voltage causing it is always a constant. Therefore, the ratio of the charge (Q) to the voltage (E) is considered to be a measure of the capacitor action, called capacitance (C). The formula for capacitance is: $C = Q/E$.

By definition, a capacitor has a capacitance of 1 farad if a 1-volt difference in potential results in the storage of 1 coulomb of charge. One coulomb represents a quantity of 6.28×10^{18} electrons (6,280,000,000,000,000,000). For practical purposes, a capacitance of 1 farad represents a fantastically large capacitance. As a practical matter, capacitors used in radio communications are measured in terms of microfarads (one-millionth of a farad, µf) and micromicrofarads (one-millionth of a millionth of a farad, µµf).

Suppose we find the capacitance of a capacitor when the charge (Q) stored is 0.001 coulomb and the voltage is 1000 volts. By applying the above formula, $C = Q/E$, we obtain 0.001/1000, or 0.000001 farad. We call this 1 microfarad.

THE FARAD -- A measure of the Storage Ability of a Capacitor



$$C\ (\text{capacitance in farads}) = \frac{Q\ (\text{charge in coulombs})}{E\ (\text{voltage in volts})}$$

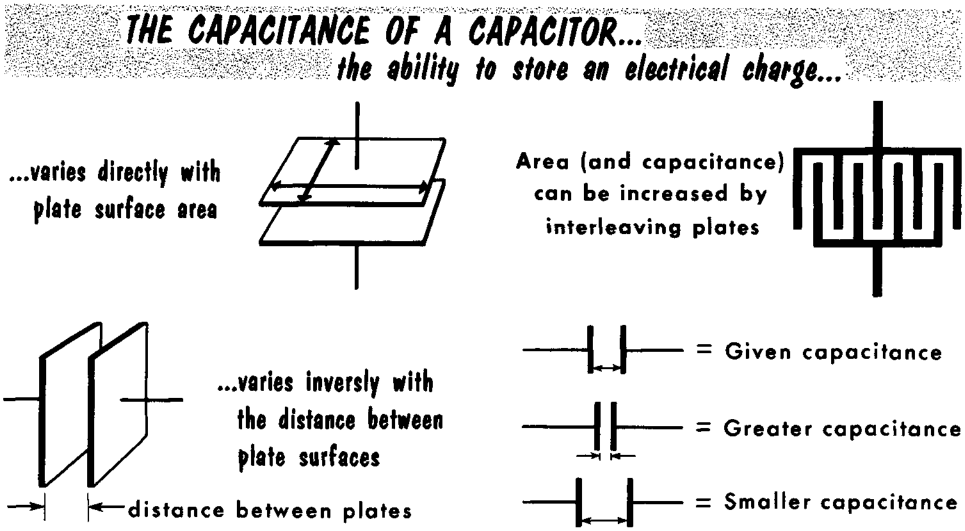
CAPACITANCE CONVERSIONS

1,000,000 MICROFARADS (µf)	=	1 FARAD (f)
1,000,000 MICROMICROFARADS (µµf)	=	1 MICROFARAD
1,000,000,000,000 MICROMICROFARADS (µµf)	=	1 FARAD
1,000,000,000,000 PICOFARADS (pf)	=	1 FARAD

$\frac{1}{1,000,000}$ f OR 0.000001 f	=	1 MICROFARAD
$\frac{1}{1,000,000}$ µf OR 0.000001 µf	=	1 MICROMICROFARAD
$\frac{1}{1,000,000,000,000}$ f OR 0.000000000001 f	=	1 MICROMICROFARAD
$\frac{1}{1,000,000,000,000}$ f OR 0.000000000001 f	=	1 PICOFARAD

500 µµf	= 0.0005 µf	50 µf	= 0.00005 f	500 µf	= 0.0005 f
10 µµf	= 0.000010 µf	0.01 µf	= 0.00000001 f	0.001 µf	= 1000 µµf
3000 µµf	= 0.003 µf	0.0047 µf	= 0.0000000047 f	0.0047 µf	= 4700 µµf
50 pf	= 0.000050 µf	0.05 µf	= 0.00000005 f	0.005 µf	= 5000 pf

Factors Determining Capacitance Area and Plate Separation



Three factors determine the capacitance of a capacitor: the area of plate surfaces; the distance between the plates; and the material used as the insulation or dielectric between the plates. The capacitance of a capacitor is related directly to the surface area of the plates. Given a fixed dielectric material and distance separating the plates, the capacitance is directly proportional to the area of the plates. The greater the area, the greater the amount of electricity (charge) which can be stored in the capacitor. This follows from the condition that the greater the area, the more the number of free electrons available for charging. Doubling the surface area, with everything else being fixed, doubles the capacitance; halving the area produces half the capacitance. We shall see that different capacitor designs give various capacitance values in the same physical space.

Assuming a capacitor with a given plate area and dielectric material, the closer the facing plate surfaces are to each other, the greater the capacitance. This is an inverse proportion. Halving the area of separation doubles the capacitance; doubling the area of separation halves the capacitance. The reason for this is that the closer the facing surfaces are to each other, the more strongly the unlike charges on the surfaces are attracted towards each other. This tends to concentrate the free electrons on the negatively charged surface nearest to the positively charged surface, thus allowing more negative charges to be crowded onto a plate or plates of a given area.

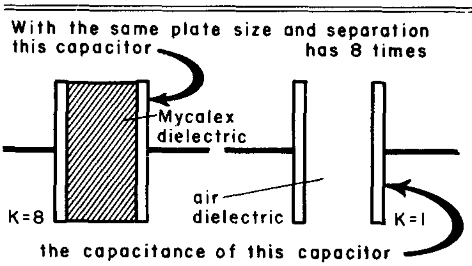
There is a limitation to the permissible closeness of the active surfaces to each other, regardless of the separating medium. If the voltage built up across the capacitor exceeds the voltage rating of the capacitor (as explained later), electrons may be pulled away from the negatively charged surfaces and leap to the positively charged surfaces inside the capacitor. This action discharges the capacitor and may destroy it.

Factors that Determine Capacitance Dielectric Constant

Given a capacitor with plates of a certain area and separated from each other by a specific distance, its capacitance is a function of the kind of material used for the dielectric. The standard of comparison is dry air, which is considered as having a dielectric constant (K) of 1. The dielectric constant of a vacuum differs so little from air that both are considered equal to unity. Dielectric constant is the ability of a material or medium to permit the establishment of electric lines of force between oppositely charged plates. Many materials will support more electric lines of force in a given space than air; these are said to have a dielectric constant greater than 1. Dielectric constants vary considerably. As examples, various types of mica have dielectric constants of from 5 to 9, and some forms of titanium dioxide have dielectric constants of up to 120. Some special kinds of chemical film deposits may have dielectric constants as high as 1000 or more. A dielectric (other than air) makes the positively charged surface of a capacitor repel more free electrons and the negatively charged surface accept more electrons than when air is the dielectric, thus increasing the capacitance.

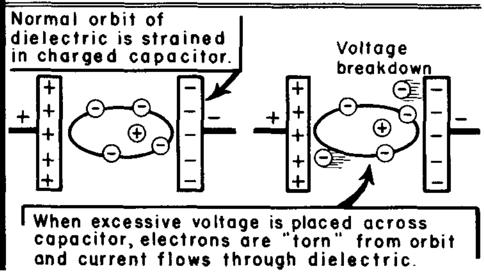
DIELECTRIC CONSTANTS

Material	K
air	1
resin	2.5
hard rubber	2.8
dry paper	3.5
glass	4.2
mica	5–9
porcelain	5.5
Bakelite	4.5–7.5
Mycalex	8
titanium dioxide compounds	90–170



VOLTAGE BREAKDOWNS

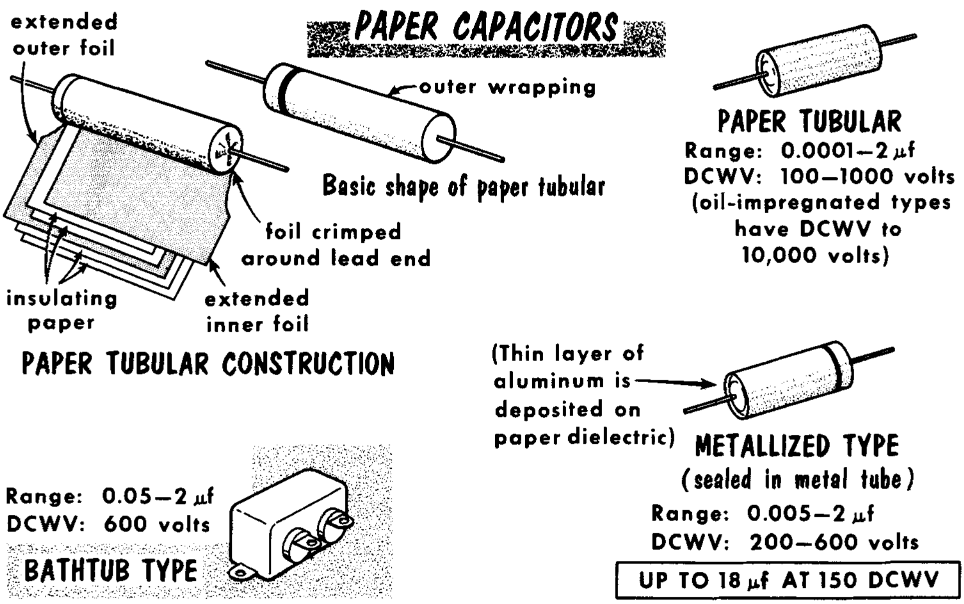
Material	Dielectric Strength (volts per 0.001 inch)
air	80
fiber	50
Bakelite	500
glass	200
mica	2000
castor oil	380
paper (beeswaxed)	1800
paper (paraffined)	1200
porcelain	750



The dielectric material deserves one further consideration – that of breakdown voltage. While the dielectric is an insulator, voltages across the plates of a capacitor may be sufficiently high to "tear" electrons out of the atomic orbits of the dielectric. When this occurs, the dielectric "breaks down," and arcing occurs between the plates through the dielectric. In many instances, this destroys the capacitor, as it is short-circuited. Thus, it is important to observe the dielectric strength of a material. A high voltage would be needed to break down a vacuum dielectric, but lower voltages could break down certain other substances. Thus, the breakdown voltage of a dielectric must be considered as well as its dielectric constant.

Fixed Capacitors – Paper Type

The simplest and most widely used form of paper capacitor consists of two strips of metal foil rolled up, with strips of paper which have been impregnated with an insulating material (dielectric) placed between them. Impregnating materials generally include various types of oils, waxes, and plastics. The type used determines the voltage, temperature, and insulation-resistance characteristics of the capacitor. When the capacitor is to be used at high working voltages, several layers of insulating paper are used.

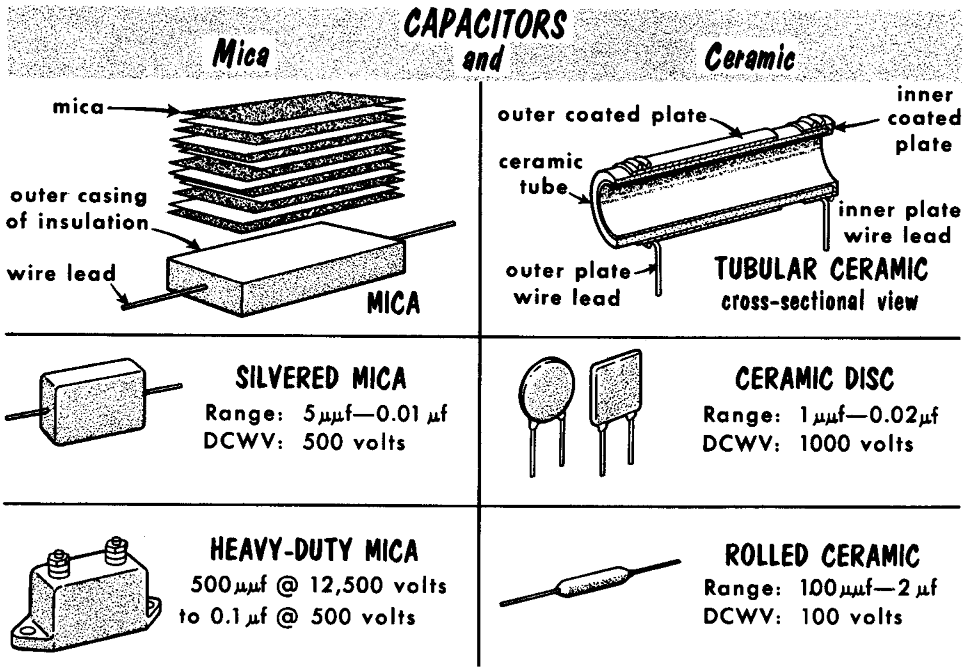


After the foil and paper strips are rolled up, the protruding ends of the foil are crimped over so that the individual layers of each strip are in electrical contact with each other. A lead is attached to each end, and an outer cover of insulating material is added. The cover is marked with the capacitance and working voltage, and a black ring is usually printed around one end to mark the terminal which is connected to the outermost layer of foil. In paper capacitors, the total capacitance is predetermined by the thickness and dielectric constant of the paper and the total of the foil plates. Capacitors are usually marked with a d-c working voltage (DCWV) which <u>must be observed</u>. It should be remembered that a-c voltages are spoken of in terms of rms (effective) value, and that the peak value is 1.41 times the rms value. Thus, when connecting a capacitor in an a-c circuit, we must always keep in mind the peak a-c voltage that will be impressed across the capacitor plates.

When paper capacitors are required to have a capacitance of over 1 μf, their physical size generally becomes too large for convenient mounting. Under such conditions, the capacitor is placed in a metal case filled with insulating material and then hermetically sealed. Units of this type are known as potted, or bathtub capacitors.

Fixed Capacitors – Mica and Ceramic

The mica capacitor consists of a number of flat strips of metal foil separated by similarly shaped strips of mica. The foil strips serve as the capacitor plates, and the mica acts as the dielectric. Alternate plates are connected together. An electrode is attached to each set of plates, and a terminal or lead wire is connected to each electrode. The entire unit is then encased in a container of plastic insulating material. An alternate construction is that of the "silvered" mica capacitor. In this unit, very thin layers of silver are deposited directly on one side of the mica, and the plates are stacked together so that alternate layers of silver are separated by alternate layers of mica. The result is the equivalent of the foil construction. Mica capacitors are available in three basic types: molded, molded-case potted, and ceramic-case potted. In addition, the "button" type mica unit is very popular.

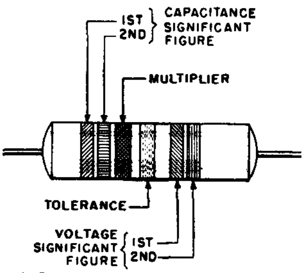


The basic construction of the ceramic capacitor consists of a ceramic disc or tube with silver or copper plates deposited on the opposite faces of the ceramic material. In the manufacturing process, electrodes are attached to the plates, leads or terminals are fastened to the electrodes, and a moisture-proof coating of plastic or ceramic is added. Ceramic capacitors are available in a number of basic shapes. The outstanding characteristic, however, is the high dielectric constant of ceramics. Steatite ceramics have a dielectric constant of 6, magnesium titanate has a K in the region of 16, and barium titanate has a K of approximately 1200. Ceramic capacitors have good stability with regard to temperature and voltage changes. The high-K ceramics provide increased capacitance without increased size.

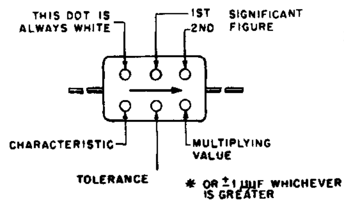
Capacitor Color Coding and Temperature Coefficient

As in the case of resistors, capacitors are also frequently color coded to indicate various capacitor characteristics. The points usually covered by color coding of capacitors are: capacitance, capacitance tolerance, and temperature coefficient. Some capacitors use color coding to indicate the d-c working voltage.

MOLDED PAPER CAPACITOR COLOR CODE



CAPACITANCE IN MICROMICROFARADS



EIA COLOR CODE (FORMERLY RETMA)

CAPACITANCE IN MICROMICROFARADS

COLOR	CAPACITANCE SIGNIFICANT FIGURE	CAPACITANCE MULTIPLIER	* TOLERANCE ± %	CHARACTERISTIC
BLACK	0	1	20 (M)	A
BROWN	1	10	1 (F)	B
RED	2	100	2 (G)	C
ORANGE	3	1,000	3 (H)	D
YELLOW	4	10,000	-	E
GREEN	5	-	5 (J)	F
BLUE	6	-	-	-
VIOLET	7	-	-	-
GRAY	8	-	-	-
WHITE	9		-	-
GOLD	-	0.1	5 (J)	-
SILVER	-	.01	10 (K)	-

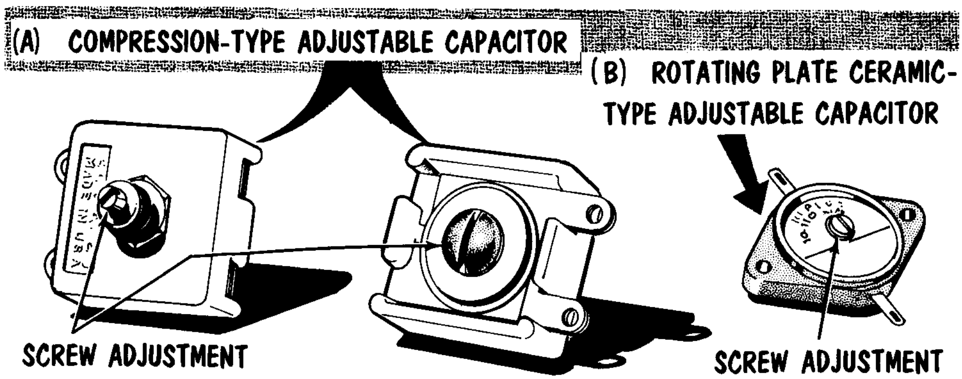
LETTER DESIGNATION FOR CHARACTERISTIC

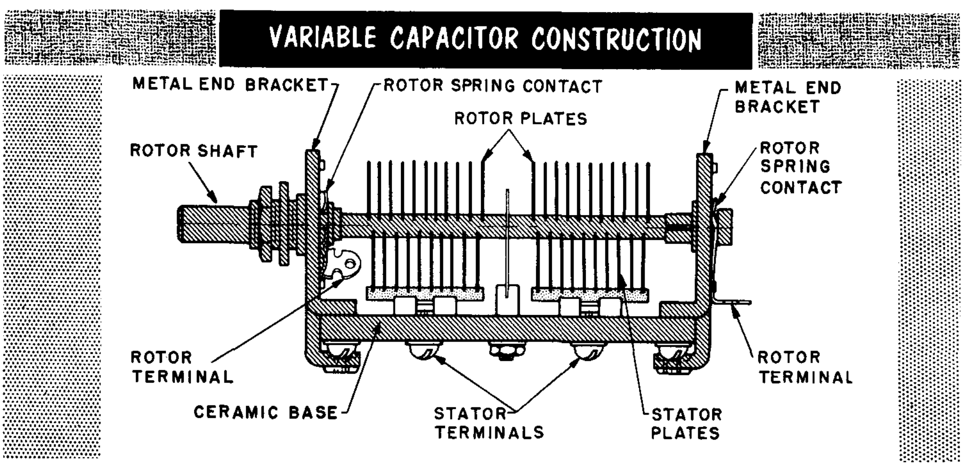
CHARACTERISTIC	TEMP. COEFF. PARTS PER MILLION PER °C	CAPACITANCE DRIFT (MAXIMUM)
A	± 1000	(5% OF NOM. CAP. + 1) μμF
B	± 500	(3% OF NOM. CAP. + 1) μμF
C	± 200	(0.5% OF NOM. CAP. + 0.5) μμF
D	± 100	(0.3% OF NOM. CAP + 0.1) μμF
E	+ 100 TO − 20	(0.1% OF NOM. CAP + 0.1) μμF

Mica- and ceramic-dielectric capacitors bear an electrical rating known as temperature coefficient. It is expressed by a number between 0 and 1300, prefixed by a minus or a plus sign (e.g., -220 or + 30). These designations state the change in capacitance (in μμf) from the nominal rating per million parts of capacitance (per μf) per degree Centigrade (ppm/°C) rise in temperature. The reference temperature is 20°C, which is the same as 68°F. The minus symbol indicates that the capacitance decreases; the plus symbol indicates that the capacitance increases. The prefix letter N indicates the same as the minus symbol, whereas the prefix NPO stands for negative-positive-zero, and indicates that the capacitance change is substantially zero over a wide range of temperature increase and decrease. For example, an N750 0.0001 = μf capacitor would have a decrease in capacitance of 750 x 0.0001, or 0.075 μμf per degree rise in temperature.

Variable Capacitors

A variable capacitor affords a continuous variation in capacitance between a fixed minimum and a fixed maximum value. While capacitance can be varied by varying plate area, distance between plates, and dielectric material, the most popular method is the varying of plate area. In this capacitor, there is a fixed set of metallic plates (called stators) mounted on some insulated base. Interleaved with these stator plates are the rotor (rotating) plates, controlled by a shaft. As the shaft is turned, the rotor plates mesh with (but do not touch) the stator plates, providing a variation in capacitor plate surface area. Usually, the entire frame is connected to the stator which serves as the grounded or common plates. Where more than one circuit is to be controlled, variable capacitors may be ganged to give simultaneous control.

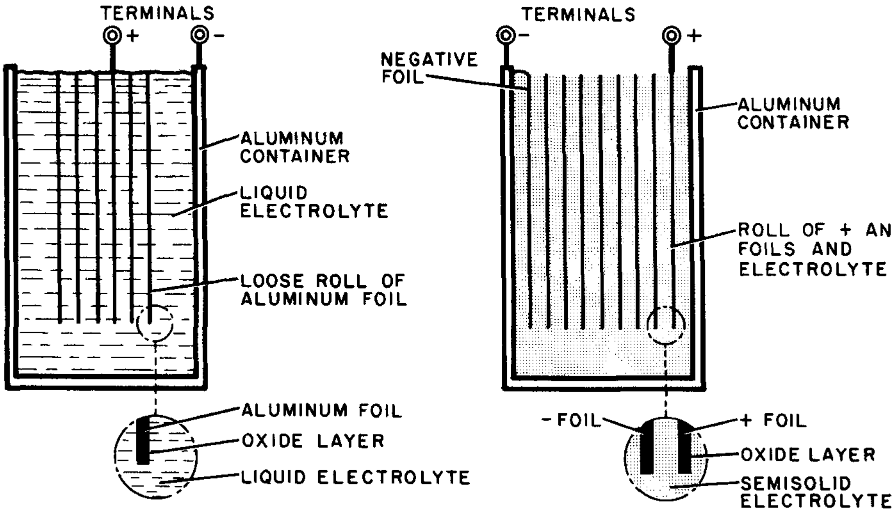




Two other types of variable capacitors are the compression type and the rotating plate. The compression type uses two flexible plates which can be compressed by a screw. The compression "squeezes" the mica dielectric, varying the distance between the plates. In the rotating-plate type, facing plates are turned so as to vary the facing plate area, and, therefore, the capacitance.

The Electrolytic Capacitor

ELECTROLYTIC CAPACITOR CONSTRUCTION



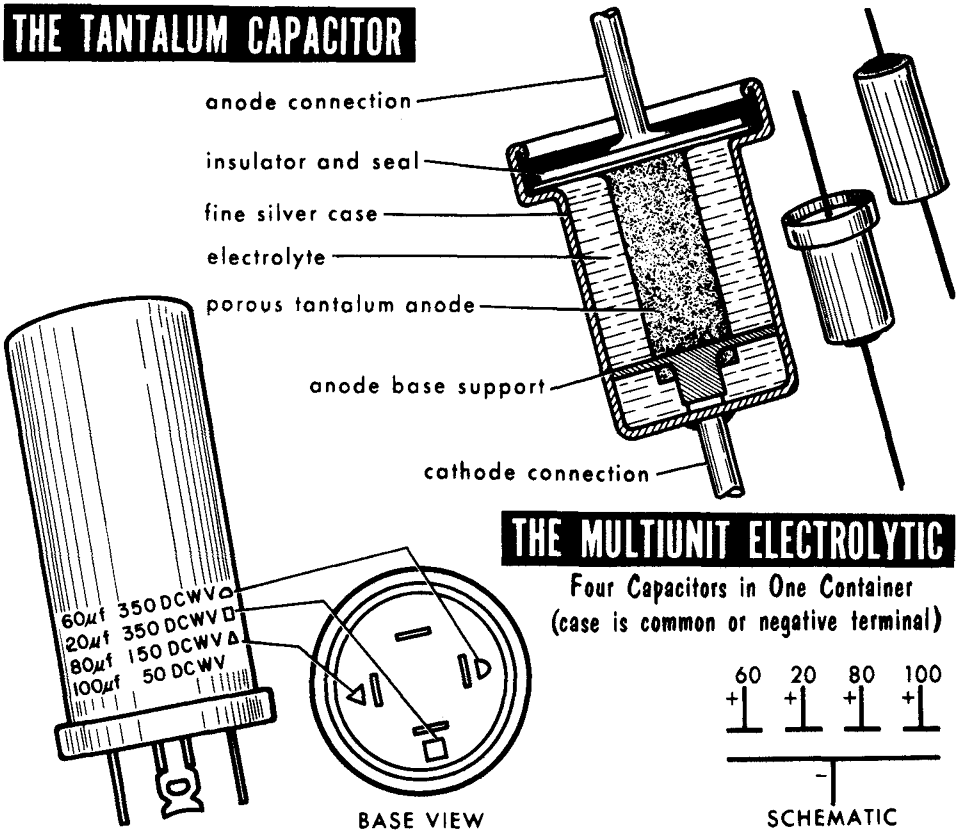
(A) WET TYPE (B) DRY TYPE

A very prominent type of capacitor used in electronic equipment is the electrolytic variety. In essence, it consists of a positive (anode) metal plate (aluminum foil or metal sprayed on cotton gauze, or a porous tantalum-oxide powder core) that is immersed in a liquid (wet type) or paste (dry type) bath known as the electrolyte. The entire capacitor is contained within a metal housing which usually serves as the negative (or cathode) terminal and as means of contact with the electrolyte, the other active surface in the capacitor. The dielectric is a very thin film (usually aluminum or tantalum oxide) which is "forced" on the metal plate.

At the time the electrolytic capacitor is being made, a d-c voltage is applied between the metal container (the negative terminal) and the metal plate (the positive, or anode electrode). It causes a relatively high current to flow in one direction inside the unit. As a result, a very thin dielectric film is formed on the outside of the positive (anode) electrode. As this film forms, the current decreases, eventually reaching a minimum. This minimum current is referred to as the leakage current. When the forming is completed, the capacitor is polarized; the metal plate is positive and the electrolyte is negative. The capacitor functions properly as long as the charging voltage has a polarity corresponding to that of the capacitor electrodes. This is a very important condition; hence, the conventional electrolytic capacitor bears polarity designations. The d-c type of unit is not suitable for charging by an a-c voltage. Another form of construction is used in a-c electrolytic capacitors. These capacitors contain two formed positive electrodes both of which act as either positive or negative electrodes, thus permitting the periodic reversal of polarity of the applied voltage.

The Electrolytic Capacitor (Cont'd)

The electrolytic capacitor offers the advantages of a very high capacitance in a small space, and at low cost. Capacitance values available in "electrolytics" extend from as low as 1 microfarad to several thousand microfarads. This huge capacitance arises from the extreme thinness of the dielectric film which, in effect, means that the active surfaces of the capacitor – the metal plate and the electrolyte – are very close to each other. Film thicknesses of as little as 0.00001 inch are commonplace. However, the thin film introduces a limitation in working voltage. D-c electrolytic capacitors of up to perhaps 100 μf have working voltage ratings up to 450 volts d-c. As the capacitance ratings increase, the working voltage ratings decrease because the dielectric film is thinner; hence, the working voltage must be reduced to prevent puncture of the dielectric film.



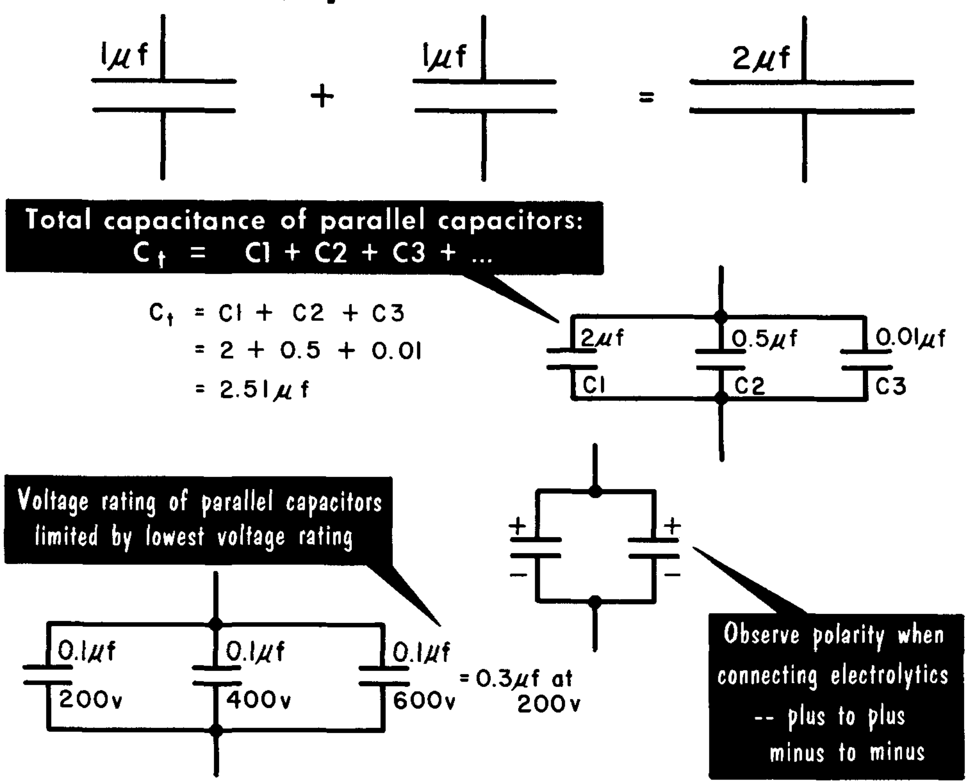
Capacitors rated at several thousand microfarads have working voltage ratings of from 6 to about 25 volts d-c. Interestingly enough, electrolytic capacitors sometimes are self-healing. Proper polarity of the applied voltage reforms the puncture in the film and "heals" the capacitor, providing the puncture is small. Operation at voltages in excess of the rating working voltage is not recommended.

Connecting Capacitors in Parallel

When capacitors are connected in parallel, the effect is to produce a total capacitance equal to the sum of all the individual capacitances. The reason for this is that, effectively, the total plate surface area of each capacitor is added, providing a larger total plate area. Since plate area is one of the factors that determines the capacitance of a capacitor, connecting capacitors in parallel increases the total capacitance. The formula for parallel capacitances is: $C_t = C1 + C2 + C3 + \ldots.$ Thus, 5-μf, 10-μf, and 15-μf capacitors connected in parallel would provide a total capacitance of 30 μf.

When capacitors are connected in parallel, the total voltage of the circuit is applied across each capacitor. Therefore, no matter how high the voltage rating of each capacitor in the parallel hookup, the unit with the lowest voltage rating effectively becomes the weakest link and limits the amount of voltage that can be applied to the parallel combination.

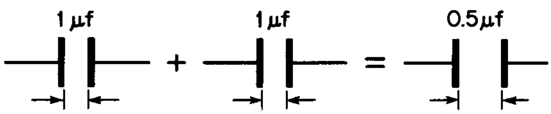
Connecting Two Capacitors in Parallel has the Effect of Increasing the Total Plate Area (capacitance increases)

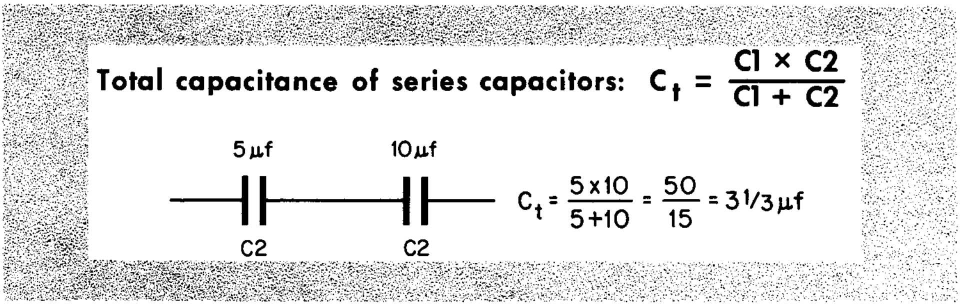


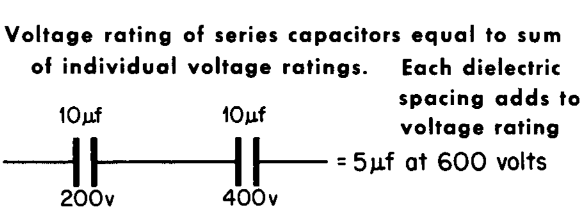
Connecting Capacitors in Series

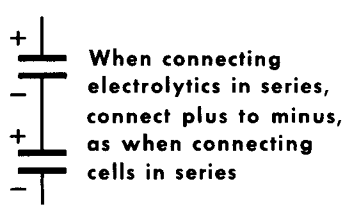
Connecting capacitors in series has the effect of reducing the total capacitance tc a value less than the lowest capacitance. This is the equivalent of connecting resistors in parallel. The reason for the reduction in total capacitance when they are connected in series is that effectively, we have added together the spacing between the plates of all the capacitors. Since the capacitance of a capacitor varies inversely with the spacing between the capacitor plates, each series capacitor added to a series string reduces the total capacitance. From this, we get the formula for the total capacitance of a series circuit: $C_t = (C1 \times C2)/(C1 + C2)$. Thus, if a 10-μf and a 15-μf capacitor were connected in series, the total capacitance would be: $C_t = (10 \times 15)/(10 + 15) = 150/25$, or 6 μf. Where more than 2 capacitors are connected in series we can use the formula $C_t = 1/[(1/C1) + (1/C2) + (1/C3) + \ldots]$.

Connecting Two Capacitors in Series has the Effect of Increasing the Distance between Capacitor Plates (capacitance decreases)









When capacitors are connected in series, the breakdown voltage of each unit is added to provide a breakdown voltage equal to the sum of each of the capacitor breakdown voltages. Two capacitors having a 600-volt and 1000-volt d-c working voltage provide a total safe working voltage of 1600 volts d-c when connected in series. Thus, series connection provides a reduction in capacitance and an increase in working voltage rating.

Capacitive Reactance (X_C)

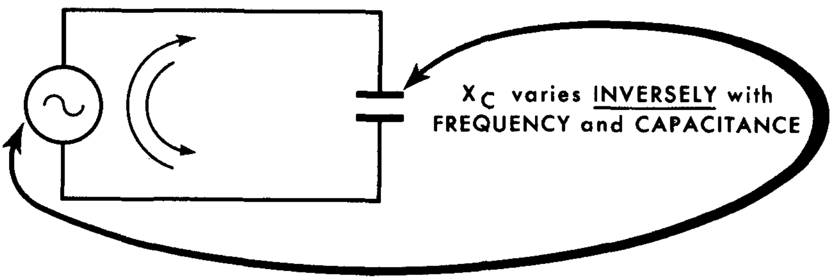
When an alternating voltage is applied to the plates of a capacitor, a certain amount of current will flow in the circuit. We recall that when a charge builds up in a capacitor, the voltage across the capacitor acts in opposition to the applied voltage. The amount of opposition that a capacitor offers to the flow of current in an a-c circuit depends upon the capacitance of the capacitor and upon the frequency of the a-c voltage source. The greater the size of the capacitor, the greater the amount of energy it can store, and the more the current that must flow to charge it. In addition, since it takes time to charge a capacitor, the lower the frequency of the a-c charging voltage, the slower will be the buildup of charge in a capacitor. The net effect of all this is to produce a certain opposition to the flow of current. This opposition in a capacitive circuit is called capacitive reactance and is measured in ohms.

capacitive reactance (X_C)

$$X_C = \frac{1}{2\pi fC}$$

in ohms — in cycles — in farads

CAPACITIVE REACTANCE is the OPPOSITION TO CURRENT FLOW offered by a Capacitor in an A-C Circuit

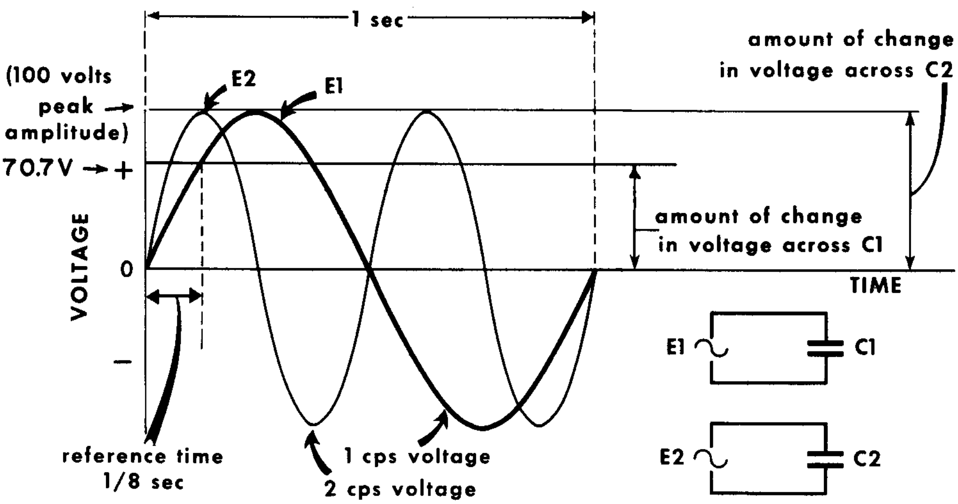


While different principles and effects are involved, the net effect of capacitive reactance on the flow of current in an a-c circuit is the same as that of inductive reactance. Whereas inductive reactance is expressed as X_L, capacitive reactance is expressed as X_C; both are measured in ohms. In addition, inductive reactance varies with inductance and frequency; capacitive reactance varies with capacitance and frequency — only inversly. The formula for computing the reactance of a capacitor is: $X_C = 1/2\pi fC$, where 2π equals 6.28, f is the frequency of the a-c source in cycles, and C is the capacitance of the capacitor in farads. From the formula, we can see an inverse relationship between capacitive reactance and frequency and capacitance. As f and C increase, X_C decreases; as f and C decrease, X_C increases.

Capacitive Reactance – Effect of Change in Frequency

To illustrate the change in capacitive reactance as frequency changes, imagine two sine waveform voltages, E1 (1 cycle) and E2 (2 cycles), each of 100 volts peak amplitude. These two frequencies are applied to two identical capacitors, C1 and C2. Voltage E1 rises from zero to peak amplitude (90°) in 1/4 second whereas E2 rises from zero to peak amplitude in 1/8 second (45°). These time intervals are arrived at by dividing the time duration of a single cycle by 4 since there are four quarter-cycles in a complete cycle. The time duration for E1 is 1 second; for E2 it is 1/2 second.

AS FREQUENCY INCREASES, CAPACITIVE REACTANCE DECREASES



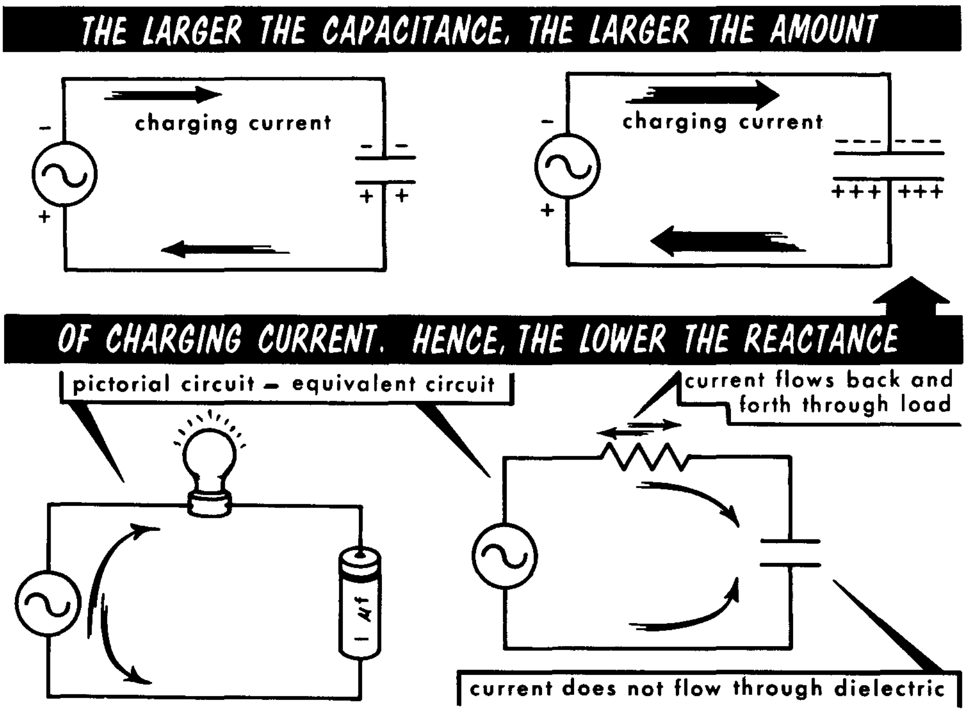
1/8 second = 90° for voltage E2 = peak × 1 or 100 volts
1/8 second = 45° for voltage E1 = peak × 0.707 or 70.7 volts

Now if we use 1/8 second as the reference time interval for both voltages, E2, being the higher-frequency voltage, changes in value more rapidly than E1. Voltage E2 rises to the peak value of 100 volts in 1/8 second, whereas E1 reaches only the 70. 7-volt level in the same amount of time. Therefore, capacitor C2 receives maximum charge, whereas C1 receives less charge. For C2 to receive more charge than C1 in the same time interval, it is necessary that more current flow into C2 than into C1. Hence, the capacitive reactance of C2 for E2 is less than that of C1 for E1.

We can translate the above action by saying that the faster the rate of change of the charging voltage – or the higher the frequency of the charging voltage – the lower the capacitive reactance of any given capacitor. Of course, the opposite is true: the lower the frequency, the higher the capacitive reactance.

Capacitive Reactance – Effect of Change in Capacitance

Capacitive reactance decreases with an increase in capacitance and increases with a decrease in capacitance. Given two capacitors, C1 (1 μf) and C2 (5 μf), both subjected to a voltage of 100 volts peak amplitude, the larger capacitance can accept more charge. For this to be true when the two capacitors are charged by the same amount of voltage in the same amount of time, the current flowing into the larger amount of capacitance must be greater than the current flowing into the smaller amount of capacitance. Hence, the capacitive reactance of C2 must be less than that of C1. From this, we see that for a given frequency, the larger the capacitance, the lower the capacitive reactance; the smaller the capacitance, the higher the capacitive reactance.



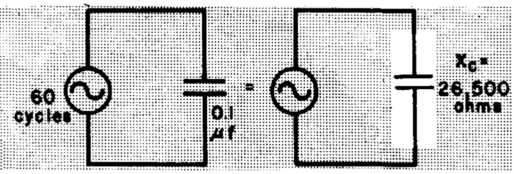
A capacitor blocks the flow of d-c but effectively "passes" a-c.

We have seen that during the charge and discharge of a capacitor, electrons flow back and forth in the circuit, first making one plate negative with respect to the other, and then making the other plate negative with respect to the first. It would seem that there is a complete closed circuit in which the current is alternating. Actually, of course, the plates of a capacitor are separated by an insulator (dielectric). No current flows through the dielectric, but since current flows back and forth from plate to plate, the current in a capacitive circuit takes on all the appearances of a closed-circuit arrangement. From this, we often use the expression that "current flows through a capacitor." Of course, this is not so – it only appears that way.

Using the Equation for Capacitive Reactance

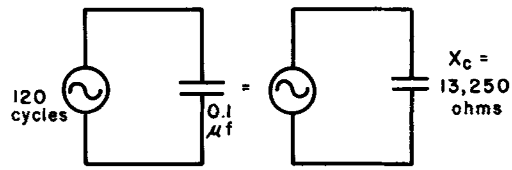
Referring again to the equation for capacitive reactance, let us solve several typical examples. While doing this, you must bear in mind that we are determining the opposition to current due only to the presence of the capacitance. The equation does not involve the resistance of the connecting wires or the actual value of the voltage applied. The illustration shows what happens when a 0.1-μf capacitor is used in a 60-cycle circuit.

$$X_C = \frac{1}{2\pi f C} = \frac{1}{6.28 \times 60 \times 0.0000001} = 26{,}500 \text{ ohms}$$



Leaving the capacitance fixed and doubling the frequency

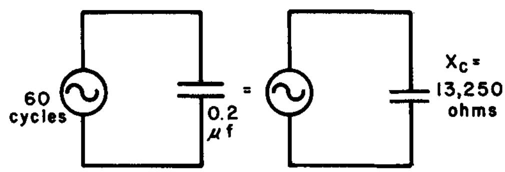
$$X_C = \frac{1}{2\pi f C} = \frac{1}{6.28 \times 120 \times 0.0000001} = 13{,}250 \text{ ohms}$$



DOUBLING THE FREQUENCY HALVES THE CAPACITIVE REACTANCE

Leaving the frequency fixed and doubling the capacitance

$$X_C = \frac{1}{2\pi f C} = \frac{1}{6.28 \times 60 \times 0.0000002} = 13{,}250 \text{ ohms}$$



DOUBLING THE CAPACITANCE HALVES THE CAPACITIVE REACTANCE

Given high values of capacitance and high-frequency voltages, capacitive reactance can fall to extremely low values, even to the point where the capacitor behaves as a virtual short-circuit to the voltage. For example, a 1-μf capacitor subjected to a voltage of 5 megacycles (5,000,000 cycles) has a capacitive reactance of

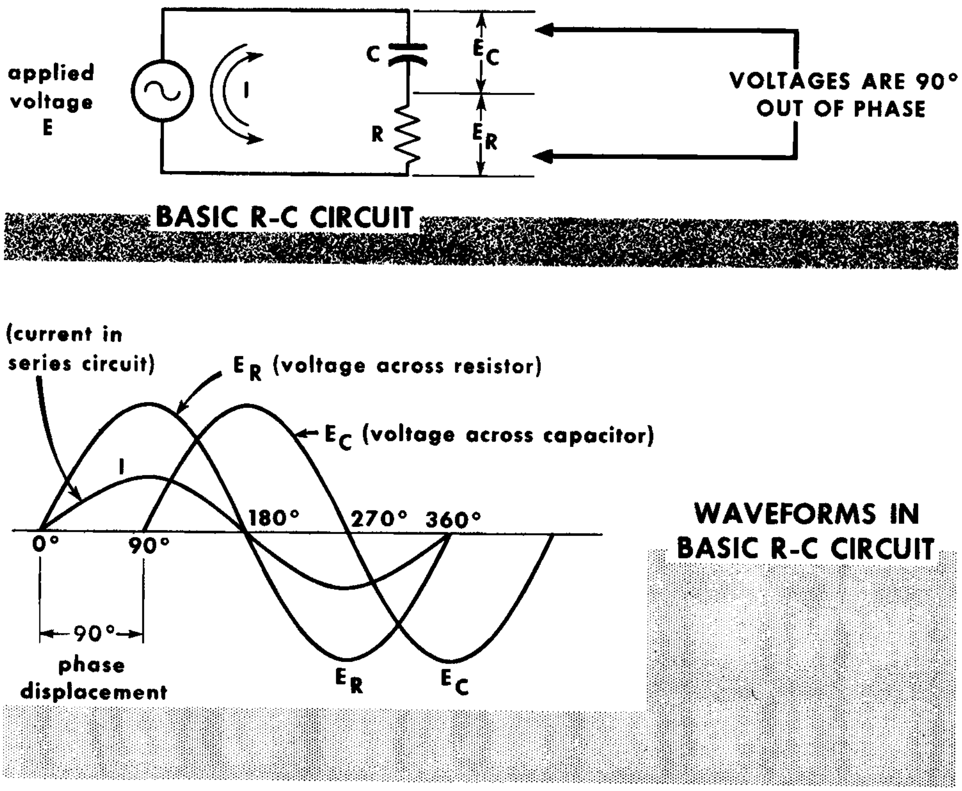
$$X_C = \frac{1}{2\pi f C} = \frac{1}{6.28 \times 0.000001 \times 5{,}000{,}000} = 0.0318 \text{ ohm}$$

On the other hand, if the capacitance value is very low, and the frequency of the applied voltage also is very low, the capacitive reactance can become so high as to behave like a virtual open circuit to the voltage. For instance, if the capacitance is 0.001 μf, and the frequency of the voltage is 5 cycles, the capacitive reactance rises to 31,847,000 ohms.

Current and Voltage in a Series R-C Circuit

In any circuit containing both capacitance and resistance, there is a 90° phase shift of current and voltage across the capacitance, and no phase shift across the resistance. As discussed in inductance, current in a series circuit is the same throughout and is, therefore, taken as the line of reference for both the capacitance and the resistance. Since the voltage across the resistance is in phase with the current through it, and the voltage across the capacitance is 90° out of phase with this same current, we can see that these two voltages are 90° out of phase with each other.

Current and Voltages in a Series R-C Circuit

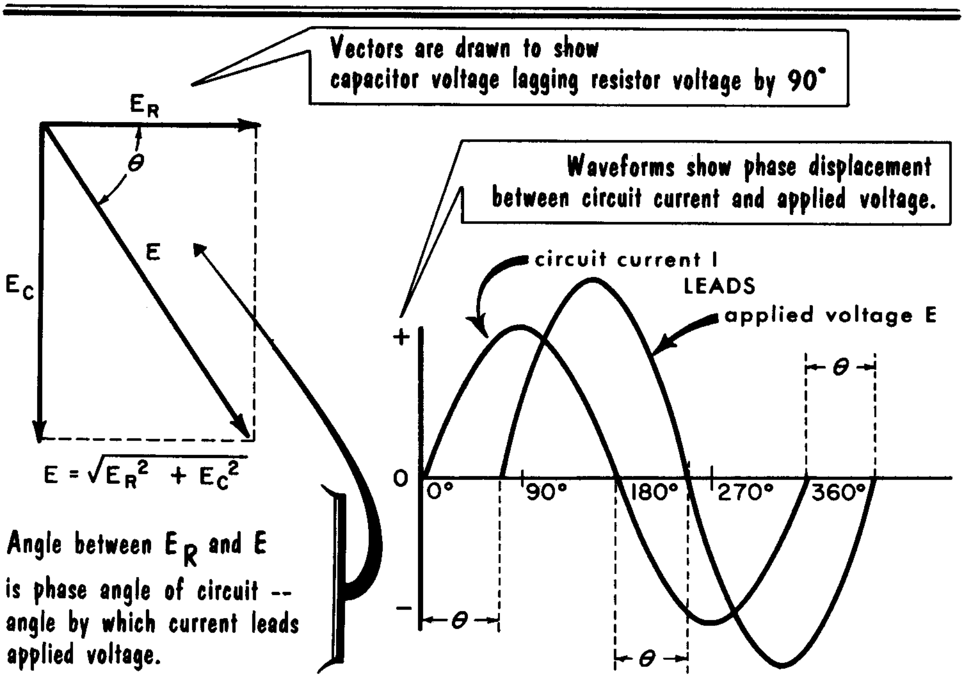


In the illustration, there is a basic series R-C circuit and the curves showing the relationship between the current and voltages across both components. The resultant voltage from the two voltage drops which are 90° out of phase is the voltage drop for the whole circuit and is, by Kirchhoff's law, equal to the applied voltage. The phase shift of the current in the circuit, measured with respect to the applied voltage, is called the phase angle of the circuit.

Current and Voltage in a Series R-C Circuit (Cont'd)

The relationship between the applied voltage, the voltage drops, and the phase angle of any series R-C circuit may be determined by means of vectors, as shown. The voltage across the resistance is plotted on the horizontal vector, and the voltage across the capacitance on the vertical vector. Since these two voltages are 90° out of phase, the angle between them is a right angle. By drawing in a parallelogram based on the two sides, the resultant vector E becomes the hypotenuse of a right triangle. By using the theorem that the square of the hypotenuse is equal to the sum of the squares of the other two sides, we get $E^2 = E_R^2 + E_C^2$, or $E = \sqrt{E_R^2 + E_C^2}$.

VECTORS AND WAVEFORMS OF CURRENT AND APPLIED VOLTAGE IN A SERIES R-C CIRCUIT

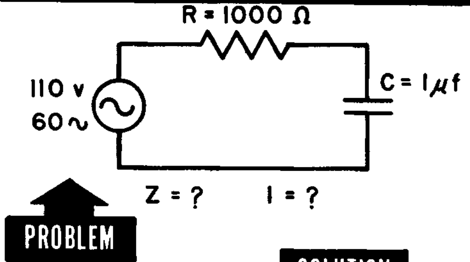


Since it is known that the current in the circuit is in phase with the voltage across the resistance, the direction of the current vector is the same as vector E_R, the voltage across the resistance. The phase angle θ then is the angle that the applied voltage E makes with vector E_R. If the voltage across the resistance is large with respect to that across the capacitance, the resultant vector will approach the horizontal and the phase angle will be small. Similarly, if the voltage across the resistance is small, the resultant vector will approach the vertical, and the phase angle will approach 90°. Hence, the presence of resistance in a capacitive circuit causes the current to lead the applied voltage by some angle less than 90°. The waveforms show the relative positions of current, voltage, and the phase angle θ.

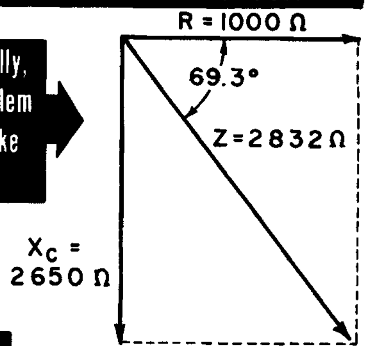
Impedance of a Series R-C Circuit

As stated in our discussion of R-L circuits, the total opposition offered by a circuit containing both a reactive element and a resistance is not the simple arithmetical sum of the reactance and the resistance. The capacitive reactance is added to the resistance in such a manner as to take into account the 90° phase difference between the two voltages in the circuit. To find the impedance of an R-C circuit, we use the same basic formula that we used in the R-L circuit; that is: $Z = \sqrt{R^2 + X_C^2}$. In short, the impedance of an R-C circuit is equal to the square root of the sum of the squares of the resistance and the capacitive reactance.

HOW TO FIND THE IMPEDANCE (Z) OF A SERIES R-C CIRCUIT



Vectorially, the problem looks like this



SOLUTION

1st find the X_C of C

$$X_C = \frac{1}{2\pi f C} = \frac{1}{6.28 \times 60 \times 0.000001} = \frac{1}{0.000377} = 2650\,\Omega$$

$$\text{Tangent } \theta = \frac{X_C}{R} = \frac{2650}{1000} = 2.65$$

look up on tangent table

$$\theta = 69.3^\circ$$

2nd find Z

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{1000^2 + 2650^2} = \sqrt{1{,}000{,}000 + 7{,}023{,}000} = \sqrt{8{,}023{,}000} = 2832\,\Omega$$

$$I = \frac{E}{Z} = \frac{110}{2832} = 0.0389 \text{ Ampere, or}$$

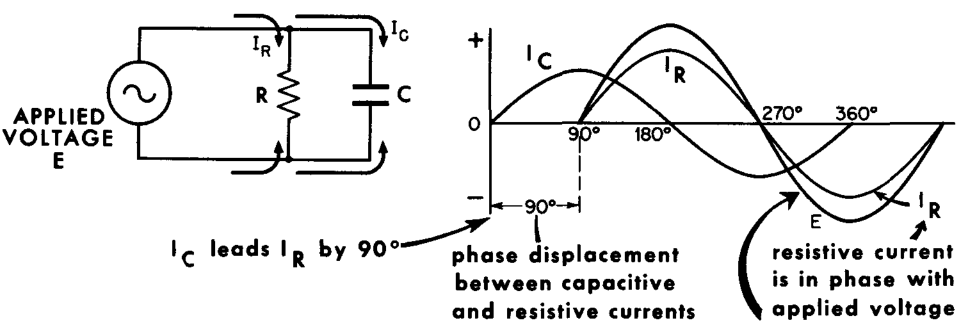
38.9 Milliamperes

The same result can be obtained by the use of vectors. Since the same current flows in C and R, the vectors can be made proportional to the resistance and the capacitive reactance. Note that the angle θ is the phase angle because the direction of the impedance vector is the same as that of the applied voltage vector. This angle may be determined by its tangent, X_C/R. The total current in an R-C series circuit can then be determined by the formula, $I = E/Z$.

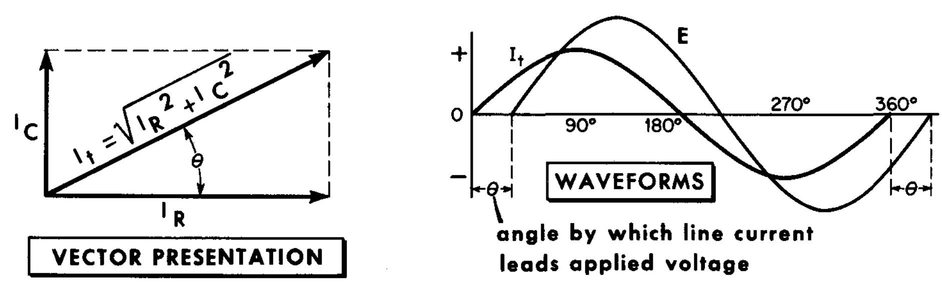
Parallel R-C Circuits

We show here capacitance C and resistance R connected in parallel across an a-c source. Since this is a parallel circuit, voltage is the same everywhere; thus, all voltages are in phase with each other. However, the current through the capacitor leads the applied voltage by 90°, and the current through the resistance is in phase with the applied voltage, as shown in the waveforms. Thus, the capacitive current leads the resistance current by 90°, and the resultant current, or total line current, is the vectorial sum of these two currents.

VOLTAGE AND CURRENTS IN PARALLEL R-C CIRCUIT



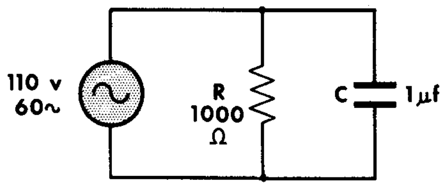
Total Current and Applied Voltage in Parallel R-C Circuit



In making vectors for this situation, the current through the resistance I_R is laid off on the horizontal vector, and the current through the capacitance I_C on the vertical vector. Because the capacitive current leads the resistive current, the I_C vector is laid off in the positive direction. The resistive current is taken as the reference vector, since it is in phase with the applied voltage and represents the direction of the applied voltage. The resultant vector I_t represents the total current in the circuit, and the angle this vector makes with the horizontal is the phase angle θ. The line current, then, is said to lead the applied voltage by the angle θ. The tangent of this angle is equal to R/X_C. The total current of this circuit is equal to: $I_t = I_R^2 + I_C^2$. Thus it can be seen that total current in a parallel R-C circuit, as in d-c circuits, is always greater than the current in either branch.

Impedance of a Parallel R-C Circuit

HOW TO FIND THE IMPEDANCE OF A PARALLEL R-C CIRCUIT



Z = ? I = ?

PROBLEM

$$I_t = \frac{E}{Z} = \frac{110}{937} = \text{0.117 ampere or 117 milliamperes}$$

The line current leads the applied voltage by the angle θ

$$\text{Tangent } \theta = \frac{R}{X_C} = \frac{1000}{2650} = .377$$

Found in the Tangent Table → = .377

$$\theta = 20.7^\circ$$

SOLUTION

1st Find X_C of C

$$X_C = \frac{1}{2\pi fC}$$

$$= \frac{1}{6.28 \times 60 \times 0.000001}$$

$$= \frac{1}{0.000377}$$

$$= 2650 \text{ ohms}$$

2nd Find Z

$$Z = \frac{R X_C}{\sqrt{R^2 + X_C^2}}$$

Vectorial Addition of R + X_C

$$= \frac{1000 \times 2650}{\sqrt{1{,}000{,}000 + 7{,}023{,}000}}$$

$$= \frac{2{,}650{,}000}{2840}$$

$$= 937 \text{ ohms}$$

Less than R or X_C

The impedance of a parallel R-C circuit may be calculated by using the same general formula as for finding the total resistance of resistors in parallel. Because we are dealing with vector quantities when we discuss R and X_C, we cannot add them arithmetically – they must be added vectorially. Thus we get the formula for R and C in parallel:

$$Z = \frac{R \times X_C}{\sqrt{R^2 + X_C^2}}$$

From this, we can see that the total impedance of a parallel R-C circuit always is somewhat less than either the resistance or the reactance.

Indirectly, the impedance also can be found by finding the total current and then using the formula Z = E/I. To find the total current, we must first find the individual currents that flow in R and in C. This is done by Ohm's law, I = E/R and I = E/X_C. Knowing these currents, the total current then can be found by adding them vectorially. For instance, if 3 amperes flow through R and 4 amperes flow through C, the vectorial sum of 3 and 4 is 5 amperes, which represents the total current ($I_t = \sqrt{I_R^2 + I_C^2}$, or $5 = \sqrt{3^2 + 4^2}$).

A capacitor is any two conductors separated by an insulating material (dielectric).

Capacitance is a property of a circuit whereby energy may be stored in the form of an electric field between two conductors (plates) separated by a dielectric.

Capacitance in a circuit opposes any change in voltage.

The action of storing electricity in a capacitor is called charging.

The action of recovering the energy stored in a charged capacitor is known as discharging the capacitor.

In a capacitive circuit, current leads the applied voltage by 90°.

The ratio of the charge (Q) to the voltage (E) is the measure of capacitive action, and is called capacitance. $C = Q/E$.

A farad is the unit of capacitance. A capacitor has a capacitance of 1 farad if a 1-volt difference in potential results in the storage of 1 coulomb of charge.

A microfarad is one-millionth of a farad; a micromicrofarad is one-millionth of a microfarad, or one million millionth of a farad.

The capacitance of a capacitor varies directly with the plate surface area and inversely with the distance between the plate surfaces.

Dielectric constant (K) is the ability of a material, or medium, to permit the establishment of electric lines of force between oppositely charged plates.

Fixed capacitors are distinguished, according to the dielectric material used, as paper, oil, mica, ceramic, and electrolytic capacitors.

Electrolytic capacitors show polarity and are used principally in high-power, low-frequency filter circuits up to 600 volts.

Capacitors connected in parallel add like resistances in series; capacitors connected in series divide according to the parallel-resistance formula.

The opposition a capacitor presents to a-c is called capacitive reactance (X_C); the formula for capacitive reactance is $X_C = 1/2\pi fC$.

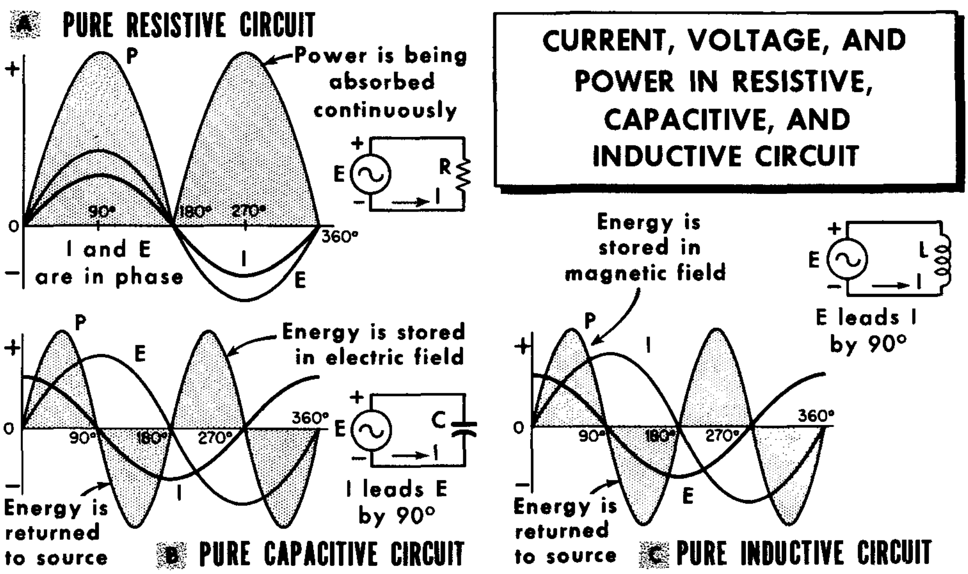
If either, or both, frequency or capacitance increases, X_C decreases, and vice versa. Impedance of a series R-C circuit is calculated by $Z = \sqrt{R^2 + X_C^2}$; impedance of a parallel R-C circuit is calculated by $Z = R \times X_C/\sqrt{R^2 + X_C^2}$.

REVIEW QUESTIONS

1. What is a capacitor? What is capacitance?
2. Describe the action of current flow in a capacitor when a d-c voltage is applied.
3. What factors determine the amount of current flow in a capacitor?
4. What is a dielectric constant?
5. Define capacitive reactance (X_C). Give the formula for it.
6. How does capacitance vary with respect to the area of the plates, the distance between them, and the dielectric constant?
7. Define the farad, the microfarad, and the micromicrofarad.
8. When does a capacitor discharge?
9. What two functions are characteristic of a capacitor?
10. Give the formulas for calculating capacitors in series and in parallel.
11. Give the formulas for calculating the impedances of series R-C circuits and parallel R-C circuits.

Resistive, Inductive, and Capacitive Circuits

The power absorbed by a resistance in a d-c circuit is expressed by P = E × I, or by I^2R and or by E^2/R. All lead to the same answer. The power absorbed by resistance in an a-c circuit is expressed by exactly the same equations. When an a-c voltage is applied to a pure resistance only, power is absorbed each instant, regardless of the direction of the current. The power consumed during a complete cycle is equal to the effective value of current multiplied by the effective value of voltage, shown simply as P = E × I. The power curve for current flowing through a resistance is positive (power is being absorbed) for each half-cycle of the current; there are two positive power loops for the complete cycle of 360°.



If an a-c voltage is applied to a pure capacitance, current flows into the capacitor during the charging half-cycle. During this interval, the capacitor absorbs energy from the voltage source and stores it in the form of an electric field. Then, during the discharge half-cycle, the capacitor returns all of the energy it has stored to the voltage source. Thus, over a complete cycle, the power absorbed by the pure capacitance is zero. When illustrated by a power curve, the power absorption intervals are shown by positive loops, whereas the power return intervals are shown by negative loops. There are two power loops (one positive and one negative) for each half-cycle of current, or four power loops for a complete cycle.

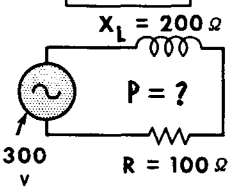
A similar situation prevails when an a-c voltage is applied to a pure inductance. Energy is absorbed during the time the current is building the magnetic field around the inductor, and power is returned to the source when the magnetic field collapses back into the inductor. As much electrical energy is returned as is absorbed; hence the net power consumed by a pure inductance during a complete cycle is zero.

Power Factor

The power consumed in a resistive a-c circuit is calculated in exactly the same manner as in d-c circuits (P = E × I). To get an equivalent answer between d-c and a-c, we use the effective (rms) values of a-c voltages and currents. However, in inductive and capacitive circuits, the simple power formula P = E × I requires further consideration. We saw that, in purely inductive and capacitive circuits, all the energy stored in the form of magnetic and electric fields was returned to the source on the 2nd quarter-cycle, and that the net power dissipated was zero. The practical inductive or capacitive circuit always contains some resistance, however little. This resistance makes the phase angle between voltage and current somewhat less than 90°, and some power will be consumed – none will be returned to the source. If we were to measure the current and voltage in an inductive circuit that contained resistance, we would not get the true power consumed by multiplying E × I, because we would be ignoring the partial power of the inductor, which is returned to the source.

When we measure E and I and then find their product, we get the apparent power consumed in the circuit. In a purely resistive circuit, the apparent power is the same as the true power. However, in an inductive or capacitive circuit, we must take into consideration the phase angle between E and I, using the formula P = E × I × cosine θ. The use of "cosine θ" adds a power factor to our calculations. The cosine of 0° is 1. Thus, in a purely resistive circuit where the phase angle between current and voltage is 0°, power is simply E × I. The cosine of 90° is 0; therefore, in a purely inductive or capacitive circuit, the power is E × I × 0, or zero.

PROBLEM



Power in an A-C Circuit is measured by

$$P = E \times I \times \text{cosine}\ \theta$$

SOLUTION

$$P = E \times I \times \text{cosine}\ \theta$$
$$= 1.34 \times 300 \times \text{cosine}\ 63.4°$$
$$= 1.34 \times 300 \times 0.447$$
$$= 179 \text{ watts}$$

A

$$Z = \sqrt{200^2 + 100^2} = 224 \text{ ohms}$$

B

$$I = \frac{300}{224} = 1.34 \text{ amperes}$$

C

$$\text{tangent}\ \theta = \frac{X_L}{R} = \frac{200}{100} = 2 \text{ or } 63.4°$$

$$\text{Power factor} = \frac{\text{True power}}{\text{Apparent power}}$$

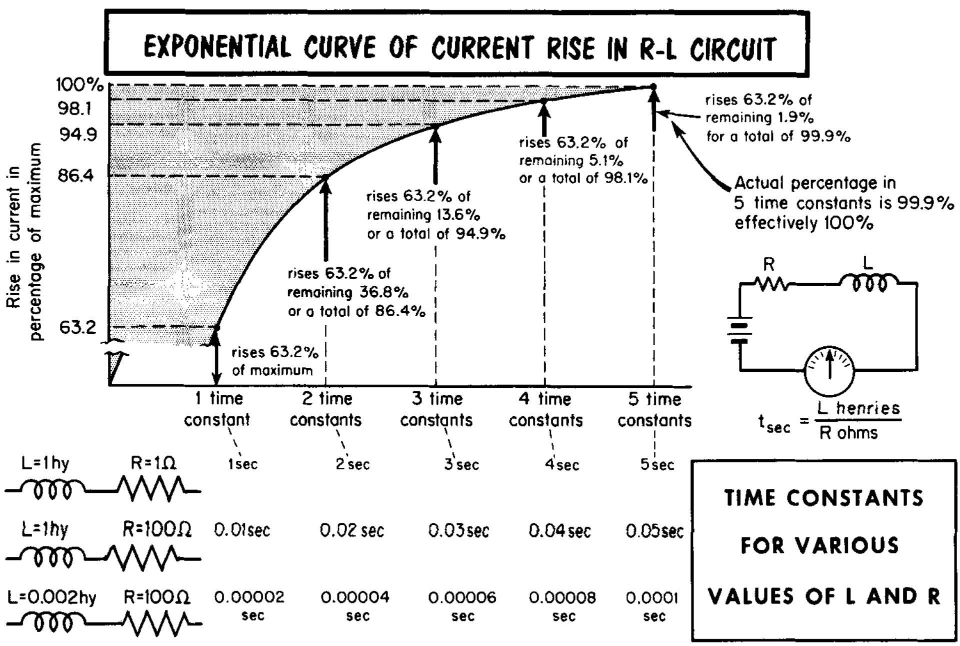
$$(1.34 \times 300) \rightarrow \; = \frac{179}{401} = 0.447$$

In all practical circuits containing C and R, or L and R, the cosine of the phase angle (θ) between E and I enables us to determine the true power consumed by the circuit. By dividing true power by apparent power, we get the power factor of the circuit – something between 0 and 1.

R-L Circuit Time Constant

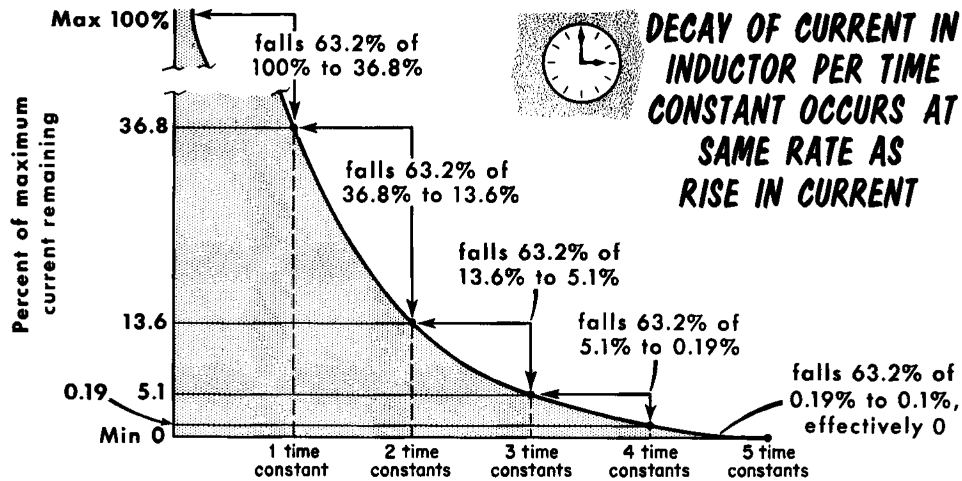
When a constant (d-c) voltage is applied to a resistance, the rise in current to a maximum value is instantaneous. If an inductance is connected in series with the resistance, time elapses while the current builds up to maximum. If we could observe the behavior of the current, we would see it rise rapidly from zero and then its rate of increase would progressively diminish. After a lapse of time, the current would reach a value which for all intents and purposes is maximum, equal to $I = E/R$. The relationship between the rise in the current to a given value and the time lapse while it is happening is determined by a term known as the time constant. Time constant is a means of comparing how rapidly the current in one R-L circuit rises to a given value relative to the current in another R-L circuit. Time constant is expressed in seconds, and is equal to the inductance L (in henries) divided by the resistance R (in ohms). The equation is: $t = L_{henries}/R_{ohms}$. If, for the moment, we assume L = 1 henry and R = 10 ohms, the time constant L/R = 0.1 second. The time constant varies in direct proportion to L and inversely with R.

The pattern of the increase in current in an R-L circuit is the same in all R-L circuits regardless of the specific values of L and R. When shown graphically, it is a singularly shaped curve known as an exponential curve. The characteristic of this pattern of change is that in a time equal to 1 time constant (1t), the current rises to 63.2% of its maximum value (regardless of what the maximum value may be). On this basis, in the numerical example given above, the current would rise to 63.2% of the maximum current in 0.1 second. The lapse of time corresponding to additional time constants permits current to rise to specific percentages of maximum (as illustrated).



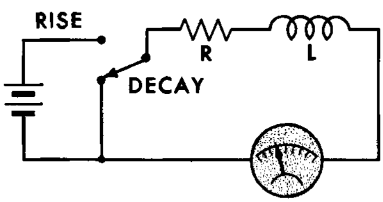
R-L Circuit Time Constant (Cont'd)

When a constant voltage is removed from a resistance, the current falls to zero instantly. When such a voltage is removed from an R-L combination, a time lapse occurs before the current decays to zero. As in the case of the rise in current, the decay in current follows an exponential curve, except that now the curve is an inverted version of the one which showed the rise in current. In a time equal to 1t, the current decreases 63.2% from the maximum; i.e., it falls to 36.8% of the maximum. In time 2t, it decays 86.4% from the maximum to a value equal to 13.6% of the maximum. As shown in the illustration, it decreases to 0.1% of the maximum; i.e., it falls 99.9% from the maximum or, effectively, to zero in 5t.



Units of Time Constant

When R is	When L is	Time constant is in
ohms	henries	seconds
megohms	henries	microseconds
ohms	millihenries	milliseconds
ohms	microhenries	microseconds



How do we calculate the momentary current? It is simple for whole-number time constants; for values in between, the chart is most convenient, as discussed later. Suppose L = 1 henry and R = 10 ohms. The applied voltage is 10 volts; t is 0.1 second. The maximum current is $I = E/R = 10/10 = 1$ ampere. If in constant 1t (0.1 second), the current rises 63.2% of maximum, it rises to 63.2% of 1 ampere or to $1 \times 0.632 = 0.632$ ampere. In 2t (0.2 second), it rises 86.4% of maximum, or to $1 \times 0.864 = 0.864$ ampere, etc., as shown on the previous page until in 5t (0.5 second), it reaches 0.999 ampere, effectively 1.0 ampere maximum. As to the decay of the current when the voltage is removed, in 1t (0.1 second) the current falls 63.2% of maximum, or decreases 63.2% of 1 ampere to 36.8% of the maximum, which amounts to 0.368 ampere.

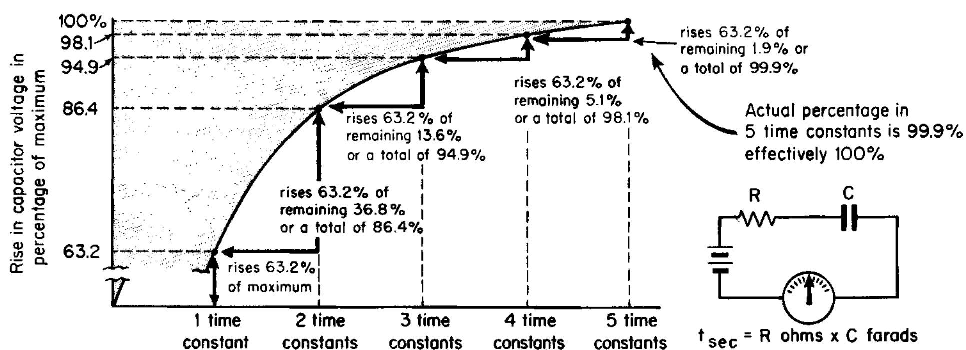
R-C Circuit Time Constant (Charging)

When a constant voltage is applied to a capacitance, the voltage built up in the capacitor by the charging current reaches the value of the charging voltage almost instantly. If a resistance (R) is connected in series with the capacitor (C), it tends to limit the amount of charging current and, in so doing, causes time to lapse while the capacitor acquires a charge. The factor that determines the rise in voltage in the capacitor of an R-C circuit relative to time is the time constant of the circuit. The R-C circuit time constant is also expressed in seconds, and equals the resistance (in ohms) multiplied by the capacitance (in farads). Expressed as an equation, it is:

$$t = R \times C$$

Imagine a C of 1 µf (0.000001 farad) in series with an R of 1000 ohms. The time constant t = 1000 × 0.000001 = 0.001 second. The applied voltage is 100 volts. The buildup of voltage in the capacitor is shown by exactly the same shaped curve as illustrated in the current rise of the R-L circuit. In 1t, the current flowing into the capacitor builds the capacitor voltage to 63.2% of the maximum (the applied voltage), or to 63.2 volts. In 2t, the capacitor voltage rises to 84.6%; in 3t, to 94.9%; in 4t, to 98.1%; and in 5t, it reaches 99.9% (effectively 100%). In voltage values, these percentages are 84.6, 94.9, 98.1, and 99.9 (or 100) volts. In 5t, the capacitor is, for all practical purposes, fully charged. Given the same applied voltage, but an R-C circuit with a t of 0.005 second, exactly the same percentages of maximum voltage would appear in the capacitor per time constant as before, but now it would require 0.005 second to build to 63.2%, and a correspondingly increased time to build to the higher voltage values.

EXPONENTIAL CURVE OF VOLTAGE ACROSS C IN AN R-C CIRCUIT



R	C	1 time constant	2 time constants	3 time constants	4 time constants	5 time constants
R=1000 Ω	C=1µf	0.001sec	0.002sec	0.003sec	0.004sec	0.005sec
R=10K	C=10µf	0.1 sec	0.2 sec	0.3 sec	0.4 sec	0.5 sec
R=2Meg	C=0.0025µf	0.005sec	0.01sec	0.015 sec	0.02sec	0.025sec

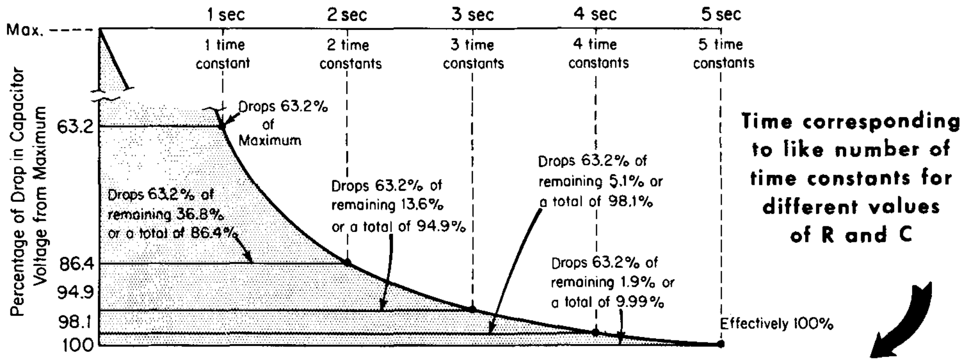
Time Constants for Various Values of R and C

R-C Circuit Time Constant (Discharging)

The action during discharge of the capacitor in the R-C circuit is the opposite of capacitor charging. The time required for complete discharge and for the capacitor voltage to fall effectively to zero is extended over that when no resistance is present in the circuit. The curve which shows the decrease of the capacitor voltage is exactly the same as the one which shows the decrease in inductor current. It is the inverted version of the curve which shows the rise in voltage across the capacitor during charging (see preceding page).

The percentage fall of the capacitor voltage from its maximum value relative to time is a function of the time constant of the circuit. You have learned that the capacitor acquires 63.2% of its maximum charge during the first period amounting to 1t. During discharge the capacitor loses .632 of its full charge in the first time interval equal to 1t. Thereafter, during the time interval equal to each succeeding time constant, it loses .632 of the charge still remaining in it.

IN THE R-C CIRCUIT, THE ACTION DURING DISCHARGE IS THE OPPOSITE OF CHARGE

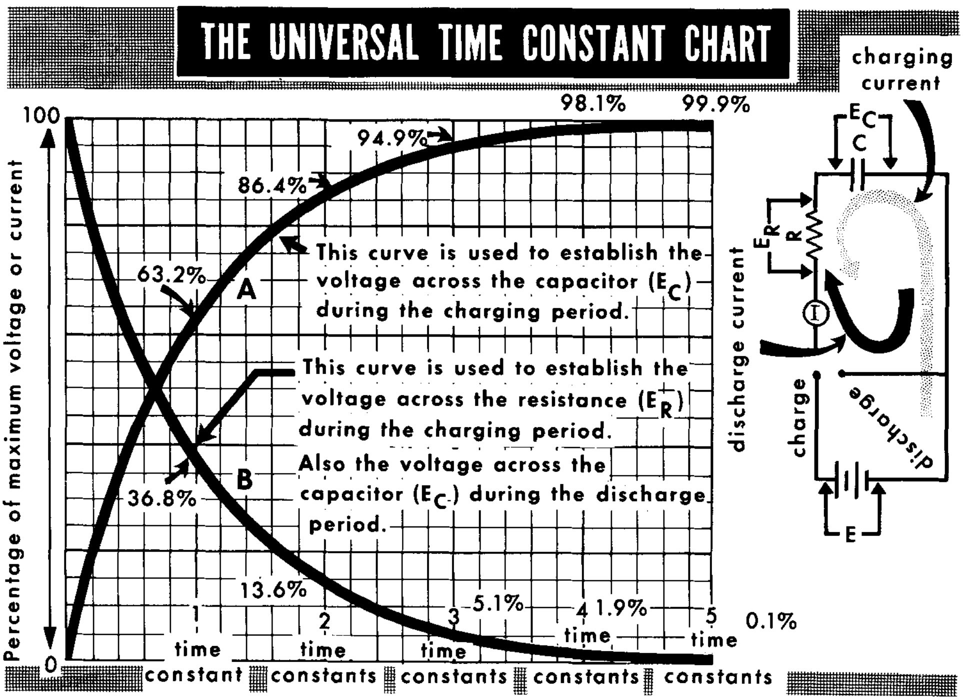


R=1000, C=1μf	0.001 sec	0.002 sec	0.003 sec	0.004 sec	0.005 sec
R=10K, C=10μf	0.1 sec	0.2 sec	0.3 sec	0.4 sec	0.5 sec
R=2 Meg, C=0.0025μf	0.005 sec	0.01 sec	0.015 sec	0.02 sec	0.025 sec

Consider the first time interval equal to 1t. Assume that the capacitor is charged to a maximum of 100 volts; it loses 63.2% of its maximum charge. If 100 volts = 100%, the capacitor loses 63.2 volts of its charge; hence, there remains 100 – 63.2, or 36.8 volts of charge in the capacitor. In 2t, the capacitor loses 63.2% of its charge, or 63.2% of the 36.8 volts that still remain in it. Thus, in time equal to 2t, the capacitor loses a total of 86.4% of the original maximum charge, leaving 13.6% or 13.6 volts in the capacitor, and so on as illustrated. At the end of 5t, the capacitor has lost 99.9% of its charge, or 99.9 of the original 100 volts. Effectively, this is considered as leaving zero voltage in the capacitor.

Applying the Universal Time Constant Chart

We have said that the pattern of current rise in an R-L circuit is the same for all values of R and L. The curve that shows the decay of current in the R-L circuit similarly suits all values of R and L. The same two curves apply to the R-C circuit. The curve that shows the rise in capacitor voltage while charging is the same as the one that shows the rise in current in the R-L circuit. The curve that shows the fall in capacitor voltage during discharge is identical with the curve that shows the decay of current in the R-L circuit. Because of these similarities, the two curves are known as Universal Time Constant Curves. They are shown on a single chart. The horizontal axis is calibrated in units totalling 5 time constants, whereas the vertical axis is divided uniformly in percentages of rise and fall of current in the R-L circuit, or rise and fall of capacitor voltage in the R-C circuit.



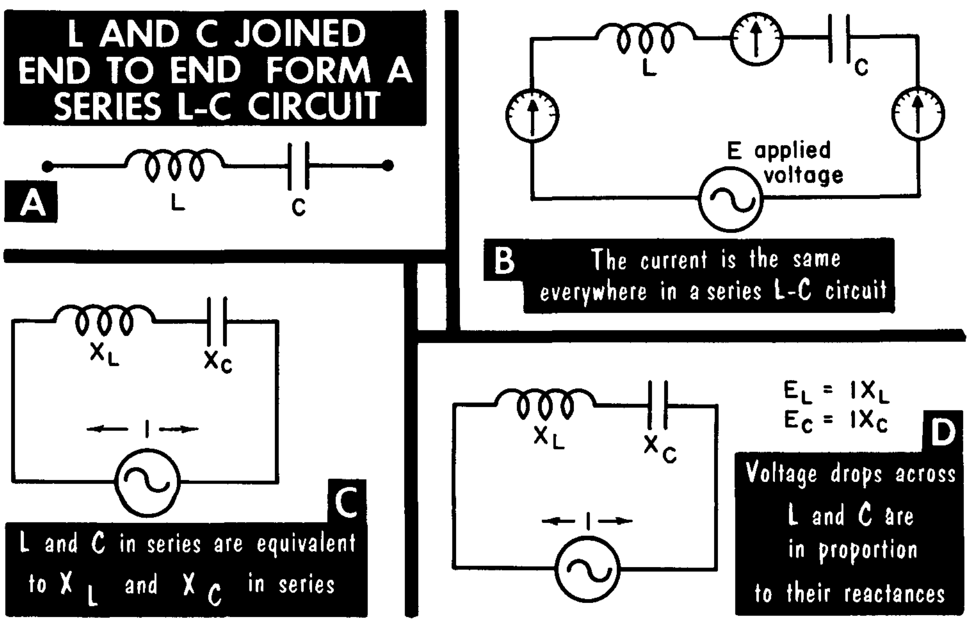
Assume that 100 volts is applied to an R-C circuit. t = 0.1 second. In 1t, the capacitor voltage will rise to 63.2 volts. What will be the voltage in 1.5t? Projecting 1.5t from the horizontal axis to curve A and its point of intersection with the curve to the vertical axis, we see that the capacitor voltage builds up to approximately 77.5% or 77.5 volts. At 0.5t, the voltage in the capacitor is just under 40% of maximum, or just under 40 volts. How low does the capacitor voltage fall in 0.4t? Using curve B, it is seen to fall 32.5% of the full charge to 67.5 volts. How low does the voltage fall in 2.6t? Using curve B, the answer is approximately 93% of maximum, leaving 7% in the capacitor. Hence, the capacitor voltage is 100 × 0.07 = 7 volts.

Relationship of L and C in Series (General)

When an inductance L and a capacitance C are joined end to end, a series L-C circuit is formed, as shown in A. Let us assume L to be a pure inductance and C to be a pure capacitance. In other words, the circuit has no d-c resistance. In practice, this electrical situation is never realized. The assumption is valid, however, because in most practical cases, the d-c resistance is so little as to be unimportant in the analysis of the circuit. Later, we shall discuss the effects of resistance.

If an a-c E is applied to a series L-C circuit, a current I will flow. Since a series circuit offers only one path for the current, the same current flows in the inductance and the capacitance – i.e., through the inductive reactance of L and the capacitive reactance of C (as in B). The characteristic that limits the current in an a-c circuit is the impedance of the circuit. In the series L-C circuit assumed to be free of d-c resistance, the circuit impedance consists of reactance only – inductive reactance X_L in series with capacitive reactance X_C (as in C).

Inductive and capacitive reactances are, with reference to their opposition to current flow, comparable to resistance, although they are different phenomena. The factor common to them is that current flow through a reactance or a resistance results in the appearance of a voltage drop across each. Current through a resistance results in voltage drop $E_R = IX_R$; current through an inductive reactance results in voltage drop $E_L = IX_L$; current through a capacitive reactance results in voltage drop $E_C = IX_C$ (see D). These voltages differ in one major respect – voltage across resistance is in phase with current; voltage across an inductance leads current by 90°; voltage across capacitance lags current by 90°.



L and C in Series (Impedance)

Imagine a circuit in which inductance L is 1 henry (negligible d-c resistance), in series with a capacitance C of 10 μf. Applied voltage E is 120 volts, and frequency f is 60 cycles. Solving for the reactances:

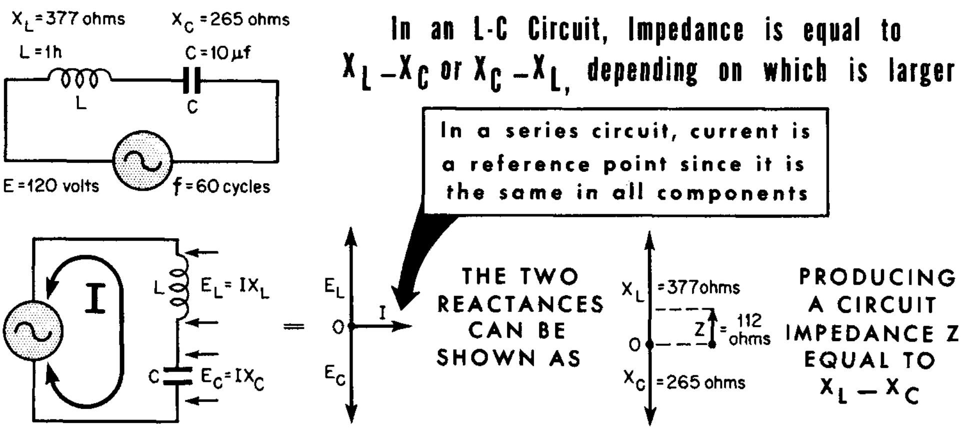
$$X_L = 2\pi fL = 6.28 \times 60 \times 1 = 376.8 \text{ or } 377 \text{ ohms}$$

$$X_C = \frac{1}{2\pi fC} = \frac{1}{6.28 \times 60 \times 0.00001} = 265.3 \text{ or } 265 \text{ ohms}$$

Now how do we establish the circuit impedance? We find the answer in a rule, "The impedance of a series L-C circuit having no resistance is equal to the difference between the ohmic values of the inductive and capacitive reactances." Therefore, for our circuit:

$$\begin{aligned} \text{Impedance } Z &= X_L - X_C \\ &= 377 - 265 \\ &= 112 \text{ ohms (inductive)} \end{aligned}$$

Reactance X_C was subtracted from reactance X_L because X_L was the greater quantity. If the situation were reversed, X_L would be subtracted from X_C.



In an L-C Circuit, Impedance is equal to $X_L - X_C$ or $X_C - X_L$, depending on which is larger

We illustrate this rule vectorially by arranging the X_L vector pointing up from a reference point, and the X_C vector pointing down from the <u>same</u> point. The vectors point in opposite directions. If we make each vector length proportional to its ohmic value, using the same scale for both, the shorter vector can be subtracted from the longer. The remainder is the difference between the two and is the <u>net reactance</u>, which we call <u>impedance</u> (Z). Since X_L in our circuit is greater than X_C, the impedance consists of inductive reactance; hence, the Z vector has the same direction as the X_L vector. This accounts for the reference, "inductive," in connection with the impedance value. More about this later.

Impedance, and Calculating Impedance Problems

The references to inductive and capacitive impedances (Z) require clarification. Where X_L exceeds X_C, Z equals $X_L - X_C$. With X_C greater than X_L, Z equals $X_C - X_L$. In either case, the remainder (called impedance) is net reactance: in one case, a certain amount of X_L; in the other, a certain amount of X_C.

For a given frequency, changes in the values of L and C result in different circuit impedances. The same is true if L and C are held constant and the frequency of E is varied. For example, assume L = 10 h (ten times that of the previous example) while C = 10 μf (as before). Frequency remains at 60 cycles. Impedance is as shown in A.

Now reduce capacitance by making C = 1 μf (a tenfold decrease), and restore L to 1 h. Frequency stays at 60 cycles. Impedance now is as shown in B.

Compare this amount of impedance with the example on page 2-98 (112 ohms inductive). Reducing C increases X_C so that it exceeds X_L. The circuit impedance therefore becomes capacitive.

Now change E from 60 to 160 cycles. Everything else stays as in the example immediately above. Impedance is as shown in C.

Changing frequency causes the two reactances to approach equality. Each individual reactance remains relatively high, but impedance has fallen to 10 ohms. In certain special cases (resonance), this net reactance can fall to zero. Lower frequency to 40 cycles, leaving everything else as above (D).

A study of these examples shows that the impedance can be low, medium, or high, and inductive or capacitive, depending on the values of L, C, and f.

CALCULATIONS OF IMPEDANCE IN SERIES L-C CIRCUITS

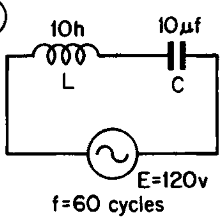
(A)

$X_L = 2\pi fL = 3768$ ohms

$X_C = \frac{1}{2\pi fC} = 265$ ohms

then $Z = X_L - X_C =$

3768 - 265 = 3503 ohms, inductive



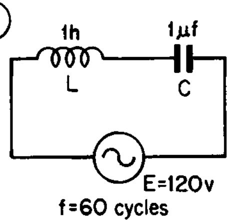
(B)

$X_L = 2\pi fL = 377$ ohms (actually 376.8Ω)

$X_C = \frac{1}{2\pi fC} = 2653$ ohms

then $Z = X_C - X_L =$

2653 - 377 = 2276 ohms, capacitive



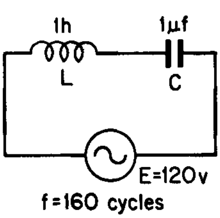
(C)

$X_L = 2\pi fL = 1005$ ohms (actually 1004.8Ω)

$X_C = \frac{1}{2\pi fC} = 995$ ohms (actually 995.2 Ω)

then $Z = X_L - X_C =$

1005 - 995 = 10 ohms, inductive



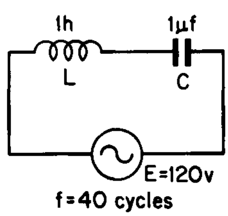
(D)

$X_L = 2\pi fL = 251$ ohms (actually 251.2Ω)

$X_C = \frac{1}{2\pi fC} = 3981$ ohms (actually 3980.8Ω)

then $Z = X_C - X_L =$

3981 - 251 = 3730 ohms, capacitive



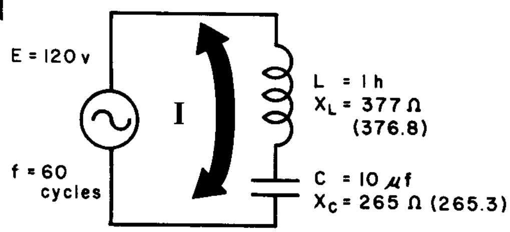
Calculating Current Problems

Given any value of circuit impedance for L and C in series, and knowing the applied voltage, the circuit current can be computed by applying Ohm's law for a-c. The equation is: $I = E/Z$, where E is the applied voltage, and Z is the circuit impedance. If you examine the equation, you can see that current varies in direct proportion to the applied voltage, and in inverse proportion to changes in Z. The latter relationship is shown in the different examples on this page.

The values of L and C, E, and f, as well as the current used in the examples, are not typical of series L-C circuits found in radio receivers. L and C usually are much smaller, as is the current. The values used here were selected because it was felt they helped to clarify the discussion.

CALCULATING CURRENT

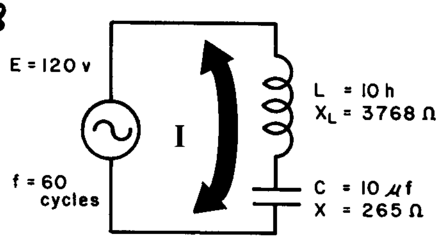
A



$$Z = X_L - X_C$$
$$= 112\ \Omega\text{, Inductive}$$
$$\text{Then } I = \frac{E}{Z} = \frac{120}{112}$$
$$= 1.07 \text{ Amperes}$$
in round numbers

B

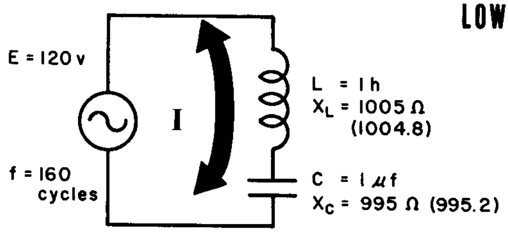
HIGH IMPEDANCE means LOW CURRENT



$$Z = X_L - X_C$$
$$= 3503\ \Omega\text{, Inductive}$$
$$\text{Then } I = \frac{E}{Z} = \frac{120}{3503}$$
$$= 0.034 \text{ Ampere, or}$$
34 milliamperes
in round numbers

C

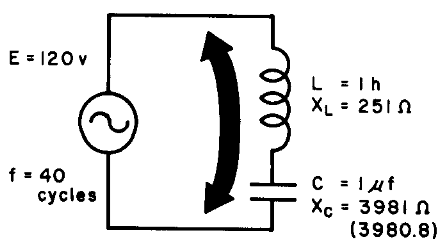
LOW IMPEDANCE means HIGH CURRENT



$$Z = X_L - X_C$$
$$= 10\ \Omega\text{, Inductive}$$
$$\text{Then } I = \frac{E}{Z} = \frac{120}{10}$$
$$= 12 \text{ Amperes}$$
in round numbers

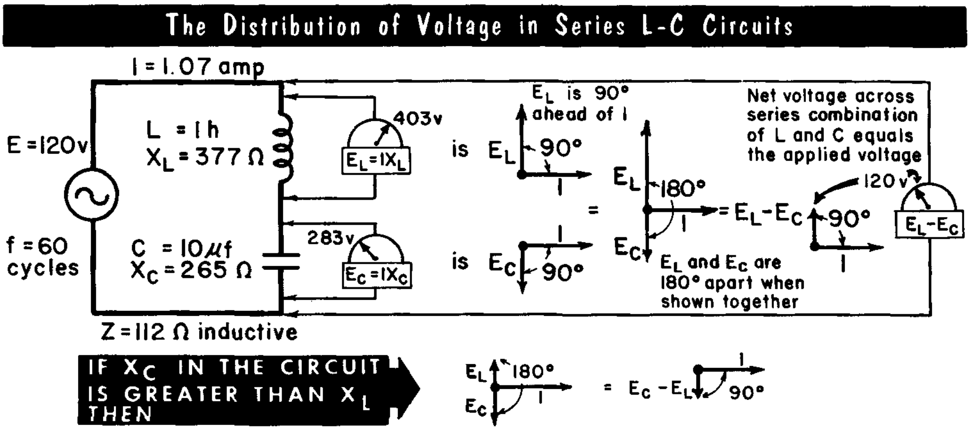
D

HIGH IMPEDANCE means LOW CURRENT



$$Z = X_C - X_L$$
$$= 3730\ \Omega\text{, Capacitive}$$
$$\text{Then } I = \frac{E}{Z} = \frac{120}{3730}$$
$$= 0.0322 \text{ Ampere, or}$$
32 milliamperes
in round numbers

Analyzing the Distribution of Voltage



In the circuit under discussion, E_L = 403 volts and E_C = 283 volts. Note that both of these are higher than the applied voltage, which is 120 volts. How can this be? It does happen in series L-C circuits; in fact, one or the other of these two voltages is always higher than the applied voltage, and, as shown in this example, both can be higher. The reason is that for any given circuit current, the individual voltage drops are determined by the individual reactances. The higher the reactance for a given current, the greater the individual voltage developed across it by the current. It is a characteristic of series L-C circuits that very high voltages can develop across L and C.

Let us analyze the voltages in the circuit discussed on page 2-98. As the result of current flow through L and C, a voltage E = IX_L appears across L, and a voltage $E_C = IX_C$ appears across C. Substituting the circuit values in the equations for E_L and E_C, we get:

$$E_L = IX_L = 1.07 \times 377 = 403 \text{ volts}$$

$$E_C = IX_C = 1.07 \times 265 = 283 \text{ volts}$$

Several significant details are associated with these two circuit voltages: each is independent of the other; each is measurable with a suitable voltmeter. Also, voltage E_L leads circuit current I by 90°, while voltage E_C lags I by 90°. If we show these two phase relationships in a single vector presentation using I as the reference vector, it is seen that E_L and E_C are 180° apart. They thus tend to offset each other in their effects on the circuit. This leads to the conclusion that the voltage present across the series L-C circuit as a whole (i.e., across the series combination of L and C), is the difference between the two voltages. In this instance, it is $E_L - E_C$, or 403 - 283 = 120 volts (the applied voltage). Since the current flowing through impedance Z develops the same voltage as the voltage difference between E_L and E_C, we can conclude that the voltage across the circuit impedance always equals the applied voltage.

Calculating the Impedance

The practical L-C circuit really is a series L-C-R circuit. The R is the inherent d-c resistance of the connecting wires and of the coil itself. The ohmic value of R equals the total d-c resistance of the circuit.

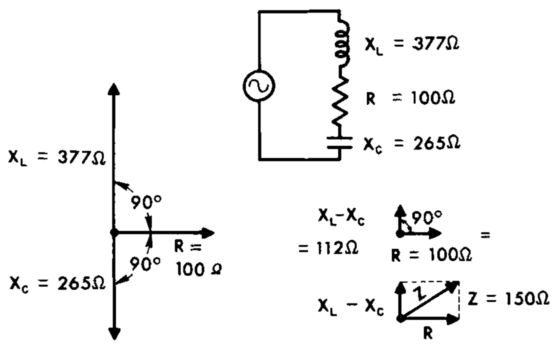
When R is a substantial quantity, the manner of determining impedance is different from when R is negligible. The net reactance of X_L and X_C must be determined, to which the circuit resistance is added. But this cannot be done by simple addition because the voltage and current relationships in the reactance differ in phase by 90°, while they are in phase in the resistance. To calculate the circuit impedance, the net reactance and the resistance must be added vectorially. We can do this using the right triangle relationship expressed by the equation $Z = \sqrt{R^2 + X^2}$, where X is the net reactance and R is the d-c resistance. Because the net reactance is the difference between X_L and X_C, the equation is changed to read $Z = \sqrt{R^2 + (X_L - X_C)^2}$, when X_L is greater than X_C, and to $Z = \sqrt{R^2 + (X_C - X_L)^2}$ when X_C is greater than X_L.

The Solution of the Impedance Z of Series L-C-R

by the Pythagorean Equation Method

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$
$$= \sqrt{100^2 + (377 - 265)^2}$$
$$= \sqrt{100^2 + 112^2}$$
$$= \sqrt{10{,}000 + 12{,}544}$$
$$= \sqrt{22{,}544}$$
$$= 150 \text{ ohms}$$

by Vector Presentation



For the problem to be solved, we refer back to a previous example: L = 1 henry; C = 10 μf; and f = 60 cycles. Here X_L was calculated to be 377 ohms and X_C 265 ohms. With R = 100 ohms, and the net reactance $X_L - X_C$ equal to 112 ohms (inductive), the circuit impedance Z equals 150 ohms. It constitutes an impedance (equivalent to an L-R circuit) in which the voltage is neither 90° ahead of the current, nor in phase with the current. The angle of lead of the voltage relative to the current is expressed by $\tan \theta = (X_L - X_C)/R$. This equals 112/100 = 1.12, or θ = 48.3°.

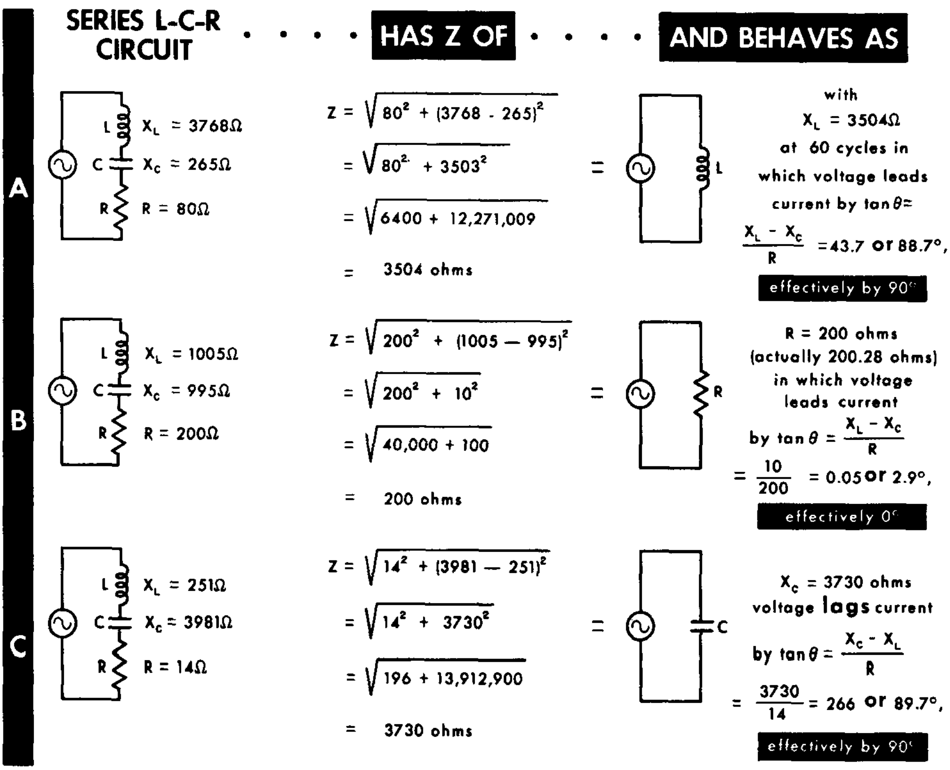
L-C-R in Series (Impedance)

The relative values of R and the net reactance of the series L-C-R circuit create a variety of impedance and circuit conditions. Imagine several circuits in which L and C are of such value as to present the indicated capacitive and inductive reactances. The resistance R of each circuit is as indicated.

When the circuit resistance is small, as in circuit A, in comparison to the net reactance, it contributes little to the circuit impedance. As a whole, the circuit behaves as if it were an inductance of such value as to present a reactance of 3504 ohms.

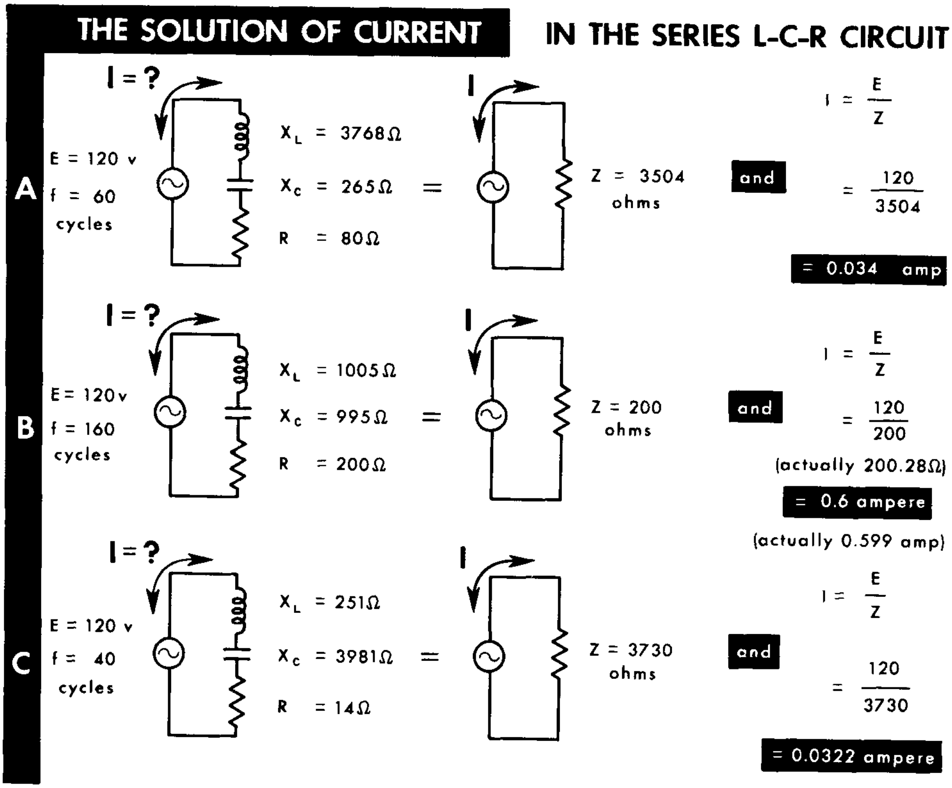
When the circuit resistance is high in comparison to the net reactance (circuit B), the impedance, for all intents and purposes, is made up of the resistance. As a whole, circuit B behaves like a resistance of 200 ohms. The greater the ratio between the net reactance and the circuit resistance, the greater the contribution of the resistance to the circuit impedance.

In circuit C too, the circuit resistance is very small in comparison to the net reactance (capacitive). Therefore, the resistance contributes very little to the circuit impedance. Circuit C behaves as though it were a capacitance with a reactance of 3730 ohms.



L-C-R in Series

Given an applied voltage E and circuit impedance Z, the circuit current I equals E/Z. The nature of Z – whether it is inductive, capacitive, or resistive – is unimportant in the calculation. It is only after the current is known, as with the voltage across the impedance, that the nature of the impedance becomes of interest. The same equation applies to all cases for Z. As you can see, the equation for current I is Ohm's law for a-c and is identical for the theoretically ideal (resistanceless) L-C circuit and for the practical L-C-R circuit. Let us apply the equation to the three examples of circuit impedance shown on the preceding page. Assume that applied voltage E = 120 volts.



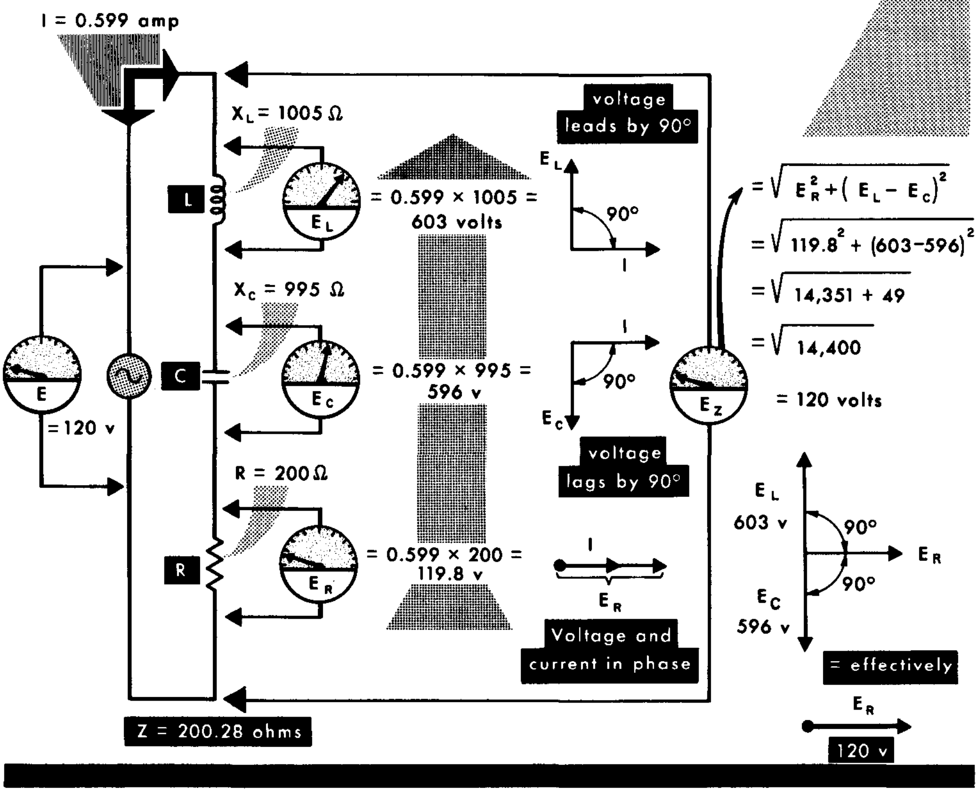
If you compare the current flowing in examples A on this page with example B on page 2-100, you will note no change in current despite the presence of a resistance of 80 ohms. The reason is that the prime control of the current is X_L; the additional current limitation imposed by the 80 ohms of resistance causes only a trifling reduction in current, which we do not show. On the other hand, if you compare example B above with example C on page 2-100, you will note that the presence of 200 ohms resistance reduces the current from 12 amperes to 0.6 ampere – a tremendous change.

L-C-R in Series (Voltages)

The voltages present across the components of the series L-C-R circuit are calculated in the same way as for the theoretically ideal (resistanceless) L-C circuit. The individual voltage drops (E_X) across the reactances are $E_X = IX$, where X stands for reactances; the voltage drop across the resistance is $E_R = IR$; and the voltage across the circuit as a whole (i.e., across the impedance Z) is $E_Z = IZ$. The illustration is of a typical case.

The addition of the two reactive voltages E_L and E_C to the resistive voltage E_R is done in the same way as the addition of reactances and resistance; that is, solving for Z. The same answer can be obtained by using vectors, but the equation method is much easier to use because many values are difficult to read from vector dimensions. When the three voltages are shown in the same vector presentation, the voltage E_R is the reference voltage. It is the voltage that is in phase with the series circuit current. The other two voltages, E_L and E_C, differ by 90° from E_R; one (E_L) leading E_R, and the other (E_C) lagging E_R by 90°. When all the voltages in the circuit are added, the voltage across the circuit as a whole is equal to the applied voltage.

THE DISTRIBUTION OF VOLTAGES IN THE SERIES L-C-R CIRCUIT



Resonant Frequency

Radio communications involves the transmission and reception of signals of selectable frequency. Such selection is possible because every combination of L and C responds better to voltages of one frequency than to voltages of other frequencies. The single frequency at which the circuit responds best is called the resonant frequency of the circuit.

Resonance occurs when the amount of inductance and the amount of capacitance in a circuit present equal amounts of reactance; i.e., $2\pi fL = 1/(2\pi fC)$, or $X_L = X_C$. Resonance in a series L-C-R circuit is, therefore, a particular condition in the circuit. Resistance R plays no part in determining the resonant frequency, although, as you will see, it does limit the amount of current at resonance and affects the behavior of the circuit off resonance. The equation used for calculating the resonant frequency of a circuit is:

$$\text{frequency of resonance, } f = \frac{1}{2\pi\sqrt{LC}}$$

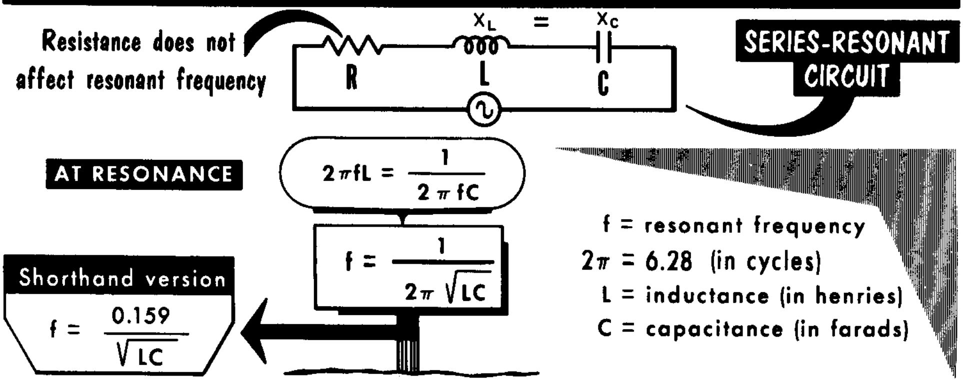
where π is a constant 3.1416, L is the inductance in henries, and C is the capacitance in farads. To illustrate the application of the equation, assume that L = 2 mh (0.002 henry) and C = 80 μμf (0.00000000008 farad). What is the resonant frequency of the circuit?

$$f = \frac{1}{6.28\sqrt{0.002 \times 0.00000000008}} = \frac{1}{6.28\sqrt{0.00000000000016}}$$

$$= \frac{1}{6.28 \times 0.0000004} = \frac{1}{0.000002512} = 398{,}089 \text{ cycles or } 398 \text{ kc (round numbers)}$$

A change in either L or C, or both, results in a change in resonant frequency, except when the product of L and C remains the same. For example, 0.002 × 80 μμf = 0.160. If L were 0.001 h (instead of 0.002), and C were 160 μμf (instead of 80), the product of 0.001 × 160 would be 0.160 – the same as above. Hence the two circuits would resonant at the same frequency – 398,089 cycles.

An L-C Circuit is Resonant when its Inductive Reactance equals its Capacitive Reactance



Determining the Impedance

Resonance is a particular condition in an L-C circuit at which $X_L = X_C$. Since the net reactance of a series L-C-R circuit is the difference between X_L and X_C, when these two reactances are equal (at resonance), the net reactance is zero. This leaves only the resistance as the current-limiting factor in the circuit. Thus, at the resonant frequency, the impedance of the series L-C-R circuit must be the lowest possible for the circuit, for at this frequency, the total circuit impedance is the circuit resistance.

Let us examine the impedance conditions at resonance. We shall use the constants mentioned on the preceding page. L = 2 mh; C = 80 μμf; and the resonant frequency is 398,089 cycles. Now we include the circuit resistance of 100 ohms. Then,

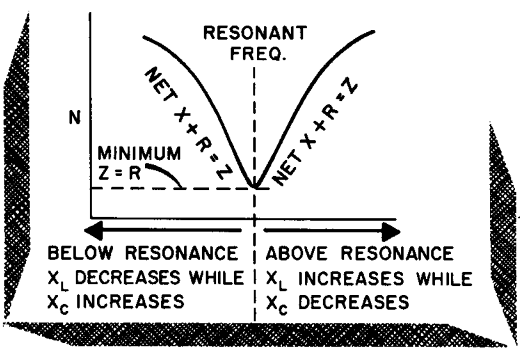
$$X_L = 2\pi fL = 6.28 \times 398,089 \times 0.002 = 5000 \text{ ohms (in round numbers)}$$

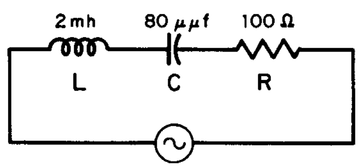
$$X_C = \frac{1}{2\pi fC} = \frac{1}{6.28 \times 398,089 \times 0.00000000008} = 5000 \text{ ohms}$$

Therefore,

$$Z = \sqrt{R^2 + (X_L - X_C)^2} = \sqrt{100^2 + (5000 - 5000)^2} = \sqrt{100^2 + 0} = 100 \text{ ohms}$$

At all other frequencies the net reactance is not zero; it has a finite value which increases (as does Z) as the operating frequency moves farther above or below the resonant frequency. We show this in the table for a range of frequencies from 100 to 600 kc. Examine the values in the Z column. The references to Z relative to frequency can be translated into a graph known as the impedance curve of the series resonant L-C-R circuit. Of course, the curve is for this particular set of circuit values.





Impedance Calculations At, Above, and Below Resonance

Frequency kc/sec	X_L ohms	X_C ohms	X ohms	R ohms	Z ohms
100	1,256	19,890	18,634	100	18,634
200	2,512	9,945	7,433	100	7,433
300	3,768	6,630	2,862	100	2,862
*398	5,000	5,000	0	100	100
450	5,620	4,410	1,240	100	1,240
500	6,280	3,978	2,302	100	2,302
600	7,536	3,315	4,221	100	4,221

*RESONANT FREQUENCY

Variations in Current

We have established that the circuit impedance of the series L-C-R circuit is minimum at the resonant frequency. For any given value of applied voltage, this means that the current is maximum. In fact, the occurrence of minimum impedance and maximum current at resonance is the identifying feature of the series resonant L-C-R circuit. Two equations can be used to compute the current: $I = E/Z$, for general application during the off resonant condition, and $I = E/R$, at the resonant frequency. Assume that E equals 30 volts at all frequencies and R equals 100 ohms. The rest of the circuit has the constants shown on the preceding page. Therefore, current at the resonant frequency of 398 kc (actually 398,089 cycles) is:

$$I = \frac{E}{R} = \frac{30}{100} = 0.3 \text{ ampere} = 300 \text{ milliamperes}$$

The table shows the frequency, impedance, and current for the circuit under discussion, over a frequency range of from 100 kc to 600 kc.

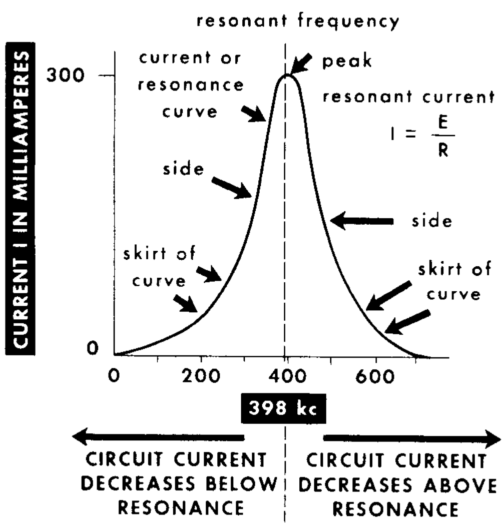
MAXIMUM CURRENT OCCURS AT ONE FREQUENCY -- THE RESONANT FREQUENCY

CHART OF VALUES

frequency (kc)	Z ohms	E(volts)	I(milliamps)**
100	18,634	30	1.61
200	7,433	30	4.0
300	2,862	30	10.48
350	1,270	30	23.6
398*	100	30	300
400	122	30	246
450	1,240	30	24.2
500	2,302	30	13.0
600	4,221	30	7.1

398 = resonant frequency

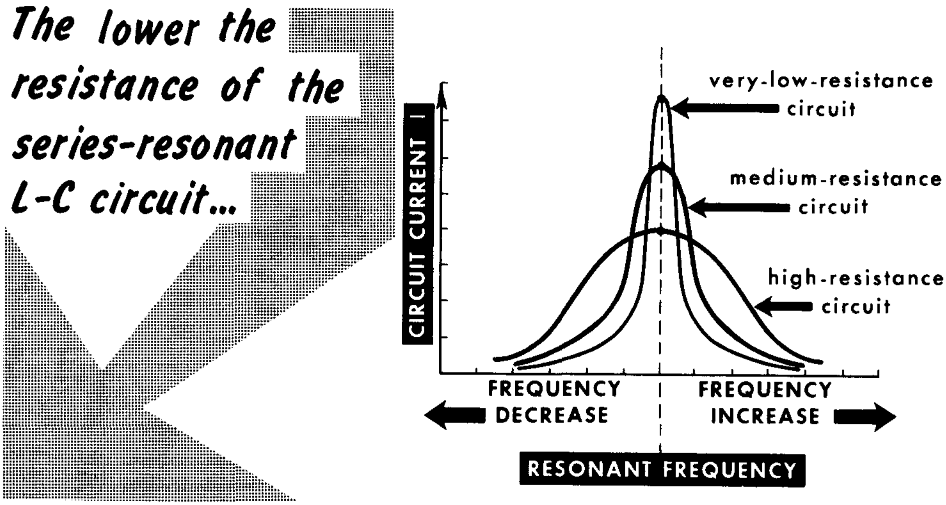
** When computing the current I, the answer will be in amperes, which must be converted to milliamperes.



The variation in circuit current with the change in frequency can be shown graphically. It is known as the current curve, or resonance curve, of the circuit. Frequency is shown along the horizontal axis, and the current amplitude is scaled along the vertical axis. Note that while maximum current flows at the resonant frequency, some current flows at frequencies off resonance. The high current at one single frequency (resonance) and reduced currents at other frequencies can be interpreted as discrimination by the circuit against frequencies other than the resonant frequency. This action of the circuit is the basis of its use.

The Effect of Resistance

We have explained the action of resistance in the series L-C-R circuit as being the factor which contributes to the control of the current. When we think of a circuit as a frequency-resonant system, we must take note of other effects of circuit resistance – that is, its effect on the circuit current or resonance curve. The lower the circuit resistance when $X_L = X_C$, the higher is the circuit current. If the current is shown graphically as a resonance curve, the peak of the current curve will be higher with lower circuit resistance. For instance, if the resistance of the circuit discussed on the preceding page were 10 ohms instead of 100 ohms, the resonant frequency current would have been 3 amperes instead of 300 milliamperes.



...the greater the change in current per unit change in frequency -- hence the SHARPER the circuit tuning.

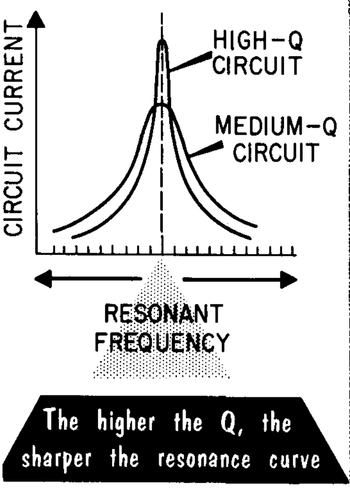
Equally important is another effect – this time on the shape of the resonance curve. The lower the circuit resistance, the greater the relative control of the current by the net reactance of the circuit as the frequency changes from resonance to off resonance. The result is that the circuit current falls rapidly as the frequency changes, and the sides of the resonance curve are steep. Conversely, the higher the circuit resistance, the less will be the relative control of the circuit current by the net reactance as the frequency is changed. The result is a broadening of the resonance curve. The steeper the sides of the resonance curve, the sharper will be the tuning of the circuit. This means that the circuit discriminates more against frequencies that are off resonance. In other words, the lower the resistance of a circuit, the sharper its performance as a frequency selector. Note the difference in peak amplitudes at the resonant frequency for the three conditions of resistance in the circuit. Note also the limited change in current for resonance and off resonance conditions with high resistance, compared with the large change in current for resonance conditions when there is very low resistance.

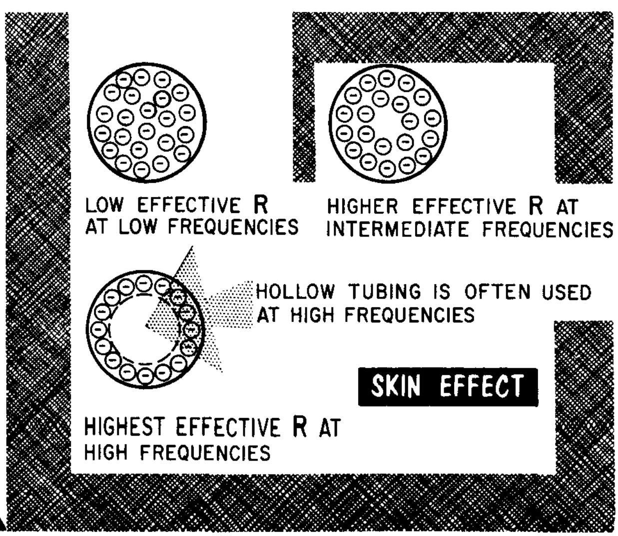
Effective Resistance and Q

All inductive windings, i.e., all coils which demonstrate the property of inductance, can be rated in terms of a figure of merit called Q. It is expressed as a number – for instance, a Q of 100. The higher the number, the more effectively the coil acts as an electrical device. When stated as an equation: $Q = X_L/R$, or $2\pi fL/R$.

In this equation, R has a somewhat different meaning than d-c resistance. Here, it stands for effective resistance, a term used to describe all forms of resistance (not reactance) which tend to retard the flow of current in a circuit. At d-c and low frequencies, the only opposition to the flow of current is the ohmic resistance of the wire. But at high frequencies, another kind of resistance – a-c resistance – appears and adds to the effective resistance of the coil, lowering its Q. This resistance is produced by skin effect. When skin effect is present, the electron flow is redistributed over the conductor cross section so as to make most of the electrons flow where they are encircled by the least number of magnetic flux lines. Because a greater number of flux linkages exists in the center of a conductor, the inductance at the center is greater than near the surface. Thus, at high frequencies, the reactance is great enough to affect the flow of current, most of which flows along the surface of the conductor. Therefore, the effective resistance is increased, since, in effect, the useful cross section of the conductor is greatly reduced.

Skin effect can be minimized by forming the conductor from a large number of small enameled wires connected in parallel at their ends, but insulated from each other throughout the rest of their length and interwoven. Each conductor will then link with the same number of flux lines as every other one, and the current will divide evenly among the strands, thus greatly increasing the useful cross section of the wire. A stranded cable like this is called a Litz conductor.





Time constant in an R-L circuit is a means of comparing how rapidly the current in one R-L circuit rises to a given value relative to the current in another R-L circuit. It is expressed in seconds and is calculated by the equation t (seconds) = L (henries)/R (ohms).

Time constant in an R-C circuit is calculated by the equation: t (seconds) = R (ohms) × C (farads).

In an R-C circuit, voltage rises to 63.2% of its maximum value and falls 63.2% from its maximum value in the first unit of time.

If X_L is greater than X_C, the circuit acts inductively and the current lags the applied voltage (E) by the phase angle θ. Tan $\theta = X_L - X_C/R$.

If X_C is greater than X_L, the circuit acts capacitively and the current leads the applied voltage (E) by the phase angle θ. Tan $\theta = X_C - X_L/R$.

If X_L is equal to X_C, the circuit is resistive and the current is in phase with the applied voltage (E).

The resonant frequency of the circuit is the particular frequency at which the circuit responds best.

The frequency at which an L-C-R circuit resonates is found by the formula $f = 1/2\pi\sqrt{LC}$.

At resonance, the reactances cancel, current is a maximum, impedance is a minimum, and the phase angle is 0°.

At resonance, the voltages across the reactances are maximum, and circuit current is maximum.

In a series L-C-R circuit, the voltages across the reactive elements are 180° out of phase and may be subtracted directly: $E = \sqrt{E_R^2 + (E_L - E_C)^2}$; $Z = \sqrt{R^2 + (X_L - X_C)^2}$.

Q is a measure of the selectivity of a circuit, and varies inversely with the resistance.

The Q of a series resonant circuit is the ratio of the inductive reactance to the effective resistance, and is equal to X_L/R.

In considering the Q of a coil, it is important to consider skin effect – an a-c resistance which at high frequencies adds to the effective resistance of a coil, causes losses, and lowers the Q.

REVIEW QUESTIONS

1. Give the formula for finding the impedance of an R-L circuit and of an R-C circuit?
2. What is the formula for finding the impedance of a series L-C-R circuit?
3. What determines the voltage drop across any single element in an a-c circuit?
4. In an L-C-R circuit, what is the phase relationship between current and voltage when X_L is equal to X_C.
5. What is the phase relationship between the voltages across the reactive elements in a series L-C-R circuit?
6. Give the formula for calculating the resonant frequency of a circuit.
7. Name the conditions present in a series resonant L-C circuit.
8. What is the relationship between current and impedance at resonance?
9. What is the Q of a resonant circuit? How is it determined?
10. What effect does the resistance in a circuit have upon the frequency of resonance?
11. In a series resonant circuit, what is the relationship between the voltage across either reactance and the applied voltage?

Branch Currents in the Basic Parallel L-C Circuit

A capacitor and an inductor connected in parallel across a voltage source make up a basic parallel L-C circuit. Since they are in parallel, the applied voltage appears across both L and C. We refer to the voltage across L as E_L and across C as E_C.

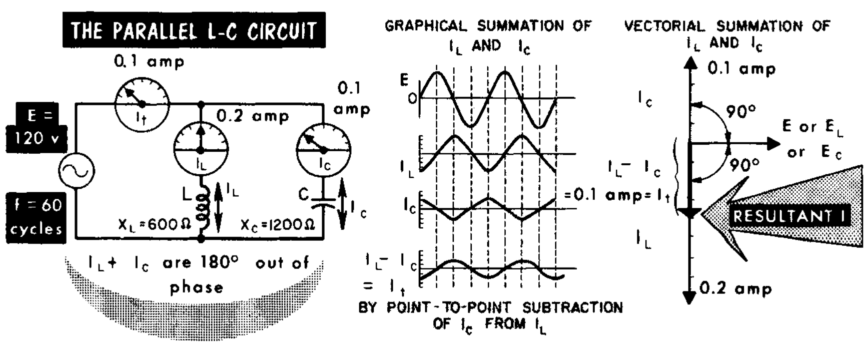
Assume that a parallel L-C circuit is made up of pure inductance and capacitance. The inductance is of such an amount as to present an X_L of 600 ohms; the capacitance is such as to present an X_C of 1200 ohms. The applied voltage is 120 volts at 60 cycles. Thus, the two branch currents are:

$$I_L = \frac{E}{X_L} = \frac{120}{600} = 0.2 \text{ ampere or 200 milliamperes}$$

$$I_C = \frac{E}{X_C} = \frac{120}{1200} = 0.1 \text{ ampere or 100 milliamperes}$$

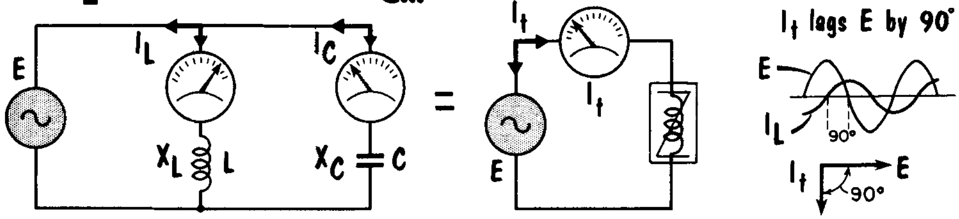
Note that the inductive branch presents the lesser amount of reactance; hence, it passes the greater amount of current.

Because the branch currents differ in phase by 180°, the line current in the parallel L-C circuit is determined by vectorial addition of the branch currents. When arranging the vectors, the applied voltage (identical to E_L and E_C) has the same phase across each branch; hence, it is suitable for use as the reference vector. With the inductive current I_L lagging E_L by 90°, the I_L vector is positioned 90° behind the voltage vector. The length of the I_L vector is determined by using any desired scale compatible with the value. The capacitive current I_C leads E_C by 90°; hence, the I_C vector leads the reference voltage vector by 90°. The scale used for vector I_C must be the same as for vector I_L. The current-voltage relationships shown establish the two currents as being 180° out of phase. The resultant of two vectors 180° out of phase is the difference between their magnitudes. So we subtract the smaller vector I_C from the larger vector I_L. The resultant is the total line current ($I_t = I_L - I_C$). The line current I_t equals 0.2 - 0.1 = 0.1 ampere. Since the inductive current I_L is the predominant current in the parallel network, the resultant current I_t has the same phase as the original I_L; that is, it lags the applied voltage by 90°.

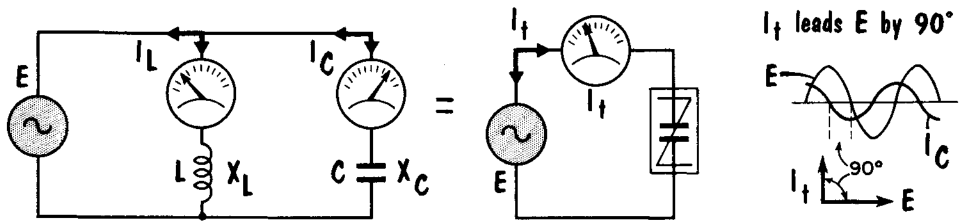


Branch Currents and Circuit Impedance

When I_L is greater than I_C, or X_L is lower than X_C... ...the circuit behaves as an inductance. Z is inductive.



When I_C is greater than I_L, or X_C is lower than X_L... ...the circuit behaves as a capacitance. Z is capacitive



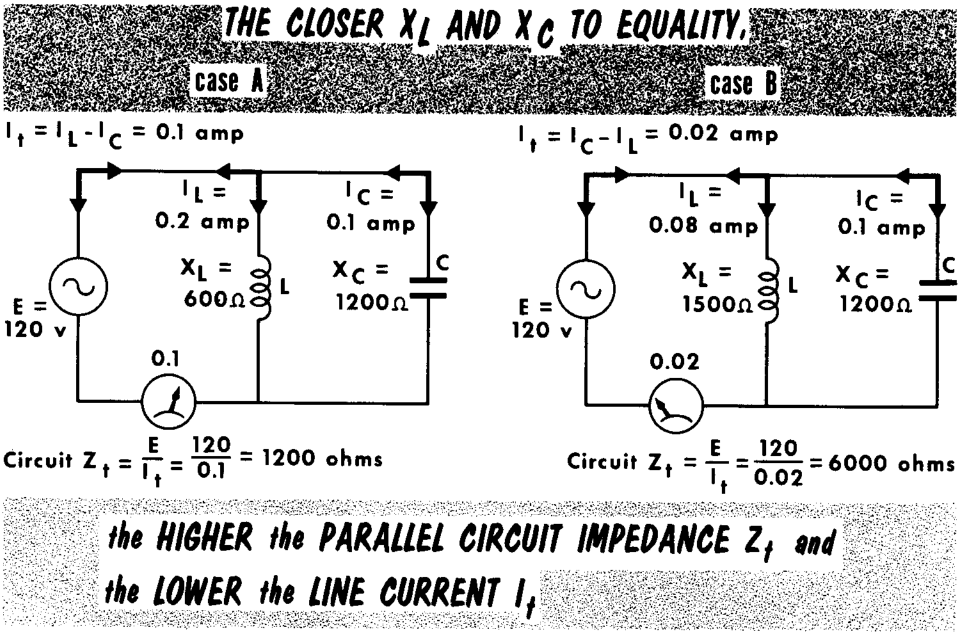
There are several interesting aspects of the parallel L-C circuit which we shall discuss, but first we must establish the total circuit impedance Z_t. Since we know that line current I_t equals 0.1 ampere and applied voltage equals 120 volts, the circuit impedance Z_t equals E/I_t. Substituting the appropriate numbers in the equation:

$$Z_t = \frac{120}{0.1} = 1200 \text{ ohms (inductive)}$$

The voltage source "looking" into the parallel circuit "sees" an impedance of 1200 ohms. The reference to <u>inductive</u> for the impedance has a meaning similar to that in the series L-C circuit; namely, the behavior of the impedance as an inductance or as a capacitance. The impedance of the parallel L-C circuit can be inductive or capacitive; this is determined by which form of opposition to the current flow is most prominent in the parallel network (assuming that X_L does not equal X_C). The predominant branch reactance is the <u>lesser</u> one, since it permits the <u>greater</u> amount of branch current to flow. Therefore, it is most prominent in the resultant line current. In this way, the lesser reactance determines the overall behavior of the circuit, as well as the phase relationship between line current and applied voltage. The line current can have two relationships relative to the applied voltage – leading or lagging. (Later on, you will learn of a third possible identity, this being the in-phase condition when the L-C circuit is <u>resonant</u> and behaves like a resistance.)

Line Current and Circuit Impedance

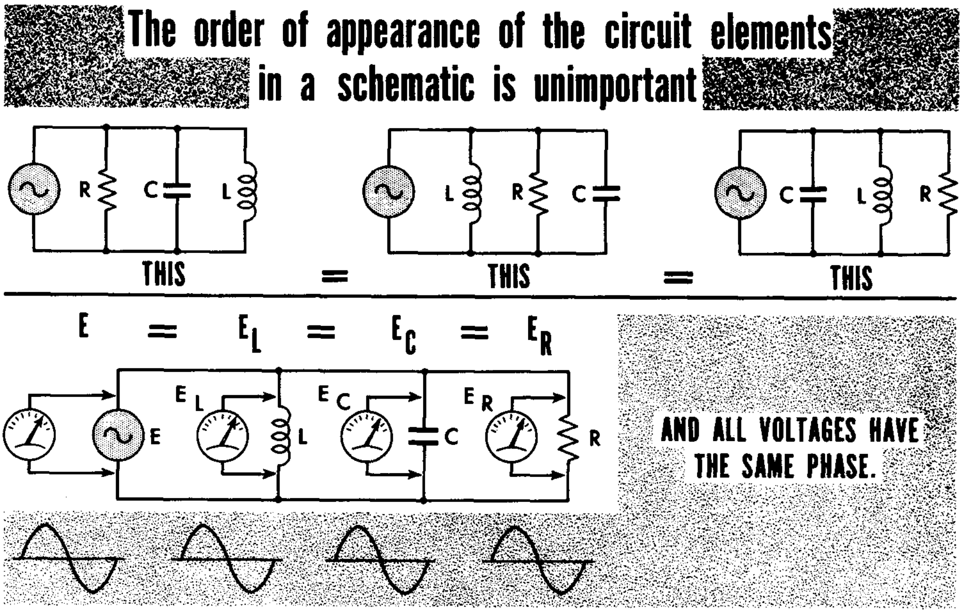
The branch and line currents have been established in circuit A and are: I_L = 0.2 ampere; I_C = 0.1 ampere; I_t = 0.1 ampere; Z_t = 1200 ohms. Line current I_t is seen to be less than one of the branch currents. This is not unusual in an a-c circuit when the parallel network consists of L and C. The currents flowing through the branches are 180° out of phase with each other; therefore, they tend to cancel in the path which carries the two currents. If the two branch currents differ greatly, as in case A, the line current I_t is less than the higher of the two branch currents; if the two branch currents do not differ by too much, the current can be less than either of the two branch currents. The example which follows illustrates this point.

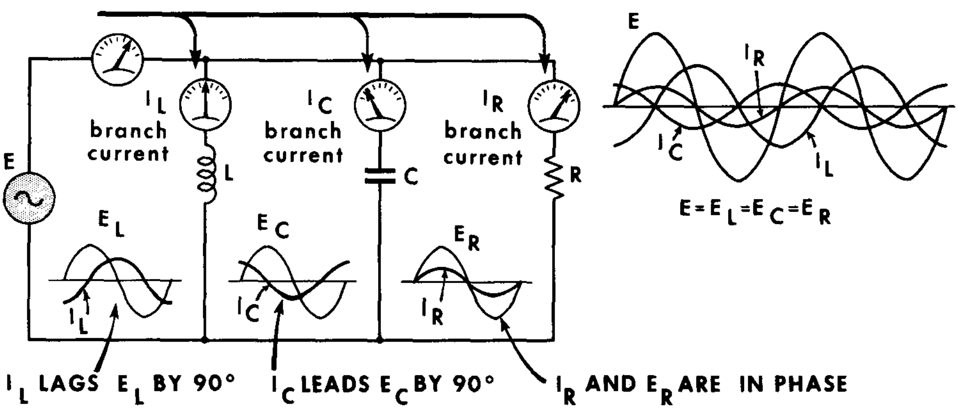


Assume a second parallel L-C circuit (B) in which E = 120 volts at 60 cycles, X_L = 1500 ohms, and X_C = 1200 ohms. Applying Ohm's law for current: $I_L = E/X_L$ = 120/1500 = 0.08 ampere and $I_C = E/X_C$ = 120/1200 = 0.1 ampere, from which $I_t = I_C - I_L$ = 0.1 - 0.08 = 0.02 ampere, an amount less than either branch current. Then the circuit impedance $Z_t = E/I_t$ = 120/0.02 = 6000 ohms. If now we study closely the constants of circuits A and B, two extremely important situations are brought to light. By comparing the two values of line current and the respective reactances in the circuits, it is seen that the closer to equality X_L and X_C are, the less the line current; the more one reactance differs from the other, the greater the line current. Since line current I_t is the denominator in the equation for the circuit impedance, the closer to equality X_L and X_C are for any given voltage, the higher the circuit impedance; the greater the difference between X_L and X_C, the lower the circuit impedance. These two electrical conditions are important to remember.

The Basic Parallel L-C-R Circuit (General)

Let us now consider an L-C circuit with parallel resistance R added. Elements L and C remain "pure" inductance and capacitance. The addition of the parallel resistance does not change the individual actions of L and C. Now $E = E_L = E_C = E_R$; therefore, all voltages are in phase. R is simply another branch across which the applied voltage appears as E_R, and through which a current I_R equal to E_R/R flows. Branch current I_R is not influenced by the presence of I_L or I_C. There is, however, a difference in the phase relationship between the voltage and current associated with R. Voltage E_R and current I_R are in phase, while voltage E_L leads current I_L by 90°, and voltage E_C lags current I_C by 90°. Thus, the resistive current is 90° out of phase with the inductive and capacitive currents, which are 180° out of phase with each other.





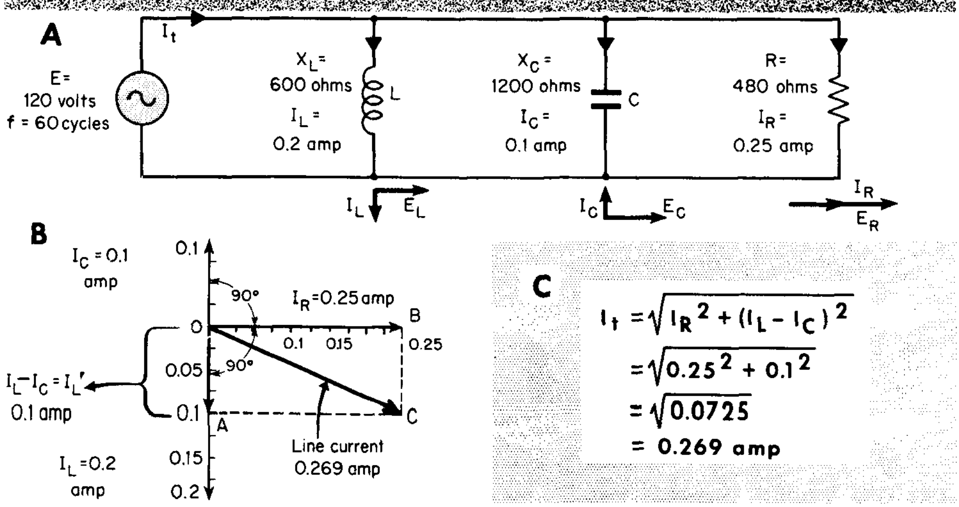
Branch Currents and Line Current

Let us say that a parallel L-C-R circuit (A) contains an inductive branch in which X_L equals 600 ohms, a capacitive branch with X_C equal to 1200 ohms, and a resistive branch with R equal to 480 ohms. The applied voltage is 120 volts at 60 cycles; hence, E_L equals E_C equals E_R equals 120 volts. $I_L = E_L/X_L = 0.2$ ampere; $I_C = E_C/X_C = 0.1$ ampere; and $I_R = E_R/R = 0.25$ ampere.

When the three branch currents are known, the line current is determined by vectorial addition (B). Because the current and voltage across the resistance are in phase, I_R is used as the reference vector. The capacitive branch current leads its voltage (E_C) by 90°; hence, it is positioned 90° ahead of the I_R vector. The inductive current lags its voltage (E_L) by 90°; hence, it is positioned 90° behind I_R. The I_C and I_L vectors are 180° apart; therefore, their resultant is the difference between them, or $I_L - I_C = 0.1$ ampere. Current I_L is greater than current I_C; therefore, the difference between them has the same direction as I_L. This difference current is 90° out of phase with I_R. The two currents can be added by the parallelogram method (B), or by the equation method (C). Completing the parallelogram and drawing the diagonal OC furnishes the answer. With all vectors similarly calibrated, the dimension of OC can be read directly as the amount of line current. The answer is 0.269 ampere. The same answer is arrived at by the equation method (C). With both inductive and resistive current present in the line current, the line current therefore lags the applied voltage by the angle θ. Using a protractor on the vector presentation shows the angle of lag to be 21.8°. Expressed mathematically, the lag of the line current is:

$$\tan\theta = \frac{I_L}{I_R} = \frac{0.1}{0.25} = 0.4 \text{ or } \theta = 21.8°$$

THE VECTOR ADDITION OF THE BRANCH CURRENTS



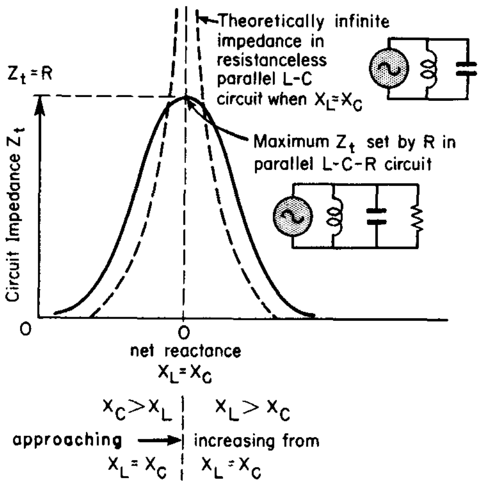
Line Current and Circuit Impedance

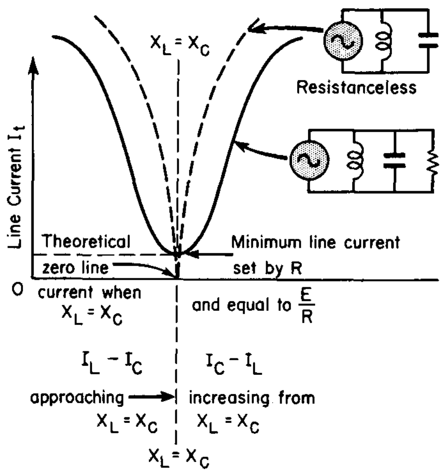
Since we know the line current (developed on the preceding page), we can determine the circuit impedance Z_t.

$$Z_t = \frac{E}{I_t} = \frac{120}{0.269} = 446 \text{ ohms}$$

Comparison of the line current with the individual branch currents shows that the line current is greater than the highest branch current. The impedance is less than the lowest ohmic value among the branches. The calculated circuit impedance Z_t = 446 ohms compared to R = 480 ohms shows this to be true. Although it is true in this instance, the circuit impedance of a parallel L-C-R circuit is not always less than the lowest ohmic value among the branches. For example, imagine I_L equal to I_C because X_L equals X_C at some frequency. Then the two reactive currents cancel each other as far as line current is concerned, and the only current appearing in the line current is the one flowing through the resistive branch. When this is true, circuit impedance Z_t equals R for all practical purposes.

THE PARALLEL R IN THE PARALLEL L-C-R CIRCUIT LIMITS THE MAXIMUM CIRCUIT IMPEDANCE AND SETS THE MINIMUM CURRENT ABOVE ZERO WHEN $X_L = X_C$





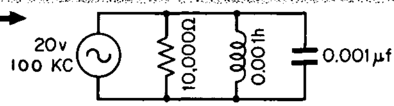
But a much more important point is that when R is in parallel with paralleled L-C, the circuit impedance can never rise higher than the ohmic value of R, even though the individual reactances X_L and X_C may be very much higher. On page 2-114, we found that when X_L and X_C approached equality, the circuit impedance of the resistanceless circuit increased greatly. With parallel R present, such an increase cannot take place. The parallel R prevents the parallel L-C circuit from presenting a very high impedance when $X_L = X_C$. This situation is important when working with parallel-resonant L-C circuits, as explained later.

Line Current and Circuit Impedance (Cont'd)

To emphasize the action of R in a parallel-connected L-C-R circuit, we examine two more sets of circuit constants.

The presence of the parallel-connected resistive branch (R = 10,000 ohms) appears to contribute very little to the operation of the circuit. When the difference between the inductive and capacitive branch currents is much greater than the resistive branch current, the R branch has very little effect on the circuit action. As the figures show with R in the circuit, the line current I_t is 0.018 ampere and is lagging the applied voltage by 83°. The circuit impedance Z_t is 1111 ohms. With R removed, the I_t would be 0.0185 ampere and lagging the applied voltage by 90° – very little difference. The circuit impedance would be 1081 ohms – again, very little difference. It would be good practice to solve for the circuit current and impedance with R out of the circuit.

R HAS LITTLE EFFECT ON THIS CIRCUIT ($I_L - I_C$ MUCH GREATER THAN I_R)



$X_L = 2\pi fL$

$= 6.28 \times 100{,}000 \times 0.001$

$= \boxed{628}$ ohms

$X_C = \dfrac{1}{2\pi fC}$

$= \dfrac{1}{6.28 \times 100{,}000 \times 0.000000001}$

$= \boxed{1594}$ ohms

$R = \boxed{10{,}000}$ ohms

$I_L = \dfrac{E_L}{X_L} = \dfrac{20}{628} = \boxed{0.031}$ ampere

$I_C = \dfrac{E_C}{X_C} = \dfrac{20}{1594} = \boxed{0.0125}$ ampere

$I_R = \dfrac{E_R}{R} = \dfrac{20}{10{,}000} = \boxed{0.002}$ ampere

$I_t = \sqrt{I_R^2 + (I_L - I_C)^2}$

$= \sqrt{.002^2 + (.031 - 0.0125)^2}$

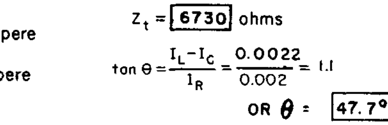
$= \boxed{.018}$ ampere

$Z_t = \dfrac{E}{I_t} = \dfrac{20}{0.018} = \boxed{1111}$ ohms

$\tan\theta = \dfrac{I_L - I_C}{I_R} = \dfrac{0.0185}{0.002} = 9.25$

OR $\theta = \boxed{83°}$

R HAS CONSIDERABLE EFFECT ON THIS CIRCUIT ($I_L - I_C$ LESS THAN I_R)



$X_L = 2\pi fL = \boxed{942}$ ohms

$X_C = \dfrac{1}{2\pi fC} = \boxed{1063}$ ohms

$R = \boxed{10{,}000}$ ohms

$I_L = \boxed{0.020}$ ampere

$I_C = \boxed{0.0188}$ ampere

$I_R = \boxed{0.002}$ ampere

$I_t = \sqrt{I_R^2 + (I_L - I_C)^2}$

$= \boxed{0.00297}$ ampere

$Z_t = \boxed{6730}$ ohms

$\tan\theta = \dfrac{I_L - I_C}{I_R} = \dfrac{0.0022}{0.002} = 1.1$

OR $\theta = \boxed{47.7°}$

Let us now change the operating conditions by raising the frequency of the applied voltage to 150 kc. It changes the values of X_L and X_C, making the difference between I_L and I_C nearly I_R. X_L almost equals X_C. Hence, the presence of the R branch displays a major effect. In the absence of R, the circuit impedance Z_t would be 9090 ohms, and the line current would lag the applied voltage by 90°. But with R present, Z_t = 6730 ohms and the line current lags the applied voltage by only 47.7°, a major change. It is evident, therefore, that the parallel R displays its greatest effect in the vicinity of, and at, resonance.

Calculating the Resonant Frequency

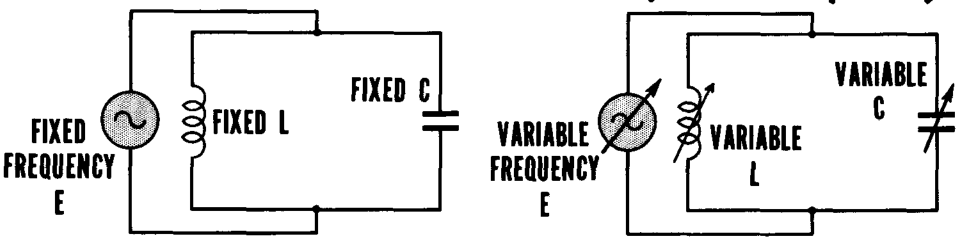
The parallel resonant L-C circuit differs from the ordinary parallel L-C circuit in one respect – resonance. Resonance occurs when the inductive reactance ($2\pi fL$) equals the capacitive reactance ($1/2\pi fC$), or $X_L = X_C$. For any given fixed amount of L and C, parallel resonance occurs at only one frequency (the same as in the series-resonant L-C circuit). The frequency of resonance is expressed by $f = 1/2\pi\sqrt{LC}$, or $1,000,000/2\pi\sqrt{LC}$. Both equations are exactly the same as used for the series-resonant circuit and both yield the same result. They differ only in the units which are used for L and C – the first is in henries and farads; the second is in microhenries and microfarads. To illustrate the application of the two equations, assume a parallel L-C circuit in which L = 100 microhenries (or 0.0001 henry), and C = 100 micromicrofarads (or 0.0001 microfarad, or 0.0000000001 farad). Using each equation for the solution, the resonant frequency (often indicated f_o) is:

$$
\begin{aligned}
f \text{ cycles} &= \frac{1}{6.28 \times \sqrt{0.0001 \times 0.0000000001}} \\
&= \frac{1}{6.28\sqrt{0.00000000000001}} \\
&= \frac{1}{6.28 \times 0.0000001} \\
&= \frac{1}{0.000000628} \\
&= 1,592,197 \text{ cycles} \\
&= 1.5922 \text{ mc}
\end{aligned}
$$

or

$$
\begin{aligned}
f \text{ cycles} &= \frac{1,000,000}{6.28\sqrt{100 \times 0.0001}} \\
&= \frac{1,000,000}{6.28\sqrt{0.01}} \\
&= \frac{1,000,000}{6.28 \times 0.1} \\
&= \frac{1,000,000}{0.628} \\
&= 1,592,197 \text{ cycles} \\
&= 1.5922 \text{ mc}
\end{aligned}
$$

Fixed L and C Can Resonate at Only One Frequency



Parallel L or C Can Be Tuned to Resonate at a Variety of Frequencies

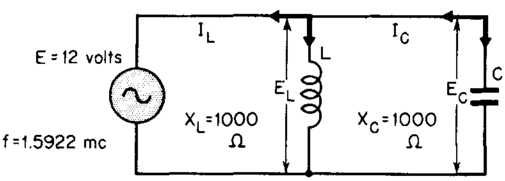
The values of L and C in a parallel-resonant circuit may be precalculated so as to be resonant to a single frequency. On the other hand, the requirement may be for resonance over a band of frequencies. This condition is provided for by making either L or C (sometimes both) variable between a minimum and a maximum amount. Each setting within the range of variation then affords a different resonant frequency. For instance, if in the example above, L were fixed at 100 μh while C were variable between a minimum of 20 μμf and a maximum of 100 μμf, the circuit could be resonated individually to all frequencies between a low of 1.5922 mc to a high of 3.56 mc.

Line Current and Impedance

You have learned that the series resonant L-C circuit offers minimum impedance at the resonant frequency. The behavior of the parallel resonant L-C circuit is the exact opposite – at resonance, the circuit impedance is maximum. When equality between X_L and X_C is reached, the respective branch currents I_L and I_C are equal. Since the line current I_t has been established as the difference between the branch currents, and the difference between two equal amounts is zero, the parallel resonant L-C circuit shows no line current. From the viewpoint of the voltage source, it is applying voltage across a circuit having infinite impedance.

Let us illustrate this situation. E = 12 volts at 1.5922 mc; L = 100 microhenries, for which X_L is 1000 ohms; C = 0.0001 microfarad, for which X_C = 1000 ohms. Then:

PARALLEL RESONANCE PRODUCES MINIMUM LINE CURRENT and MAXIMUM IMPEDANCE



When $X_L = X_C$

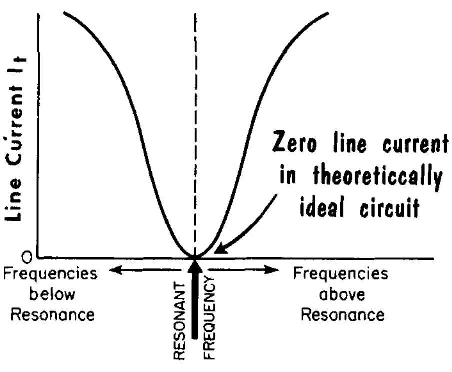
$1000 = 1000$

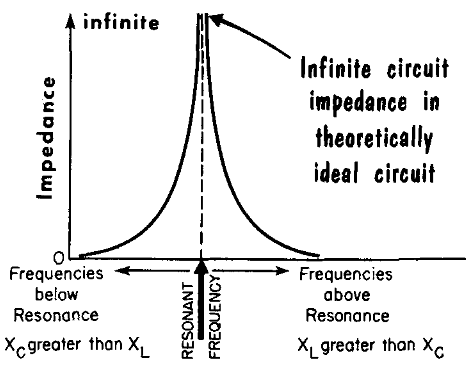
and $I_L = \frac{E_L}{X_L} = \frac{12}{1000} = .012$ amp

$I_C = \frac{E_C}{X_C} = \frac{12}{1000} = .012$ amp

$I_t = I_L - I_C$
$= 0.012 - 0.012$
$= 0$

$Z_t = \frac{E}{I_t}$
$= \frac{12}{0}$
= INFINITE



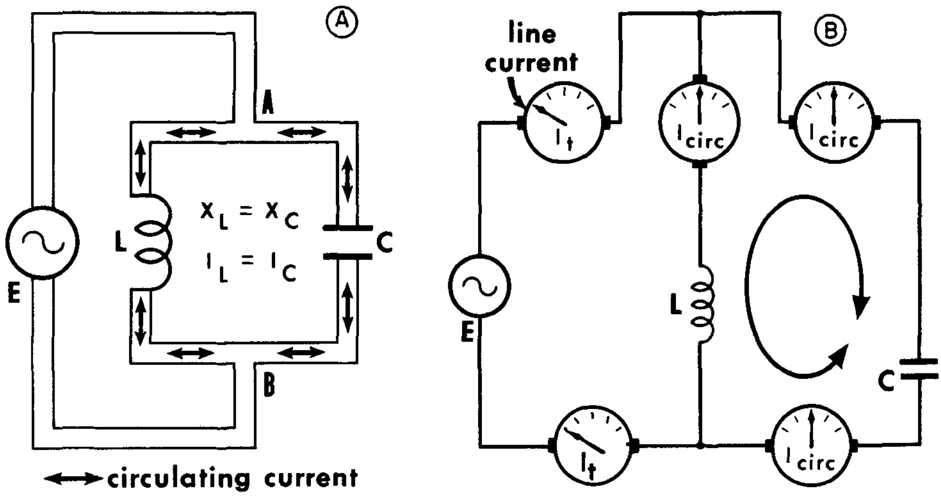


The ideal case of zero line current and infinite circuit impedance in the resonant parallel L-C circuit is not realizable in practice; nevertheless, we assume that they are attainable because they enable us to establish clearly the limiting conditions for later comparison with the practical circuit. There is similarity in behavior between the two. If we show a graphic picture of the change in line current with change in frequency for the ideal case, the line current is seen to be zero at resonance, and increases for frequencies on both sides of resonance. The graph of circuit impedance vs. frequency shows infinite impedance at resonance and reduced impedance off resonance. With equality between the branch reactances and branch currents, infinite impedance is interpreted as infinite resistance.

Circulating Current

The infinite impedance of the resistanceless parallel resonant L-C circuit should be understood as being the impedance "seen" by the voltage source as it "looks" into the parallel L-C network. The voltage source "sees" an open circuit. Also, the zero line current condition should not be mistaken for zero current conditions inside the parallel network. Interestingly enough, a significant current flow situation prevails inside the parallel L-C circuit. Let us examine the current conditions inside the parallel L-C network at resonance.

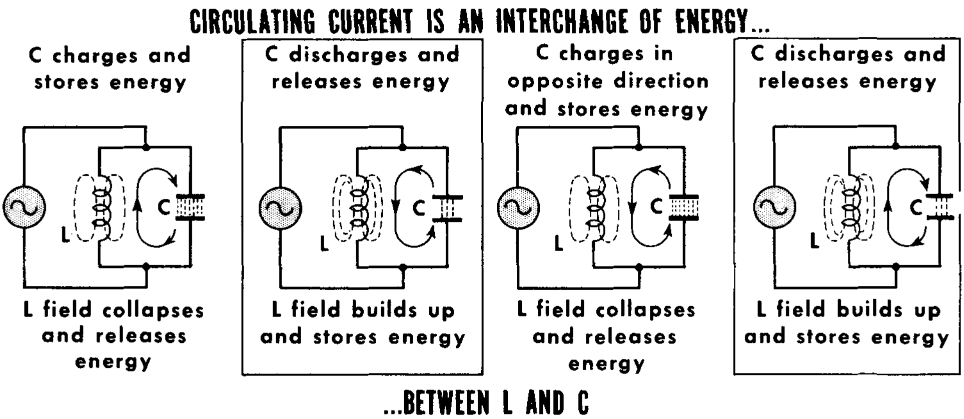
The circuit is resistanceless, $X_L = X_C$, and $I_L = I_C$. The two branch currents are 180° out of phase as the result of their 90° lag and lead relationships with their respective voltages, E_L and E_C. Examination of the flow of the two branch currents shows that they move in opposite directions through their respective branch elements. When the polarity of the applied voltage changes, the two branch currents reverse their directions of flow. Now, if we take points A and B as references, and examine the directions of the two branch currents, they are seen to have like directions. All the current which flows into A moves away from A; all the current which flows into B moves away from B. In other words, the two branch currents have become one and the same current, as far as to-and-fro circulation through L and C inside the parallel-connected circuit is concerned. In fact, as far as the circulating current is concerned, the parallel L-C network is really a series circuit, since there is only one path for the circulating current. The circulating current is equal to E_L/X_L or E_C/X_C.

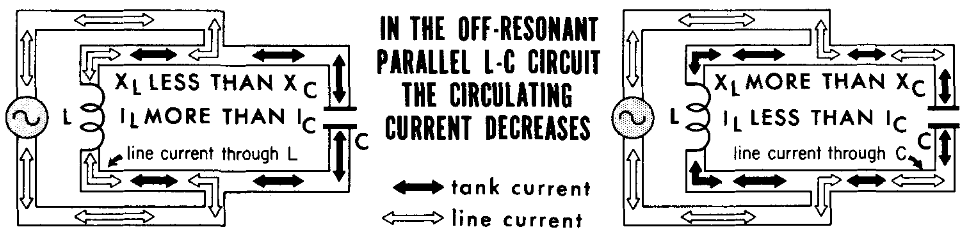


The CIRCULATING CURRENT in the PARALLEL RESONANT CIRCUIT is EVERYWHERE THE SAME in the L CIRCUIT and in the C CIRCUIT

Circulating Current (Cont'd)

What is the importance of the circulating current? The parallel resonant L-C circuit is sometimes called a "tank" or storage circuit. The circulating current charges C, momentarily storing energy in it. When C discharges, the discharge current flows into L and builds a magnetic field, in which electrical energy is stored. When the magnetic field collapses, the current again flows into C, recharging it. Discharging again builds a magnetic field around the coil, thus effecting an interchange of electrical energy between C and L of the parallel resonant L-C circuit. This energy is maximum at resonance; hence, maximum energy is available for transfer to another circuit or to be kept within the circuit for a purpose.





What happens when the circuit is not resonant? Several actions occur. Assume X_L to be greater than X_C. Then I_C is greater than I_L. Suppose that I_C = 200 ma and I_L = 50 ma. We have learned that the line current I_t equals the difference between the two, or in this case 200 - 50 = 150 ma. Circulating current also flows in the nonresonant state of the circuit, but now it is equal to the lesser of the two branch currents – in this case, to I_L = 50 ma. An equal amount of the greater branch current (50 ma) becomes part of the circulating current. The remaining 150 ma of the capacitive branch current is the line current, and flows through the parallel L-C circuit via the capacitive branch. If the situation were reversed and I_L equalled 200 ma and I_C equalled 50 ma, the circulating current would be 50 ma while the line current of 150 ma would flow through the circuit via the inductance. The farther away from resonance, the less the circulating current and the greater the line current.

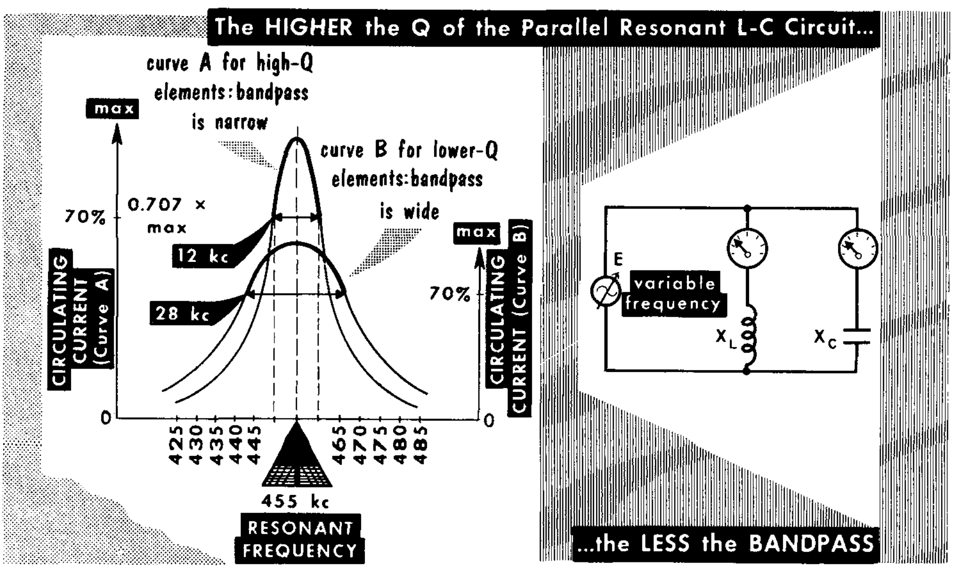
Resonance Curve

Like the series resonant L-C circuit, the performance of the parallel resonant L-C circuit can be portrayed by a resonance curve. A convenient way of preparing this curve is by plotting the circulating current vs. frequency. A current meter is used in each branch circuit.

Current at the resonant frequency is determined first. Both meters will indicate the same amount of current. As the frequency of the applied voltage is lowered, X_L decreases while X_C increases. The inductive branch current thus increases, whereas the capacitive branch current decreases. Since the circulating current has the value of the lesser branch current, the X_C branch meter is used as the current indicator. The lower the frequency relative to resonance, the lower will be the indication on the X_C branch meter. Since the inductive branch current exceeds the capacitive branch current, the parallel current as a whole behaves like an inductance.

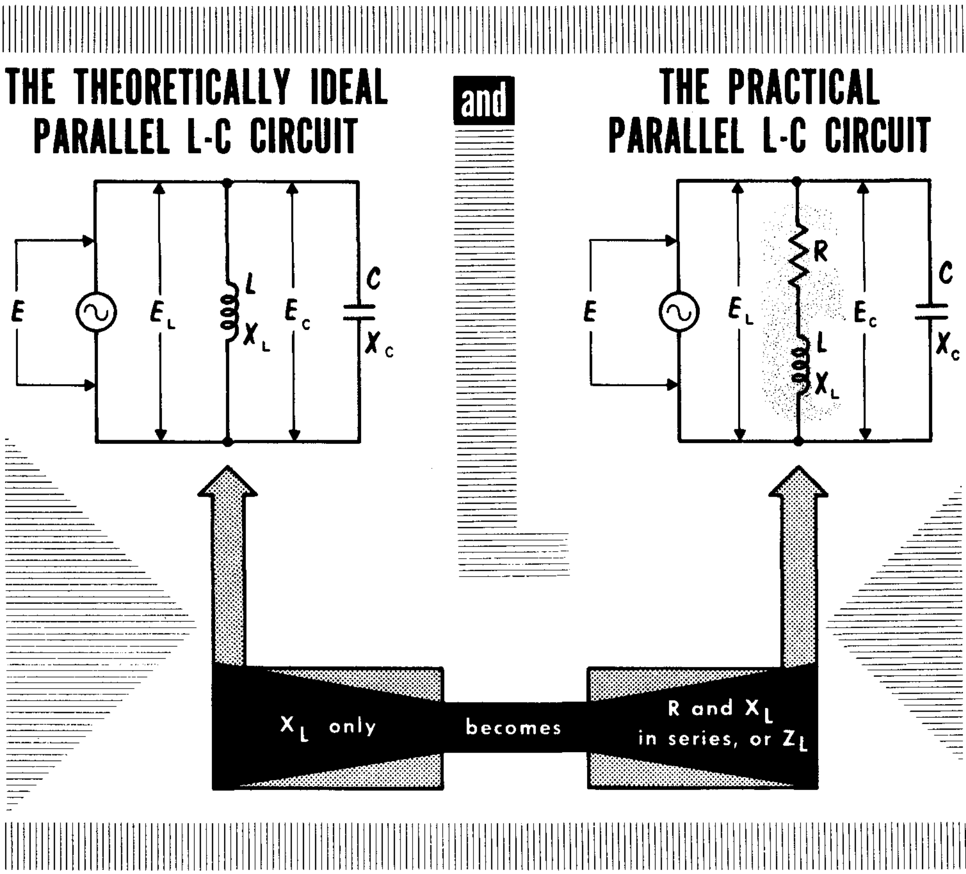
Then, the circulating current is plotted as the frequency is increased above resonance. As frequency increases, X_C decreases and X_L increases. The capacitive branch current now increases, whereas the inductive branch current decreases. The indication on the meter in the X_L branch is plotted for a range of frequencies above resonance. Since the capacitive branch current exceeds the inductive branch current, the parallel circuit behaves like a capacitance.

As in the series resonant circuit, the parallel resonant circuit also affords a selective frequency bandpass. It is the band of frequencies embraced by this curve at a level corresponding to a circulating current of 70% (70.7% exactly) of the maximum circulating current. The higher the circuit Q, the steeper the sides and the narrower the bandpass. The lower the Q, the broader the sides and the wider the bandpass.



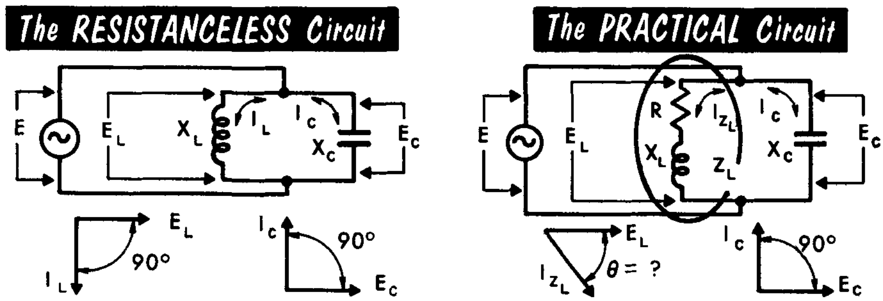
Comparison with Theoretically Ideal Parallel L-C Circuit

The practical parallel L-C circuit differs from the theoretically resistanceless version in one major respect: the presence of resistance. It exists in the inductance, in the capacitance, and in the interconnecting wires. Of these resistance sources, only the resistance contained in the inductance is important, so we disregard the others.



The resistance is treated as if it were a resistance in series with the coil L. The capacitive branch is assumed to consist of X_C only, whereas the inductive branch consists of R in series with X_L. In low-frequency circuits, L could have a great many turns; hence, R could be relatively high. In high-frequency circuits, L usually is small and therefore R is small; in fact, R is often so small that it is completely negligible. Yet we must recognize that R is present, especially (as you will see) when $X_L = X_C$ during resonant conditions. In the theoretically resistanceless circuit, when $X_L = X_C$, the line current is zero and the circuit impedance is infinite. In practical systems using parallel resonant L-C circuits, X_L can equal X_C, yet there is a finite amount of line current flowing due to the resistance in the circuit.

Branch Currents



A

In the theoretically resistanceless parallel resonant circuit, the inductive branch current I_L lags the inductive branch voltage by 90°, the capacitive branch current I_C leads its voltage E_C by 90°, and the line current I_t is zero. In the practical version of the circuit (with R in series with L), the capacitive branch current still leads its voltage E_C by 90°, but the inductive branch no longer is only reactance X_L. Now the branch is impedance Z_L made up of R and X_L in series.

In the study of the behavior of the series R-L circuit, we found that while the current is everywhere the same, the voltage across R is in phase with the current, but leads the current in X_L by 90°. When R is not negligible, the current in the series circuit lags the voltage across the series combination by some amount less than 90°. Applying these conditions to the practical parallel resonant L-C circuit shows that the two branch currents are not 180° out of phase. Resonance occurs, nevertheless because $X_L = X_C$. To determine the inductive branch current, we must first establish the impedance of the circuit. Assume that E = 12 volts, X_L = 775 ohms, R = 300 ohms (a completely unrealistic amount, but one which will illustrate the point), and X_C = 775 ohms. Then:

Inductive branch Z_L

$$Z_L = \sqrt{R^2 + X_L^2} = \sqrt{300^2 + 775^2} = \sqrt{90{,}000 + 600{,}625} = \sqrt{690{,}625} = \boxed{831 \text{ OHMS}}$$

Inductive branch current I_{Z_L}

$$I_{Z_L} = \frac{E_L}{Z_L} = \frac{12}{831} = 0.0144 \text{ ampere} = \boxed{14.4 \text{ MA}}$$

Capacitive branch current I_C

$$I_C = \frac{E_C}{X_C} = \frac{12}{775} = 0.0154 \text{ ampere} = \boxed{15.4 \text{ MA}}$$

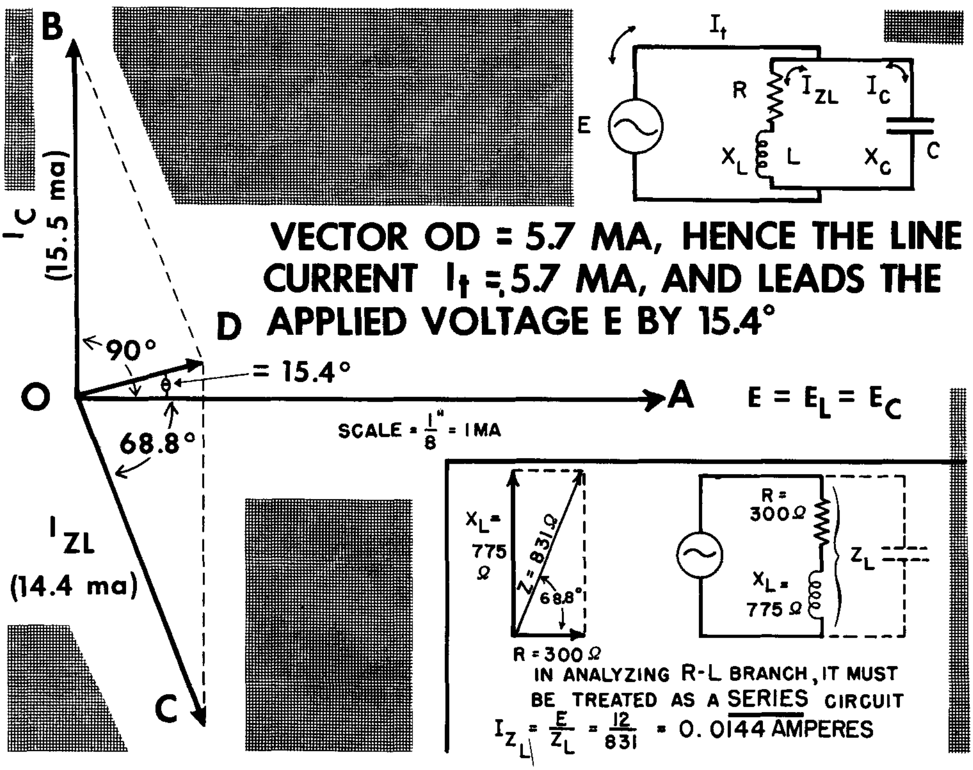
B

$$\tan\theta = \frac{X_L}{R} = \frac{775}{300} = 2.58$$ which from trigonometric table = **68.8° CURRENT LAGGING**

We now know that Z = 831 ohms; I_{ZL} = 0.0144 ampere (14.4 ma) while I_C = 0.0154 ampere (15.4 ma). The phase angle between the inductive branch current I_{ZL} and its voltage E_L is 68.8°.

Line Current

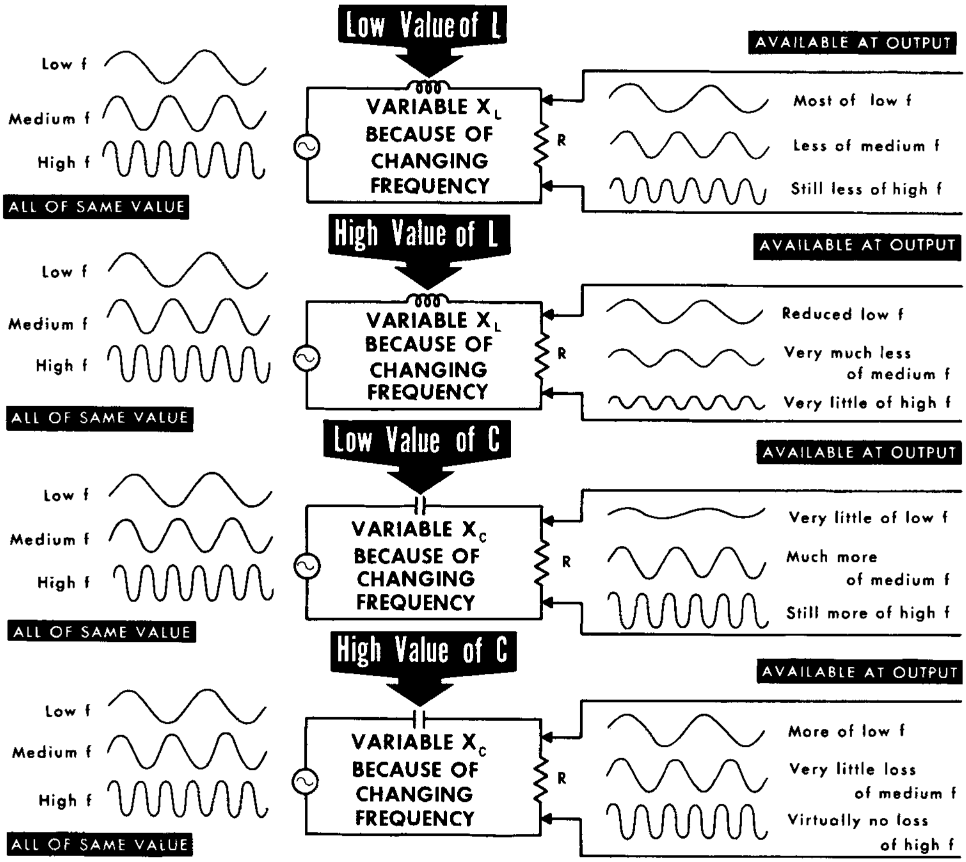
To determine the line current I_t in the practical parallel resonant circuit, we must resort to trigonometric means and use vectors. If we show the answers derived on the preceding page in a vector presentation, the voltages $E = E_L = E_C$ form the reference vector OA. Its length is arbitrary. Then we position the capacitive branch current vector; OB = 15.5 ma, 90° ahead of the reference voltage vector. The length of the capacitive current vector conforms with whatever scale is decided upon. In this case, it is 1/8 inch = 1 ma. The same scale is used for all currents, thereby permitting direct reading of current values from the dimensions of the lines. Using a protractor, we locate the inductive branch current I_{ZL} vector; OC = 14.4 ma, at 68.8° behind the voltage vector. This completes one portion of the branch. Because of the difference in the phase relationships between I_C and E_C, and I_{ZL} and E_L, we cannot subtract I_{ZL} from I_C and assume that the arithmetical difference is equal to the line current. To establish the line current, we must lay down two additional sides – side CD equal and parallel to vector OB, and side BD equal and parallel to vector OC. Now, the diagonal OD represents the line current. By measuring its length, vector OD equals 5.7 ma.



Vector OD is located above the reference voltage vector OA; therefore, I_t leads applied voltage E by angle AOD. Using a protractor, we find it to be 15.4°.

Effects of L and C at Different Frequencies

To aid in understanding the action of the filter circuits, it is best to review some of the effects of C and L at different frequencies. The diagrams show the effect of series-connected small and large values of inductance and capacitance on low-, medium-, and high-frequency voltages.



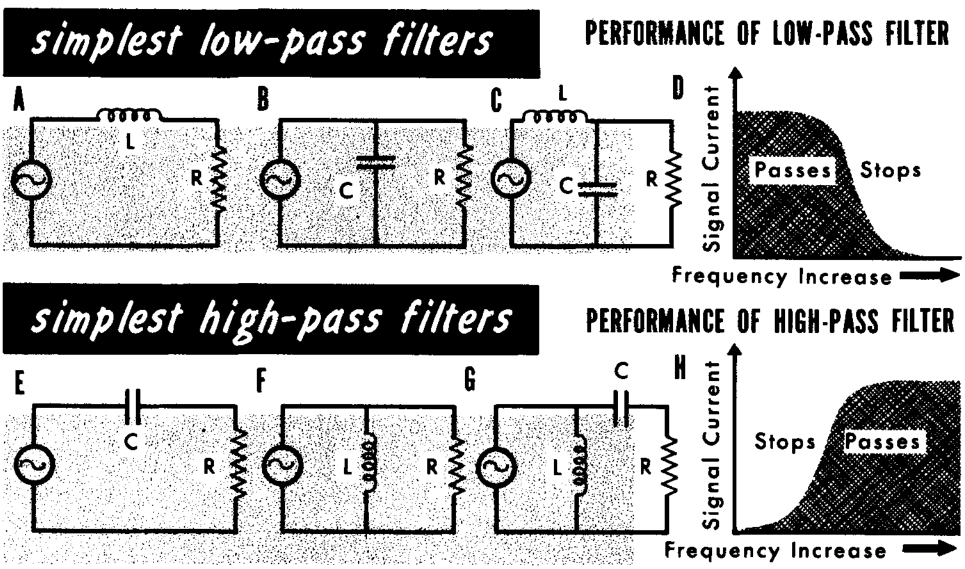
The higher the frequency of the applied voltage for any given value of L, the larger will be the voltage drop across L; hence, the less the signal voltage available at the output. The higher the frequency for any given value of C, the less will be the voltage drop across C and the higher the available signal voltage at the output.

L and C are used in various ways in different kinds of filters. Sometimes they form resonant circuits; sometimes they form networks which will pass a wide band of frequencies and reject others, or reject a wide band of frequencies and pass others. Examples appear on the following pages, with more detailed discussions in subsequent volumes dealing with receivers and transmitters.

Low-Pass and High-Pass Filters

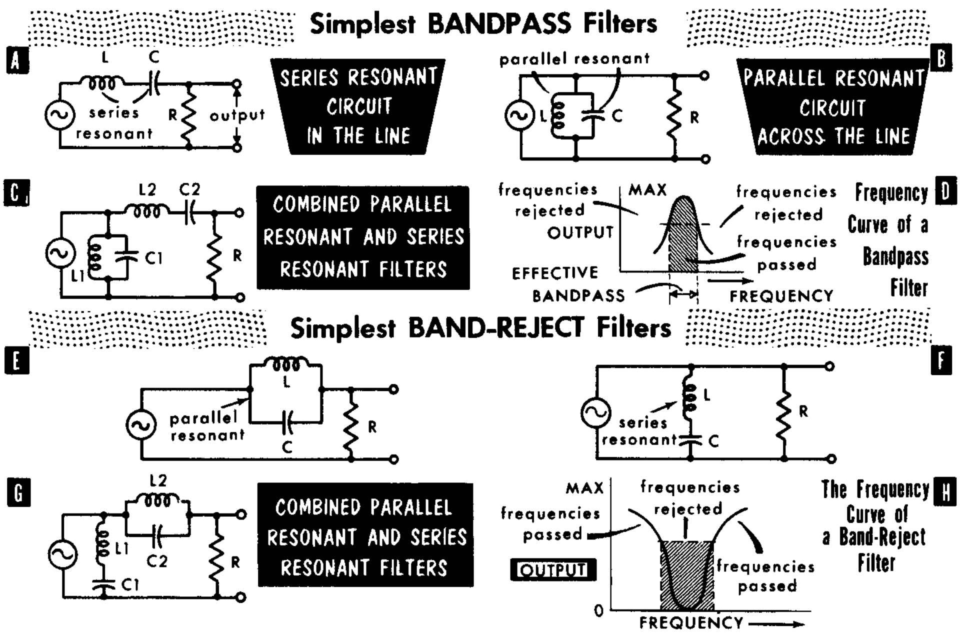
A filter circuit that passes all the low frequencies below a particular frequency, and rejects all higher frequencies, is called a low-pass filter. A filter circuit that passes all the high frequencies above a particular frequency and rejects the lower frequencies is called a high-pass filter.

The simplest form of low-pass filter is an inductor in series with the line as in A, or a capacitor connected in parallel with the line as in B. The inductor presents a low X_L at the low frequencies but a high X_L at the high frequencies. The capacitor presents a high X_C at the low frequencies and progressively less X_C as the frequency increases. When L and C are combined, they form a low-pass filter as shown in C, with their performance curve shown in D. The low reactance of L at low frequencies provides an easy path for the signal. At the same time, the shunt capacitance presents a high impedance to the low-frequency signal currents; therefore, very little is lost across C. At high frequencies, the high reactance of L presents increasing opposition to the flow of signal currents. At the same time, the progressively decreasing reactance of C at high frequencies offers an easy bypass path for the currents.



The simplest form of high-pass filter would be either C in series with the line (as in E) or an inductor across the line (as in F). When combined (G), the high reactance of C at low frequencies offers high opposition to their path. The low reactance of L at low frequencies will effectively bypass them. At high frequencies, the reactance of C is low, and it readily passes these signal currents. On the other hand, the increasing reactance with increasing frequency of L minimizes signal-current bypass through the coil. The net result is to pass all the higher frequencies readily to the load, but to reject the lower frequencies (H).

Bandpass and Band-Reject Filters



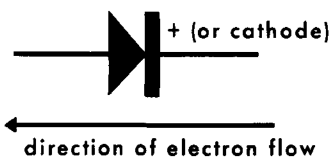
Bandpass filters permit passage of a narrow band of frequencies while rejecting all other undesired frequencies. The simplest form of bandpass filter would be a series resonant circuit in the line as shown in A. Another simple bandpass filter using a parallel resonant circuit across the line is shown in B. Combining the two as in C provides a more effective bandpass filter. The series resonant circuit offers low impedance and readily passes the desired frequencies, while offering high impedance and blocking undesired frequencies. The parallel resonant circuit offers high impedance to the desired band of frequencies, preventing any bypass action; the undesired frequencies find the parallel resonant circuit a low-impedance path and are effectively bypassed through it. The characteristic curve of a bandpass filter is shown in D.

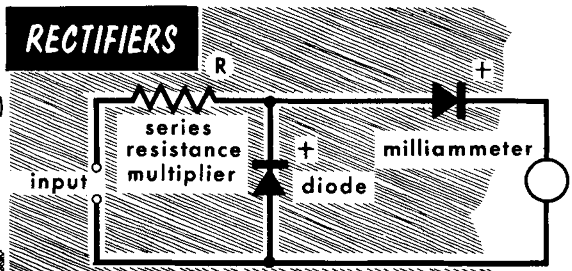
Band-reject filters are used to block the passage of a narrow band of frequencies while passing all other frequencies. The simplest form of band-reject filter would be a parallel resonant circuit in the line as shown in E. A simple band-reject filter using a series resonant circuit across the line is shown in F. Combining the two as in G provides an effective band-reject filter. The parallel resonant circuit offers high impedance to the desired band of frequencies to be rejected, while offering a low-impedance path to all other frequencies. The series resonant circuit across the line offers a low impedance bypass path to the band of frequencies to be rejected, while offering a high-impedance path to all other frequencies. The characteristics curve of a band-reject filter is shown in H.

Characteristics and Operation

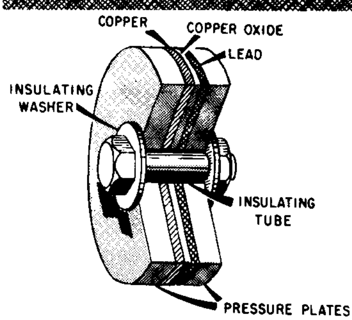
The D'Arsonval moving-coil meter used for measuring voltages and currents in Volume I is basically a d-c meter. If we apply a-c to it, one half-wave tries to make it read in the normal way; the other half-wave tries to make it read backward. As the meter pointer does not have time to move back and forth so rapidly, it either stands still or vibrates rapidly around zero. However, the D'Arsonval movement can be used to measure a-c if we first change the a-c to d-c. This can be done through the use of rectifiers or diodes. These are electrical devices that have a special characteristic — they permit current to flow through them in one direction (low resistance), but not in the other (high resistance). While many kinds of rectifiers can be used, a-c ammeters and voltmeters most often use the copper-oxide rectifier. Adding a rectifier circuit to the D'Arsonval movement gives us an a-c meter.

Symbol for Rectifier

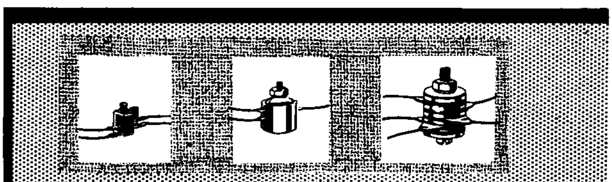




Half-Wave — the Simplest Rectifier Meter



Copper-Oxide Type Dry-Plate Rectifier



Typical Copper-Oxide Instrument Rectifiers

Copper-oxide rectifiers generally provide good rectification for a-c up to about 20,000 cycles. They are constructed of a series of copper discs clamped together flat in a stack. On one side of each disc is a coating of copper oxide which forms a layer of this material between adjacent discs. Washers made of lead are clamped against each oxide surface to improve the contact, and the complete assembly is held together by an insulated bolt running through the center. Two types of rectifier circuits are used in meters to a great extent – the half-wave and full-wave types.

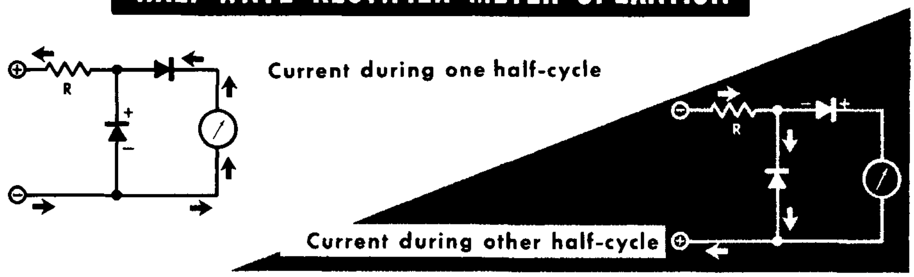
Characteristics and Operation (Cont'd)

In the half-wave type (which uses two rectifiers), one half-wave is bypassed through one rectifier, while the other half goes through the meter. The meter pointer will not have time to follow the fluctuations, so it will average out the current that flows through it.

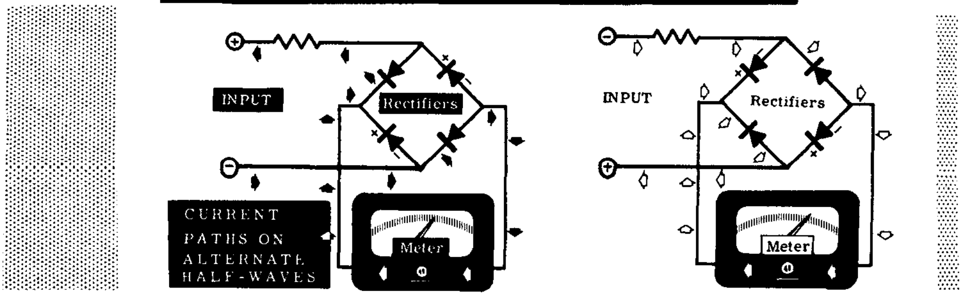
During one half-wave, no current flows through the meter, while during the other half-wave, it follows half a sine wave in form. The average of half a sine wave is 0.637 of peak value. However, during half the time, no current goes through the meter; therefore, the average over the whole time will be half of 0.637 or 0.3185. (If an ordinary d-c meter movement is used in this circuit to measure an alternating voltage, a 1-volt peak – 2 volts peak to peak – voltage will give a reading of only 0.3185 volt.) In the full-wave bridge-type rectifier, the meter gets both halves of the wave, and it will read 0.637 of the peak voltage if a regular d-c meter is used. To provide practical a-c voltage scales, multiplier resistors must be used in series with the meter movement and the rectifier, with appropriate shunts connected across the meter for current measurements.

The electrodynamometer-type wattmeter discussed in Volume I is equally useful for the measurement of a-c power. Making use of the voltage across a circuit as well as the current flow through it, this type wattmeter is ideal for measuring the actual or true power in an a-c circuit. Since the torque on the moving coil is proportional to the applied power, the "cosine θ," or phase angle between the current and voltage, is automatically taken into consideration.

HALF-WAVE RECTIFIER METER OPERATION



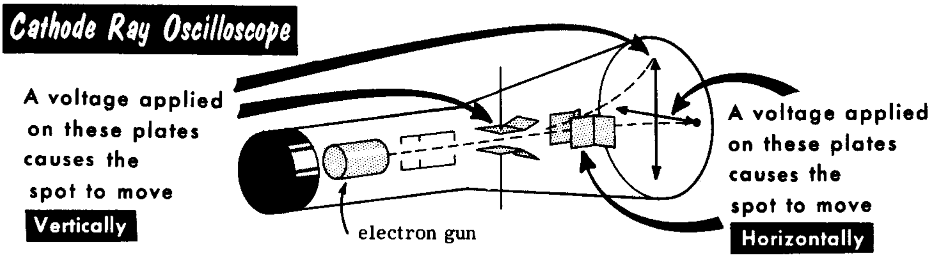
FULL-WAVE RECTIFIER METER OPERATION

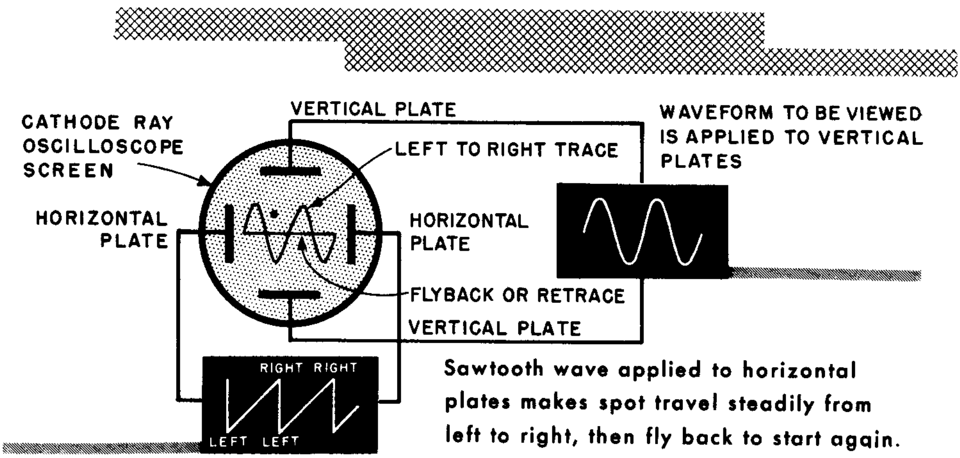


Measuring Waveforms

Waveforms can actually be examined by means of an instrument called the "oscilloscope." It uses a special tube in which a beam, or "pencil" of electrons, is focused to a point on a fluorescent screen that glows with the impact of the electrons. Two pairs of deflecting plates bend the beam in accordance with the voltages applied to them.

Applying different voltages to the plates at the sides of the beam will move the spot formed on the fluorescent screen sideways, and voltages applied to the upper and lower plates will deflect the beam up or down. If different fluctuating voltages are applied to both pairs of plates, the spot will trace a pattern on the screen representing the combined effect of the two voltage fluctuations.





If the fluctuation applied to the horizontal plates follows a "sawtooth" waveform, the spot will move steadily from left to right across the screen, and then rapidly return to its starting point. By using this waveform as a "time-base," in this way, the voltage applied to the vertical plates traces its own waveform. In this way, we can view electrical waveshapes and, by use of a scale with the scope, be able to measure actual values of voltages in circuits under examination.

In a parallel resonant circuit, the branch currents cancel, the line current is a minimum, impedance is a maximum, and the phase angle is 0°.

In the parallel L-C circuit, at frequencies above resonance, X_L is greater than X_C, I_C is greater than I_L, and the circuit acts capacitively; at frequencies below resonance, X_C is greater than X_L, I_L is greater than I_C, and the circuit acts inductively.

The Q of a parallel resonant circuit is the ratio of the current in the tank (I_L or I_C) to the current in the line.

A parallel circuit is resonant if $X_L = X_C$ and $I_L = I_C$.

The parallel resonant circuit may be used as a bandpass or band-rejection circuit.

In a parallel L-C circuit, the branch currents are 180° out of phase and can be subtracted directly: $I_t = I_L - I_C$.

The current in either branch of a parallel L-C circuit may be greater than the line current.

A tank circuit can be used to simulate the properties of either a capacitor or an inductor.

In a parallel L-C circuit, the closer X_L and X_C are to equality, the higher the circuit impedance Z_t and the lower the line current I_t.

In the parallel resonant circuit, the circulating current is the same everywhere in the L and in the C circuit.

Circulating current in a parallel resonant L-C circuit is an interchange of energy between the inductance and the capacitance.

The higher the Q of a parallel resonant L-C circuit, the narrower the bandpass.

Bandpass filters are used to permit the passage of a narrow band of frequencies while rejecting all other undesired frequencies; band-reject filters are used to block the passage of a narrow band of frequencies while passing all other frequencies.

A-C meters use the same D'Arsonval movement as is used in d-c meters, except that a rectifier circuit is added to convert a-c to d-c.

The most commonly used rectifier in a-c ammeters and voltmeters is the copper-oxide rectifier, which provides rectification up to about 20 kc.

REVIEW QUESTIONS

1. What is a parallel resonant circuit?
2. Give the conditions present in a parallel resonant circuit.
3. Describe the circulating current in a parallel resonant L-C circuit.
4. What is the formula for the impedance of a parallel resonant circuit?
5. What is a tank circuit and why is it so named?
6. How does a parallel L-C circuit act at frequencies above and below the resonant frequency? Why?
7. In a parallel resonant circuit, what relationship exists among the branch currents, the line current, and the impedance?
8. What is the nature of the impedance of the tank circuit at resonance?
9. In a parallel L-C circuit, what is the phase relationship between the branch currents?
10. Define bandpass and band-rejection filter circuits.
11. Basically, how do a-c meters differ from d-c meters?
12. What two types of rectifier circuits are principally used in meters?

NATURAL TRIGONOMETRIC FUNCTIONS

Angle	*Sine*	*Cosine*	*Tangent*	*Angle*	*Sine*	*Cosine*	*Tangent*
0°	0.000	1.000	0.000				
1°	.018	1.000	.018	46°	.719	.695	1.036
2°	.035	0.999	.035	47°	.731	.682	1.072
3°	.052	.999	.052	48°	.743	.669	1.111
4°	.070	.998	.070	49°	.755	.656	1.150
5°	.087	.996	.088	50°	.766	.643	1.192
6°	.105	.995	.105	51°	.777	.629	1.235
7°	.122	.993	.123	52°	.788	.616	1.280
8°	.139	.990	.141	53°	.799	.602	1.327
9°	.156	.988	.158	54°	.809	.588	1.376
10°	.174	.985	.176	55°	.819	.574	1.428
11°	.191	.982	.194	56°	.829	.559	1.483
12°	.208	.978	.213	57°	.839	.545	1.540
13°	.225	.974	.231	58°	.848	.530	1.600
14°	.242	.970	.249	59°	.857	.515	1.664
15°	.259	.966	.268	60°	.866	.500	1.732
16°	.276	.961	.287	61°	.875	.485	1.804
17°	.292	.956	.306	62°	.883	.470	1.881
18°	.309	.951	.325	63°	.891	.454	1.963
19°	.326	.946	.344	64°	.899	.438	2.050
20°	.342	.940	.364	65°	.906	.423	2.145
21°	.358	.934	.384	66°	.914	.407	2.246
22°	.375	.927	.404	67°	.921	.391	2.356
23°	.391	.921	.425	68°	.927	.375	2.475
24°	.407	.914	.445	69°	.934	.358	2.605
25°	.423	.906	.466	70°	.940	.342	2.747
26°	.438	.899	.488	71°	.946	.326	2.904
27°	.454	.891	.510	72°	.951	.309	3.078
28°	.470	.883	.532	73°	.956	.292	3.271
29°	.485	.875	.554	74°	.961	.276	3.487
30°	.500	.866	.577	75°	.966	.259	3.732
31°	.515	.857	.601	76°	.970	.242	4.011
32°	.530	.848	.625	77°	.974	.225	4.331
33°	.545	.839	.649	78°	.978	.208	4.705
34°	.559	.829	.675	79°	.982	.191	5.145
35°	.574	.819	.700	80°	.985	.174	5.671
36°	.588	.809	.727	81°	.988	.156	6.314
37°	.602	.799	.754	82°	.990	.139	7.115
38°	.616	.788	.781	83°	.993	.122	8.144
39°	.629	.777	.810	84°	.995	.105	9.514
40°	.643	.766	.839	85°	.996	.087	11.43
41°	.656	.755	.869	86°	.998	.070	14.30
42°	.669	.743	.900	87°	.999	.052	19.08
43°	.682	.731	.933	88°	.999	.035	28.64
44°	.695	.719	.966	89°	1.000	.018	57.29
45°	.707	.707	.000	90°	1.000	.000	∞

GLOSSARY

Alternating Current (a-c): Electric current which moves first in one direction for a fixed period of time and then in the opposite direction for the same period of time. Ac changes in value continuously and reverses direction at regular intervals.

Ampere-Turns: The unit of magnetomotive force. Equal to the number of amperes of current flowing in a winding, multiplied by the number of turns in the winding.

Autotransformer: A transformer in which part of the primary winding serves as the secondary or in which part of the secondary winding is also in the primary. It has good voltage regulation under varying load conditions.

Bandpass Filter Circuit: A filter circuit which passes a desired narrow band of frequencies while rejecting all other undesired frequencies.

Bandwidth: The number of cycles that receive approximately the same amplification in an amplifier.

Band-Rejection Filter Circuit: A filter circuit which rejects a desired narrow band of frequencies while passing all other desired frequencies.

Capacitance (C): That property of an electric circuit which tends to oppose a change in voltage.

Capacitive Reactance (X_C): The opposition offered by a capacitance to alternating current. Measured in ohms. $X_c = 1/(2\pi fC)$.

Capacitor: Any two conductors separated by a dielectric.

Copper-Oxide Rectifier: A rectifier made up of discs of copper, coated on one side with cuprous oxide. Allows current flow in one direction and opposes current flow in the other direction.

Counter EMF: A voltage produced by a changing current and which at every instant opposes the change of current that produces the voltage.

Dielectric: Any insulating or nonconducting material. Air, mica, glass, paper, oil, and rubber are common dielectrics.

Dielectric Constant: The ratio of the ability of a given material to establish electric lines of force between two conductors, as compared to dry air.

Distributed Capacitance: Stray or random capacitance that exists between connecting wires, between components located physically near to each other, and between different parts of a given component.

Eddy Currents: Small circulating currents (power losses) set up by the induced voltage in any conductor carrying alternating currents.

Electrolytic Capacitor: A type of fixed capacitor which shows polarity, and is used principally in relatively low-frequency filter circuits at voltages up to 600 volts.

Electromagnet: A coil of wire, usually wound, or a soft-iron core, which produces a strong magnetic field when current is sent through the coil.

Farad: The unit of measurement of capacitance. One million microfarads (μf) equals one farad.

Frequency: The number of complete cycles per second that an alternating current undergoes.

Galvanometer: A sensitive instrument used to measure small voltages and currents.

Henry: The unit of measurement of inductance. A thousand millihenries (mh) equals one henry. A million microhenries (μh) equals one henry.

High-Pass Filter: A type of filter which offers little opposition to the passage of high frequencies, and high opposition to the passage of low frequencies.

Hysteresis Losses: Energy lost in the core of a transformer by the constant reversing of the alternating current.

Impedance (Z): Opposition to the flow of alternating current that results from any combination of resistance, inductive reactance, and capacitive reactance, or any two of these factors.

Induced EMF: A voltage produced when a current-carrying conductor is moved through a magnetic field and cuts across the lines of force, or when the magnetic field is moved across the conductor.

Inductance (L): That property of an electric circuit or component which opposes any change in current.

Inductive Reactance (X_L): The opposition offered by an inductance to alternating current. Measured in ohms. ($X_L = 2\pi fL$).

Kirchhoff's Current Law: States that the sum of all the currents flowing to a point in a circuit must be equal to the sum of all the currents flowing away from that point.

Kirchhoff's Voltage Law: The sum of all the voltage drops around a closed circuit is equal to the applied voltage.

Left-Hand Rule for Motors: A means of showing the relative directions of magnetic field flux, current flow in a conductor, and motion of the conductor through the field.

Low-Pass Filter: A type of filter that offers little opposition to the passage of low frequencies, and high opposition to the passage of high frequencies.

Mutual Induction: Production of an alternating voltage that occurs when two coils are placed close to one another in such a manner that the magnetic flux set up by one coil links the turns of the other coil.

Parallel-Resonant Circuit: A circuit in which an inductor and a capacitor are connected in parallel and have such values that at the resonant frequency the inductive reactance and the capacitive reactance are equal. Line current is at a minimum.

Peak Voltage: The highest instantaneous voltage attained in a circuit in a given period of time. Equal to 1.414 times the rms value.

Peak-to-Peak Voltage: For any alternating waveform, the total potential difference between maximum voltage amplitudes of opposite polarities.

Phase: The time difference between any point on a cycle and the beginning of that cycle.

Phase Difference: The time difference between any two cycles.

Q: A measure of the "quality" of a circuit. Varies inversely with the resistance of the circuit. Equal to X_L/R.

Resonant Frequency: The single frequency at which $X_L = X_C$ in a circuit.

Right-Hand Rule for Generators: A means of showing the relative directions of magnetic field flux, motion of a conductor through the field, and of the current induced in the conductor.

RMS (Root-Mean-Square) Value: The effective value of an alternating voltage or current. Equal to 0.707 of maximum or peak value. Corresponds to the equivalent d-c value which produces the same heating effect.

Series-Resonant Circuit: A circuit in which an inductor and a capacitor are connected in series and have such values that at the resonant frequency the inductive reactance and the capacitive reactance are equal. Current is at a maximum.

Skin Effect: The name given to the tendency of high-frequency (r-f) currents to concentrate at the surface of a conductor. Caused by counter-emf's induced in the center of a conductor carrying high-frequency currents which forces them to travel at the surface.

Step-Down Transformer: One in which the voltage induced in the secondary is less than that applied to the primary.

Step-Up Transformer: One in which the voltage induced in the secondary is greater than that applied to the primary.

Tank Circuit: Any resonant circuit (usually applied to parallel circuits).

Transformer: A device which by electromagnetic induction converts an a-c input voltage higher or lower than the input voltage.

Turns Ratio: A comparison of the number of turns in the primary winding of a transformer to the number of turns in the secondary winding.

INDEX TO VOL. II

(Note: A cumulative index covering all six volumes in this series is included at the end of Volume VI.)

INDEX

basic radio

by MARVIN TEPPER

Electronic Services Division
Raytheon Company

Author of
FUNDAMENTALS OF RADIO TELEMETRY

VOL. 3

JOHN F. RIDER PUBLISHER, INC., NEW YORK

Copyright September 1961 by John F. Rider Publisher, Inc.

All rights reserved. This book or any parts thereof may not be reproduced in any form or in any language without permission of the publisher.

Library of Congress Catalog Card Number 61-11229

Printed in the United States of America

Fifth Printing, 1968

PREFACE

The purpose of this book is to fill a need for a text stating in plain, everyday language, what many people consider a complex technical subject. A technical subject need not be complex. Careful filling in of the background with essential information, and then leading step by step to the final explanation, provides a logical method of explaining the most difficult subject.

It would be impossible to cover in a single book or series of books, the immense scope implied in the word *electronics*. However, an understanding of radio circuits serves as a foundation for advanced study in all fields of electronics, such as television, radar, computers, etc. For teaching radio, the all-important basic tool of electronics, most available textbooks are woefully inadequate. One type contains information so brief as to acquaint rather than instruct. Another type is based on the premise that teaching a student to design a circuit is the best method of having him understand that circuit's operation.

Basic Radio represents the neglected middle ground. It is a course in radio communications, as distinct from a general course in electronics. The text deals with the circuitry and techniques used for the transmission and reception of intelligence via radio energy. Assuming no prior knowledge of electricity or electronics, the six volumes of this course "begin at the beginning" and carry the reader in logical steps through a study of electricity and electronics as required for a clear understanding of radio receivers and transmitters. Illustrations are used on every page to reinforce the highlights of that page. All examples given are based on actual or typical circuitry to make the course as practical and realistic as possible. Most important, the text provides a solid foundation upon which the reader can build his further, more advanced knowledge of electronics.

The sequence of *Basic Radio* first establishes a knowledge of d-c electricity. Upon this is built an understanding of the slightly more involved a-c electricity. Equipped with this information the reader is ready to study the operation of electron tubes and electron tube circuits, including power supplies, amplifiers, oscillators, etc. Having covered the components of electronic circuitry in Volumes 1 through 3, we assemble these components

in Volume 4, and develop the complete radio receiver, AM and FM. In Volume 5 we recognize the development of the transistor, and devote the entire volume to the theory and circuitry of transistor receivers and semiconductors. The last volume of the course, Volume 6, covers the long-neglected subject of transmitters, antennas, and transmission lines.

No prior knowledge of algebra, electricity, or any associated subject is required for the understanding of this series; it is self-contained. Embracing a vast amount of information, it cannot be read like a novel, skimming through for the high points. Each page contains a carefully selected thought or group of thoughts. Readers should take advantage of this, and study each individual page as a separate subject.

Whenever someone is presented with an award he gives thanks and acknowledgement to those "without whose help . . ." etc. It is no different here. The most patient, and long-suffering was my wife Celia, who typed, and typed, and typed. To her, the editorial staff of John F. Rider, and others in the "background", my heartfelt thanks and gratitude for their assistance and understanding patience.

MARVIN TEPPER

Malden, Mass.
September 1961

CONTENTS

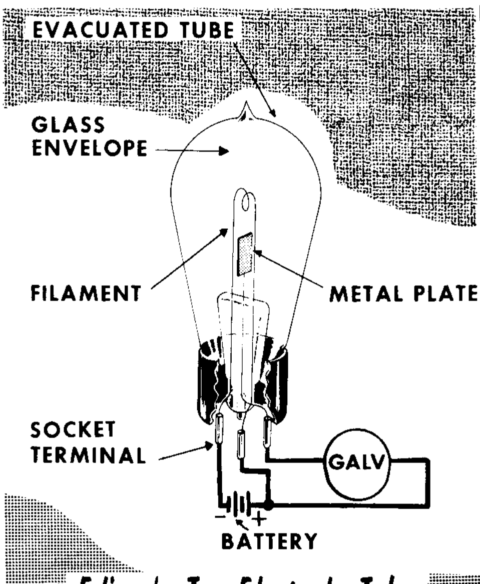
VOL. III — BASIC RADIO

Development of the Electron Tube

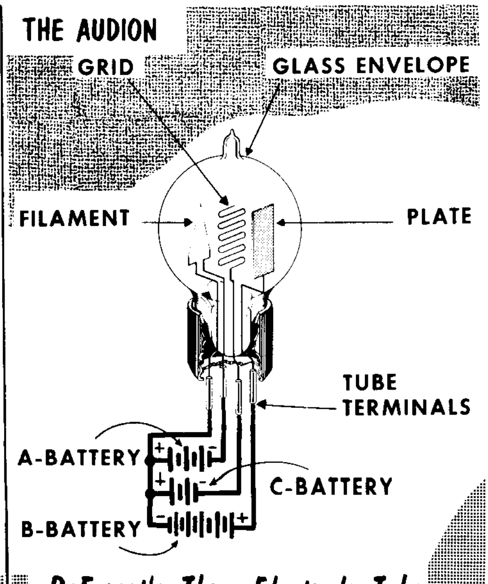
In Volumes 1 and 2, we studied d-c and a-c electricity. In Volume 3, we learn about the electron tube and its associated circuitry, which takes us into the field of electronics. The electron tube has made possible the transmission and reception of music and speech over great distances, and its many applications have given birth to the communications industry. Although the electronics field has become quite vast, we shall be concerned primarily with radio communications.

It can be said that the modern electron tube began with the phenomenon of the Edison Effect. Thomas A. Edison experimentally inserted a metal plate into the same glass bulb with a carbon wire filament, and connected it to the positive side of the battery used to heat the filament. Despite the open-circuit condition prevailing between the heated filament and the plate, Edison still measured some current flow on the galvanometer (sensitive ammeter) connected in series with the plate. He was unable to explain this situation, which came to be called the Edison Effect.

In 1899, Sir J. J. Thompson presented his electron theory as an explanation of the Edison Effect. Thompson said that electrons were emitted by the heated filament as a result of operating it at incandescence, or white heat. He said, further, that these electrons, because of their negative charge, were attracted to the positively charged plate. These electrons formed an electron current that bridged the filament-plate gap. Later development led to J. A. Fleming's "valve," a two-electrode, improved version of Edison's device. Finally, the three-electrode tube called the audion was invented by Lee De Forest in 1907; this gave the real impetus to the growth of the electronics industry by its ability to perform as an amplifier in many circuits.



Edison's Two-Electrode Tube

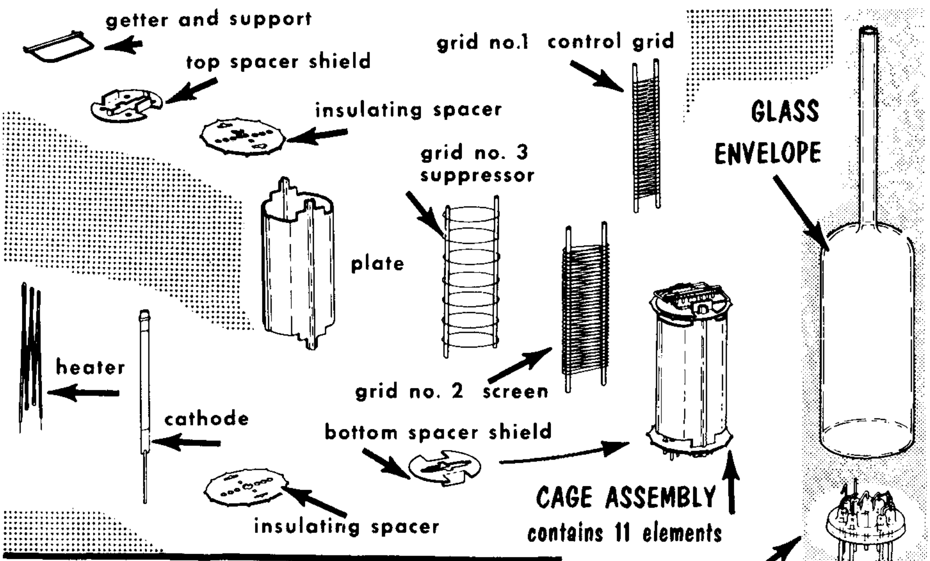


DeForest's Three-Electrode Tube

Electron Tube Construction

Before discussing the operation of the various tube types, such as the diode, triode, tetrode, and pentode, we should become familiar with the basic construction of the electron tube. Because the vast majority of tubes used in radio communications are highly evacuated, we shall refer to them by their more popular name: vacuum tubes.

The most common material used in the construction of a vacuum tube envelope is glass. Many tube envelopes are made entirely of glass, and even in the so-called "metal" tubes, the electrode leads pass through a glass bead sealed into an eyelet. Some glass tubes, such as octal types, are fitted into a plastic base for convenience in handling. The electrodes in a vacuum tube are supported by insulators such as mica and a variety of ceramics. The electrodes themselves are commonly made from metals such as nickel, copper, aluminum, molybdenum, and tungsten.

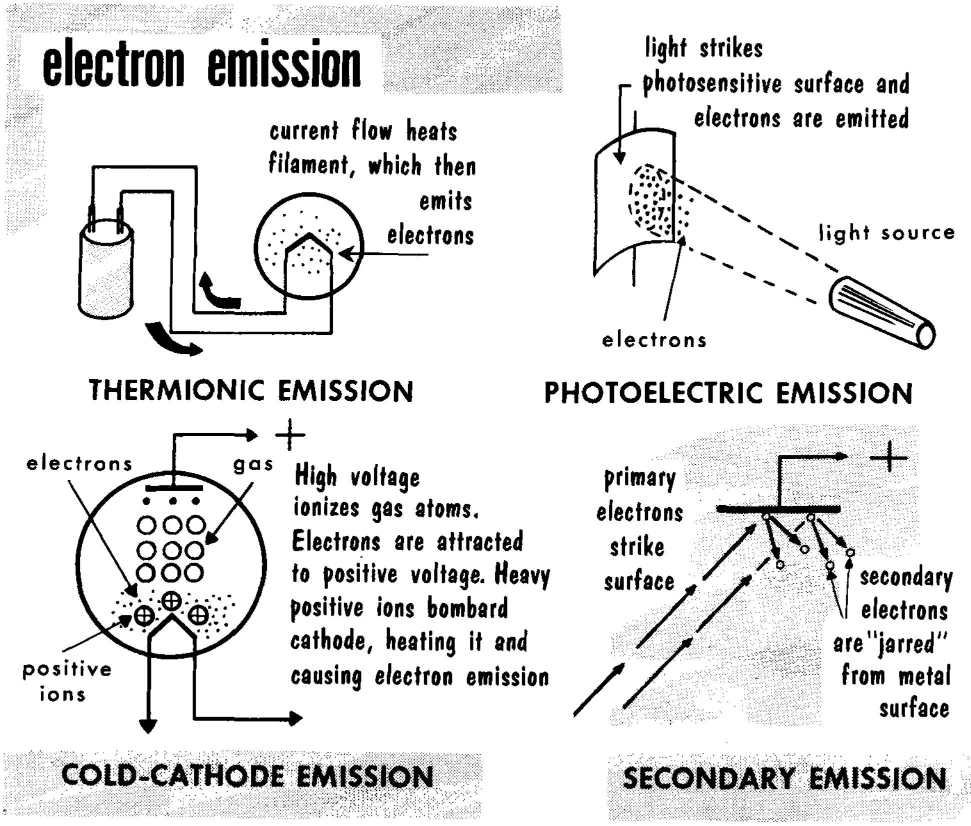


the parts of a miniature pentode

An important consideration is the creation of a high vacuum. Should there be a significant amount of air in the tube, the filament would burn up. To obtain the necessary vacuum, special vacuum pumps are used to bring the pressure in the vacuum tube down to less than 1/1,000,000,000 that of normal atmospheric pressure.

After a tube is evacuated and sealed, a magnesium or barium getter placed in the tube is used to remove residual gases by combining with them. The getter is flashed by an r-f induction process, which often leaves a visible silvery deposit on the wall of the tube.

Types of Electron Emission

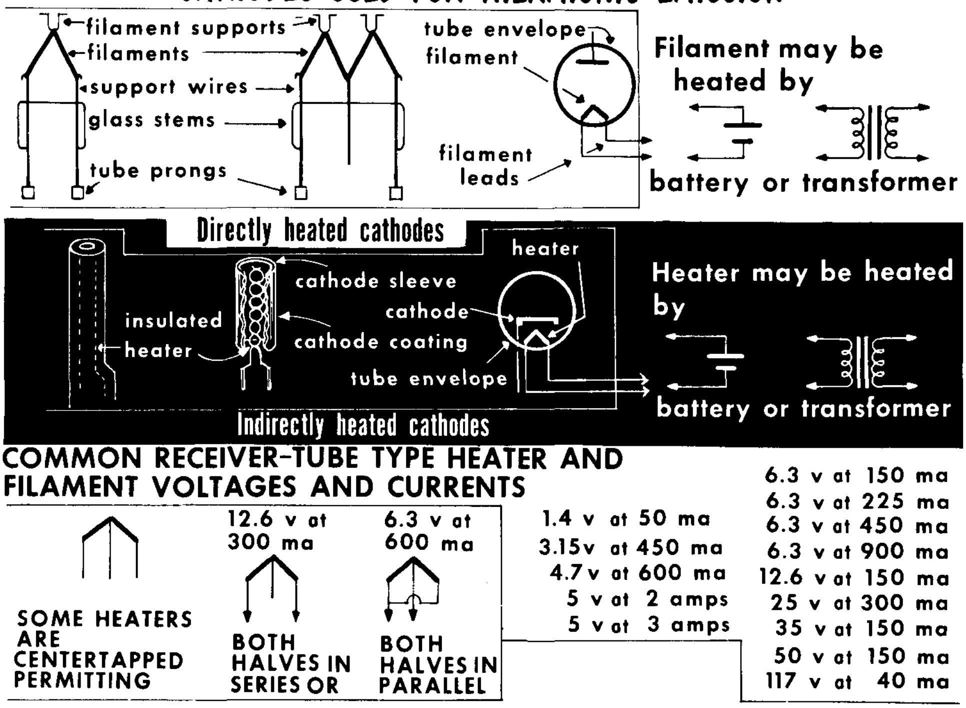


In the discussion of the electron or vacuum tube, our starting point is logically the source of electrons – the filament or cathode. When the cathode is cold, these electrons roam freely within the cathode material, but generally do not fly off the metal into the surrounding air. The surface of the metal forms a "barrier" which prevents electrons from leaving. One way in which we can force electrons to break through the surface barrier is to increase their energy of motion. If the electron energy in the cathode material can be increased sufficiently, the electrons "boil off" the cathode surface in much the same way as water vaporizes at sufficient temperature. Of the several methods which exist for speeding up the movement of electrons, four are most frequently used. Most important is the heating of the cathode material which causes thermionic emission. The heating, which may be either direct or indirect, brings about electron emission. The electrons are liberated from the cathode surface which is covered with a special chemical coating such as alkaline-earth oxides. A second method (used with certain substances) is to project light onto the emitting surface to produce photoelectric emission. A third technique is to bombard the cathode with ions, producing cold-cathode emission, and the fourth, called secondary emission, is produced by fast-moving electrons striking a surface and releasing other electrons.

Thermionic Emission

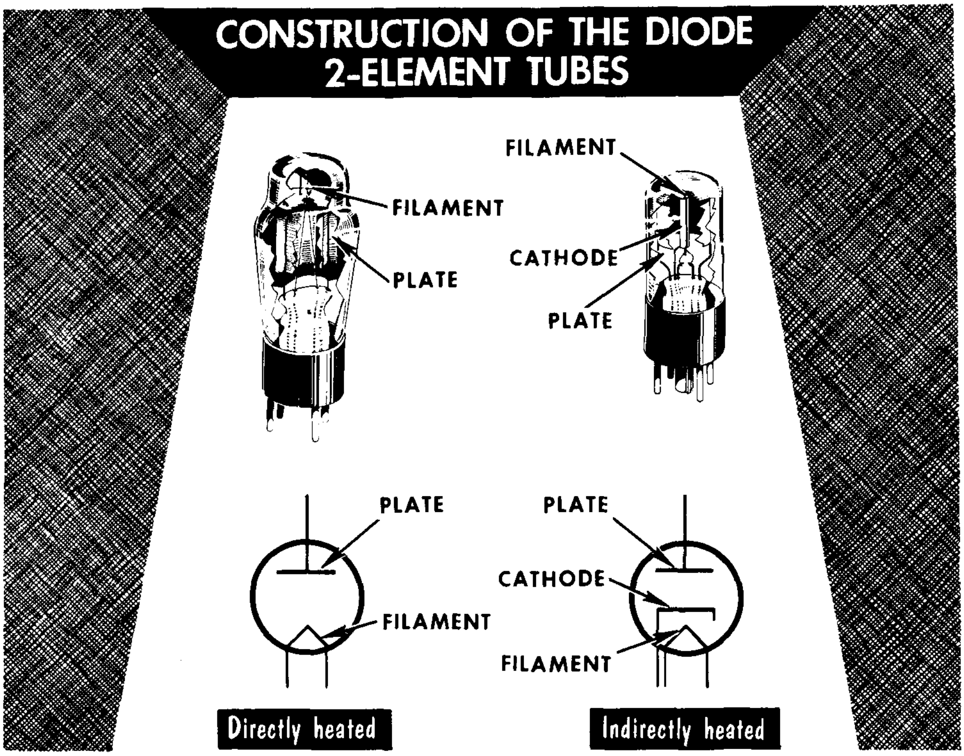
In a vacuum tube, thermionic emission is obtained when the cathode is heated inside an evacuated envelope. This is accomplished by using the heating effect produced by an electric current passed through a conducting wire or heater. Cathodes may be heated directly or indirectly. In a directly heated cathode (called a filament, or filament-cathode), the heated element is the actual emitter of electrons. In an indirectly heated cathode (or heater-cathode), the heater is used only to heat the cathode to an emission temperature and does not contribute to the thermionic emission of the tube. Filaments are usually made of pure tungsten or thoriated tungsten, or have a coating of alkaline-earth oxide.

CATHODES USED FOR THERMIONIC EMISSION



Filament-type tubes are seldom used in radio receivers except in power supply applications; they do, however, find wide use in the higher-power tubes used in radio transmitters. Pure tungsten filaments are excellent emitters where high values of thermionic emission are desired, but they require extremely high operating temperatures of about 2500°C. Thoriated-tungsten filaments provide somewhat less emission, but operate at lower temperatures of about 1900°C. The most efficient electron emitters are the oxide-coated filaments and cathodes. Here, the emission takes place from the oxide coating, which provides excellent emission for low-power tubes at temperatures of about 1000°C.

The Diode

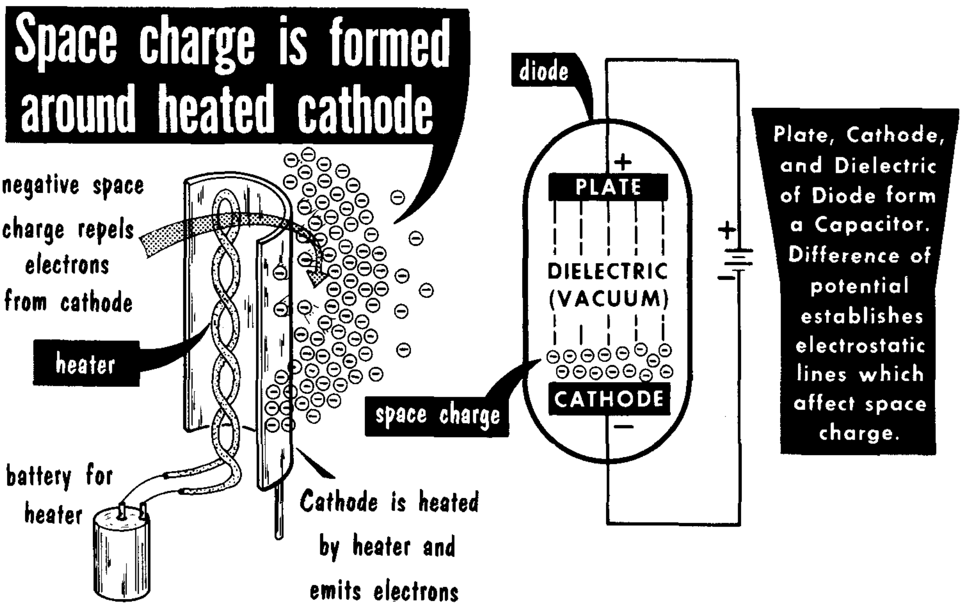


The simplest type of electron tube is the diode. It consists of two elements, or electrodes: one is the emitter of electrons and the other the collector of electrons. Both electrodes are enclosed in a glass or metal envelope which is then evacuated. Later, we shall study diodes that contain a certain amount of gas. In calling the diode a two-element tube, we consider the electron emitter as a single element, despite the fact that it might be a filament or a heater-cathode combination. The electron collector is the plate, and is sometimes referred to as the anode. We shall see that when the plate electrode is positive with respect to the cathode, it will attract electrons emitted from the cathode.

The primary function of the diode is rectification, about which we shall study later in this volume. In this function, the diode appears in two general constructions – signal diodes and power diodes. The signal types are generally small both in physical size and in their ability to handle currents and voltages. The power diode is usually relatively large, is built to handle high voltages and currents, and may become exceedingly hot during operation. Some diodes have a single cathode and plate, others have a single cathode and two plates, and still others have two cathodes and two plates. Where two plates are used in a single tube envelope, the tube is referred to as a duodiode. The plate material must be able to withstand relatively high temperatures and usually is metal such as nickel, iron, or molybdenum.

Space Charge

Heating a filament or cathode in an evacuated envelope produces electron emission. But what happens to the emitted electrons? The first group of emitted electrons gathers in the space surrounding the cathode. The following group of electrons is then repelled by the like (negative) charges of the electrons already out in space, and start to return to the cathode. A third group of emitted electrons prevents their return by repelling them outward. The result of this action is a dense cloud of electrons in the space around the cathode, with fewer electrons further away from the cathode. This electron cloud is called the space charge.

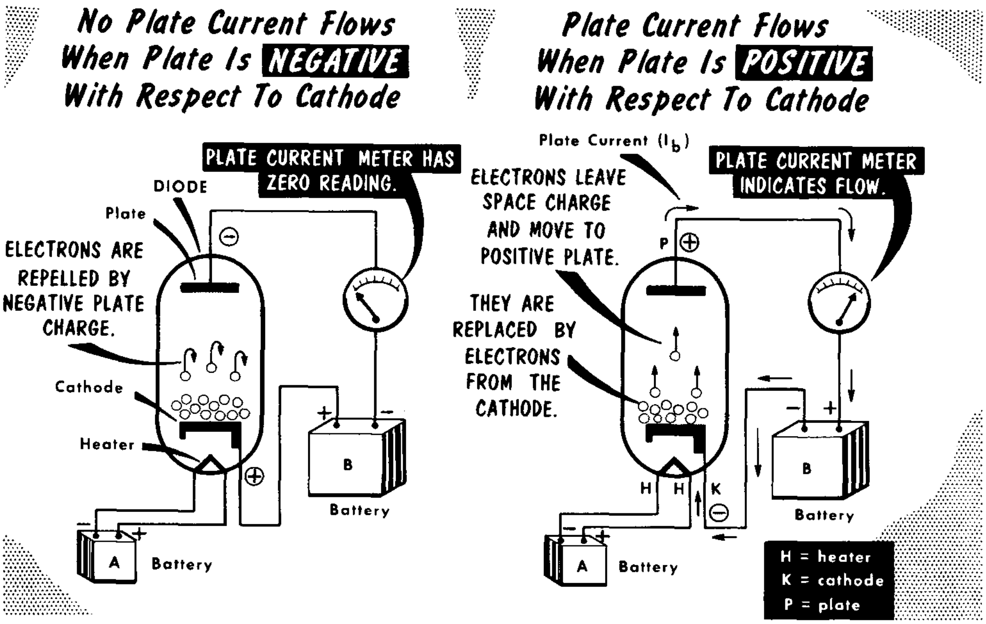


At first, it might seem that the electrons will continue to be emitted indefinitely. However, the negative space charge soon becomes so strong that it repels any additional emitted electrons back toward the cathode. This represents the maximum amount of electrons that could be emitted from the cathode unless the cathode temperature were raised. A temperature increase would cause more electrons to be released until a new equilibrium point were reached. As we shall see, the space charge is very useful in that it acts as a reservoir of electrons.

We have said that the diode is basically a two-element tube containing a cathode and a plate. From our study of capacitors, we see that these two elements, separated by an insulator (the vacuum in the tube), form a capacitor. If we place a difference of potential between cathode and plate, electrostatic lines will be established across the dielectric and will influence the space charge. If the plate is made positive with respect to the cathode, electrons from the negative space charge will be attracted to it.

Plate Current

Now that the electrons are in the form of a space charge, the problem is to put them to use. If another electrode is placed in the enclosed vacuum and given a positive charge with respect to the cathode, electrons will be attracted to it from the space charge. (The reduction of the space charge resulting from this permits more electrons to be emitted from the cathode.) A continuous positive voltage on the second electrode (called the anode or plate) can be maintained by connecting the positive terminal of a battery to it, while connecting the negative terminal to the cathode. This results in a continuous flow of electrons through the tube. These electrons then move to the positive terminal, through the battery, and finally complete the circuit by returning to the cathode.

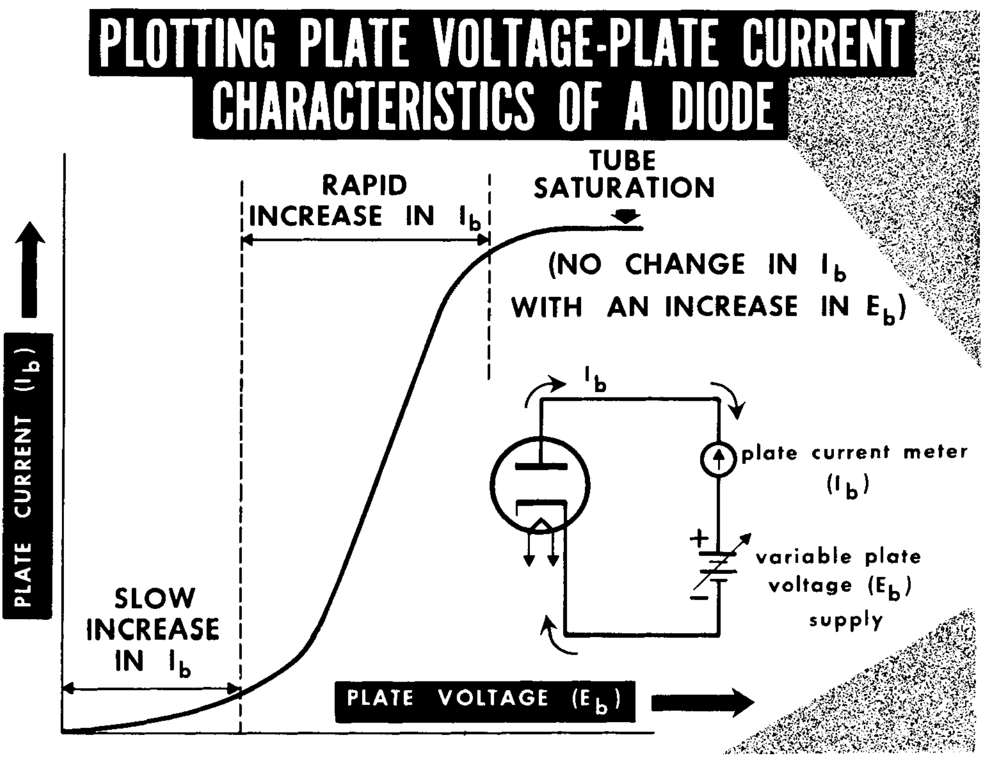


The electron flow through the vacuum is called the plate current (I_b). The voltage applied between plate and cathode is called the plate voltage (E_b), and the voltage source the B supply. (The voltage source used to heat the heater or filament is called the A supply.) We can control the flow of plate current in a diode two ways: by varying the cathode temperature; and by varying the difference of potential between plate and cathode (plate voltage). Cathodes are designed to operate most efficiently at one particular temperature, so variation of cathode temperature must be ruled out in favor of plate voltage variation. Control of plate current is relatively simple to maintain: increasing plate voltage increases plate current and decreasing plate voltage decreases plate current. If the negative terminal of the B supply is connected to the plate and the positive terminal to the cathode, the plate will repel all the emitted electrons, and there will be zero plate current.

Plate-Voltage, Plate-Current Curves

An important relationship in electron tubes exists between the voltage applied to the plate and the resultant plate current when the cathode temperature is held constant. If we place an ammeter in the plate circuit of the diode and slowly vary the voltage on the plate from zero to some high value (depending on the tube type), we notice a variation in plate current. As the plate voltage increases from zero to a slightly positive value, the amount of plate current (I_b) begins to increase gradually. As plate voltage (E_b) further increases, the plate current begins to increase at a fairly regular amount for a given change in plate voltage. Finally, a point is reached where increases in plate voltage produce less and less change in plate current. This point is called "saturation," and the plate is taking all the electrons available at the cathode for that given cathode temperature. At this saturation point, further increases in plate voltage produce virtually no change in plate current. Raising the cathode temperature produces more electron emission, but only at the expense of overworking the heater element.

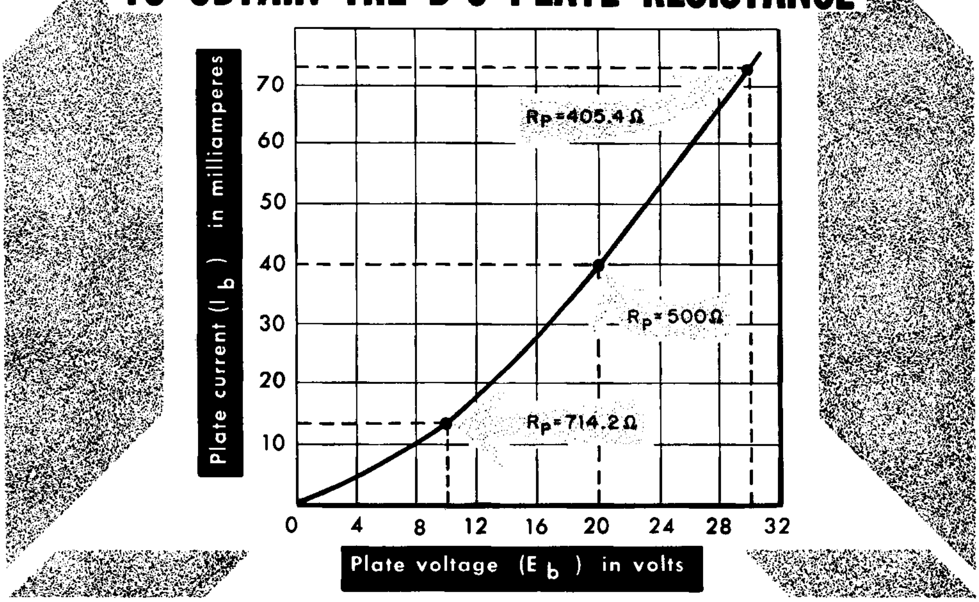
If we were to make a graph by plotting the current value for all the plate voltage values, we would end up with a characteristic curve of the tube. This E_b-I_b curve provides, at a glance, the entire operating characteristic. Static curves generally show operation with no "load" or voltage variations in the tube circuit; dynamic curves generally show tube operation when there is a load and voltage variations in the tube circuit.



Diode Plate Resistance – D-C

The d-c plate resistance of the diode is opposition to the flow of plate current offered by the tube when a d-c voltage is applied to the plate. From Ohm's law, we can develop the formula for d-c plate resistance, R_p. Using E_b and I_b for tube voltage and current, we get $R_p = E_b/I_b$. The d-c resistance of a diode depends upon many things, such as the size of the electrodes, temperature of the cathode, or distance between electrodes. We can calculate this resistance from the plate voltage-plate current characteristic of a tube. In the drawing, we see that when the plate voltage is 10 volts, 14 milliamperes (0.014 amp) of plate current flow. Since $R_p = 10/0.014$, the plate resistance at this point is 714.2 ohms. Going to 20 volts on the plate produces a plate current of 40 ma, and a diode plate resistance of 20/0.040, or 500 ohms. Taking one further reading at a plate voltage of 30 volts shows that the plate current has now risen to 74 ma, for a tube plate resistance of 30/0.074, or 405.4 ohms.

USING A TYPICAL DIODE STATIC CHARACTERISTIC TO OBTAIN THE D-C PLATE RESISTANCE

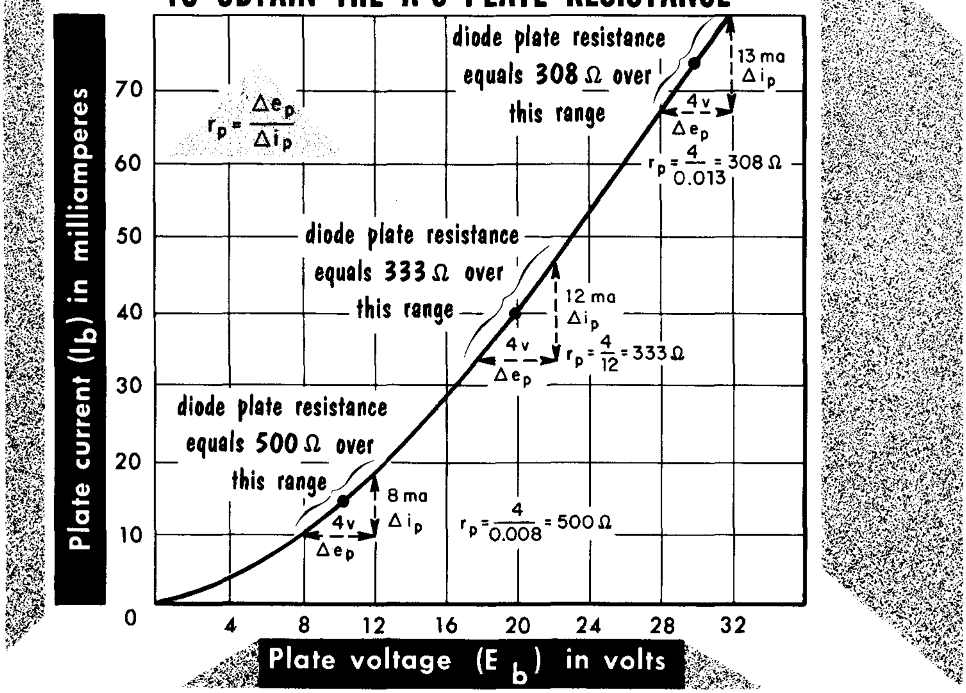


From these figures, we see that the resistance offered by the diode to the flow of plate current is not constant, as in a conventional resistor. The characteristic shows that the diode resistance decreases as the plate voltage increases, and increases as the plate voltage is decreased – it behaves nonlinearly. If the I_b-E_b curve were a straight line, the resistance of the diode would be constant at all points on the curve.

Diode Plate Resistance – A-C

We can consider the a-c plate resistance of the diode as the resistance of the path between cathode and plate to the flow of an alternating current inside the tube. We use the same curve for finding the diode a-c resistance as we did for the d-c resistance. However, the a-c plate resistance (r_p) is the ratio of a small change in plate voltage to the small change in plate current that it produces. Written in the style of Ohm's law, $r_p = \Delta e_p / \Delta i_p$. The Greek letter delta (Δ) means "a small change in." Thus, to find the a-c resistance of a diode, we consider changes in plate voltage and changes they produce in plate current.

USING A TYPICAL DIODE STATIC CHARACTERISTIC TO OBTAIN THE A-C PLATE RESISTANCE



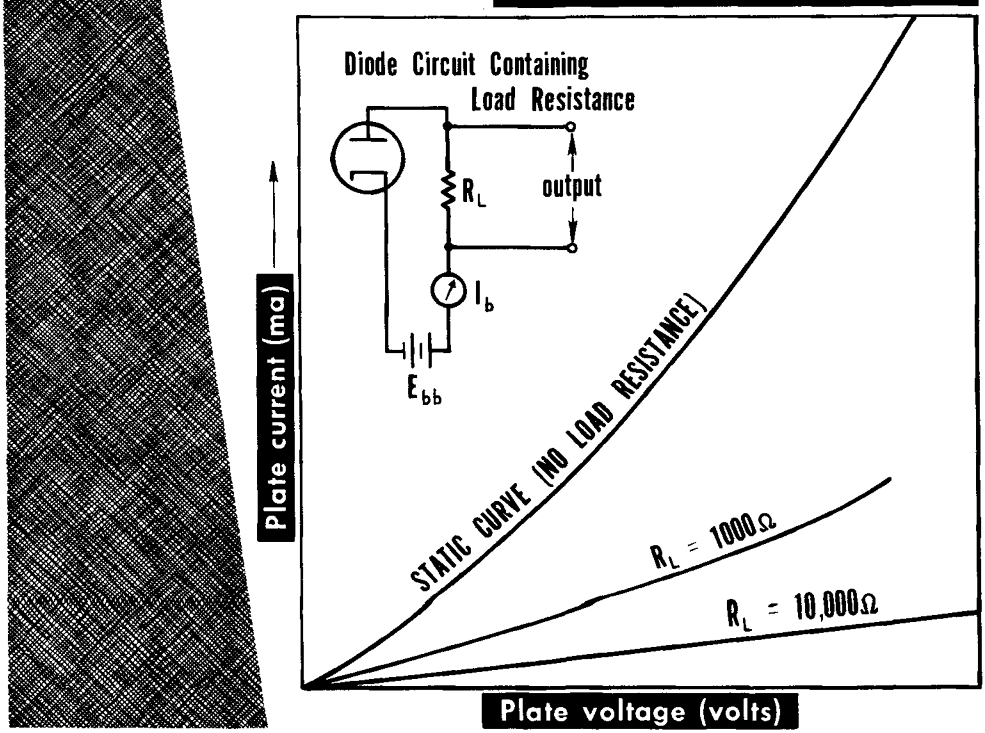
If we change the plate voltage from 8 volts to 12 volts, there is a change of 4 volts. With 8 volts on the plate, 10 ma of plate current flowed. With 12 volts on the plate, 18 ma of plate current flowed. Thus a change of 4 volts (from 8 to 12) on the plate produced a change of 8 ma (from 10 to 18) of plate current. The a-c plate resistance over this portion of the curve is then: $r_p = 4/0.008$, or 500 ohms. Using the same points on the curve as in d-c plate resistance and varying the plate voltage 2 volts above and below these points, we get a-c plate resistances of 333 ohms and 307 ohms over these ranges. From this, we can see that when the plate resistance of a tube is mentioned, it is always with respect to a particular point on the tube's characteristic curve.

Static and Dynamic Diode Characteristics

So far, we have discussed static conditions in the diode, where no load was in the circuit. Much more important are the dynamic conditions of the diode circuit. For a diode, or any tube, to perform its normal function, its external circuit must contain a load. It is through this load that the diode current flows outside the tube, and the voltage drop developed across this load then represents the output of the tube. The load resistance is given the symbol R_L.

With no load resistance in the circuit, virtually all circuit resistance consists of the diode plate resistance. However, when an external load resistance is added, the total opposition to plate current flow includes the tube plate resistance plus the load resistance. When the load resistance is many times greater than the plate resistance, it is primarily R_L that determines the shape of the characteristic curve. Note that the larger the load resistance, the less the curvature in the characteristic. As we shall learn later, a linear or straight dynamic characteristic is important because it provides proportionality between changes in plate voltage and the accompanying changes in plate current. This in turn provides freedom from distortion – important in many diode circuits.

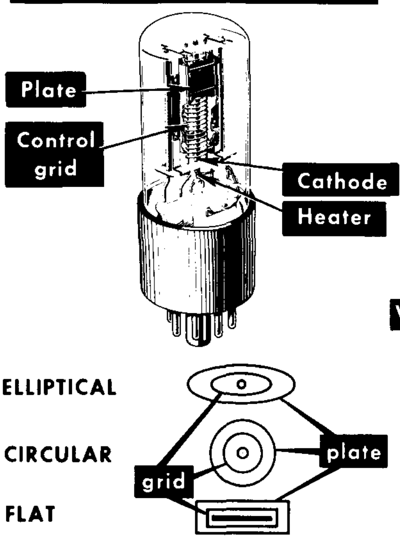
DYNAMIC CHARACTERISTIC OF DIODE BECOMES LINEAR AS LOAD RESISTANCE INCREASES



The Triode

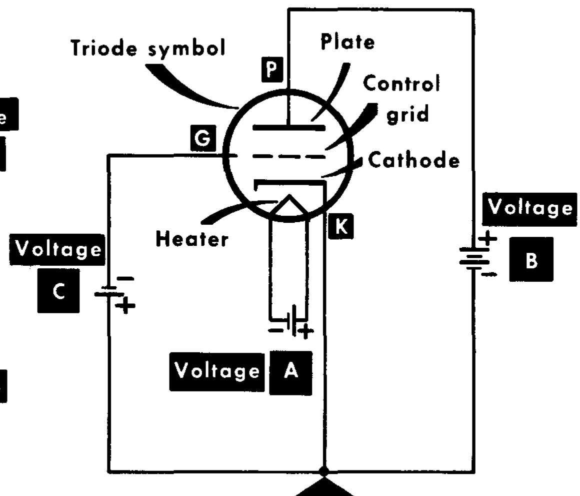
The invention of the triode was an extremely important advance in electronics because it permitted electronic "amplification." Basically, the triode is nothing more than a diode to which has been added a third electrode – a grid. Physically, the grid is a ladder-like structure of metal. In most instances, it has a helical form and consists of a number of turns of fine wire wound in the grooves of two upright metal supports. The physical dimensions of the grid determine many of the operating characteristics of the triode.

CONSTRUCTION OF A TYPICAL TRIODE



Cross-sectional top view of triode constructions

VOLTAGES APPLIED TO THE TRIODE CIRCUIT



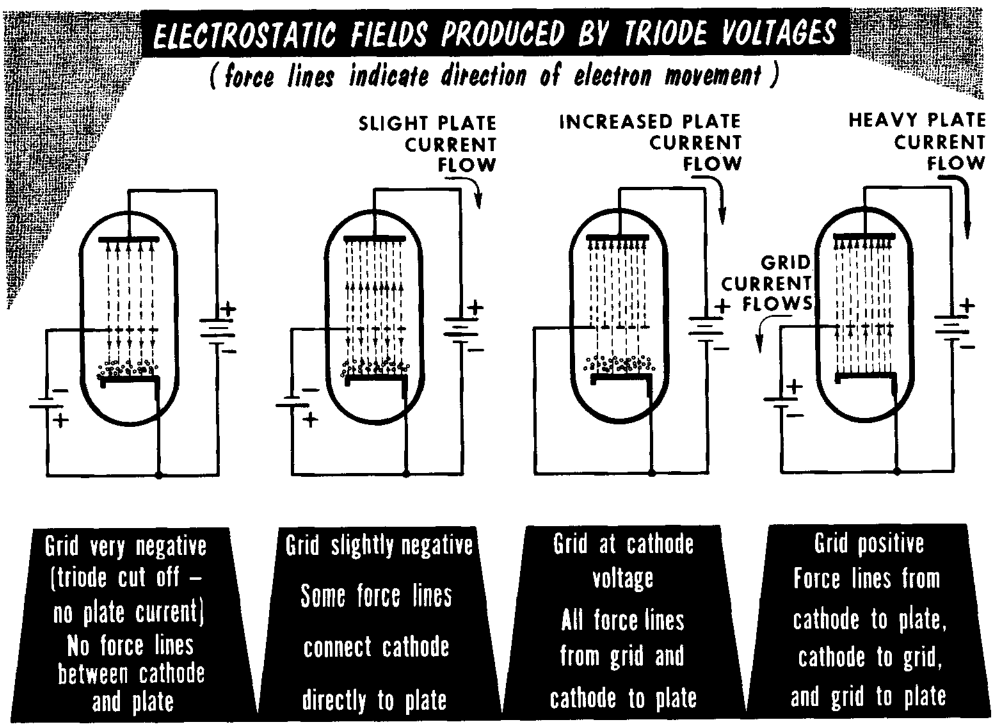
all grid and plate voltages are stated with respect to cathode

The basic function of the grid is to control the movement of electrons between the cathode and plate, thereby controlling the amount of plate current flowing in the tube. It is for this reason that this third electrode is given the name control grid. The control grid may be wound around the cathode in the space between the cathode and plate, or merely exist as a screen-like mesh between parallel cathode and plate surfaces. Because it is so very close to the cathode, voltages on the control grid exert much more influence on plate current than do voltages on the plate. Since the cathode is generally taken as the reference point, all plate and grid voltages are measured with respect to the cathode. D-c plate voltages are known as "B" voltages, d-c grid voltages (called "bias") as "C" voltages, and heater, or filament voltages as "A" voltages.

Electrostatic Field in Triode

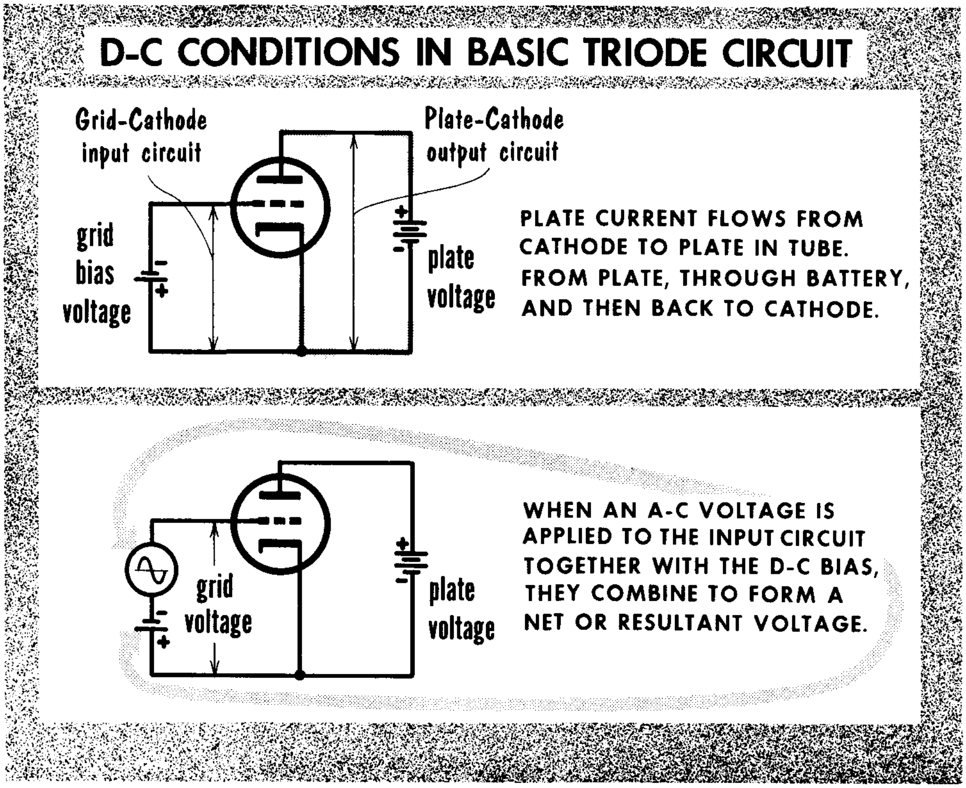
The control grid in the triode serves primarily as an electrostatic shield between the plate and cathode, and allows some, but not all, of the electrostatic field from the plate to get to the cathode. When the grid is at the same potential as the cathode, it exerts no electrostatic influence, and the triode acts just as a diode. However, when the control grid is made negative with respect to the cathode, it sets up an electrostatic field in opposition to the cathode-plate field. The action of a negative grid is to repel electrons from the cathode back to the cathode. At the same time, the positive plate acts to attract electrons from the cathode to the plate. The net effect is a compromise, depending upon the relative grid and plate voltages. Since the grid is so close to the cathode, a relatively small grid voltage has as much influence as a much larger plate voltage.

When the grid is made slightly negative, some electrons from the cathode are repelled back, but many others pass through the wire mesh grid structure and are attracted to the positive plate. It is possible that, for a given plate voltage, the grid voltage can be made sufficiently negative to cut off the flow of plate current completely. On the other hand, the grid (bias) voltage could be made so positive as to produce plate current saturation, with further increases in grid voltage having no effect on plate current. When the grid is positive with respect to cathode, some electrons are attracted to the grid structure and produce a flow of grid current. The control grid is most commonly operated at a voltage slightly negative with respect to the cathode.



Steady-State Condition of a Triode Circuit

The basic triode circuit consists of the grid-cathode portion called the input and the plate-cathode portion called the output. In this circuit, all input voltages or signals are applied between grid and cathode; all output voltages or signals appear between plate and cathode. Thus, we see that the cathode is the element common to both input and output circuits. The voltage applied to the plate or anode of the tube is positive with respect to the cathode and, in typical circuits, may vary from 100 to 300 volts. In most receiver circuits, the control grid is kept at a voltage slightly negative with respect to the cathode. Hence, it acts to limit or control the amount of electron flow between cathode and plate.

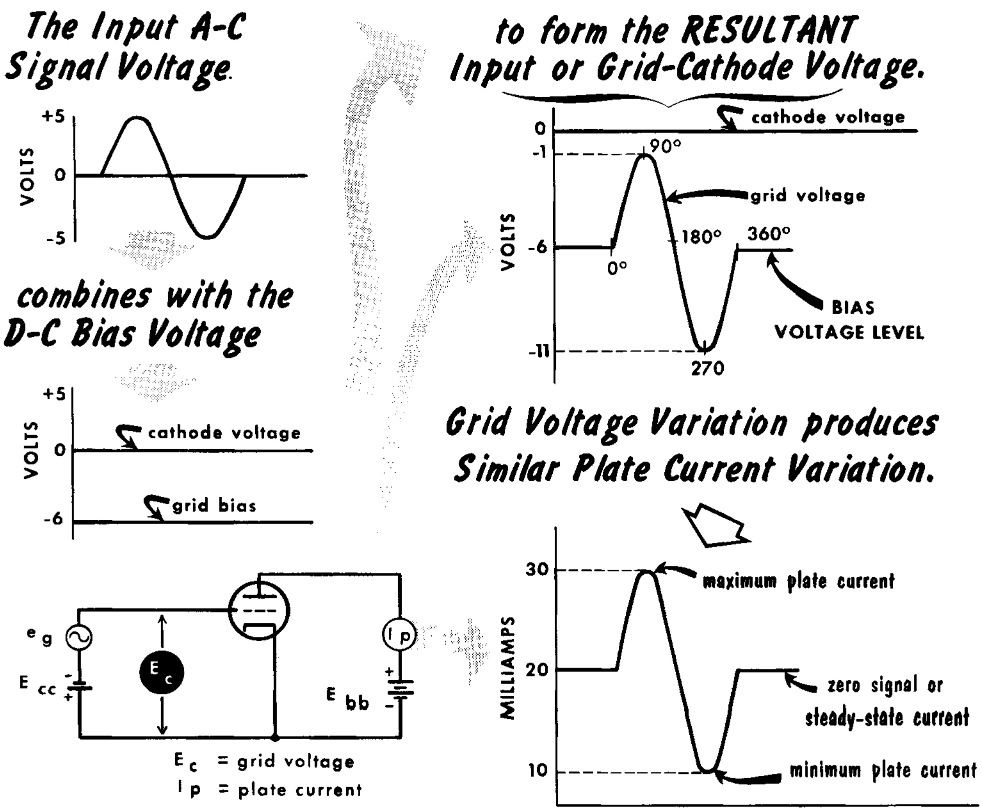


The slight negative voltage applied to the control grid is called the bias voltage. For the moment, we shall assume that the plate and bias voltages are obtained through the use of batteries. With no input voltage applied to the circuit, a small steady plate current will flow. The amount will depend upon the values of positive plate voltage and negative grid voltage. This is often referred to as the steady-state condition of the circuit – when no external voltage is applied to the input circuit. These conditions change when an a-c voltage (signal) is applied to the input circuit.

Effect of Grid Voltage on Plate Current

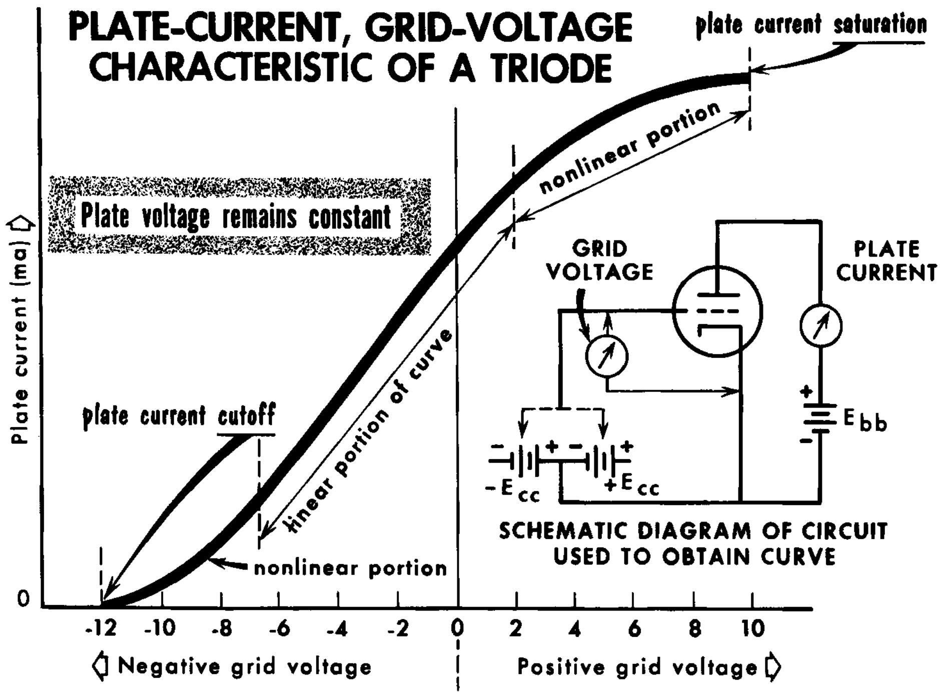
Let us now apply an a-c voltage to the input circuit in addition to the fixed grid bias voltage. We shall assume that the a-c voltage has a peak of 5 volts, and that the steady-state d-c grid bias is -6 volts. At 0°, or the beginning of the input cycle, the 6-volt negative bias permits a plate current flow of 20 milliamperes (ma). At 90° of the input cycle, the a-c component of the grid voltage has risen to its maximum positive value, +5 volts. This voltage, added to the steady -6 volts, produces a net grid voltage of -1 volt with respect to the cathode. The grid voltage is still negative, but now it is only slightly negative, and plate current increases to its maximum of 30 ma.

From 90° to 180°, the input signal returns to zero, which added to the 6-volt negative bias produces a grid bias of -6 volts. The plate current thus drops back to its steady-state value of 20 ma. From 180° to 270°, the input voltage increases to its maximum negative value, -5 volts, which is added to the fixed bias, producing a total grid voltage of -11 volts. This large negative grid voltage reduces the plate current to 10 ma. From here, the plate current rises back to its steady value of 20 ma as the input voltage also rises back to its 360° value of zero volts. From this, we see that the plate current waveform follows the grid voltage, increasing as the grid is made less negative and decreasing as the grid becomes more negative.



Triode Plate-Current, Grid-Voltage Curve

The relationships among the various voltages applied to the triode and the effects they have on the plate current are very important. As in the case of the diode, these relationships are shown through the use of characteristic curves. Basically, plate current in a triode is determined by the grid and plate voltages, assuming that the cathode remains at a constant temperature. We shall consider the grid voltage first, and assume that the plate voltage remains constant. We can then construct a plate-current, grid-voltage (I_b-E_g) curve by varying the grid voltage first in a positive direction from zero, and then in a negative direction. The plate current at various grid voltages is then plotted. This curve is called the static plate-current, grid-voltage characteristic, because it represents the tube behavior under no-load conditions. The plate and grid voltage are simply those of their respective supply voltages, and there are no voltage drops across load resistors.

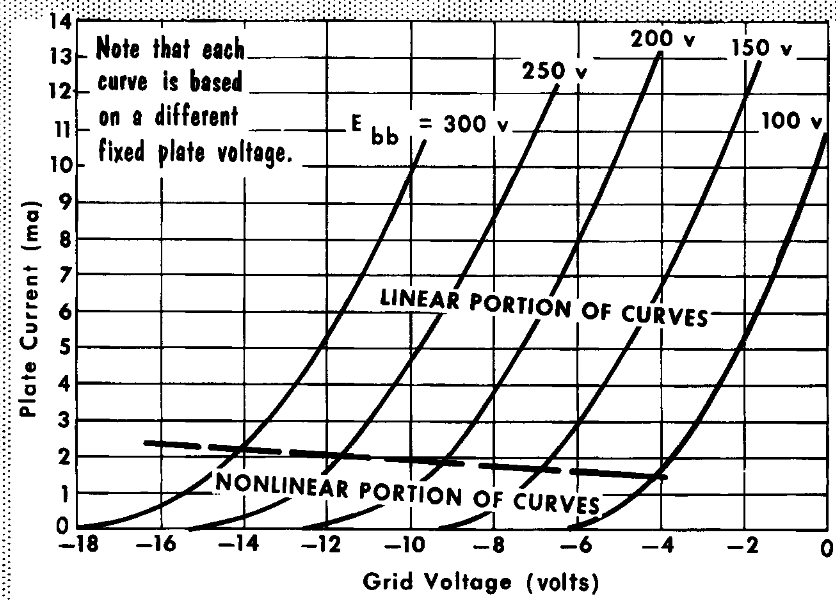


Note that the curve shown drops to zero at -12 volts. This point is known as cutoff and represents the minimum negative grid voltage needed to reduce plate current to zero. As the grid is made less negative, plate current begins to flow. The rise is gradual at first, then more rapid over the linear portion of the curve. Finally, the curve reaches a point where further increases in positive grid voltage produce no further increases in plate current. This is known as plate current saturation. We will see that most amplifiers operate on the linear portion of the curve between the cutoff and saturation areas.

Family of Grid Characteristic Curves

A single plate-current, grid-voltage characteristic curve furnishes important information, but is of limited value. A number of such curves shown on the same scale for different values of plate voltage give much more information concerning the effects of various grid voltages on plate current. Such curves, plotted on a single graph, comprise a "family" of characteristic curves. As a rule, the grid family of characteristic curves does not involve the positive region of grid voltage because, for most triode applications, the grid is not driven positive with respect to the cathode. The family of curves shown are for the 6J5 triode. These curves are made as just described, except that the plate voltage was changed for every series of grid voltage variations. Note the close similarity among the general shapes of the characteristic curves: each curve has a linear portion and a nonlinear portion.

TYPICAL GRID FAMILY OF CHARACTERISTIC CURVES



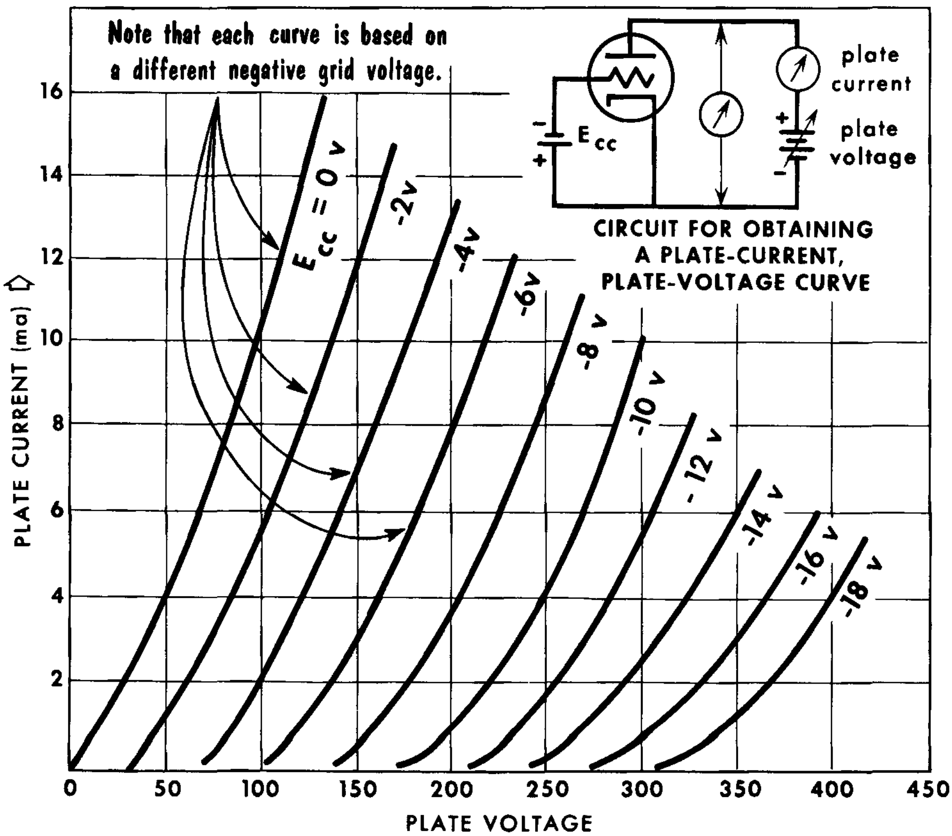
From this family of grid curves, we can see that the higher the positive plate voltage, the higher the negative grid voltage required to cut the tube off (reduce plate current to zero). In addition, the higher the plate voltage, the more plate current will flow for any given grid voltage. For purposes of amplification, we should remember that proportional changes (ratio of plate current to grid voltage) occur only over the linear (straight) parts of the curve. The greatest change in plate current per unit change in grid voltage occurs along the straight (rather than the curved) portion of the characteristic.

Triode Plate-Current, Plate-Voltage Curves

As in the case of the grid family of curves, we can make use of a plate family of curves. For every curve in this group, a particular grid voltage is held constant while the plate voltage is varied. In every instance, the plate current range shown for a particular fixed grid voltage starts at the point along the plate voltage axis where the negative grid voltage causes approximate plate current cutoff. In general, the grid family and plate family of curves furnish the same information; however, in somewhat different forms.

Where the grid family displays the plate current for small changes or increments of grid voltage and fixed differences in plate voltage, the plate family displays the plate current for small increments of plate voltage and fixed differences of grid voltage. These family graphs present the relationship between the different triode electrode voltages under static conditions. As we shall soon learn, these various curves are of tremendous value in understanding the operation of tubes under varied conditions. Virtually every tube type has its own set of grid and plate characteristics which are often shown in tube manuals and technical data sheets issued by tube manufacturers.

TYPICAL PLATE FAMILY OF CHARACTERISTIC CURVES



Amplification Factor

Vacuum tubes have special figures of merit called tube constants. As we shall learn, these constants are used for tetrodes and pentodes, as well as for triodes. The three most commonly used tube constants are the amplification factor, symbolized by the Greek letter μ, plate resistance r_p, and mutual conductance g_m. Understanding their meaning and significance permits us to look at tube specifications and determine important design and operating characteristics of electron tubes.

Amplification factor tells us of the relative influence of grid voltage as compared to plate voltage on the amount of plate current flow in an electron tube. We could say that the amplification factor of a triode is equal to the ratio of a change (Δ) in plate voltage to a change in grid voltage in the opposite direction, under the condition that the plate current remain constant. Another way of stating it is to say that μ is equal to the ratio of a change in plate voltage to a change in grid voltage that results in the same change in plate current. As a formula, we say:

$$\mu = \frac{\Delta e_p}{\Delta e_g}$$

or a change in plate voltage divided by change in grid voltage (plate current remaining constant). For example, suppose that a 1-ma plate current change could be produced by a change in grid voltage of 0.1 volt, and a 1-ma plate current change was produced by a change in plate voltage of 10 volts. Such a tube would have an amplification factor of 10/0.1, or 100. This is a relatively high μ, and such a tube is called a high-μ tube. Low-μ tubes have an amplification factor of 10 or so. Between 10 and 30, we have the medium-μ tubes; above 30, a tube is considered to have a high μ. As can be seen, the μ of a tube can vary slightly, depending upon which portion of the curve is used.

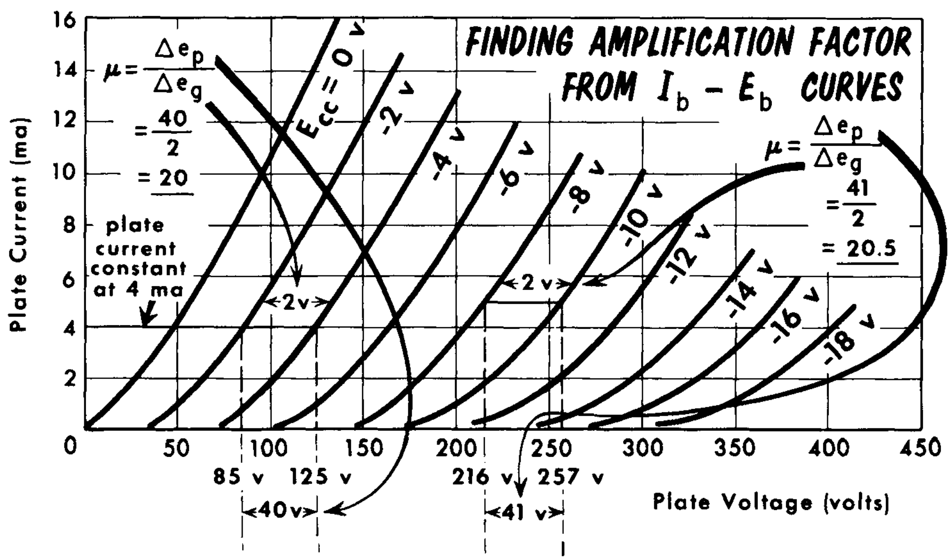


Plate Resistance

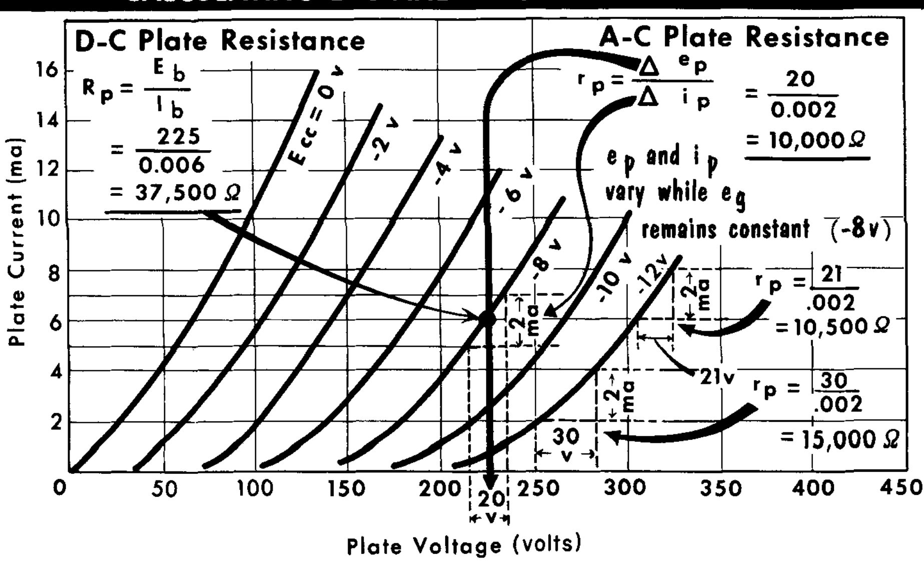
The plate resistance of a tube describes its internal resistance – that is, the opposition to electron flow between cathode and plate inside the tube. The d-c plate resistance expresses the steady-state resistance. This occurs when fixed voltages are on the control grid and plate. Under such conditions, the d-c plate resistance (R_p) can be determined by Ohm's law. We can thus say:

$$R_p = \frac{E_{\text{(d-c plate voltage)}}}{I_{\text{(d-c plate current)}}}$$

We can find R_p by taking any point on any of the plate family of curves, and then projecting down to the plate voltage and across to the plate current axes. We then solve by $R_p = E_b/I_b$.

Finding a-c plate resistance (r_p) is somewhat more involved, because this involves changes in plate voltage and plate current. We start by taking any point on any of the plate family of curves. We project a line horizontally from the curve – this represents a change in plate voltage. We then project a line upward vertically back to the curve – this line represents a change in plate current. By using the a-c equivalent of Ohm's law for plate resistance ($r_p = \Delta e_p/\Delta i_p$), we get the answer. Note that R_p concerns itself with a particular operating point, whereas r_p represents an operating range. Note also that the higher the applied plate voltage for a particular grid voltage, the lower is the a-c plate resistance. This is because the curves become steeper as they advance toward the higher plate voltages. This means that for a particular increase in plate voltage, there will be a disproportionately greater increase in plate current. The importance of r_p will be seen in our study of amplifiers.

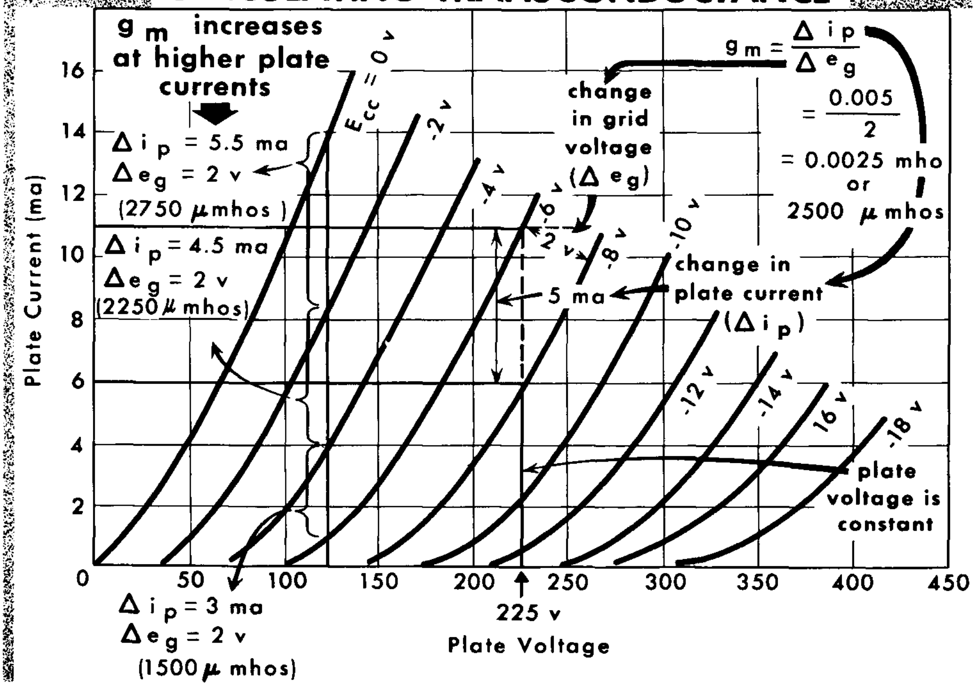
CALCULATING D-C AND A-C PLATE RESISTANCE



Transconductance (Mutual Conductance)

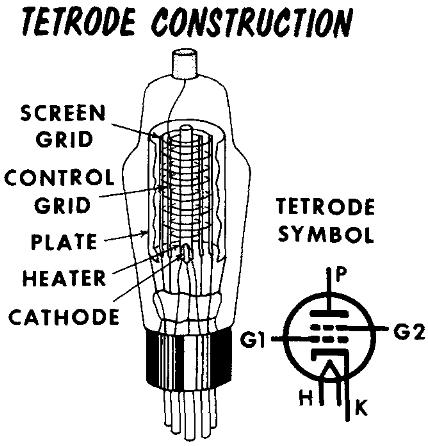
The third important tube constant discussed here is transconductance, also referred to as mutual conductance. By definition, transconductance is the quotient of a small change in plate current divided by the small change in the control grid voltage producing it, under the condition that all other tube voltages remain the same (in a triode this would refer to the plate and heater voltages). As an equation, transconductance, $g_m = \Delta i_p/\Delta e_g$. Transconductance is a measurement of conductance rather than resistance. As such, it is measured in mhos, mho being the reversed spelling of ohm. In practical tube circuitry, the mho is a rather large unit of measurement. To make it more workable, we use the micromho (μmho), one-millionth of a mho.

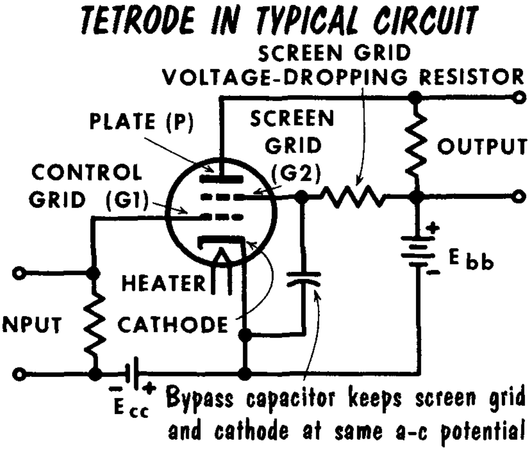
CALCULATING TRANSCONDUCTANCE

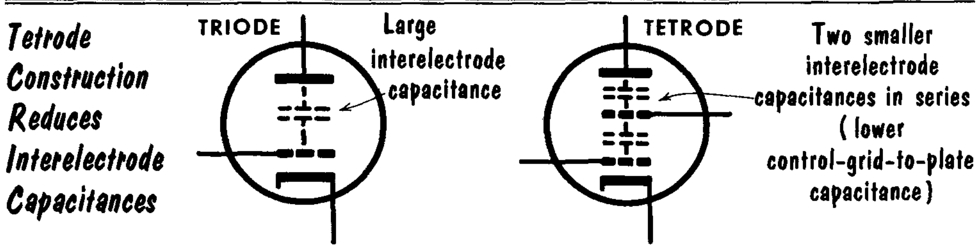


If a change of 1 volt on the control grid produces a change of 1 ma in plate current, our formula for transconductance ($g_m = 0.001/1$) would show that such a tube had a g_m of 0.001 mho, or 1000 μmhos. Transconductance is an excellent measure of how "good" a tube is. A tube having a high transconductance is capable of furnishing greater signal output than a tube with a low g_m, assuming that the same circuitry and voltages are applied to both tubes. The transconductance of most vacuum tubes varies from about 2000 to 9000 μmhos. To calculate g_m from the plate family of curves, we project along a constant-voltage line from one grid voltage curve to another. The difference in grid voltage divided by the difference in plate current, projected across to the plate current axis, gives us the g_m over this operating range.

The Tetrode







In the triode tube, there are two metallic structures very close to each other – the plate and the control grid. Since a vacuum exists between them, we have a capacitor – two conducting surfaces separated by a dielectric. As we shall learn in our study of electron tube circuitry, the capacitance between the plate (output circuit) and the control grid (input circuit) can become extremely troublesome at high operating frequencies and produce undesirable "feedback." To prevent this, a second grid is inserted between the control grid and the plate to act as a "screen" between the two. Appropriately, this second grid is called the screen grid. An electron tube containing both a control grid and a screen grid is called a tetrode. The tetrode has a finely wound control grid surrounding the cathode or emitter of electrons which, in turn, is surrounded by a coarser screen grid at a considerably greater distance out from the cathode. The screen grid, then, is surrounded by the plate.

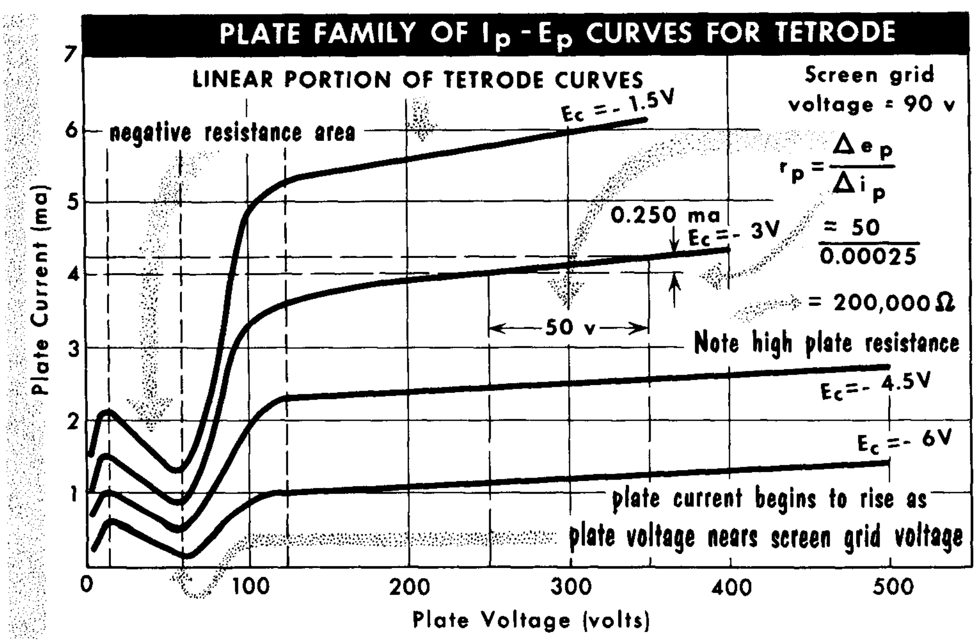
To serve as an effective electrostatic shield, the screen grid is usually at a signal or a-c potential equal to that of the cathode, so that no voltage exists between these two electrodes. This is not to be confused with the d-c potential on the screen grid, which is very often approximately the same as the positive plate voltage. To obtain a "zero" a-c potential on the screen grid, a capacitor is connected between the screen grid and "ground" or cathode to act as a short circuit to the signal voltages.

Tetrode Characteristics

When the proper voltages are applied to the tetrode, electrons are attracted from the cathode to the plate. The screen grid, being positive with respect to the cathode, also attracts electrons. However, because of the comparatively large space between the screen grid wires, most of the electrons attracted by the screen grid pass through it to the plate. Thus the screen grid produces a strong electrostatic force that attracts electrons from the cathode, leaving the plate with very little electrostatic force on electrons emitted from the cathode. This results in an important effect: as long as the plate voltage is higher than the screen voltage, the plate current depends primarily on the screen grid voltage. As we shall learn, because the plate current in this tube is largely independent of plate voltage, it is possible to obtain much higher amplification with a tetrode than with a triode.

From the plate family of characteristic curves, we note that after an initial early rise, plate current decreases until the plate voltage is equal to the screen grid voltage. Following this, plate current increases sharply, and finally levels off slightly, having a small linear increase with plate voltage.

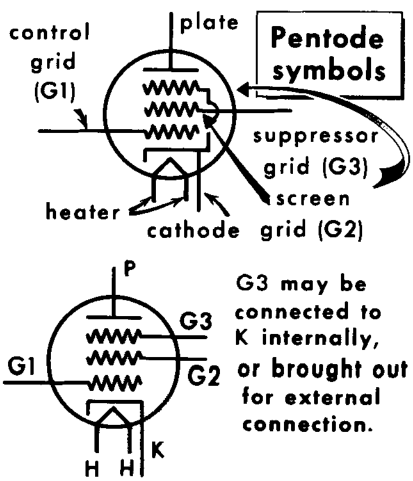
"Secondary emission" effects are brought about when electrons are dislodged from the plate by bombardment from regular cathode emission. Dislodged electrons are then attracted to the screen grid, resulting in a loss of plate current. These effects produce the decrease in plate current at low plate voltages and, in effect, give the tetrode a negative resistance over this range. The amplification factor and plate resistance of a tetrode are considerably higher than that of a triode; transconductance is not too high. Tetrodes are seldom used in radio receivers; they do find use, however, in transmitters.

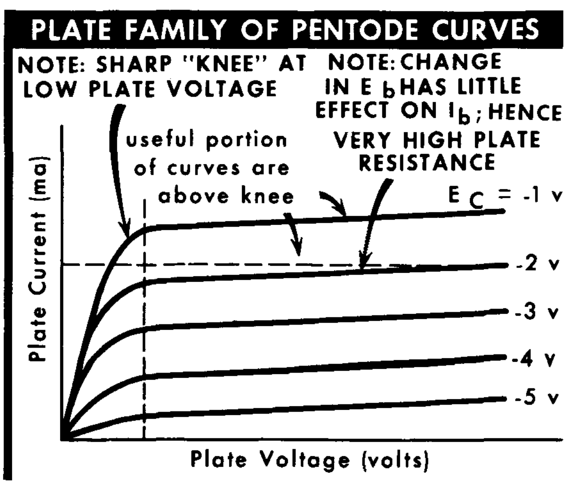


The Pentode

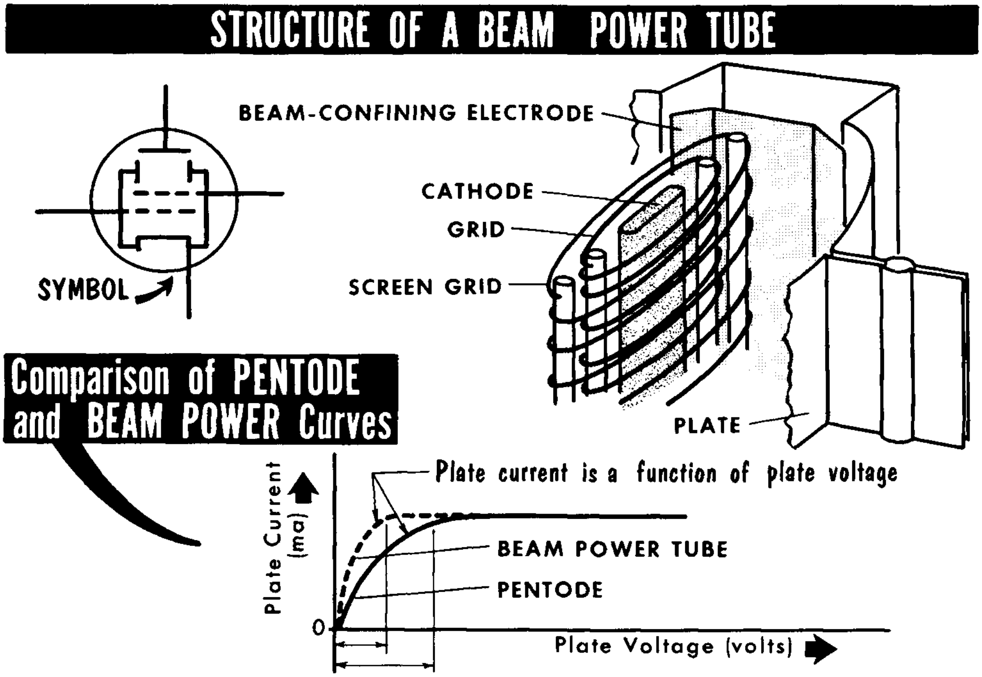
The pentode is a five-element electron tube containing a cathode, a plate, a control grid, a screen grid, and a suppressor grid. The suppressor eliminates or suppresses secondary emission of electrons from the plate, thus removing the major drawback of the tetrodes. In addition, the capacitance between the control grid and plate now consists of three capacitances in series: there is capacitance between control grid and screen grid, screen grid and suppressor grid, and suppressor grid and plate. This reduces the value of the plate-to-control-grid capacitance still further, enabling the pentode to operate at still higher frequencies.

The suppressor grid is a coarse wire mesh placed between the screen grid and plate. It is usually connected to the cathode, which places it at a negative potential with respect to plate and screen grid. Because it is placed beyond the screen grid, its action does not interfere with that of the screen grid in attracting electrons. Moreover, because its wires are relatively widely spaced, it does not interfere with the fast-moving electrons which travel to the plate. When electrons striking the plate cause secondary emission, the negative potential of the suppressor grid repels them back to the positively charged plate, suppressing the secondary emission. Note that for a typical pentode, when the plate voltage is below 100 volts, there is no dip in the curve, which merely shows an increase in plate current with an increase in plate voltage. Above approximately 50 volts on the plate, the curves are relatively flat. This indicates that changes in plate voltage above this value have very little effect on plate current flow. Since the plate current is relatively independent of plate voltage, the two important factors to consider are screen grid voltage and control grid voltage. Screen grid voltage is fixed; hence, the control grid emerges as the major factor in controlling plate current flow. The amplification factor of pentodes may exceed 400, and the plate resistance is often in excess of 1 megohm. Transconductance of pentodes compares favorably with that of triodes and tetrodes, often being in the area of 5000 μmhos.





Beam Power Tubes



In many ways, the beam power tube is a cross between a tetrode and a pentode. It is capable of handling high power levels, an ability obtained from that part of its design which concentrates the plate current electrons into sheets or beams of moving charges. In the beam power tube, the screen and control grids are wound in helical form so that each turn of the screen grid is shaded from the cathode by a turn of the control grid. It is this construction which causes electrons emitted from the cathode to be formed into beams, and reduces the amount of wasteful screen current flow. The beam-forming plates, used to confine the electron beam, are connected internally. Because the screen grid and plate are operated at approximately the same d-c potential, an effect equivalent to a space charge is developed in the space between the screen and plate. This effect is often called a "virtual cathode," and repels secondary electrons dislodged from the plate back to the plate, allowing high plate current efficiency. In some beam power tubes, the beam-forming plates are replaced with a conventional suppressor grid.

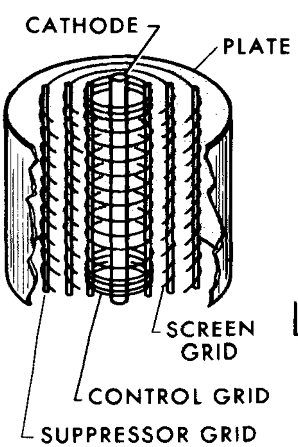
Note that the plate current of the beam power tube rises much more rapidly than that of the pentode. This shows that the region in which the plate current is primarily a function of the plate voltage is much smaller in the beam tube; that is, the plate current becomes independent of plate voltage at much lower values of plate voltage. This characteristic enables the beam power tube to handle much more power at lower values of plate voltage than an ordinary pentode. The beam power tube is extremely popular for the power output stages of radio receivers.

Variable-Mu Tubes

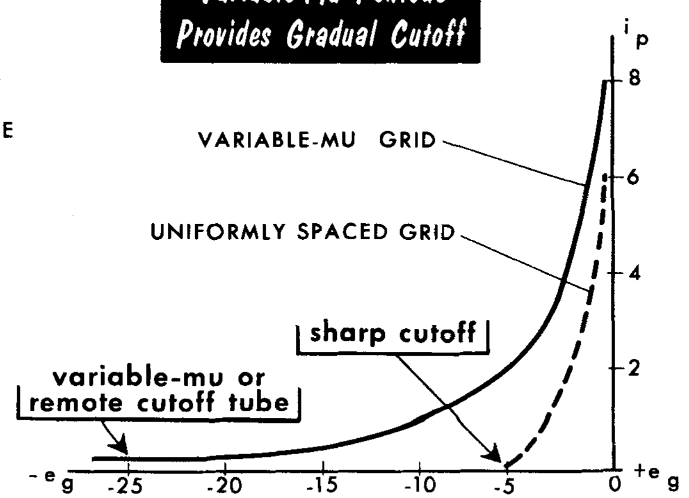
The amplification factor of a tube has been described as being equal to $\Delta e_p / \Delta e_g$, with i_p remaining constant. This characteristic of a tube is largely determined by the geometry of the tube; that is, the shape and placement of the electrodes. Thus, we have gone on the basis that amplification factor is a relatively fixed characteristic of a particular tube. A fixed amplification-factor tube represents a problem when large signals are to be handled, since the grid bias voltage is driven highly negative at times, and the tube goes close to or into cutoff. To minimize distortion in large signal inputs to electron tubes, special kinds of high-amplification tetrodes and pentodes are used. These are known as variable-mu tubes, and they differ from ordinary tubes in the construction of their control grids.

In these tubes, the grid wires are unequally spaced. The turns are closer together at the top and bottom of the winding, and wider at the center. This form of control grid construction produces a tube which does not have a constant gain. Instead, its amplification changes with the value of grid voltage applied to the control grid. At low values of bias, the grid operates in the normal manner. As the control grid is made more negative, the effect of the closely spaced grid wires becomes greater, and the electron flow from the space charge in this region is cut off completely. The center of the grid structure also displays a greater effect, but still allows electrons to advance to the screen grid and plate. The overall reduction in plate current, therefore, is gradual. Eventually, with sufficient negative voltage on the grid, all parts of the grid winding act to cut off the plate current, but the negative grid voltage required to attain this is perhaps three to four times as much as for the conventional tube operated at similar screen and plate voltages. Thus, variable-mu tubes are used where it is desired to control transconductance by varying the control grid potential of the tube.

Grid Structure of Variable-Mu Pentode

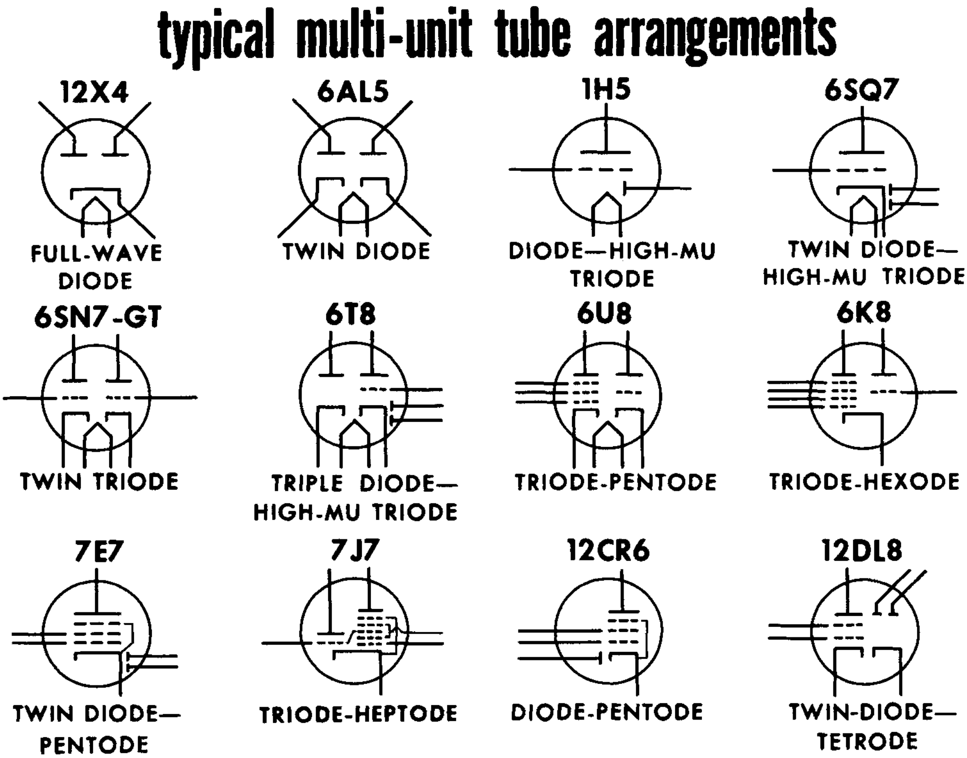


Variable-Mu Pentode Provides Gradual Cutoff



$i_p - e_g$ CURVE FOR NORMAL AND VARIABLE-MU TUBES

Multigrid and Multi-unit Tubes



Our discussion of tubes so far has been related to "conventional" types such as the diode, triode, tetrode, and pentode. These tube constructions serve virtually all needs in electron tube receiver circuitry. However, one more must be considered for completeness – the heptode (7-elements), or pentagrid tube. This tube contains a plate, five grids, and a cathode (as in the 6BY6), or a directly heated filament (as in the 1R5). A tube of this type will very often have more than one signal input applied to it. (The same is true of the six-grid 7A8.)

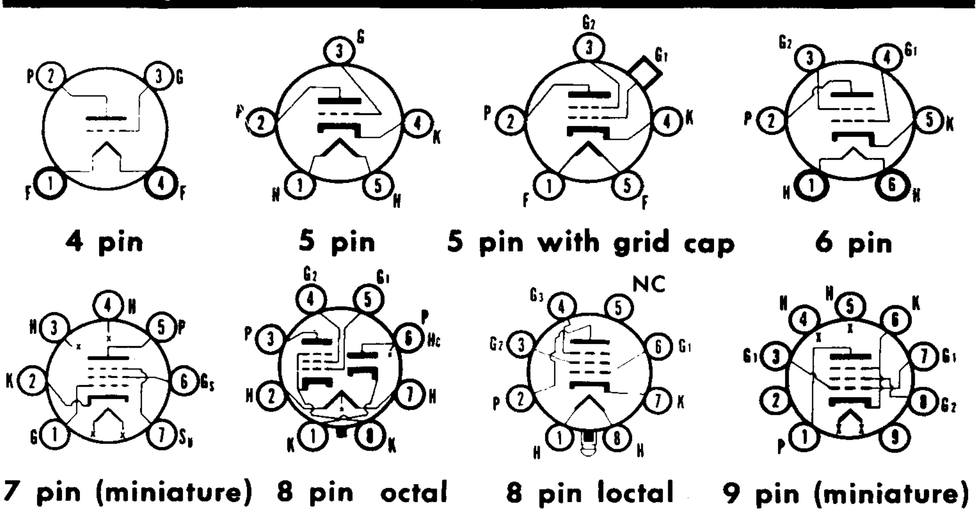
Recent years have seen the combining of tube units or sections in one tube envelope. This permits circuit economy and compactness, since two or more circuits operate from one tube type. In some instances, the tube units are completely independent, having separate cathodes. In other instances, two tube units may operate from a single (common) cathode. Perhaps the simplest multi-unit tube is the twin diode 6AL5, containing merely two separate diodes. Extremely popular is the twin triode 6SN7. An example of complexity in multi-unit tubes is the 6K8 triode-hexode, which contains a triode and hexode (four grids) in one envelope. When more than one grid is used in a tube, they are usually numbered G1, G2, G3, etc., beginning with the grid closest to the cathode. In addition, tubes are often described by a characteristic or function (low-mu triode, pentagrid converter, etc.).

Vacuum Tube Designation and Basing

Vacuum tubes have undergone several changes in standardizing their type styles. In the early days of radio, tube types such as VT1 and 201A were used. This was followed by a numbering system, such as types 40 or 43. Since the mid-thirties, however, a more formalized system has been developed in which the first number of a tube type indicates the heater or filament voltage. The letters and numbers that follow give an indication of the tube's function and number of useful elements. However, in recent years, the heavy flow of new tube types made much of this designation inexact. For instance, the 35C5 requires a heater voltage of 35 volts; the 6BZ7 requires a heater voltage of 6.3 volts; the 1U4 requires 1.4 volts for its heater; and the 117Z6-GT requires a heater voltage of 117 volts. In all instances, a tube manual should be referred to in order to determine the exact heater voltage and current requirements. Some tube types may use suffixes such as -G, -GT, -GTA, and -GTB to indicate later modifications in the tube structure, but the tube characteristics remain essentially the same.

Tube manuals show the tube basing as it would be seen while holding the tube upside down. All pins are numbered in basing diagrams, with the numbers reading in a clockwise direction. In some instances, special notations are made where necessary. For instance, some tubes have an internal shield; this is indicated because it may be necessary to "ground" the pin to which the shield is connected.

typical basing diagrams and notations



F = filament
G = grid
P = plate
H = heater

H_c = heater centertap
K = cathode
NC = no connection

Electron tubes are classified as diodes, triodes, tetrodes, or pentodes according to the number of elements they contain.

Four important types of electron emission are thermionic, secondary, photoelectric, and cold-cathode.

Oxide-coated materials for cathodes are most commonly used because of their high emission rate.

The important features of a vacuum tube can be seen from its characteristic curves. A number of curves drawn on the same graph are known as a family of curves.

Two types of resistance are present in a diode—d-c and a-c plate resistance.

Static characteristic curves generally show operation with no load or voltage variations in the tube circuit; dynamic characteristic curves generally show operation under load and with voltage variations in the tube circuit.

The purpose of the control grid in triodes is to control the movement of electrons from the space charge to the plate.

Amplification in a triode occurs when a stronger signal is delivered to the output circuit than is received in the input.

An electrostatic field exists between the cathode (space charge) and the control grid in a triode. The direction of this field is such as to pull electrons to the plate when the grid is made positive.

In a triode, the plate is positive with respect to the cathode.

The main advantages of a tetrode over a triode are that it has lower interelectrode capacitance between control grid and plate, and it provides higher amplification.

Plate current saturation is reached when no further increase in plate current occurs as the grid is made more positive.

The main purpose of the screen grid is to reduce interelectrode capacitance between the plate and control grid, thereby preventing self-oscillation.

Pentodes use a suppressor grid to reduce the effects of secondary emission.

Secondary emission occurs when electrons from the space charge strike the plate with sufficient force to dislodge secondary electrons from the plate.

Multi-unit tubes contain more than one set of elements in a single envelope.

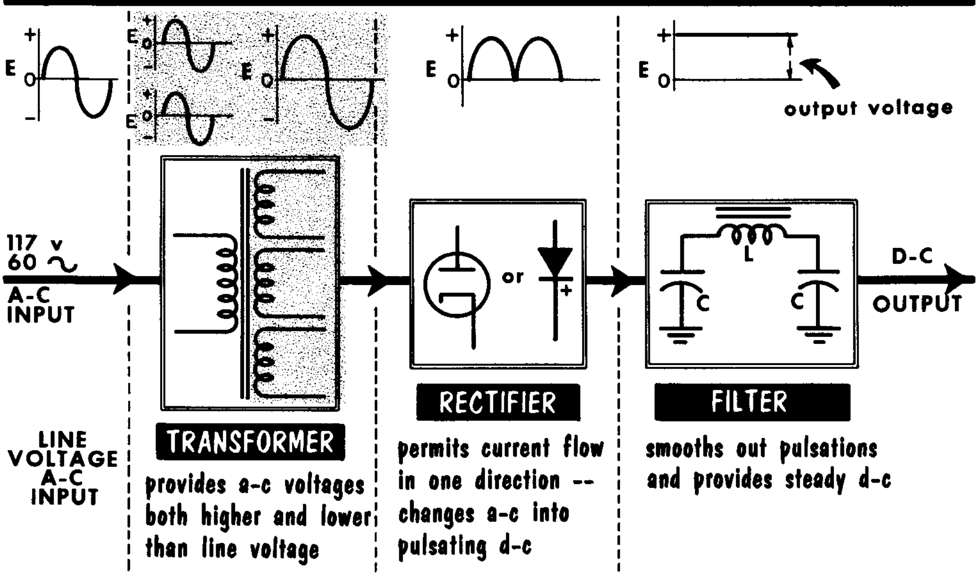
REVIEW QUESTIONS

1. Name and explain the action of four important methods of accomplishing electron emission.
2. What is the difference between directly and indirectly heated cathodes?
3. What is emission saturation, and how can it be overcome?
4. Describe the two methods used to control plate current.
5. Explain the difference between static conditions and dynamic conditions when using graphs to illustrate tube characteristics.
6. Describe the action of a diode in rectifying a sine-wave signal.
7. Explain the following tube constants: (a) amplification factor; (b) plate resistance; (c) transconductance.
8. What is the relationship between the value of the plate load and the tube's plate resistance in (a) triode tubes, (b) pentode tubes?
9. Why is there a minimum limit for the plate voltage of a tetrode?
10. What is the purpose of a suppressor grid in a pentode?
11. How does a beam-power tube eliminate the need for a suppressor grid?
12. What is the phase relationship between grid and plate current signals?

Electronic Power Supplies

For the proper operation of radio circuits, various a-c and d-c voltages must be applied to the electron tubes used. As we have seen, the plates and screens of these tubes require d-c voltages – perhaps as much as 400 volts. The heaters or filaments can generally use a-c or d-c voltages, although, as we shall see, a-c voltages are more practical. These voltages may be as high as 117 volts in radio receivers. In certain circuits, the negative d-c

Basic Arrangement of Transformer-Type Power Supply



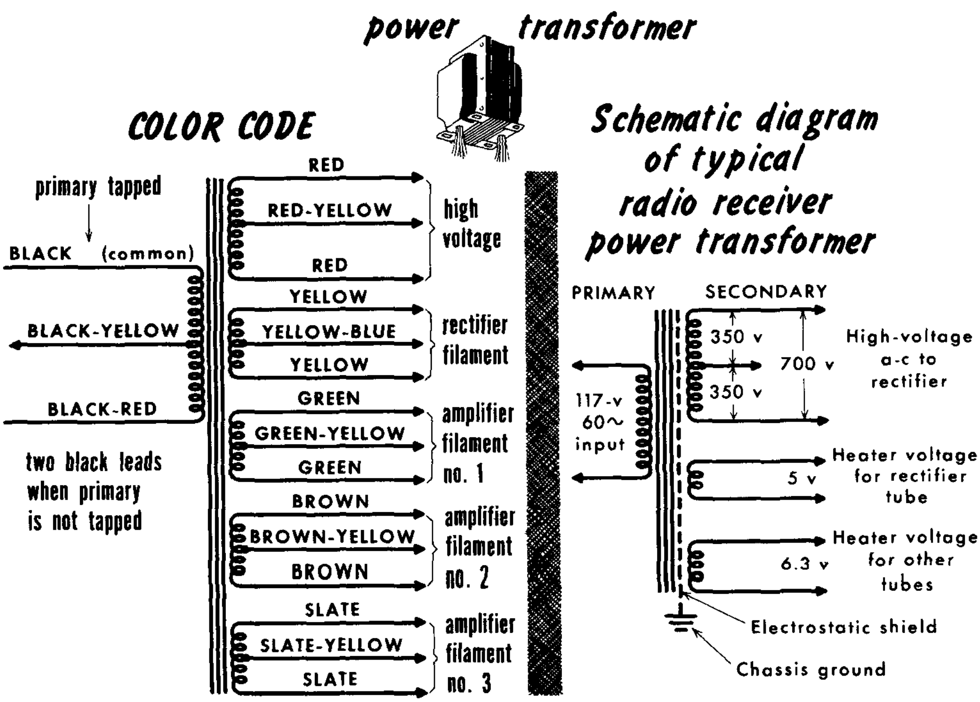
bias applied to the control grids of tubes is supplied by the power supply. In the early days of radio, these voltages were supplied by batteries – an A battery for the filaments, a B battery for the plate and screen voltages, and a C battery for the grid voltage. The modern electronic power supply has eliminated the need for these batteries (except in some portable equipment), and supplies all the necessary voltages required by a radio receiver.

Basically, an electronic transformer-type power supply for radio receivers requires an input of 117-volt 60-cycle a-c, a relatively standard power frequency and voltage. This input is fed into the input of a power transformer which has both step-up and step-down windings. The power transformer supplies a-c voltages to a rectifier – either tube, metallic, or semiconductor. The output of the rectifier feeds into a filter which, in turn, provides a steady d-c voltage output. In addition, various a-c voltages are taken directly from the secondary of the power transformer. The commonly used a-c – d-c or transformerless power supply does not require a power transformer. In normal operation, it is not unusual for the case of a power transformer to run warm, and even slightly hot.

Power Transformers

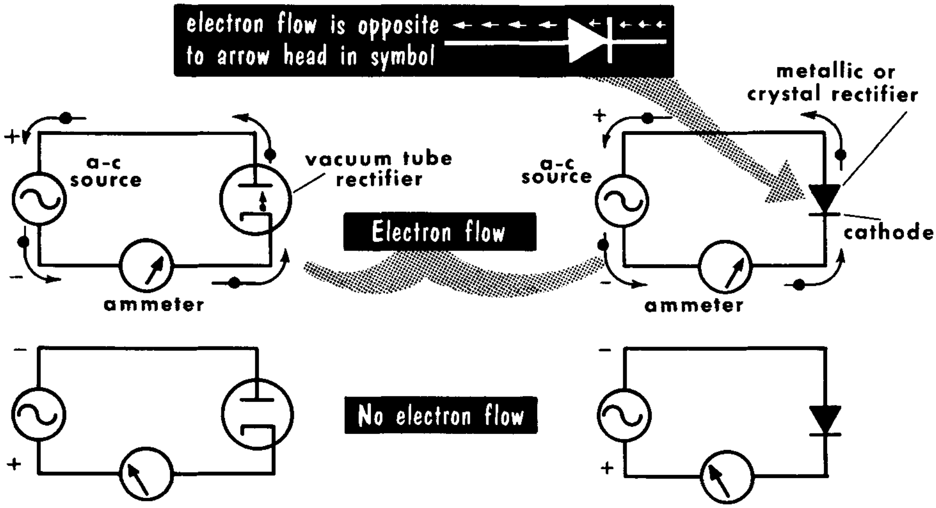
The purpose of the power transformer is to increase or decrease the a-c input voltage to the values required by the rectifiers and tube heaters of the receivers in use. Usually, the power transformer consists of a primary winding, a high-voltage secondary winding, and a number of low-voltage windings which supply power to the various tube filaments. Since most vacuum tube rectifiers, such as the 5U4, 5Y3, and 5V3, require 5-volt filament voltage, one secondary on most power transformers is rated at 5 volts and 2 or 3 amperes. The ratings of the other filament windings are determined by the number and type of tubes to be heated by this transformer. The most popular heater voltage is 6.3 volts; however, the current rating varies considerably, with popular tube current ratings of 225 ma, 300 ma, and 450 ma. The various heater voltages are obtained through step-down windings from the conventional 117-volt a-c input primary winding.

The high-voltage winding (usually centertapped) is a step-up winding, and commonly provides a-c voltages as high as 800 volts. Many power transformers have an electrostatic shield between the primary and secondary which is bonded to the transformer case; this in turn, is grounded to the receiver chassis. The electrostatic shield prevents high-frequency disturbances in the power line from being fed into the power supply. A current rating is commonly given for the high-voltage secondary winding. For example, a typical power transformer specification might read: "700 volts centertapped at 200 ma, 5 volts at 3 amps, 6.3 volts at 5 amps." Modern power transformers are mounted in steel cases and impregnated.



Rectification

RECTIFIERS PROVIDE A LOW-RESISTANCE PATH TO ELECTRON FLOW IN ONE DIRECTION AND A HIGH-RESISTANCE PATH IN THE OTHER DIRECTION



The function of a radio receiver power supply is to provide d-c and a-c voltages for the operation of various circuits. The a-c voltages are obtained directly from the power transformer, which steps up or steps down the line voltage as required by the circuits. The d-c voltages, however, are produced only through a process of rectification and filtering. A rectifier is a device that permits electron flow in one direction and not in another. Since the beginning of radio, the most popular rectifier has been the diode vacuum tube.

The diode vacuum tube makes an excellent rectifier. It permits electron flow from cathode to plate when the plate is positive with respect to cathode and does not permit electron flow when the plate is negative with respect to cathode. Thus, when an a-c voltage is applied across the plate and cathode of a diode vacuum tube, current flows through the tube only during one-half of the a-c cycle. Although this current flow has the waveform of the positive alternation of the a-c voltage, it nevertheless is now direct current – it flows in one direction only. As we shall learn later, the filter smooths these fluctuations into a steady d-c power supply output.

There are other types of rectifiers also. Recent years have seen the growth in popularity of the selenium rectifier, a metallic device that has a high ratio of forward-to-back resistance. That is, it offers little resistance to current flow in the opposite direction. Still more recent is the use of germanium and silicon crystal rectifiers. The use of a single rectifier permits rectification of half the input cycle. By using more than one rectifier, it is possible to make use of both halves of the input cycle for conversion of a-c to d-c.

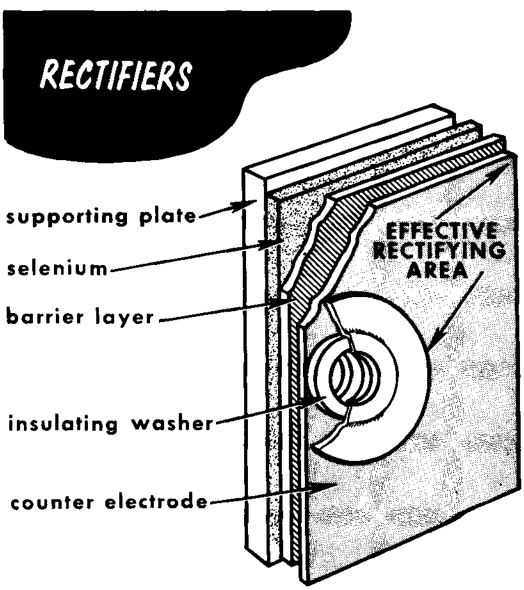
Metallic and Semiconductor Rectifiers

The selemium rectifier is a metallic (dry-disc) rectifier. The basic selenium "cell" is limited in the value of voltage and current it can handle. To overcome this limitation, selenium rectifiers consist of stacked cells. Stacking the cells in series increases the value of voltage that can be applied to them; stacking the cells in parallel increases the value of current that can flow from them.

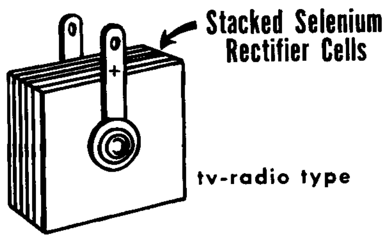
The basic selenium cell consists of a microscopically thin layer of crystalline selenium between two conductors. The supporting plate has the layer of selenium deposited on it, which covers all but the inner and outer edges. The selenium-coated plate is heat treated to form a thin barrier layer. An insulating washer at the center prevents the counter electrode from short-circuiting to the supporting plates.

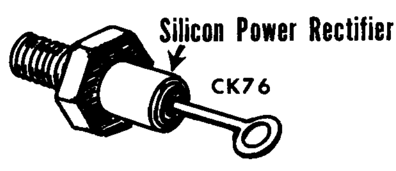
The action of the barrier layer and selenium is to provide a low-resistance path which readily allows electron flow when the selenium-coated supporting plate (anode) is made positive and the barrier coating of the counter electrode (cathode) is made negative. Reversing the polarities will produce an extremely high-resistance path, providing virtually an open circuit and no current flow. This action is identical to that discussed for diode vacuum tubes.

In addition to selenium rectifiers, the semiconductor action of germanium and silicon (explained in Volume 5) provides diodes capable of handling the voltages and currents used in radio circuits. Practical selenium rectifier stacks can handle applied voltages up to 400 volts (rms), with current ratings as high as 500 ma (or 0. 5 ampere).



Sectional view of a Selenium Rectifier Cell





Symbol for Selenium, Silicon, and Germanium Rectifiers

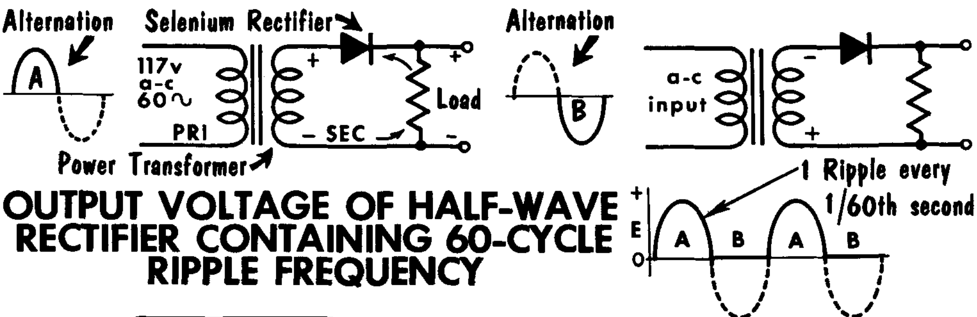
The Half-Wave Rectifier

The basic circuit to convert a-c to d-c is the half-wave rectifier. The output of the transformer secondary is a sine wave voltage, applied to a load in series with a rectifier. During alternation A, which makes the plate positive, the diode conducts. Current flowing through the load resistor develops an IR or voltage drop which represents the output voltage. During alternation B, which makes the plate negative with respect to the cathode, the diode does not conduct.

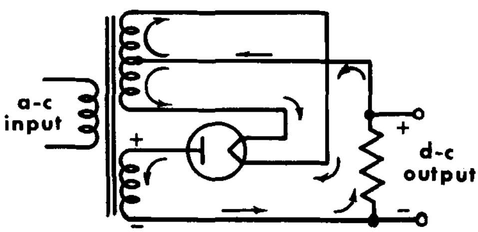
The output voltage taken across the load resistor is a pulsating wave of one polarity only. It is called pulsating d-c. The output voltage pulses once for each cycle. The pulses cause the d-c output voltage to be rippled once each cycle, producing a 60-cycle ripple frequency.

Different rectifiers are designed to handle various values of voltages and current. These values determine their ratings. The alternating voltage rating per plate (rms) is the highest value a-c voltage that can be applied between anode and cathode. The peak inverse voltage rating is the maximum voltage that can be applied between anode and cathode when the rectifier is not conducting. The peak inverse voltage of the half-wave rectifier is equal to the peak value across the transformer secondary.

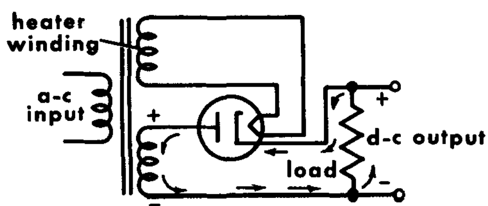
The peak plate current rating in a vacuum tube rectifier represents the maximum value of plate current the cathode can supply. It is an instantaneous value and cannot be handled for any length of time. The load current or output current rating is that value of current which the rectifier can deliver to a load during continuous service.



OUTPUT VOLTAGE OF HALF-WAVE RECTIFIER CONTAINING 60-CYCLE RIPPLE FREQUENCY



VACUUM TUBE RECTIFIER USING DIRECTLY HEATED CATHODE

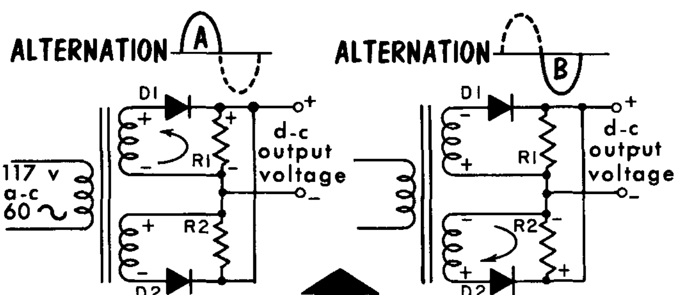


VACUUM TUBE RECTIFIER USING INDIRECTLY HEATED CATHODE

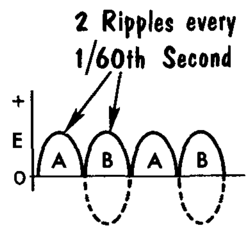
The Full-Wave Rectifier

A full-wave rectifier basically contains two half-wave rectifiers. During alternation A, which places a positive voltage at the anode of diode D1, it conducts. Current flowing through load resistor R1 develops an IR drop which is the d-c output voltage. Diode D2, having a negative voltage at its anode, does not conduct. During alternation B, which places a positive voltage at the anode of diode D2, it conducts. Current flowing through load resistor R2 develops an IR drop which represents the d-c output voltage. Diode D1, having negative voltage at its anode, does not conduct. The output voltage contains two pulses for each cycle, producing a 120-cycle ripple frequency. The average output voltage is higher than that of a half-wave rectifier because the applied a-c voltage is rectified and used for both alternations.

In a practical full-wave rectifier, a single centertapped secondary winding of the transformer replaces the two individual windings. The two load resistors are replaced with one common load resistor. To obtain the same value of output voltage in a full-wave rectifier, each half of the transformer secondary winding has the same value as the single winding used with the half-wave rectifier. The peak inverse voltage rating of a full-wave rectifier is equal to the peak voltage across the entire secondary winding. Thus, it equals twice that of the half-wave rectifiers.



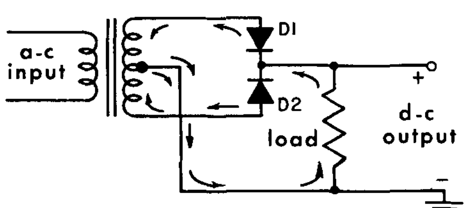
Two Half-Wave Rectifiers combined to form a Full-Wave Rectifier



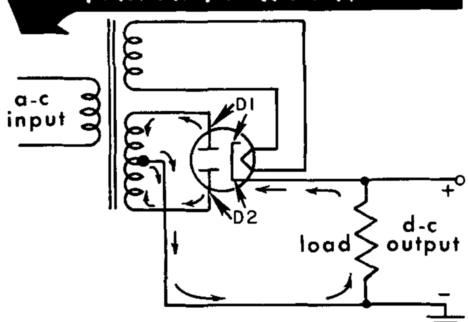
OUTPUT VOLTAGE OF A FULL-WAVE RECTIFIER CONTAINING 120-CYCLE RIPPLE FREQUENCY

Practical Full-Wave Rectifiers

Diagrams show current flow during entire cycle. Only 1/2 transformer secondary carries current at any given time.



USING A CENTERTAPPED SECONDARY WINDING AND COMMON LOAD RESISTOR

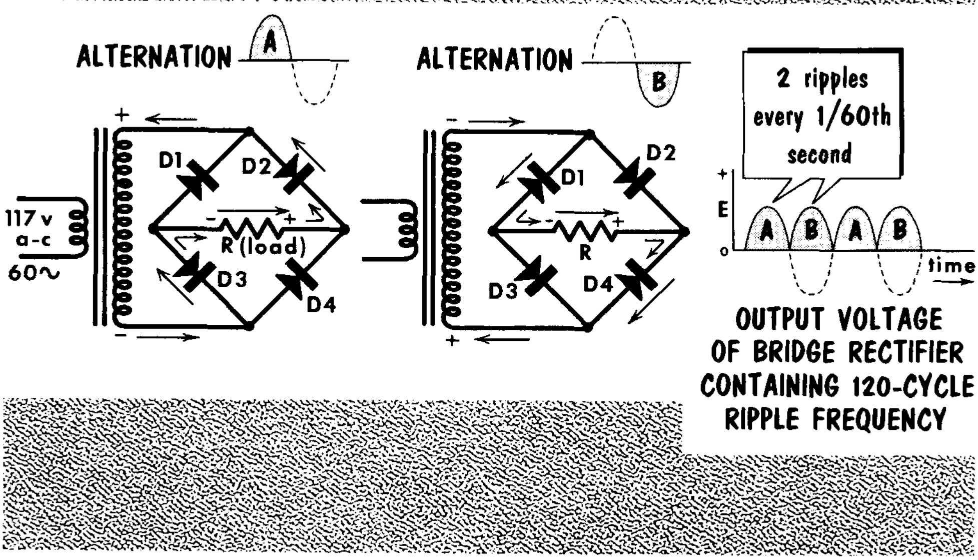


USING A DUO-DIODE VACUUM TUBE RECTIFIER

The Bridge Rectifier

Basically, a bridge rectifier contains two full-wave rectifiers. During alternation A, which places a positive voltage at the anode of diode D2 and a negative voltage at the cathode of D3, they conduct. Current flowing through the load resistor develops an IR drop which represents the d-c output voltage. Diode D1, having a positive voltage at its cathode, and diode D4, having a negative voltage at its anode, do not conduct. Alternation B, which places a positive voltage at the anode of diode D4 and a negative voltage at the cathode of D1, causes them to conduct. Current flowing through the load resistor develops an IR drop which represents the d-c output voltage. Diode D3, having positive voltage at its cathode, and diode D2, having negative voltage at its anode, do not conduct.

TWO FULL-WAVE RECTIFIERS COMBINED TO FORM A BRIDGE RECTIFIER

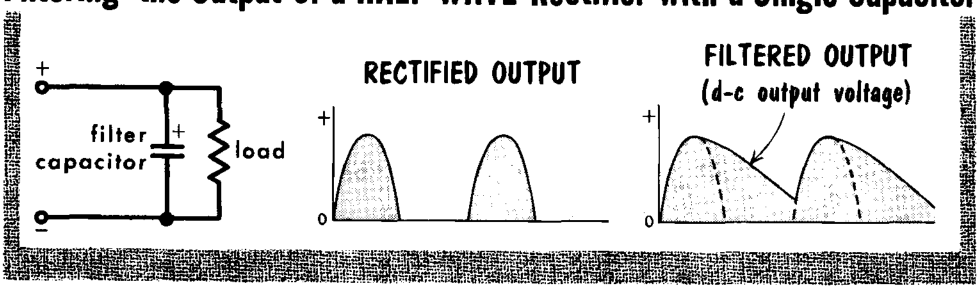


The output voltage, as in the full-wave rectifier, contains two pulses for each cycle, producing a 120-cycle ripple. A bridge rectifier requires only one untapped transformer secondary winding to provide the same value output voltage as the full-wave rectifier. The peak inverse voltage of a bridge rectifier is divided by two rectifiers; thus, it is equal to one-half that of the full-wave rectifier. Bridge rectifiers are used more in transmitter power supplies than in receivers, and will be studied in greater detail later. However, this circuit occasionally is found in receiver power supplies, and it is quite popular in instrument-type a-c rectifiers.

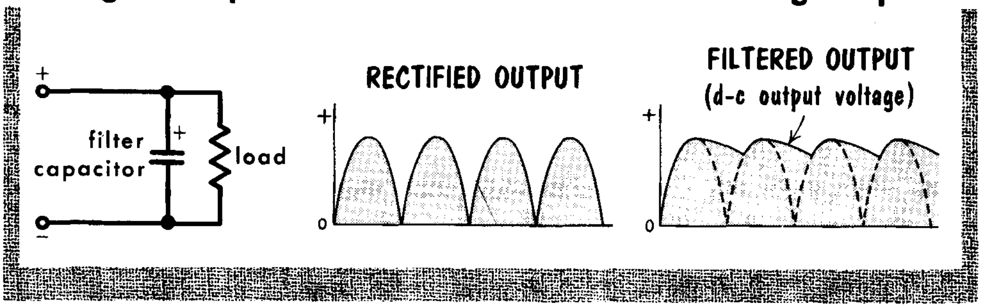
Capacitance Filters

The output of a rectifier (half-wave or full-wave) is not a pure d-c voltage, but a pulsating d-c voltage containing a ripple component. Filter circuits remove the ripple and smooth the voltage to obtain pure d-c. The simplest type of filter is the shunt capacitor connected across the load and rectifier output. We recall that a capacitor opposes any change in voltage across its terminals (in this case, the load) by storing up energy in its electrostatic field whenever the voltage tends to rise, and converting this stored energy back into current flow whenever the voltage across its terminals tends to fall.

Filtering the Output of a HALF-WAVE Rectifier with a Single Capacitor



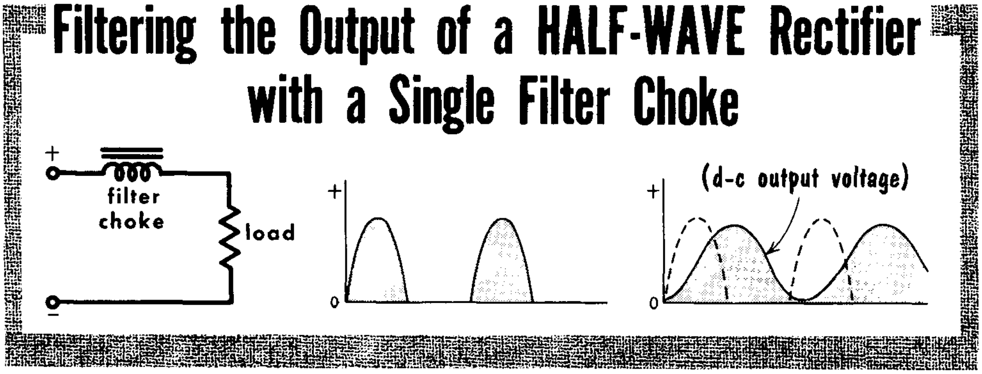
Filtering the Output of a FULL-WAVE Rectifier with a Single Capacitor

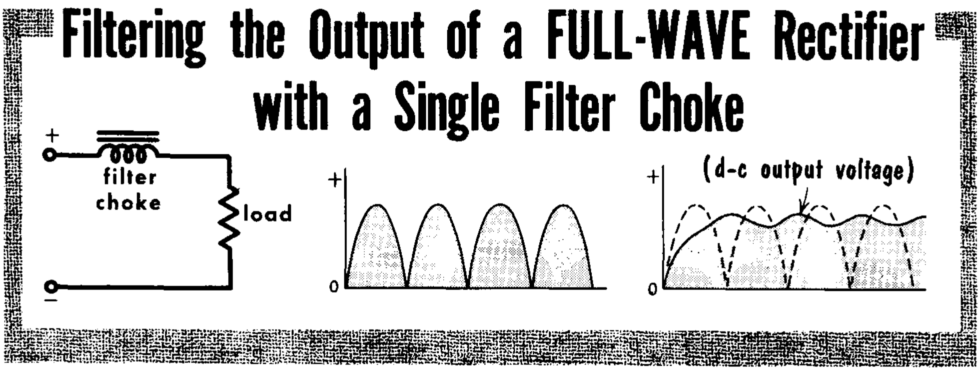


The illustrations show the action of the capacitor filter when either half-wave of full-wave rectifier output voltages are applied to the filter. In both cases, the capacitor charges up to the peak voltage of the rectifier output during the time that current pulses are delivered to the filter and load. When the rectifier output drops off to zero, the capacitor cannot discharge instantaneously; its voltage falls off slowly as it discharges through the load. During the next rectifier current pulse, the capacitor is charged again to the peak voltage, and the cycle is repeated. The only difference between the half-wave and full-wave action is that the capacitor discharges more between current pulses of the half-wave rectifier. Thus, the d-c output voltage of the latter averages less than that of the full-wave rectifier, which almost equals the peak voltage. Because of its poor regulation (large output voltage drops for small load current increases), the simple capacitor filter is not used with rectifiers which supply a large load current.

Inductance Filters

The action of a single inductor (choke coil), placed in series with the rectifier output (either half-wave or full-wave), is shown in the illustration. This filter is usually used in combination with shunt capacitors, as we shall see later. Essentially, any inductor opposes a change in the amount of current flowing through it by storing up energy in its magnetic field when the current





tends to increase, and by taking energy away from the field to maintain current flow when the voltage across the inductor tends to decrease. Hence, by placing a choke coil in series with the rectifier and load, changes in the amount of rectifier output current and output load voltage are minimized. Another way of examining this action of the series inductor is to consider that the coil offers a very low resistance (that of the winding alone) to the passage of d-c, while at the same time offering a high impedance to the passage of fluctuating or alternating currents. The d-c passes through, and the ripple is largely reduced.

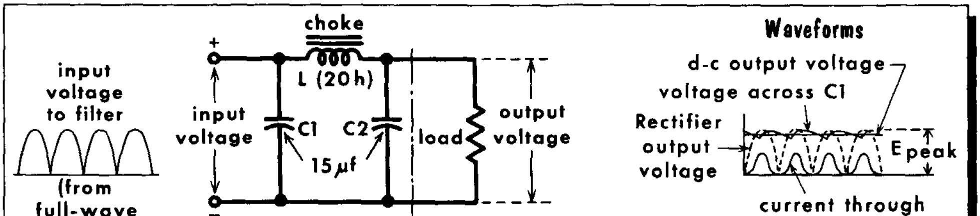
As seen in the output waveforms, the load current through the filter (and, hence, the load voltage) lags 90° behind the rectifier output voltage, and never reaches its peak value. The average d-c output voltage from a single inductor is less than that from the capacitor filter, but the voltage does not fall off as rapidly with increases in load current as with a capacitor filter.

The Capacitor-Input Filter

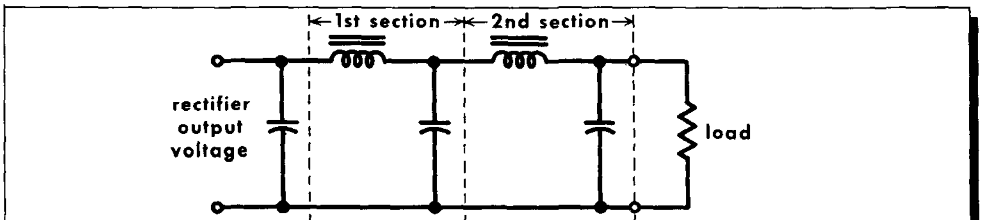
Very effective filtering action can be obtained by combining the actions of a shunt capacitor and a series inductor in a capacitance-inductance filter. Various combinations of such filters exist, and in all of them, the action of the capacitors is to resist changes in output voltage by charging and discharging, as required, while the inductors oppose any changes in the amount of the load current by the action of their associated magnetic field. Equivalently, the shunt capacitors may be thought of as forming a low-impedance path from B+ to B- for the ripple voltage, and an essentially infinite impedance to d-c. The series inductors (choke coils) can be considered as offering a low impedance to the passage of d-c and a very high impedance to the ripple current. The typical capacitor-input filter shown in (A) is also called a pi filter, because of its schematic arrangement. An important characteristic of this circuit is its high output voltage at low current drain because the voltage is almost equal to the peak value of the transformer secondary voltage applied to each tube. As the load current increases, the output voltage falls off rapidly and may become less than the effective value of the applied a-c voltage. In other words, we say that the regulation of the capacitor-input filter is poor.

In the circuit shown, the current pulses flowing through the rectifier tubes and filter charge input capacitor C1 up to the peak voltage of the transformer secondary. Between current pulses, the voltage across C1 falls off somewhat, but never reaches zero. With the additional filtering provided by the coil and capacitor C2, the d-c output voltage becomes essentially constant. The ripple voltage may be decreased further by increasing the value of C1, or by adding another filter section as in (B). In general, capacitor-input filters are used when low d-c power is desired, as in radio receivers.

The CAPACITOR-INPUT Filter



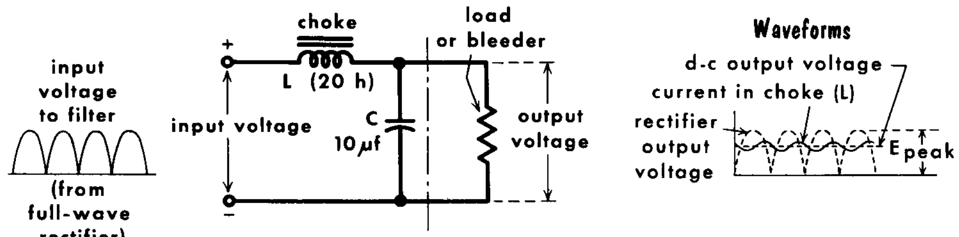
(A) Single-Section Capacitor-Input Filter



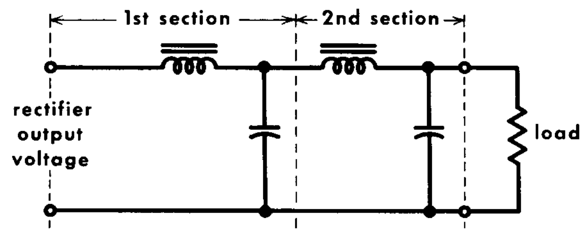
(B) Two-Section Capacitor-Input Filter

Choke-Input Filters

The CHOKE-INPUT Filter



(A) Single-Section Choke-Input Filter



(B) Two-Section Choke-Input Filter

A typical choke-input filter, consisting of a 20-henry series inductor and a 10-microfarad shunt capacitor is shown in (A). When the load draws no current, the d-c output voltage of the choke-input filter is nearly equal to the peak value of the a-c voltage applied to the rectifier, just as in the case of a capacitor-input filter. This is so because, in the absence of a load current, no voltage drop is developed across choke coil L; therefore, output capacitor C charges up to the peak value. However, if even small load current is drawn, the d-c output voltage drops rapidly to some lower value and then remains fairly constant over a wide range of load current values. The initial sharp drop occurs because the series inductor prevents the capacitor from charging to the peak voltage when a load current is drawn. After this initial drop, there is good voltage regulation of the choke-input filter.

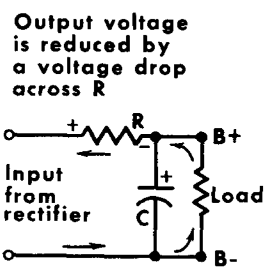
The d-c output voltage across C and the load is fairly constant, as shown in the diagram, and its value is somewhat less than the peak a-c voltage, depending upon the load current drawn. The ripple in the d-c load current through the choke can be reduced considerably by increasing the value of the inductance. In some instances, a "swinging choke" is used that varies its inductance according to the load. The inductance of a swinging choke is high at low load currents, and drops off with increasing load currents. As with capacitor-input filters, improved filtering action may be obtained by using a two-section filter shown in (B). Choke-input filters are used whenever the amount of d-c power required is large.

R-C Filters and Bleeder Resistors

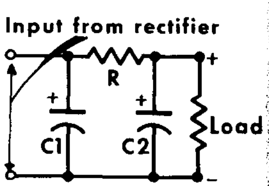
When load current requirements are small and a small d-c voltage drop across the filter is permissible, the inductance of a capacitor-input filter may be replaced with a series resistance. The resulting R-C filter is not as effective as an inductive filter, since the series resistor offers as much impedance to d-c as to the ripple voltage. The advantage of the R-C filter is that a resistance is much less expensive than an inductance. In a typical R-C filter, C1 is made sufficiently large to present a very low impedance to the ripple frequency, while at the same time offering practically infinite impedance to d-c. The ripple voltage therefore prefers the shunt path through C1, and the d-c is forced through R, developing a voltage drop across it. Most of the remaining ripple is shunted through C2.

The output voltage of a power supply is often developed across a bleeder resistor. The idea is to achieve better voltage regulation – that is, to prevent changes in current drain in the receiver from changing the power supply output voltage. The bleeder current is a steady continuous drain, lowering the amount of change and providing a steadying effect on the power supply current drain. A bleeder can also be used as a voltage divider by tapping the resistor at different points to provide voltages of different values. Each tap should have a bypass capacitor from the tap to common to prevent interactions between circuits fed by each tap. The bleeder also acts as a safeguard when the receiver is turned off by dissipating the charge stored in the filter capacitors. In addition, when the receiver is turned on, the heaters of the tubes do not warm up immediately, and the circuits draw very little current. The voltage of the power supply may rise to abnormally high values under such no-load conditions, causing component breakdowns. The constant bleeder load prevents this.

R-C FILTERS

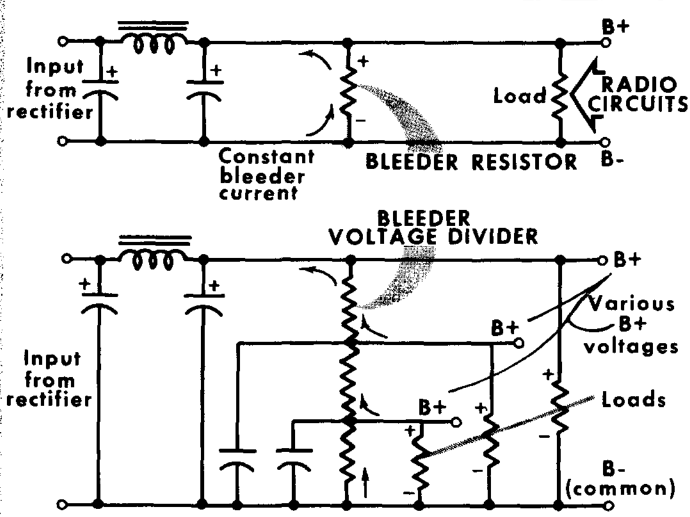


Basic R-C Filter



Capacitor-Input Filter

USING BLEEDER RESISTORS



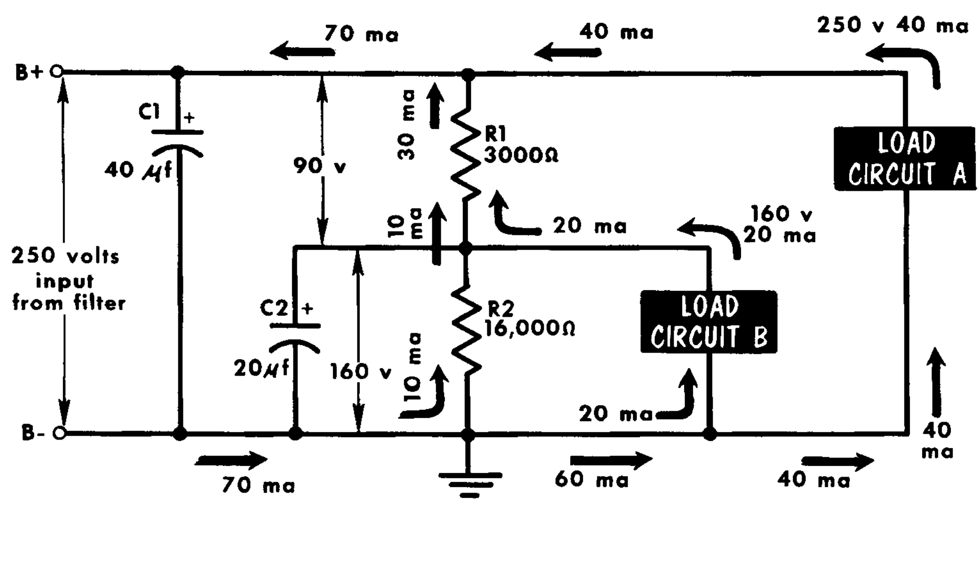
Tapped Bleeder Resistor Provides Various B+ Voltages

Typical Voltage Divider

A voltage divider connected across the output of a power supply and tapped at a number of points, can provide a selection of different values of output voltage. In the circuit shown, the total output voltage available from the power supply is 250 volts. The maximum output current is 70 ma. Capacitor C1 is the output filter unit, and C2, usually of lower value, provides extra filtering across R2. Load circuit A requires the full 250 volts and draws 40 ma; load circuit B requires only 160 volts and draws 20 ma. Thus, circuits A and B require a total of 60 ma for proper operation. Since the voltage divider or bleeder current should be approximately 10% of the total current, we shall select 10 ma (for convenience) as the bleeder current. Since only 10 ma is to flow through R2, and the voltage required across circuit B is 160 volts, R2 is equal to E/I, or 160/0.01 = 16,000 ohms. The voltage across R1 must then be 250 volts minus 160 volts, or 90 volts. We know the current through R1 must equal the 10-ma bleeder current plus the 20-ma load current from circuit B, or a total of 30 ma. With this information, we find that the resistance of R1 = E/I, or 90/0.03 = 3000 ohms.

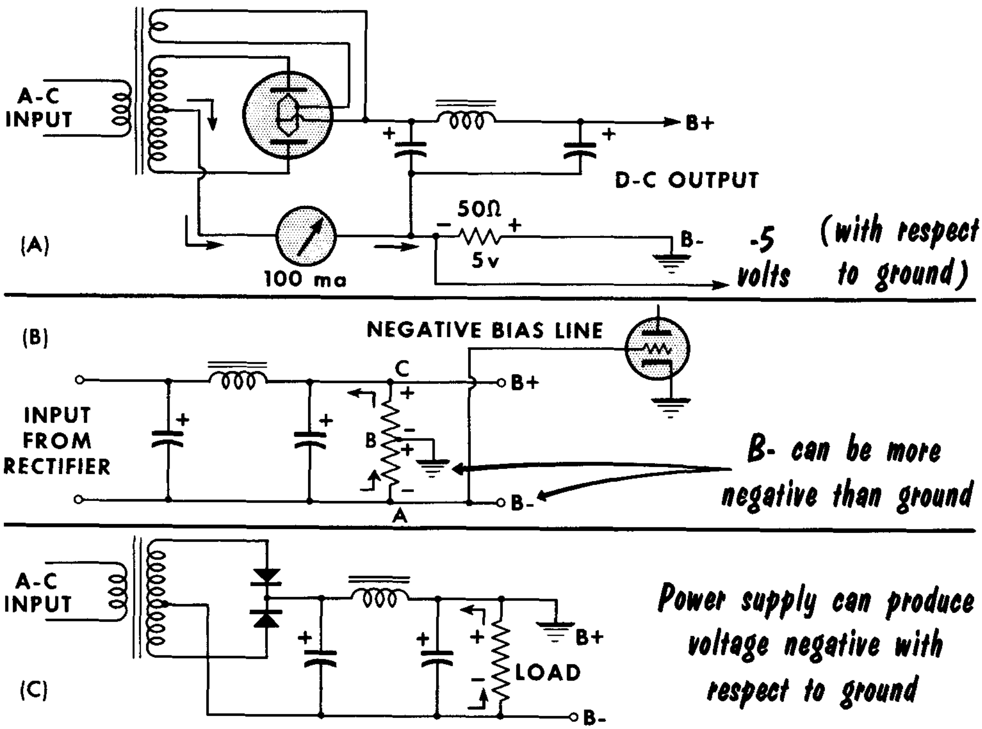
The total resistance of the voltage divider will then be R1 + R2, or 19,000 ohms. With no loads connected across the voltage divider, the bleeder current through it would be a steady value determined by the resistance of the circuit and the voltage across it. Thus, under no-load conditions, the bleeder current would be I = E/R = 250/19,000 = 13.2 ma. The power dissipated by R1 would equal E x I, or 90 x 0.03 = 2.7 watts. To avoid overheating, a resistor having about <u>twice</u> the power rating should be used. In this case, a 5-watt resistor would suffice for R1. Since R2 passes 13.2 ma under no-load conditions, its power rating equals $P = I^2R = 0.0132^2 \times 16,000 = 2.79$ watts. Once again, a 5-watt resistor can be used for R2.

VOLTAGE DIVIDER USED AS BLEEDER ACROSS POWER SUPPLY OUTPUT



Negative Voltage from Power Supplies

So far, in discussing electronic power supplies, we have considered the output to be some voltage that is positive with respect to chassis or ground. In certain instances, it is desirable to have a voltage that is negative with respect to ground, such as when a large negative grid bias is required. In dia-



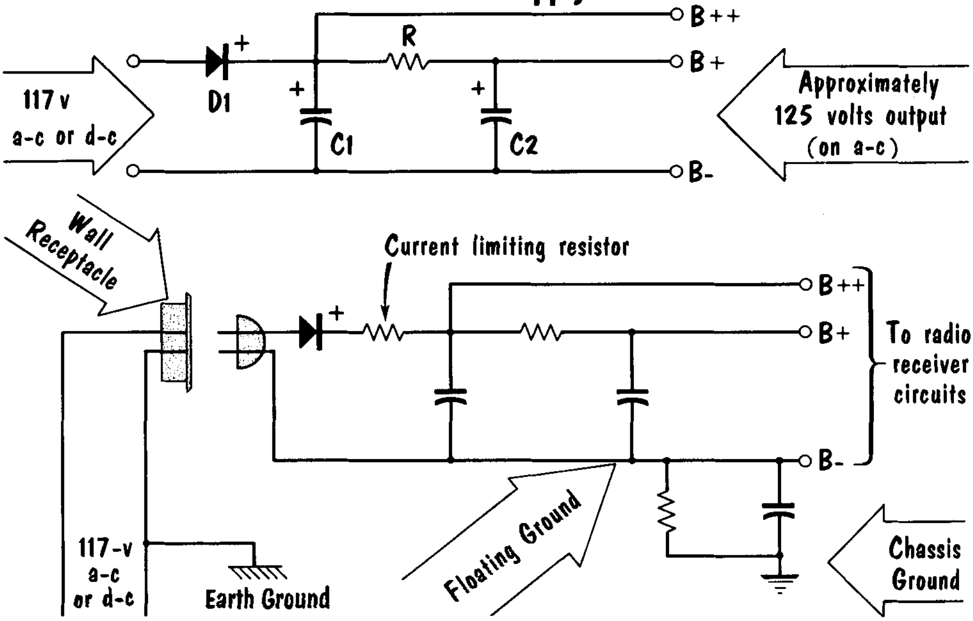
gram A, we see how a negative voltage may be obtained by placing a resistor in series with the centertap lead of a full-wave rectifier. The current passing through the resistor produces a voltage drop that makes the centertap side negative with respect to ground. This voltage reduces the total amount of B voltage available.

In diagram B, we place a ground tap on the voltage divider or bleeder resistor. Point B, or ground, is positive with respect to A, but negative with respect to C. Thus, in this circuit, B- is actually more negative than ground with respect to B+. The voltage developed between A and B can be applied to a tube as bias voltage.

Diagram C shows how a conventional power supply can be made to provide a negative voltage. By grounding the positive side of the load, all other voltages are negative with respect to this ground. From this, we can see that any power supply can be made to deliver a negative voltage, a positive voltage, or both.

A-C – D-C Power Supplies

Basic A-C–D-C Power Supply



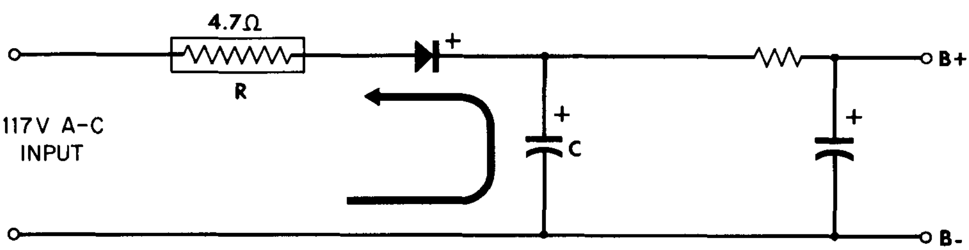
The popularity of table model radio receivers has led to the development of the economical transformerless power supply. It was found that a simple rectifier circuit could be built that would satisfy the requirements of both an a-c and a d-c line voltage input. With an a-c line input, the circuit becomes a simple half-wave rectifier; with a d-c input, the anode of the diode must be connected to the positive side of the line. The diode then acts merely as a conductor, permitting the dc to be applied directly to the circuits. A capacitor-input R-C type filter is generally used to provide maximum output voltage. Very little trouble is encountered with the 60-cycle ripple or "hum," since these receivers have a poor response at this low frequency. The filter capacitors used have a very high capacitance, usually from 20 to 80 μf. The higher B voltage at the input to the filter can be applied to circuits that do not require a completely ripple-free voltage.

An a-c – d-c filter is potentially dangerous. Notice that one side of the power line is connected directly to the power supply. Thus, if the chassis of the receiver were connected to the "hot" or ungrounded side of the power line, and a person were grounded by a damp floor (by contact with a cold water pipe, radiator, etc.), when he touched the chassis he would actually be placing himself across the power line. To prevent a lethal shock, the negative or B- terminal in these power supplies is isolated from the metal chassis, and is called a floating ground. A capacitor is usually placed between the floating ground and chassis ground to prevent hum pickup. It is often paralleled by a high-value resistance to provide a leakage path for static charges.

The Fusible Resistor

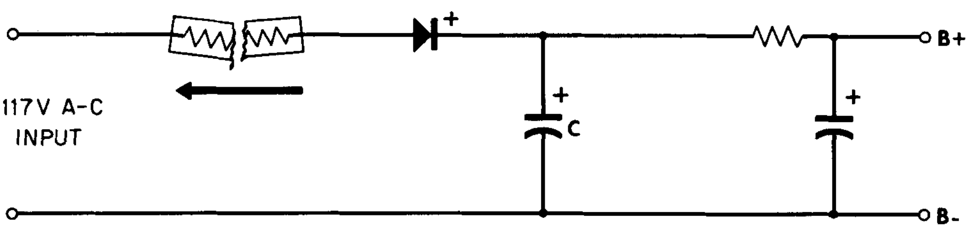
Selenium rectifiers should always be protected from surge overload by a series resistor of low value such as R. An overload may occur as follows: suppose that the a-c cycle is approaching its peak at the instant that the on-off switch is turned on. Also, suppose further, that its polarity is such as to permit current to flow in the forward direction through the selenium rectifier, causing C to charge. Without R in the circuit, the resistance in series with C is very low, so that this capacitor charges almost instantaneously. To bring a capacitor of this size – 100 μf – up to full charge in a very small fraction of a second requires an extremely heavy current, all of which flows through the selenium rectifier. A current of this magnitude flowing through the selenium rectifier even for a short time is very likely to destroy it. The introduction of R into the circuit limits the current to a safe value by increasing the time required to charge the capacitor. Modern receivers use fusible resistors, thus including two protective measures in one component.

THE FUSIBLE RESISTOR



limits charging current through rectifier

and MELTS



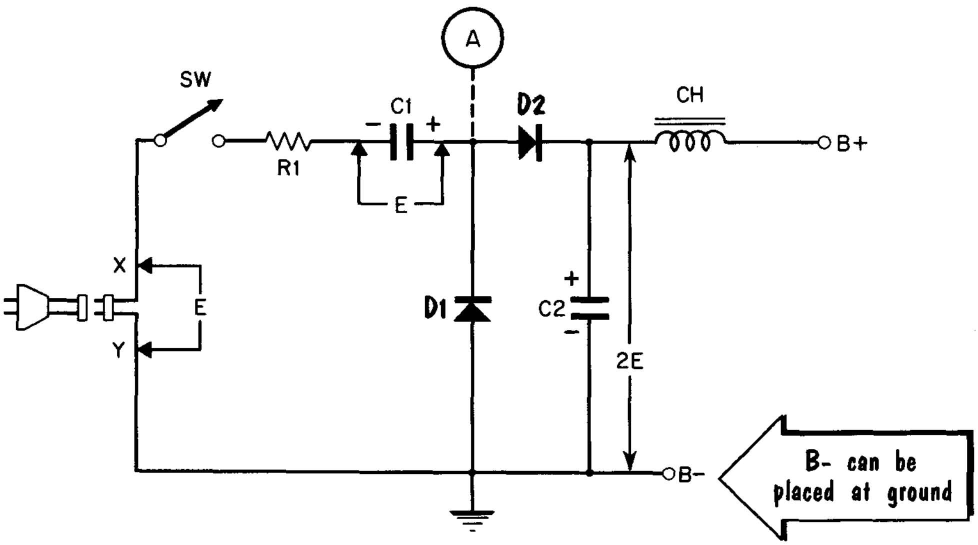
when current through rectifier is excessive.

Half-Wave Voltage Doubler

While radio circuitry can be designed around a wide range of B+ voltages, there are limitations to the quality of sound if the circuits must work from relatively low voltages. Thus, in many receivers, it has become necessary for the circuit designer to make use of a voltage doubler when slightly higher B+ voltages are wanted. A frequently used circuit is shown.

We can see how this circuit operates by assuming that line voltage E is applied to the circuit, with point X being negative with respect to point Y. This

A VOLTAGE DOUBLER SELENIUM RECTIFIER POWER SUPPLY



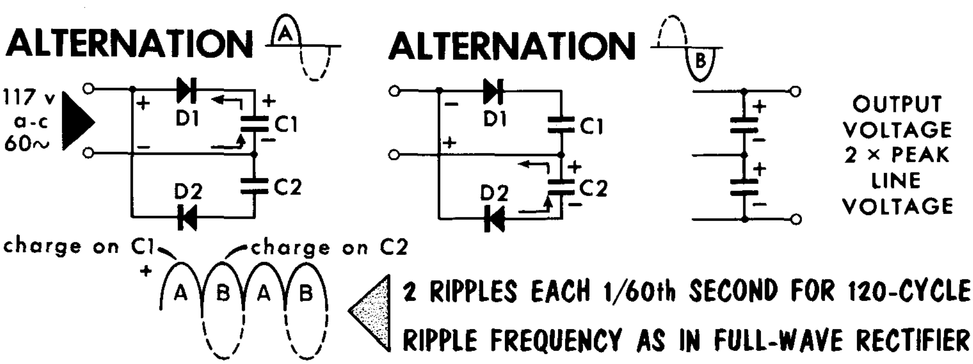
C2 ACTS AS INPUT FILTER TO FORM A CAPACITOR INPUT FILTER

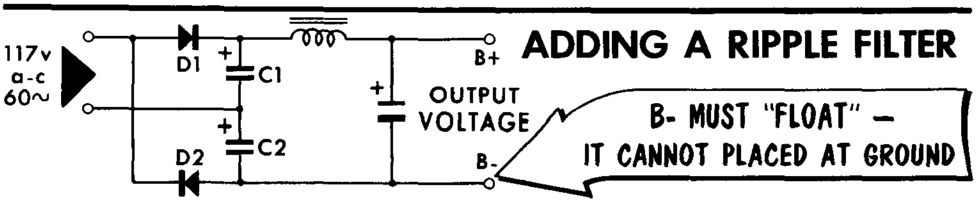
would cause current to flow through the switch, resistor R1, capacitor C1, and selenium rectifier D1, back to point Y. In the process, capacitor C1 would charge approximately to the value of E, as shown. Thus, point A on the diagram would be at E volts with respect to point X. On the next half-cycle, point X would be E volts positive with respect to point Y, or ground. Thus, since point X is E volts positive with respect to ground, and point A is E volts positive with respect to point X, then point A is 2E volts positive with respect to point Y or ground. It is at this time that capacitor C2 can charge up to the full value of point A through selenium rectifier D2. With this voltage doubler arrangement, the difference in potential between B+ and B- is approximately equal to twice that of the input line voltage.

Full-Wave Voltage Doubler

Voltage doublers operate on a-c voltages only. Since they can raise the 117-volt line voltage to approximately 250 volts without the use of a power transformer, they are often referred to as transformerless power supplies. The basic action is to charge two capacitors, each to the peak value of the applied a-c voltage, and to discharge them in series. Assuming a line voltage of 117 volts, the peak value is 1.4 times larger, or 164 volts. With two capacitors in series, each charged to 164 volts, the voltage available becomes 328 volts. With circuit losses, the output voltage averages approximately 250 volts.

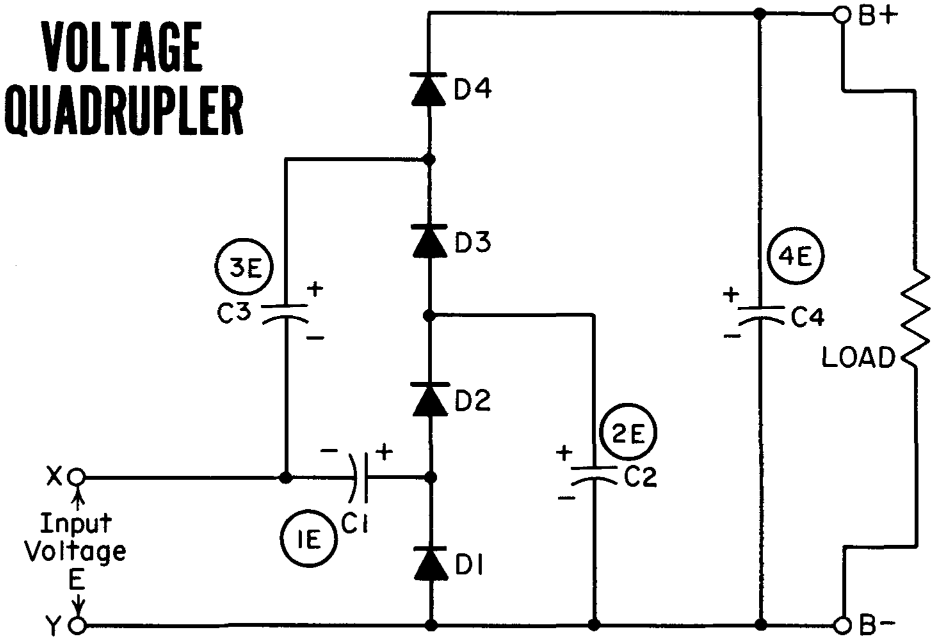
During alternation A, the polarity of the voltage applied to the full-wave douber is such as to have the cathode of diode D2 positive, and the tube does not conduct. The anode of Dl is positive, and it conducts. The current flowing from the negative terminal piles electrons on the negative plate of C1. This drives the electrons off the positive plate of the capacitor through conducting Dl to the positive side of the a-c line. Capacitor C1 charges to the peak value of the applied voltage. On alternation B, the current flows through D2 and charges C2 to the peak value of the applied voltage in the same manner. With C1 and C2 in series, their voltages add. The output voltage will be approximately twice the value of the peak applied voltage. To complete the circuit, additional filters are added. Capacitors C1 and C2 act as the input capacitors, and only the filter choke and output capacitor need to be added. The circuit is called a full-wave doubler. The ripple component is 120 cycles. To ensure that each capacitor receives an equal charge, the capacitors must be alike. A typical value is 20 μf. One side of the 117-volt a-c line is always at ground potential. Because of this, the B- output cannot be placed at ground potential. Doing so will short C2, if the bottom line is grounded, or the entire circuit, if the top line is grounded.





C1 AND C2 ACT AS INPUT FILTER TO FORM A CAPACITOR-INPUT FILTER

Voltage Quadrupler



By combining two half-wave voltage doublers a voltage quadrupler circuit is obtained. During the 1st alternation (X-, Y+), D1 conducts to charge C1 to the peak value of the line voltage (E) with the polarities indicated. On the 2nd alternation the negative side of the line (Y) charges C2 with the polarity shown. Diode D2 conducts to add the charge of C1 to that of C2, charging C2 to twice the peak voltage, or 2E. On the 3rd alternation electrons flow from the negative side of the line (X) to charge C3. The charge on C3 is coupled through D3 to place it in series with the charge on C2, causing C3 to be charged to 3E. On the 4th alternation the negative side of the line (Y) will charge C4, which when coupled through D4 is placed in series with the charge on C3. The total charge on C4 is four times the peak value of the line voltage, 4E.

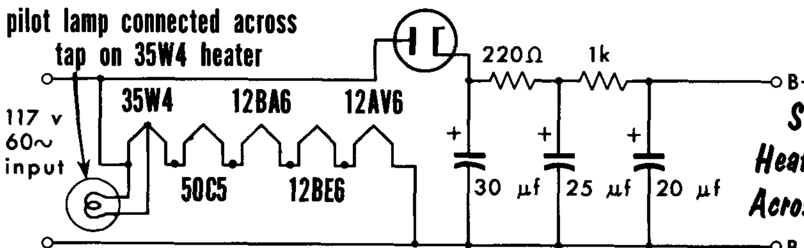
Having reached this steady operating condition, current flow during each half cycle will be only that required to maintain the charge on the capacitors, this in turn is determined by the value of the load. In practice, the quadrupler presents the technical economical limit of voltage multiplication. The regulation becomes progressively poorer and the attained output voltage drops off rapidly with even small load current increases.

The voltage stress across the capacitors becomes increasingly great, requiring costlier units with high-voltage ratings. Also, since the peak current through the rectifier must be limited to the rated values, the possible load current that can be supplied to the tubes becomes less with increasing multiplication. Finally, the cathodes of the tubes must be well isolated from each other, since they are at different potentials with respect to each other. Despite these objections, voltage-multiplying circuits using junction diodes are occasionally used for low-current applications.

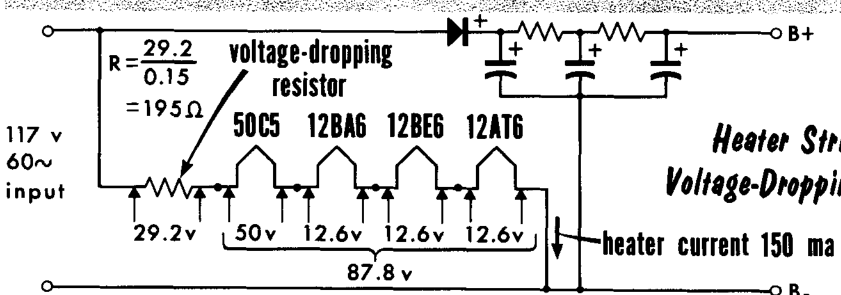
Heater Circuits for A-C – D-C Receivers

Since a-c–d-c receivers operate directly off the 117-volt power line, special consideration must be given the electrical connections of the tube heaters. By far, the most popular arrangement is the "series string," in which all the tube heaters of a receiver have the same current rating, and are connected in series. An ideal situation was reached when the five-tube a-c–d-c receivers were developed using tubes having heater voltages of 50, 35, 12.6, 12.6, and 12.6 volts, respectively. This totalled 122.8 volts. Connected directly across a 117-volt line, this series string operated very well, with the remaining 5.8 volts being distributed across the various heaters. Actually, this is not critical since most power lines vary slightly in voltage. Throughout the United States, line voltages may range from as little as 110 to as much as 125 volts.

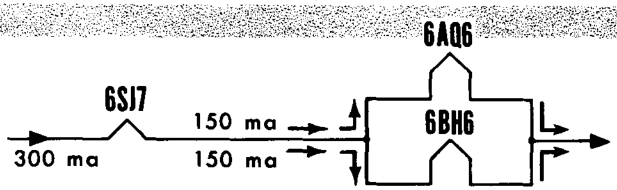
In some series-string heaters, the total voltage required for the heaters may total less than 117 volts. For instance, using a 50C5 (50 volts), 12BA6 (12.6 volts), 12BE6 (12.6 volts), and 12AT6 (12.6 volts) requires a total heater voltage of 87.8 volts. We must then drop 29.2 volts. Using Ohm's law and finding the heater current from a tube manual, we get R = 29.2/0.150 or 195 ohms. In actual practice, a 200-ohm resistor would be used. The power dissipated would equal I^2R, or P = $0.150^2 \times 200$, or 4.5 watts. To protect the receiver completely, a 10-watt resistor should be used. A disadvantage of this heater arrangement is that 4.5 watts are actually being wasted.



Series-String Heaters Connected Across Power Line



Heater String Uses Voltage-Dropping Resistor

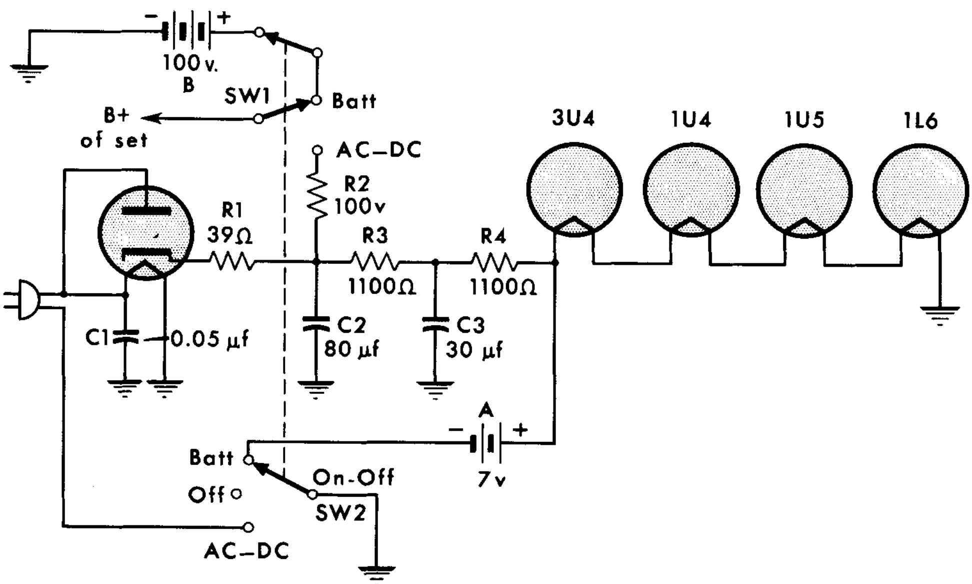


Heaters May Be Paralleled to Conduct Current from Higher-Current Heater

Three-Way Portable Power Supplies

Three-way portable radio receivers are popular because they can be used in the home on a-c or d-c line voltages, and outdoors on batteries. Special tubes using filament-type cathodes have been designed for economical battery operation with 1.4-volt filaments. Often, the power output tube has a filament that is centertapped and can be operated at 2.8 volts in series, or its two sections can be connected in parallel and operated at 1.4 volts. A typical three-way power supply is shown in the illustration. A vacuum-tube rectifier supplies power when the batteries are not in use. Switch 1 selects a-c, d-c, or battery operation. This opens or closes the B-battery supply, and selects either battery power or rectifier power for B+ circuits in the receiver. Since switch 1 is ganged to switch 2, a single control takes care of all the switching. Switch 2 completes the A-battery circuit or the a-c – d-c line operation.

When the switch is turned to "battery" position, the B battery supplies power through switch 1 to the B+ line of the set, and the A battery supplies 7 volts

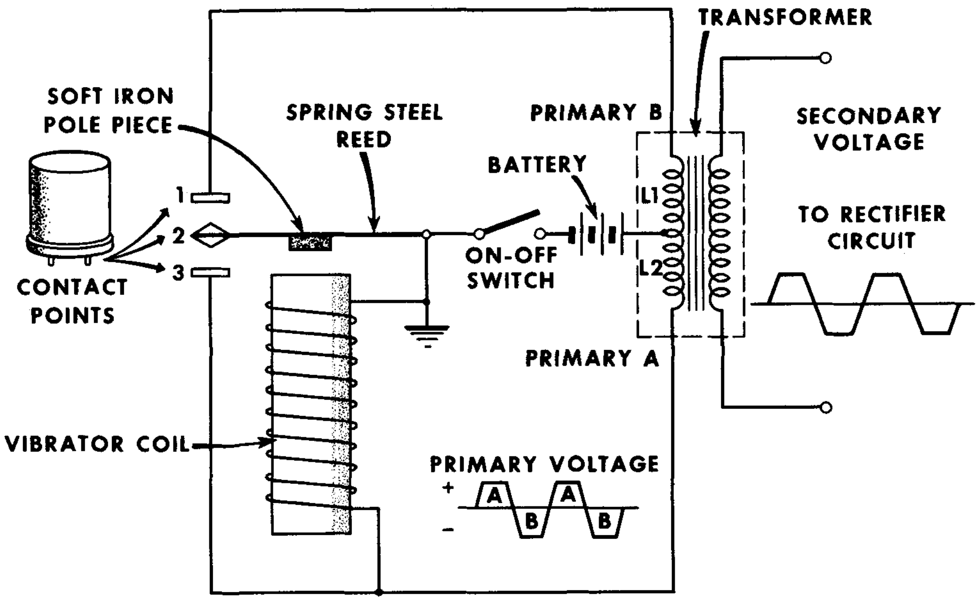


TYPICAL THREE-WAY PORTABLE POWER SUPPLY

to the series-filament circuit. When the switch is placed in the a-c – d-c position, the half-wave rectifier supplies about 100 volts dc from its cathode. The rectified current passes through current-limiting resistor R1, where the lines divide. The main B+ line is connected across filter resistor R2 to the B+ line of the set. Another branch of the rectifier current passes through the double-filter voltage-dropping network of R3 and R4, dropping the output to 7 volts for the series-filament circuits. Capacitor C1 is a small noise-filtering unit which bypasses noise impulses in the power line, preventing them from interfering with the radio circuitry.

Auto Radio Power Supplies

THE NONSYNCHRONOUS VIBRATOR



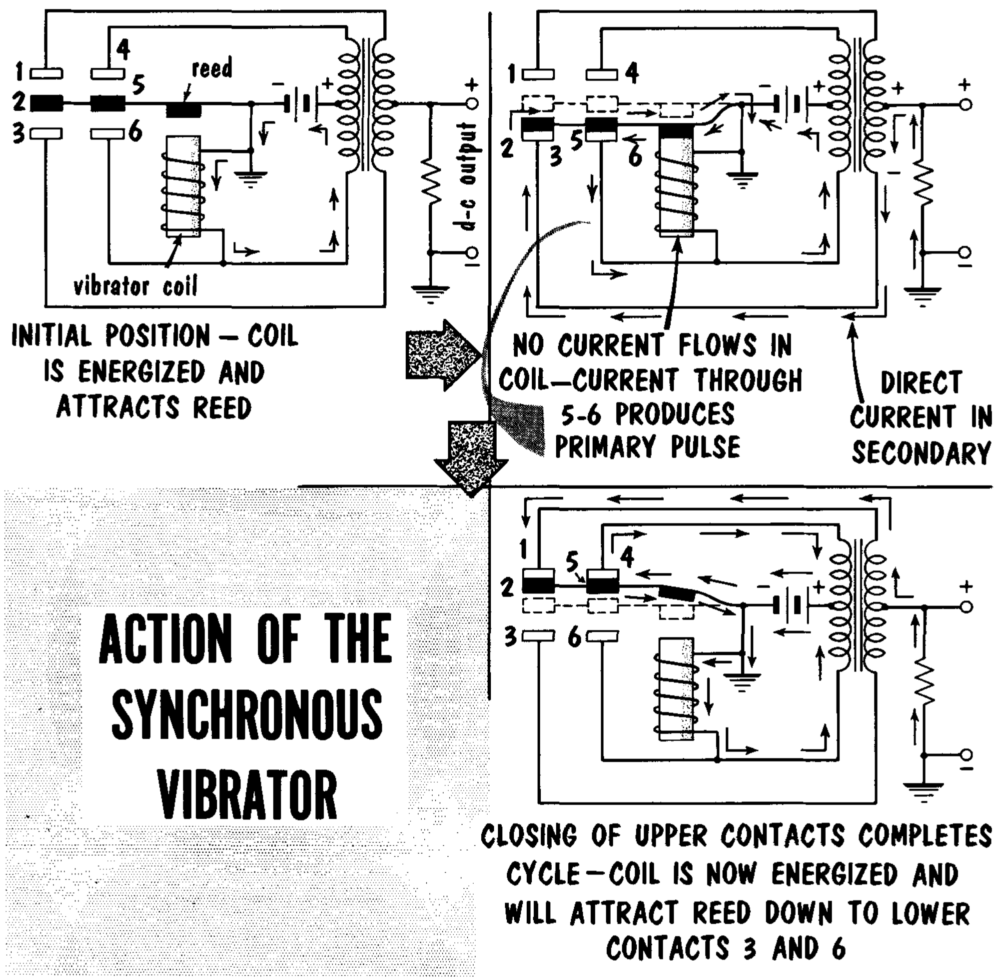
Radio receivers installed in automobiles receive all their power from the 6- or 12-volt storage battery used for supplying the electrical system. Although some modern auto receivers use tubes operating from 12 volts on the plates and screen grids, together with 12-volt heaters, most automobile radios require a special power supply that converts the 6- or 12-volts dc into an a-c voltage. This voltage can then be stepped up or down, as required, for conventional receiver operation. The important device used for making the conversion is called a vibrator. The vibrator may be compared to a two-position switch. The vibrator reed is series-connected with the battery and the transformer primary centertap.

Closing the switch produces current flow through the vibrator coil, primary winding L2, and the battery. This energizes the vibrator coil and closes contact points 2-3 by magnetic attraction of the soft-iron pole piece on the reed to the core of the vibrator coil. The instant contact is made between points 2 and 3, a direct short circuit is created across the vibrator circuit. This de-energizes the coil, and the reed springs away from contact 3 to make contact 1-2. In this position, current flows through primary winding L1, through the battery, and through contact 1-2. However, just as contact 2-3 was opened, current began to flow through the vibrator coil, energizing it, and attracting the reed back to contact 3. Thus, the 1-2 contact was made as the result of the inertia or spring of the reed as it was released from contact 3. Each time contact was made, current would flow through part of the primary winding, first in one direction and then in the other. This produces somewhat of a square-wave output from the secondary.

The Synchronous Vibrator

The important feature of the synchronous vibrator is that it is self-rectifying; that is, it provides a d-c output without requiring a separate rectifying element such as a tube. Note that initially, the reed rests between two sets of contacts. When the vibrator circuit is closed, contact is made between points 2-3 and 5-6 as the result of energizing of the vibrator coil. The closing of contacts 5-6 shorts out the coil, de-energizing it. This releases the reed, and as it moves upward, it contacts points 1 and 4. In both instances when contacts are made, the current flowing through contacts 5-6 and 5-4 provides a current pulse in the primary. This is coupled to the transformer secondary by induction.

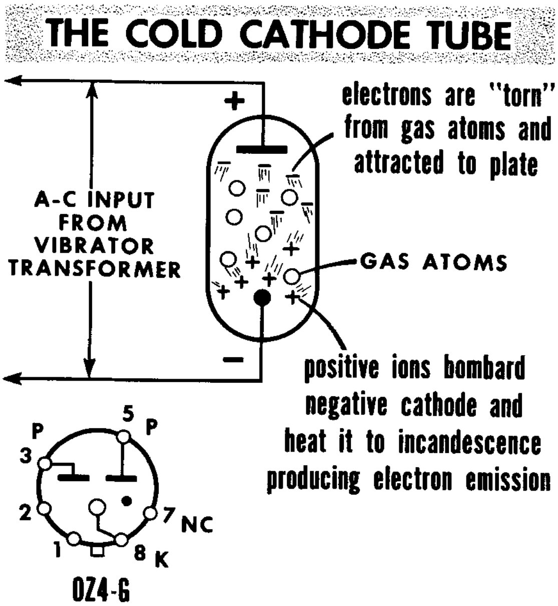
By the switching action of contacts 2-3 and 2-1, acting as a mechanical switch, the polarity of the induced voltage in the secondary is such as to produce a d-c voltage across the output circuit.

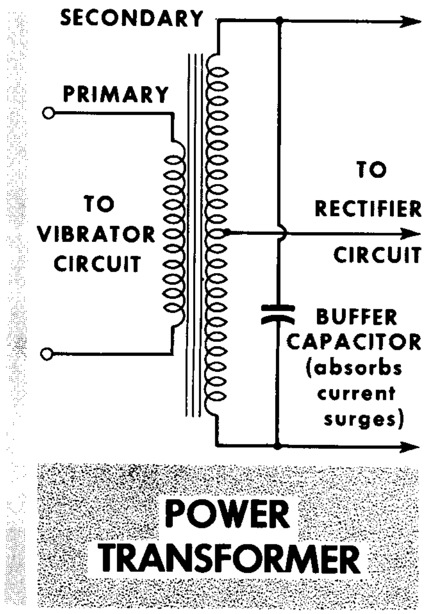


ACTION OF THE SYNCHRONOUS VIBRATOR

The Cold-Cathode Rectifier and Buffer Capacitor

Two important components of auto radio power supplies are the cold-cathode rectifier and the buffer capacitor. The most commonly used cold-cathode or ionic-heated rectifier is the OZ4-G, which depends upon gas ionization for its operation. This tube contains an inert gas at low pressure. When the plate of the OZ4-G is made positive with respect to cathode, it attracts electrons from the gas atoms in the tube, creating positive ions that are in turn at-



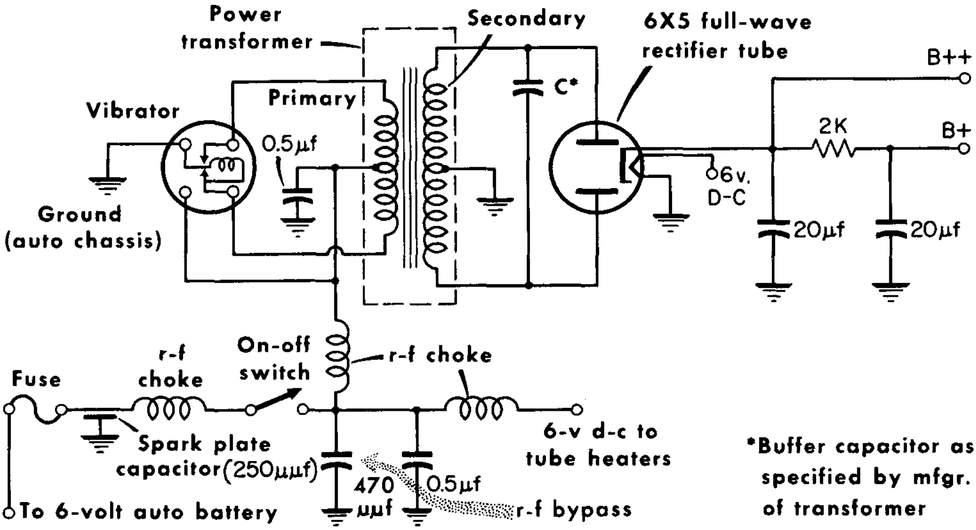


tracted to the negative cathode. The relatively large ions bombard the cathode to an extent that the cathode temperature rises quickly to incandescence, and emits electrons. Thus, this tube makes no use of conventional heater current. A minimum flow of current must be maintained at all times to keep the cathode at proper operating temperature for electron emission.

When the points of a vibrator break the circuit, they are subjected to a sudden heavy surge of current due to the inductive kickback voltage of the windings. This heavy current would soon destroy the points unless something were done to minimize its effects. The function of the buffer capacitor is to absorb this surge of current, and so protect the circuit components. The value of the buffer is critical. If the capacitance is too high, excessive current will flow that could damage the vibrator contact points; if the capacitance is too small, it does not completely absorb the current surge, and there will be arcing at the contact points. To handle the high values of back emf, buffer capacitors must have a high working-voltage rating of approximately 1500 volts. To prevent a short-circuited buffer capacitor from ruining the power transformer, a resistor is often placed in series; it also acts as a current limiter.

Auto Radio Rectifier and Filter Circuits

A TYPICAL AUTO RADIO POWER SUPPLY



The heater supply for a car radio requires a vacuum-tube rectifier with low heater current drain such as the 6X4 type. It is a vacuum rectifier tube using an indirectly heated cathode and a 6-volt low-current drain (600 ma) filament. Another such tube is the type OZ4-A, a cold-cathode gas-filled rectifier tube requiring no filament current.

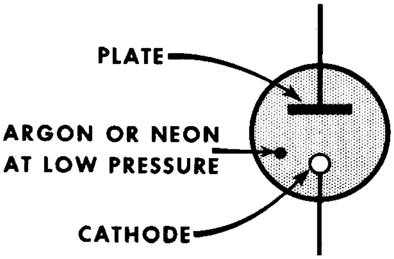
A major problem in auto radios is noise pickup. The auto radio uses the same battery as the generator, directional signal lights, ignition, etc. This noise comes in on the hot (ungrounded) lead used to supply power to the auto radio. Another source of noise is the auto radio antenna. To prevent pickup of ignition noise, the antenna is placed free and clear above the car. The metal body of the car acts as a shield between the noise and the antenna. The antenna lead-in is protected by the use of a shielded conductor. The radio itself is constructed of steel or aluminum, tightly enclosed, thus shielding the radio circuits.

Special filter circuits are used before the current flowing through the hot lead reaches the centertap of the primary or the tube filaments. A popular filter is called the spark plate, a specially constructed mica capacitor of approximately 250 µµf. The chassis is one plate of the capacitor; a metal plate separated from the chassis by an insulating layer of mica is the other plate of the capacitor. The construction of the spark plate makes it very effective in filtering the high-frequency noise generated in the ignition system. In addition, numerous r-f chokes, wound of heavy wire to pass high current, provide low values of inductance and sufficiently high reactance to high-frequency noise voltages to keep them out of the power supply and the tube filaments. Bypass capacitors of 0.5 µf are used as additional filters.

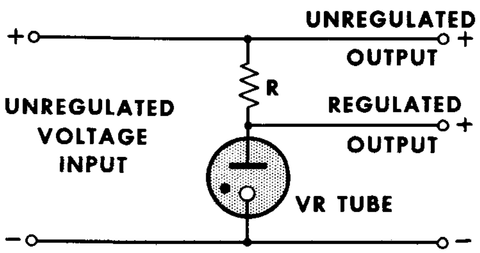
Gas Tube Voltage Regulation

One of the most commonly used voltage regulators is the glow-tube or gaseous-discharge regulator. This is a two-element cold-cathode tube filled with one of the rare gases, such as neon, argon, or helium. Voltage regulator tubes (called VR tubes) are frequently used in power supplies where it is necessary to maintain a constant output voltage in spite of wide changes in load current. The VR-75 (OA3), VR-105 (OB2), and VR-150 (OD3), are examples of these tubes, and provide regulation at specific values of voltage. The VR numbers give the rated constant voltage which occurs across the terminals of each tube for a range of current drain. In actual practice, for example, the OB2 maintains a constant voltage of about 108 volts for current

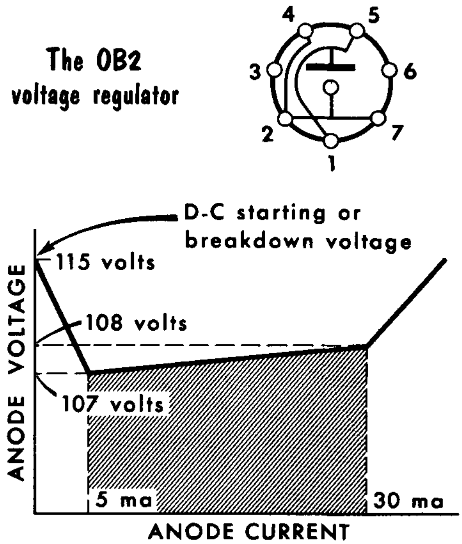
Symbol for a cold cathode gas diode



A simple voltage regulator circuit using a VR tube



VOLTAGE REGULATOR



Regulation 1 volt for current varying from 5 ma to 30 ma. D-c starting voltage about 115 volts

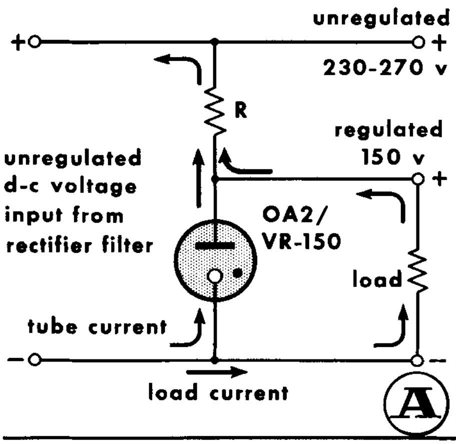
variations through it of from 5 to 30 ma. The characteristics of these tubes are determined by the electrode material, the type and pressure of the gas, and the placement and size of the electrodes.

If we look at the plate voltage vs plate current characteristic of the OB2, we notice that there is a slight change in voltage over a wide range of current. This voltage drop is referred to as the regulation of the tube. In the OB2, the regulation is 1 volt over the 5- to 30-ma range. That is, the voltage varies from about 107 to 108 volts. From this characteristic, we see that the internal resistance of the VR tube decreases as the applied voltage increases. This makes it possible for the VR tube to maintain a constant voltage across the load as the load current varies within the rated limits of the VR tube.

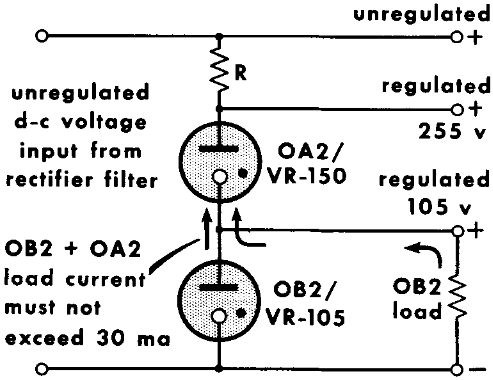
Voltage Regulators

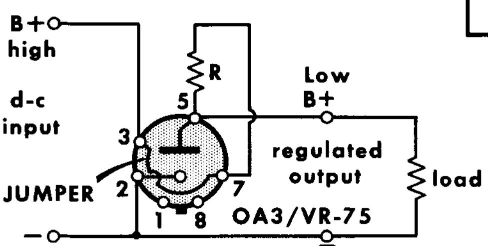
The degree of ionization in a VR tube varies with the amount of current flow through it. When a large current flows through the tube, the gas is highly ionized and the internal impedance of the tube is low; when a small current flows, the gas is ionized to a much lesser extent and the tube impedance is high. The product of the current through the VR tube and the internal imped-

VOLTAGE-STABILIZING CIRCUIT



VR TUBES CONNECTED IN SERIES PROVIDE REGULATION OF HIGHER VOLTAGES





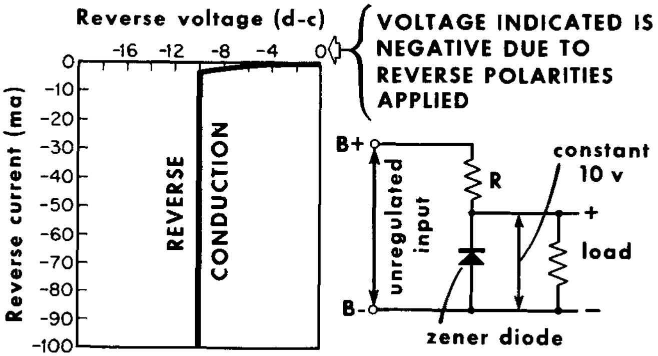
Jumper breaks power-supply circuit from load if VR tube is removed. Otherwise, full B+ would be applied to load, damaging components.

ance, which represents the voltage drop across it, remains practically constant over the operating range. An important requirement in the operation of glow-tube regulators is that a starting voltage somewhat higher than the value of the voltage at which the tube is rated be present across the tube electrodes before the tube will function.

In basic circuit A, a VR tube is connected in series with a resistor R across the output of a filter. Series resistor R limits the current flow through the tube so that its maximum rated value is not exceeded. When the unregulated B+ voltage drops below normal, the voltage across the VR tube drops, the gas in the tube becomes less ionized, the plate resistance increases, and less current flows through the tube and R. The reduced current lowers the IR drop across R, dividing the voltage so that 150 volts is once again across the VR tube. When the applied voltage rises above normal, the VR tube allows more current flow, its plate resistance decreases, and current flow through R increases. This increases the IR drop across R, and the 150 volts across the VR tube maintains a fixed 150-volt drop across it.

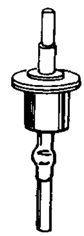
Zener Diode Voltage Regulation

The zener diode is similar in appearance to the silicon or germanium rectifier diode. Most often, it is made of silicon and has a very high back resistance. When a reverse voltage is applied to this diode, virtually no reverse current flows. However, at a certain reverse voltage point, the zener diode breaks down completely, and the back resistance drops to a very low value.



Typical Zener Diode curve

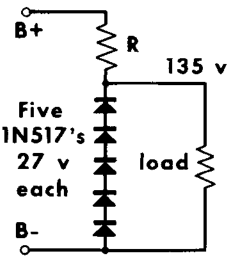
Basic Circuit



A hermetically sealed Zener Diode

ZENER DIODE REGULATION

Several diodes may be series-connected for regulation of higher voltages.

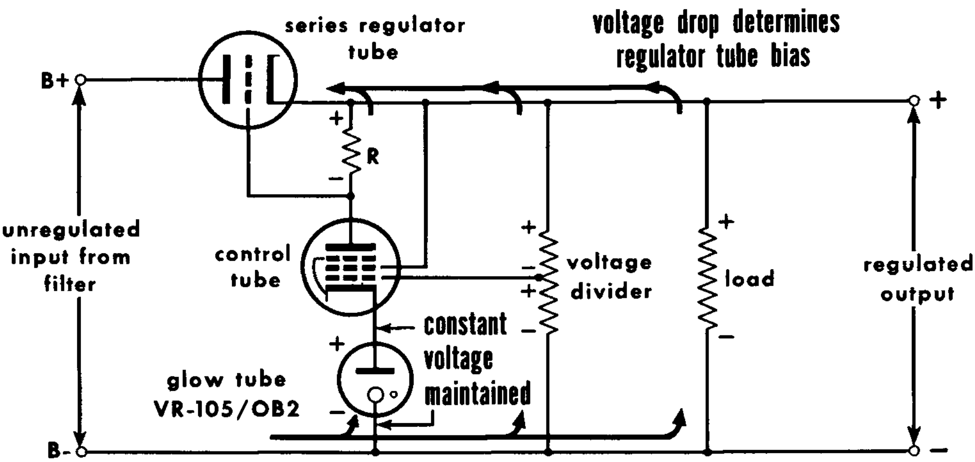


When this occurs, the reverse current increases very rapidly. The effect of a rapid increase in current, together with a rapid decrease in resistance, produces an almost constant voltage drop across the diode. Thus, when biased in a reverse direction, zener diodes can be used as voltage regulators.

Beyond the breakdown or zener voltage, the zener diode exhibits the characteristics of a gas-voltage regulator, and can be considered an equivalent. To use the zener diode in a VR circuit, positive voltage is applied (through a series resistor) to the cathode. (This is opposite to the normal application of voltage to a diode rectifier.) The current flowing through R equals the sum of the current through the diode plus the load current. When the B voltage drops below normal, the voltage across the diode drops. This increases the diode resistance, and less current flows through the diode and R. The reduced current lowers the IR drop across R, dividing the output voltage so that 10 volts, for example, is again across the diode. When the output voltage rises above normal, the diode permits more current flow, its resistance decreases, and the current through R increases. This increases the IR drop across R, and the 10-volt output across the diode is maintained. The load, connected in parallel across the diode, has a fixed 10-volt drop maintained across it.

Electron-Tube Regulator

A TYPICAL VOLTAGE REGULATOR CIRCUIT



Because an electron tube can be considered as a variable resistor, a very efficient and effective voltage regulator circuit can be built. In the circuit shown, the grid bias of the regulator is obtained by the IR drop across R. Current flows from B- and divides through the load, voltage divider, and through the VR tube, control tube, and R. The total current flows through the regulator tube. The voltage drop across the VR tube maintains a steady voltage on the control-tube cathode, and the grid is connected to a voltage divider tap. The grid – cathode voltage of the control tube is equal to the difference between the voltage drop across the VR tube and that across the lower section of the voltage divider. The voltage divider tap is set to provide the proper voltage for biasing the control tube.

If the load voltage tends to rise, either from an increase in the input voltage from the filter or because of a decrease in load current, voltage across the voltage divider and load also rises, as does voltage at the tap of the voltage divider. Voltage at the grid of the control tube thus becomes more positive, with the cathode voltage remaining constant due to the action of the VR tube. The positive-going grid produces an increase in plate current in the control tube which causes a larger IR drop across plate load R, and the grid of the regulator tube becomes more negative. The negative-going grid of the regulator tube increases its plate resistance and reduces the current flow through the tube. This increases the voltage drop across the regulator tube, and the voltage across the load is reduced to its correct value. When the applied voltage drops below normal, the voltage at the grid of the control tube drops in proportion, reducing plate current in the control tube. The reduced IR drop across R produces a less negative grid bias on the regulator tube. Its plate resistance decreases, and plate current increases. The voltage drop across the regulator tube decreases, and voltage across the load increases to its correct value.

A power transformer usually has a primary winding, a high-voltage secondary winding, and several low-voltage windings which supply the various tube filaments.

The diode vacuum-tube rectifier allows electron flow from cathode to plate when the plate is positive with respect to the cathode. When the plate is negative with respect to cathode, electron flow is retarded.

The peak inverse voltage rating of a rectifier is the maximum voltage that can be applied when it is not conducting.

The half-wave rectifier is the basic circuit used to convert ac to dc.

Practical full-wave rectifiers use a single, centertapped secondary winding of the transformer in place of two individual windings.

The output voltage of a full-wave rectifier contains two pulses for each cycle, producing a 120-cycle ripple frequency.

Capacitance-inductance filters provide very effective filtering action.

Capacitor-input filters are used when the amount of d-c power required is small; choke-input filters are used when the d-c power requirement is large.

For better voltage regulation, a bleeder resistor is used to prevent receiver current drain changes from changing the power supply output voltage.

In a-c – d-c receiver power supplies, a floating ground is used to prevent a lethal shock by isolating the B- terminal from the metal chassis.

The synchronous vibrator provides a d-c output without requiring a separate rectifying component such as a tube.

Cold-cathode or ionic-heated rectifiers depend upon gas ionization for their operation.

Voltage-regulator tubes are frequently used in power supplies where it is necessary to maintain a constant output voltage despite wide load current changes.

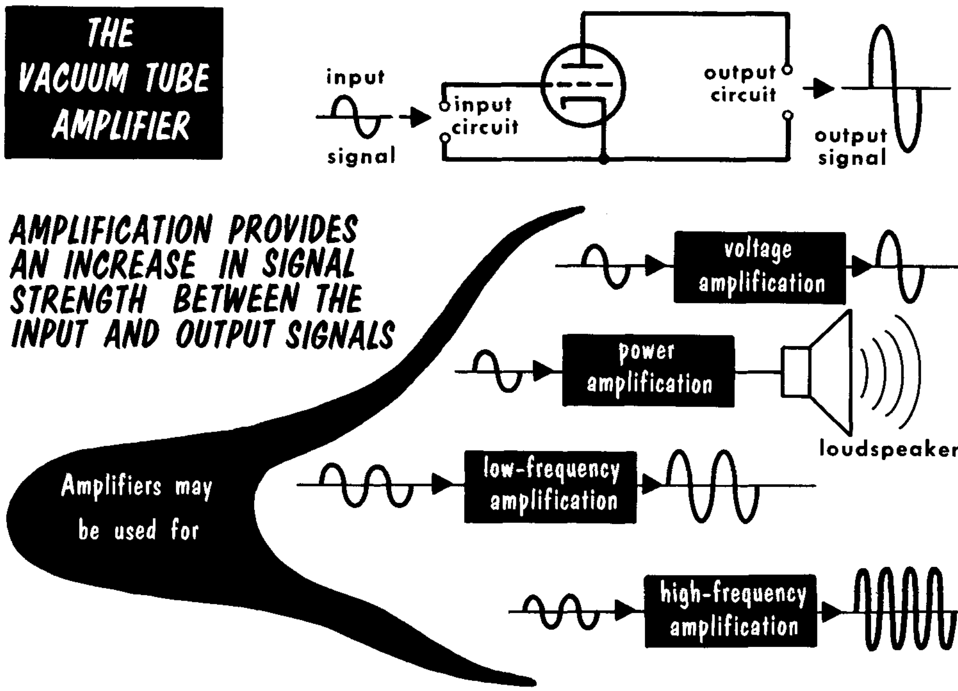
The degree of ionization in a VR tube varies with the amount of current flow through it.

Zener diodes have a very high back resistance, but at a certain reverse voltage point, they break down and the back resistance drops to a very low value.

REVIEW QUESTIONS

1. Describe the operation of a half-wave rectifier; of a full-wave rectifier.
2. What is the output voltage ripple frequency of a half-wave rectifier? Of a full-wave rectifier?
3. What are the advantages and disadvantages of gas-filled rectifier tubes?
4. Describe the operation of a bridge rectifier using selenium rectifiers.
5. How are capacitor-input and choke-input filters used in relation to d-c power requirements?
6. What are the advantages of using a bleeder resistor in a power supply?
7. Describe the action of a full-wave voltage doubler; a half-wave voltage doubler.
8. Why is a floating ground used in an a-c – d-c power supply circuit?
9. What is the main advantage of a synchronous vibrator power supply?
10. Why is a buffer capacitor used in an auto radio power supply?
11. What property of a gas-filled tube enables it to be used as a voltage regulator?
12. Describe the action of a zener diode used as a voltage regulator.

Amplification



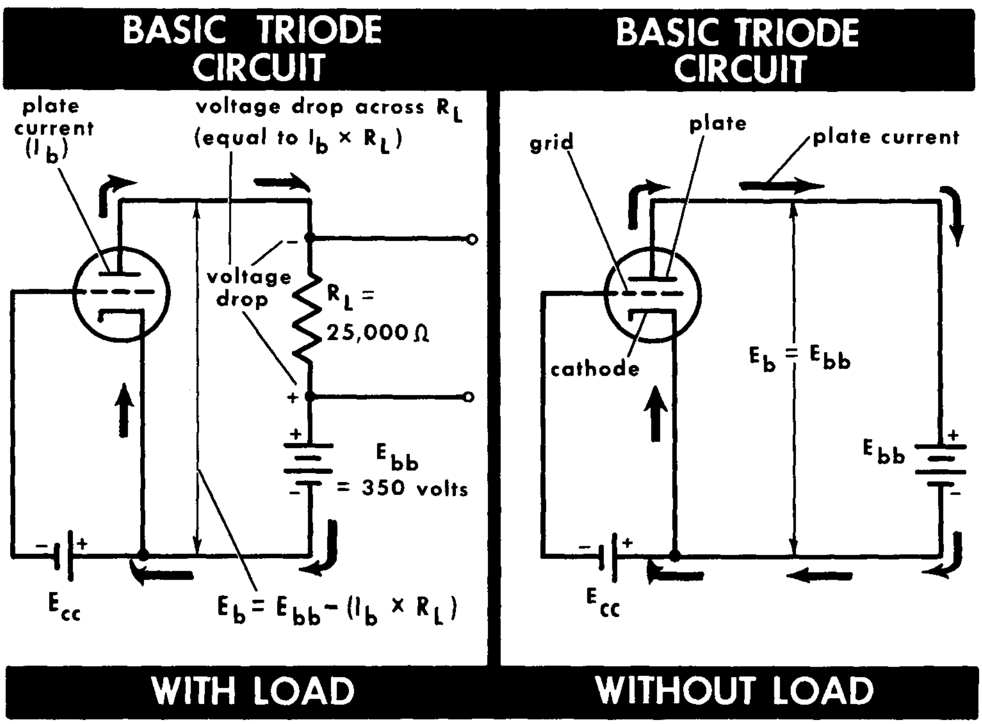
We have studied the use of the electron tube as a rectifier in power supply circuits. Now, we shall study the electron tube in its most important application – that of an amplifier. The more common expression "vacuum tube" will be used, since all electron tubes used for amplification of electrical signals are of the vacuum type. It is the use of the vacuum tube as a device for amplification that has made radio broadcasting and communications possible.

There are many ways of classifying amplifiers. In each instance, the vacuum tube itself must not be thought of as the complete amplifier, but rather as an amplifying device which, together with appropriate associated circuitry, can produce an amplified version of the input signal in its output circuit. Vacuum tube amplifiers are often classified in various categories according to the type of operation. Generally, however, there are two classifications that are most commonly accepted. The first is in terms of voltage and power. Voltage amplifiers are designed to receive small input voltages and to put out large-amplitude versions of the input signal. Power amplifiers are designed to deliver into their output circuit signal power that can be used by a special device, such as a loudspeaker or an antenna circuit in a transmitter. The second popular classification of amplifiers is in terms of frequency. The general grouping here is low frequency and high frequency. We shall study the differences in tubes and tube circuitry when we amplify low- and high-frequency signals. To begin our study, we shall return to our discussion of vacuum tube characteristics.

Dynamic Characteristics of the Triode

Thus far, we have not seen a triode do any useful work. We have varied the voltages applied to the grid and plate of the tube, and have observed the amount of plate current flowing through the tube to the plate, and from there, through the external circuit and back to the cathode. But we have not seen any results that this current has produced. In fact, we have only studied the behavior of a triode under static conditions and have developed a number of interesting families of characteristics, which up to now have served no useful purpose. In order that a vacuum tube be of any practical use, a load resistance must be inserted in its plate circuit. Once a load resistance is present, the plate current will develop a voltage drop across it which may be transferred to the input of another tube, or the tube can be used directly to do useful work. If an input signal is applied to the grid of a tube the plate current will create an amplifier version of this signal – or output signal – across such a load resistor.

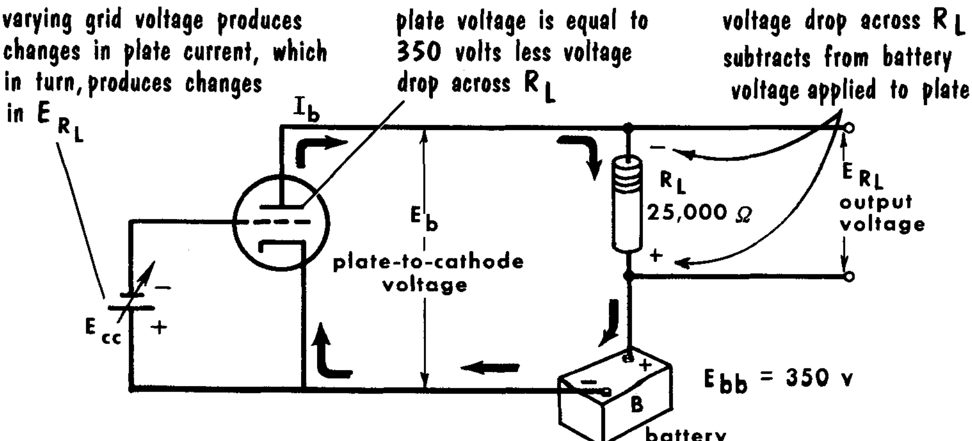
The presence of a load resistor in the plate circuit of a triode gives rise to the so-called dynamic characteristics of a tube, which are the actual conditions of operation used in practice, and hence, far more important than the static characteristics we have studied. The dynamic characteristics are thus a graphical portrayal of tube behavior under load. We have illustrated a basic triode circuit with a load inserted into the plate circuit, assigning typical values to the plate voltage and load resistor. For comparison, the same circuit is shown without a load.



Voltage Drop Across Load

The load resistor R_L is in series with the plate supply voltage E_{bb} and the tube itself. Consequently, the electrons on their way back to the tube's cathode must flow through this load resistor and develop a voltage drop across it. This is known as the output voltage E_{RL}. By Ohm's law, since $E = I \times R$, the voltage drop across R_L (E_{RL}) is then $I_b \times R_L$. You remember that the sum of the voltage drops around a series circuit must equal the source voltage. Evidently, then, the plate voltage supply E_{bb} must equal the sum of the plate-to-cathode voltage E_b, plus the voltage drop across the load resistor E_{RL}. In other words, the plate voltage across the tube is the difference between the plate supply voltage E_{bb} and the voltage drop across E_{RL} (equal to I_bR_L). This is a very important relationship to remember. It shows that the plate voltage (E_b) decreases as the plate current increases, since the plate supply voltage E_{bb} and the load resistor R_L are both fixed in value. This is the main difference between the static condition of operation, where the plate voltage equals the plate supply or battery voltage ($E_b = E_{bb}$), since no load is present, and the dynamic condition of operation with a load, where the plate voltage is the difference between the plate supply voltage and the voltage drop across the load $[E_b = E_{bb} - (I_bR_L)]$.

As an example, imagine first that E_{cc} is adjusted to such a high negative value as to cut off I_b. With no plate current flow, there is no voltage drop across R_L, and the plate voltage equals E_{bb} or 350 volts. Now assume that the bias is changed to permit 1 ma of plate current to flow through the plate circuit. We find that the plate voltage now is $350 - (0.001 \times 25,000)$, or 325 volts. In this case, the voltage drop across R_L is relatively low, and the internal drop across the plate resistance is high (325 volts). Imagine now that the grid bias is adjusted to a less negative value, to allow as much as 12 ma of plate current to flow through the circuit. The plate voltage is now $350 - (0.012 \times 25,000)$, or 50 volts. For this operating condition, we see that most of the voltage drop appears across R_L, and very little across the internal resistance of the tube.



VOLTAGE DISTRIBUTION IN A TRIODE CIRCUIT

The Load Line

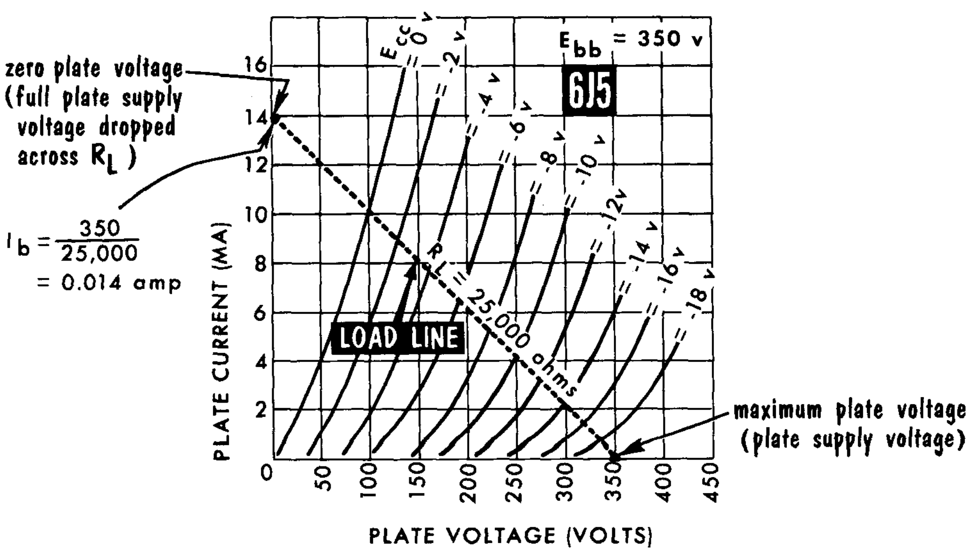
The effect of a plate load connected in a triode circuit can be predicted in advance. A graphical representation of the load, known as a load line, can be added to the static plate characteristics of the tube. The load line shows the voltage distribution of the plate supply voltage, the voltage across the load, and the plate voltage for different values of plate current and grid voltage. The load line is usually constructed by joining two points – one on the plate-current axis (corresponding to zero plate voltage), the other on the plate-voltage axis. This point corresponds to a condition when the tube is at cut-off (zero plate current), and the entire plate supply voltage E_{bb} equals the plate voltage. The point on the plate-current axis corresponds to the current flow when the entire plate supply voltage is the voltage drop across the load, resulting in zero plate voltage.

For a 6J5 triode tube using a 25,000-ohm plate load and a plate supply of 350 volts, a typical load line can be drawn on the same graph as the static plate characteristics of the tube. The point on the plate-voltage axis is the 350-volt plate supply voltage (zero plate current). Location of the point on the plate-current axis requires the use of Ohm's law. The plate supply voltage (350 volts) is the theoretically maximum voltage drop across this 25,000-ohm resistor. Ohm's law is used to find the value of the plate current:

$$I = E/R = 350/25,000 = 14 \text{ ma}$$

The second point is then marked at the intersection of 0 volts and 14 ma. Using a straight edge, a line is drawn connecting the two intersections; this line is the load line.

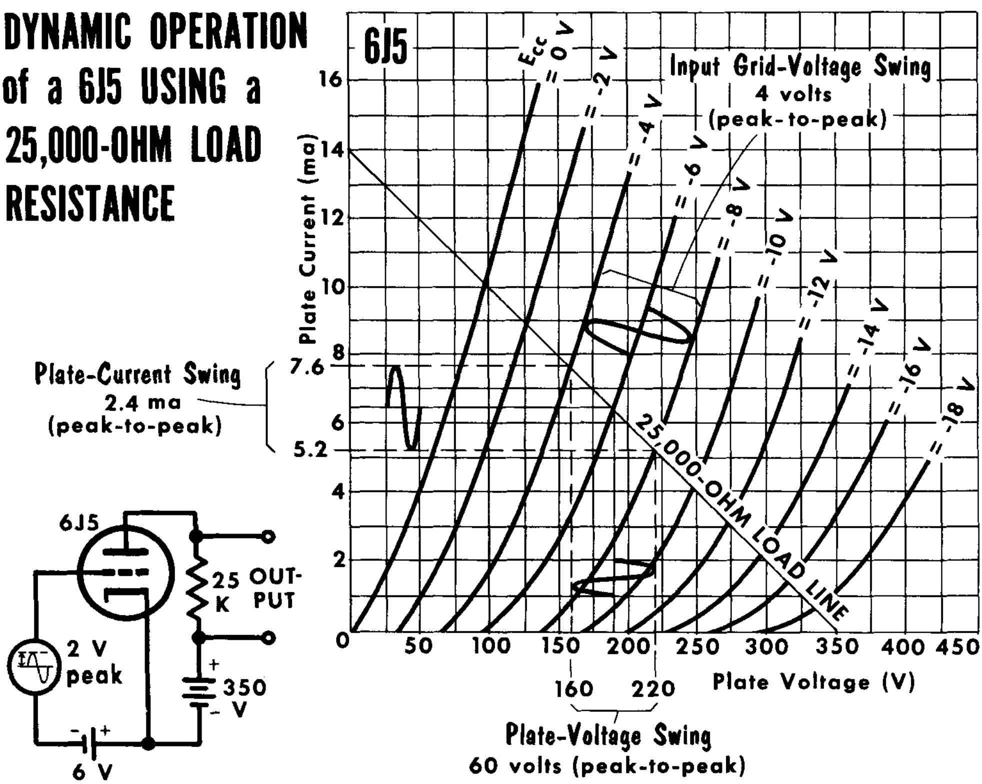
25,000-Ohm Load Line Constructed on a Plate Family of Curves



Using the Load Line

A load line can be used to find the value of plate voltage for a specific value of plate current, or the value of plate current for a specific value of grid voltage. For example, with a grid voltage of -6 volts, we check its intersection with the load line and find that the plate voltage is 190 volts and the plate current 6.4 ma.

We can go further, and observe plate current and voltage during a complete cycle of input voltage. With the 6J5 biased at -6 volts, a 2-volt peak input voltage swings the grid bias up to -4 volts and down to -8 volts. Projecting toward the plate current axis, we see that during this time, plate current rises from 6.4 ma at -6 volts bias, to approximately 7.6 ma at -4 volts bias, and decreases to 5.2 ma at -8 volts bias. At the same time, the plate voltage moves from 190 volts at -6 volts bias down to 160 volts at -4 volts on the grid, and up to 220 volts at -8 volts on the grid. Thus, we see a remarkable thing. By use of the load line, we have a "picture" of plate voltage and plate current variations at every instant of the input cycle. For any given grid voltage variation, we can predict plate current and plate voltage variations. We thus "see" the dynamic operation of a tube under a given set of conditions. For different values of load resistance, the load line would take different positions, and the same input voltage would produce different plate current and plate voltage variations.



Dynamic Transfer Characteristics

Although the load line is very important when it is added to the static plate family, it does not tell the story of dynamic operation as conveniently as does the static grid family with the effect of the load superimposed. If this is done, the resulting plate-current, grid-voltage characteristic is known as the dynamic transfer characteristic. We show the static plate family with the 25,000-ohm load line and the static grid family, but with the dynamic characteristic added. Although this dynamic curve can be developed directly by the appropriate measurements, we have taken the easy way of simply transposing the information onto the static grid family. Both figures have a common plate current (vertical) axis, but the horizontal axis for A is the grid voltage, while for B, it is the plate voltage. It is, therefore, simply necessary to plot the plate current values for any particular grid voltage from the load line of B onto the corresponding plate-current, grid-voltage points of the graph in A. Thus, we obtain the dynamic transfer characteristic for a 25,000-ohm load.

Notice that the dynamic characteristic is much less steep and less curved than the static plate-current, grid-voltage curves. The insertion of a load in the plate circuit has resulted in straightening out the static characteristics, and has made them more linear than before. This is important in relation to the amount of distortion that occurs during operation.

CONSTRUCTION of DYNAMIC TRANSFER CHARACTERISTIC for 6J5 from PLATE and GRID FAMILIES of CURVES (Using 25,000-ohm load)

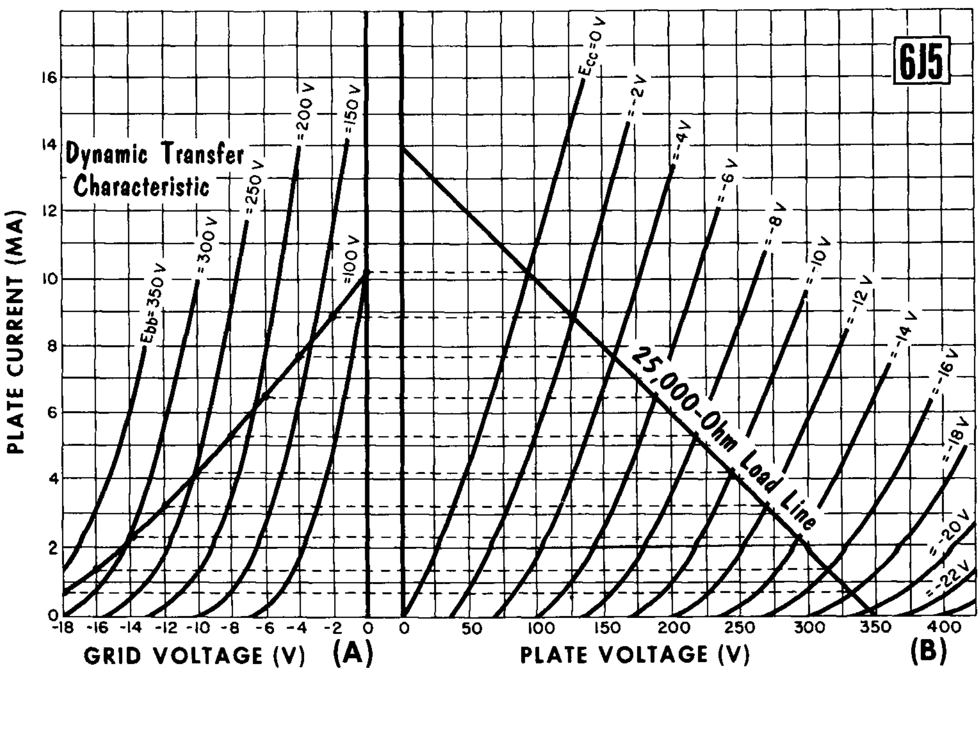
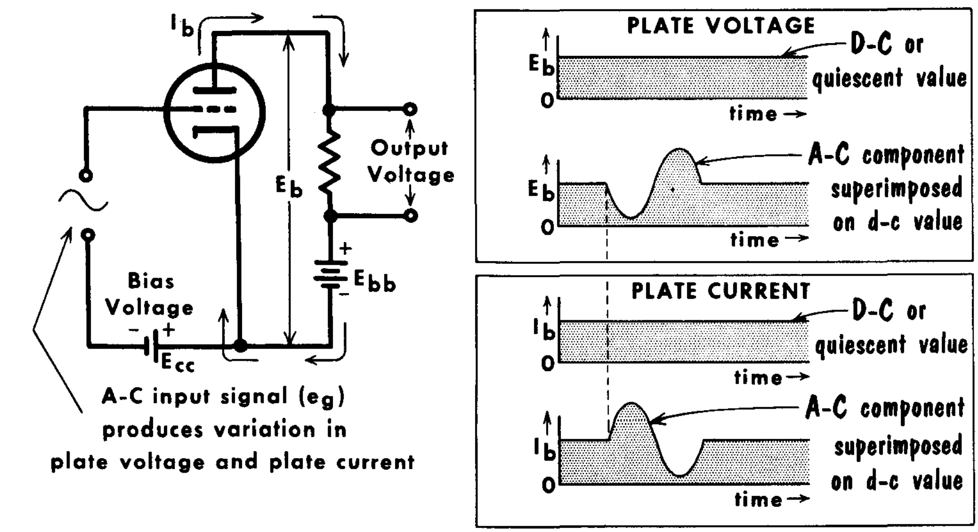


Plate Voltage and Current Components

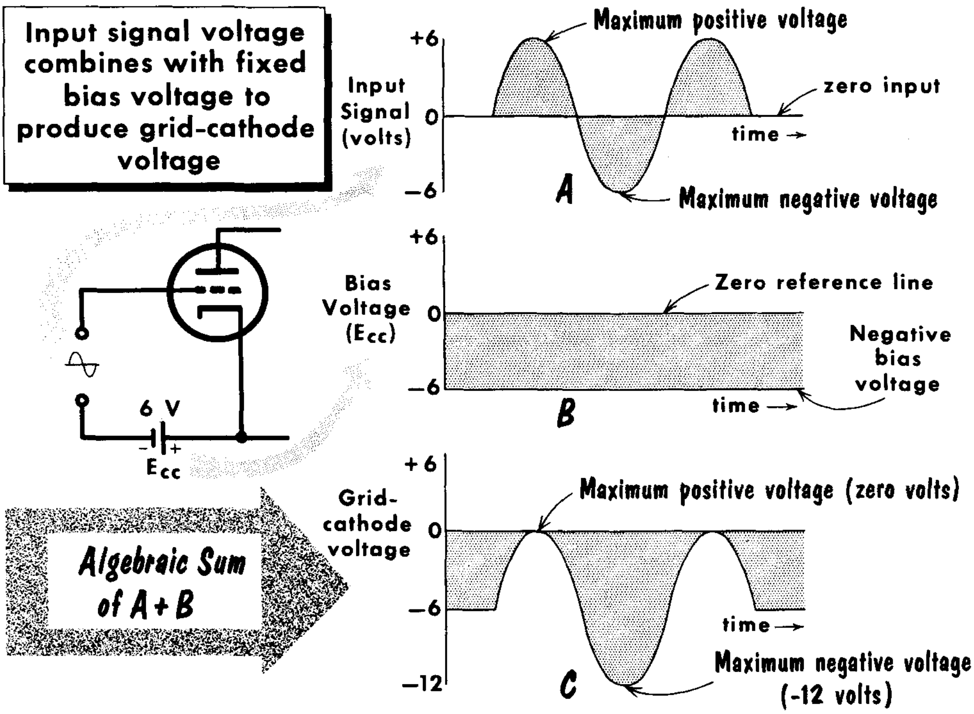
Thus far, we have considered the fundamental triode circuit operated with d-c potentials, although we have, on occasion, varied these potentials in a more or less mechanical manner to observe the effect on the plate current. In most applications, however, the triode is operated with an alternating voltage (usually called the exciting or signal voltage) applied to the grid circuit, in addition to the d-c grid bias voltage. The effect of this is to vary the grid-to-cathode voltage of the tube and cause a corresponding variation in the plate current. The plate current variations, in turn, generate a varying voltage across the load resistor, the so-called output voltage of the tube. In order to understand this dynamic amplifying process, we shall have to modify our thinking toward an alternating-current viewpoint of the triode tube. Actually, the triode is no different from the diode in that it is capable of passing a current in only one direction, from cathode to plate, and only when the plate is positive with respect to the cathode. Hence, the varying plate current and voltages of the tube are all unidirectional, and they never reverse to negative polarity.

The proper way to consider the varying plate current and voltages, with a grid signal voltage present, is to imagine them composed of two components. One component is the d-c or quiescent value of the current or voltage for a fixed grid bias, with no signal voltage present in the grid circuit. Superimposed on this d-c component is a second component of the current or voltage under consideration, namely, the varying or alternating component caused by the exciting voltage or signal in the grid circuit of the tube. This last point is important and often misunderstood. The flow of plate current is direct current. However, because it rises and falls about a center or zero signal value, it has an a-c component. Frequently, current of this type is referred to as pulsating dc.

A-C and D-C PLATE-CIRCUIT COMPONENTS



Grid Bias

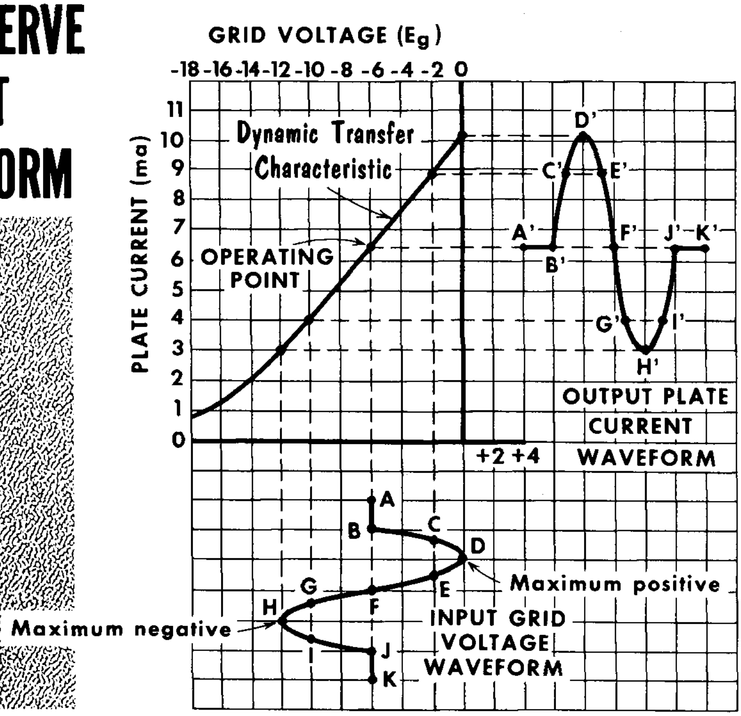


We have seen that the a-c signal voltage is inserted in series with the grid bias battery E_{CC}. The reason for this is simple. In its basic application as an amplifier of tiny signal voltages, the triode is operated to consume no power in the grid circuit, because generally, no power is available from the extremely weak radio signals. This is one of the features of a tube – it can be purely voltage operated in the grid circuit, although power may be available from the plate circuit. To consume no power in the grid circuit, it is essential that no grid current flow. To avoid the flow of grid current, the tube must be operated at a negative grid voltage, or at least, at a voltage which never rises above zero to positive values, since under those conditions, grid current would flow and power would be consumed.

This is the real purpose of the negative grid bias – to prevent the control grid voltage from ever rising to positive values which results in grid-current flow. We show an a-c sine-wave signal voltage for grid excitation which rises to a positive peak of +6 volts, and has a negative peak of -6 volts. In series with this a-c voltage, we have applied a d-c grid bias of -6 volts. (From now on, we shall always reserve the term bias for the d-c grid voltage.) The total instantaneous voltage acting between grid and cathode of the tube (e_c) is the algebraic sum of the a-c signal voltage and the d-c grid bias. The bias voltage of -6 volts has been represented by a straight line, 6 volts below the zero-voltage reference line. Note that at no time does the grid swing positive with respect to the cathode.

Operating Point

DYNAMIC TRANSFER CHARACTERISTIC is used to OBSERVE OUTPUT WAVEFORM



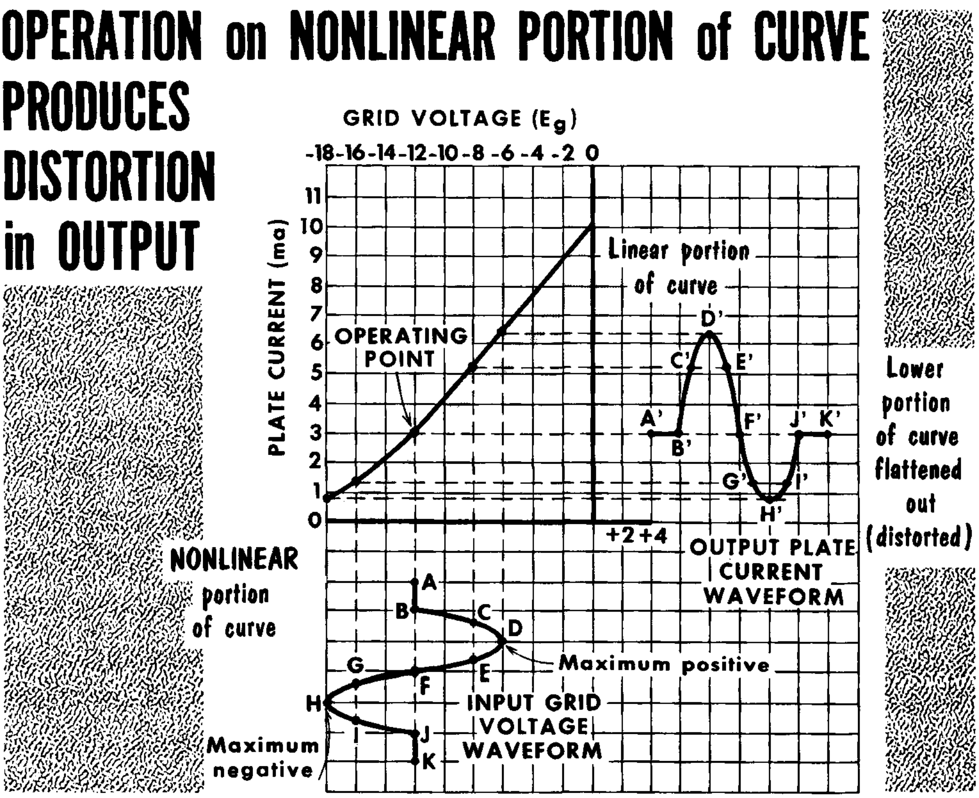
To demonstrate the method of predicting the plate-current behavior from the dynamic transfer characteristic, it is necessary first to establish an operating point on the characteristic curve. This is determined by the amount of fixed grid bias applied to the tube. The bias establishes a steady value of plate current which exists for a zero-input signal voltage, and is generally referred to as the quiescent or d-c value of plate current. In the diagram, we have repeated the dynamic transfer characteristic for the 6J5 with a 25,000-ohm load resistor and 350-volt plate supply voltage. This curve portrays graphically the variations in output plate current produced with a varying input grid voltage. The curve actually shows the behavior of the plate circuit for a given input signal voltage and a fixed operating point.

If we use -6 volts for the operating point, the a-c signal having a positive peak of +6 volts will not drive the grid positive; hence, no grid current is drawn. With -6 volts bias, corresponding to zero voltage input, we obtain a plate current of 6.4 ma from the dynamic characteristic. By plotting the input grid voltage swing against the curve, and then projecting point by point to the plate current axis, we obtain a pattern of the plate current waveform. Since the plate current flows through R_L, the voltage drop across R_L is an accurate reproduction of the grid input voltage.

Linearity of Output Waveform

For low distortion and a faithful reproduction of the input waveform, the plate current changes must be linear; that is, they must be directly proportional to the grid voltage changes. If we examine the dynamic transfer characteristic, we see that it is quite linear over the major portion; however, the lower left-hand portion curves somewhat and is not linear. On the preceding page, we projected the input grid voltage against the linear portion of the curve and thus obtained an exact replica of the input grid voltage in the form of the output plate current waveform. We shall now move our operating point down to -12 volts. In so doing, the input grid voltage reaches a maximum positive value of -6 volts, and a maximum negative value of -18 volts. In addition, we see that to the left of the operating point, the grid voltage is projected against the nonlinear portion of the curve.

When we make a point-by-point projection, it can be seen that the curvature of the dynamic transfer characteristic has produced a distorted output plate current waveform, with the negative peak flattened out. This flattening represents signal distortion. Thus, by properly locating the operating point on the linear portion of the dynamic transfer characteristic (keeping in mind the peak-to-peak swing of the input grid voltage), the output plate current waveform can be made an exact replica of the input grid voltage.



Calculating Amplification

We have not yet shown how much the simple triode amplifier has amplified the input signal voltage in the grid circuit. Although our plate current waveform looks bigger than the grid voltage waveform, this is no indication of the amount of amplification, since we cannot compare the amplitude of a current with that of a voltage. To obtain the correct amplification, we must compare the output voltage developed as a result of the drop across R_L with that of the input or signal voltage applied to the grid. Hence, we can define the amount of voltage amplification, also known as voltage gain, as the ratio of the output voltage to the input voltage.

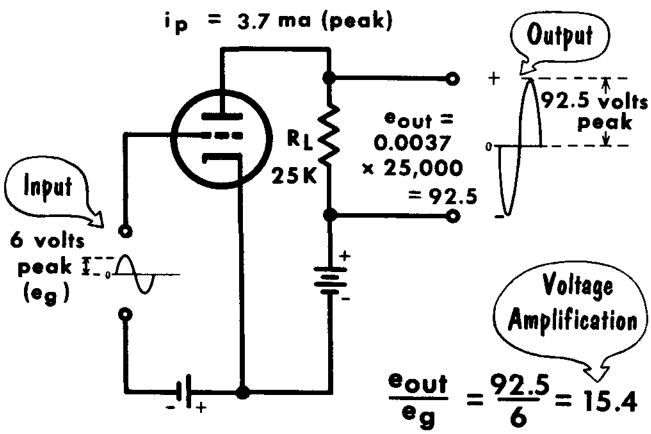
You will note that the equation deals strictly with instantaneous values of a-c quantities. There are two ways we can determine the output voltage (e_{out}). One way is to multiply the a-c plate current (i_p) by the load resistance (R_L); in other words, the a-c voltage across the load ($e_{out} = i_p \times R_L$). In our example, the plate current rises from its quiescent or d-c value of 6.4 ma for zero signal to a maximum of 10.1 ma for the 6-volt positive peak of the a-c input signal. The peak value of the a-c plate current (i_p), then, is the total

$$A = \frac{e_{out}}{e_g}$$

A = voltage amplification, or voltage gain

e_{out} = instantaneous value of a-c output voltage across R_L

e_g = instantaneous value of input signal

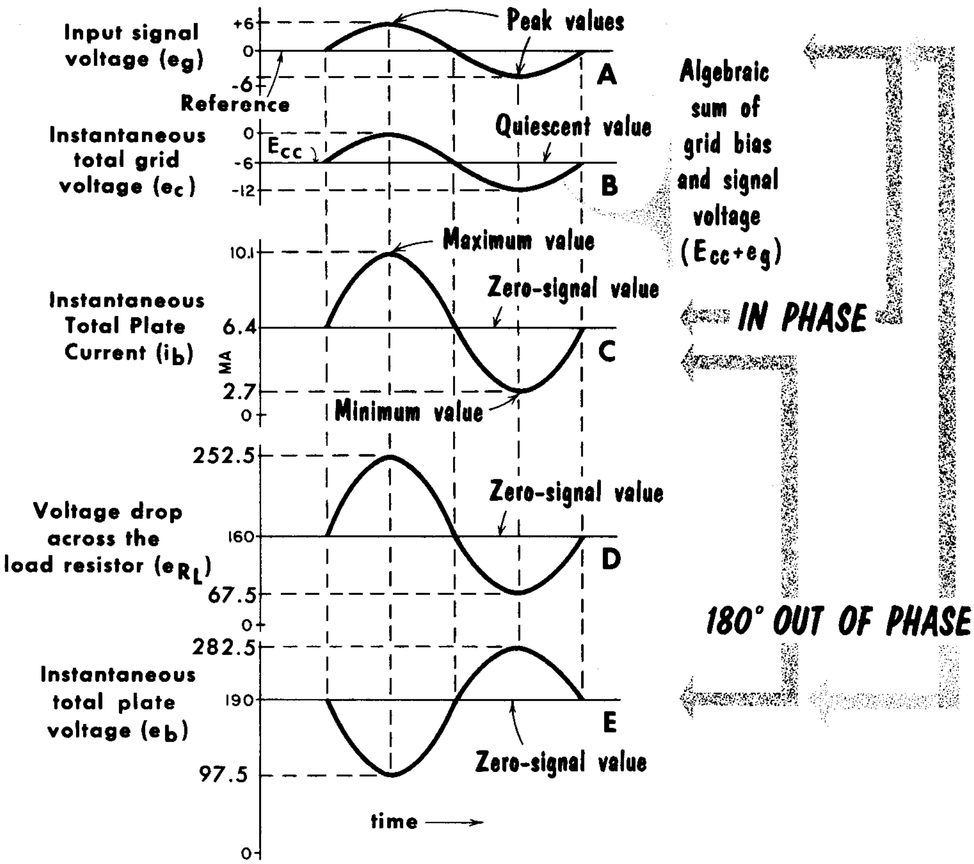


change in the plate current, which is 3.7 ma (10.1 - 6.4 = 3.7). The peak output voltage value is 92.5 volts. Thus, we see that the voltage amplification is 15.4, which means that any value of the input voltage will be multiplied by a factor of 15.4 because of the tube's amplification. (We have used peak values of the output to the input voltage. Actually, any two corresponding points of the output and input voltage wave could have been compared.)

The second way of determining the output voltage is directly from the load line, and does not involve any calculations whatsoever. Referring back to page 3-64, we note that the peak grid input voltage is 2, and the peak plate voltage swing is approximately 30. Hence, a direct reading from the graph would indicate a gain of approximately 15.

Current and Voltage Phase Relationships

PHASE RELATIONSHIPS IN TRIODE AMPLIFIER



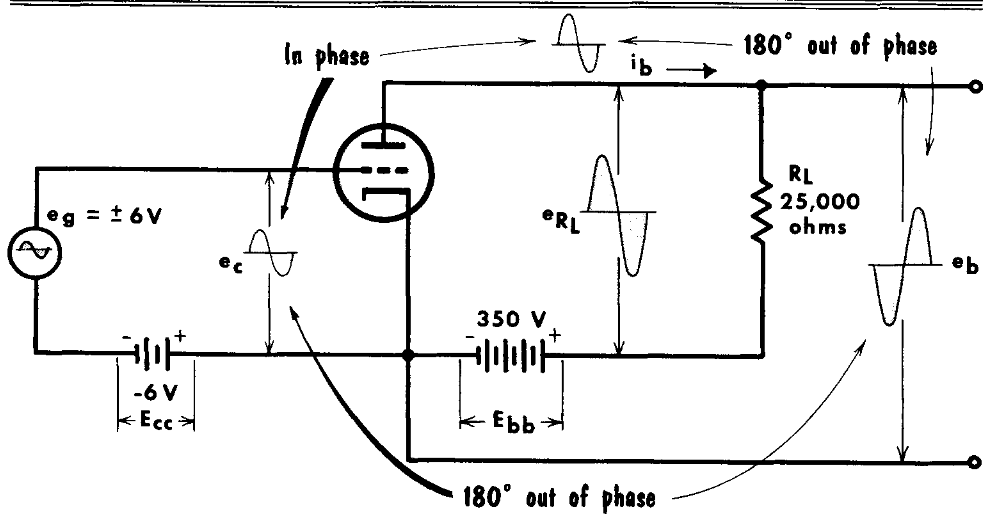
We have seen how the signal applied to the grid of a triode is amplified by the tube, causing a magnified reproduction of the input signal to appear in the output (plate circuit). We have shown further, that the plate current – hence, the output voltage – faithfully duplicates the input signal waveform, provided that the operating point of the tube is properly chosen. We have not yet considered the timing, or relative phase, between the various input and output voltages and currents.

We see five sine waves which depict the phase relationships in a triode amplifier circuit. Our previous example for -6 volts bias, 350 volts plate supply, and 25,000 ohms load resistance has been chosen again for continuity, but the phase relations are true regardless of the particular values of the voltages and currents. The dashed vertical lines passing through the waveforms compare corresponding points at the same instant in time for each of the waveforms.

Current and Voltage Phase Relationships (Cont'd.)

As you can see, the phase relations are entirely different for waveform E on the preceding page, which represents the instantaneous total plate voltage (e_b), existing between plate and cathode of the tube. You will remember the equation $e_b = E_{bb} - i_b R_L$, which shows that for a fixed plate supply voltage (E_{bb}), the instantaneous total plate voltage (e_b) decreases as the plate current (i_b) and the total voltage drop across the load, increase. This is so, you recall, because with increasing plate current and load voltage drops, less of

Current and Voltage Waveforms in a Basic Triode Amplifier



the supply voltage is available at the tube's plate. On the other hand, the lower the plate current, the smaller is the voltage drop across the load; hence, more plate voltage is left over from the fixed supply voltage. It is evident, therefore, that the total plate voltage is in an opposing or out-of-phase relation to the plate current and the input signal. This is brought out by the shape of curve E.

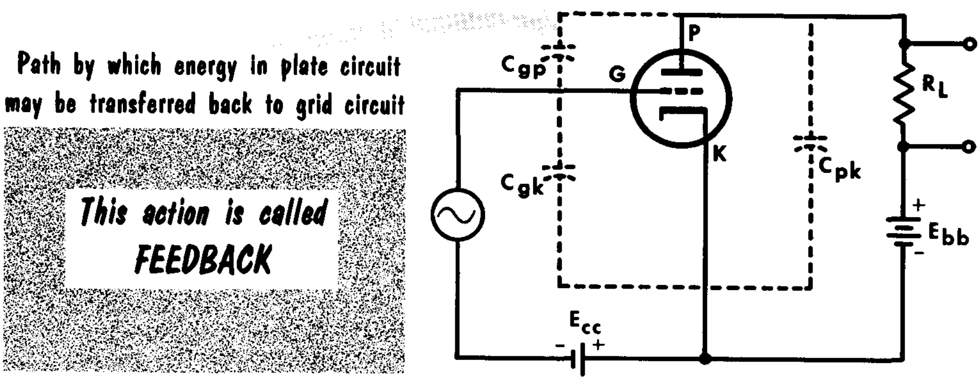
From page 3-64, we see that the quiescent value of the plate voltage (for zero signal) is 190 volts for a bias of -6 volts and a load of 25,000 ohms. When the grid voltage rises to 0 volts for a signal of +6 volts, the plate voltage falls to a minimum value of 97.5 volts, while the drop in grid voltage to -12 volts for a signal of -6 volts produces a rise in total plate voltage to its maximum value of 282.5 volts. Thus, whenever the signal voltage, e_g, is at its maximum positive value, the plate voltage e_b is at its minimum value, and vice versa. It appears as if the plate voltage has been shifted by one half-cycle, or 180° with reference to the grid voltage. We can conclude by stating that the plate current is in phase with the grid voltage, but the plate voltage is 180° out of phase with the grid voltage. This is generally true for all types of vacuum tubes that have a control grid.

Interelectrode Capacitances

Although it appears at first glance that all the electrodes in a triode tube are well isolated from each other, and that they cannot influence each other except through the flow of plate current, this is not quite correct. You remember that we discussed the electrostatic fields existing between the charged electrodes of a triode, such as the fields between plate and cathode, plate and grid, and grid and cathode. You may also recall from elementary electricity that an electrostatic field between any two charged metal plates is the equivalent of an electrical capacitor capable of holding a certain charge. Thus it is evident that definite capacitances exist between all the metal electrodes of a triode which, however tiny, do affect the operation of a tube.

The illustration indicates the capacitances existing between the metal electrodes – the so-called interelectrode capacitances. The most important of these is the capacitance between the control grid and the plate of the tube. There is also capacitance between the control grid and cathode and between the plate and the cathode. The values of these capacitances are very small, generally in the order of 2 to 10 μμf. At low audio frequencies (between 20 and 15,000 cycles), the effect of these tiny capacitances is almost negligible.

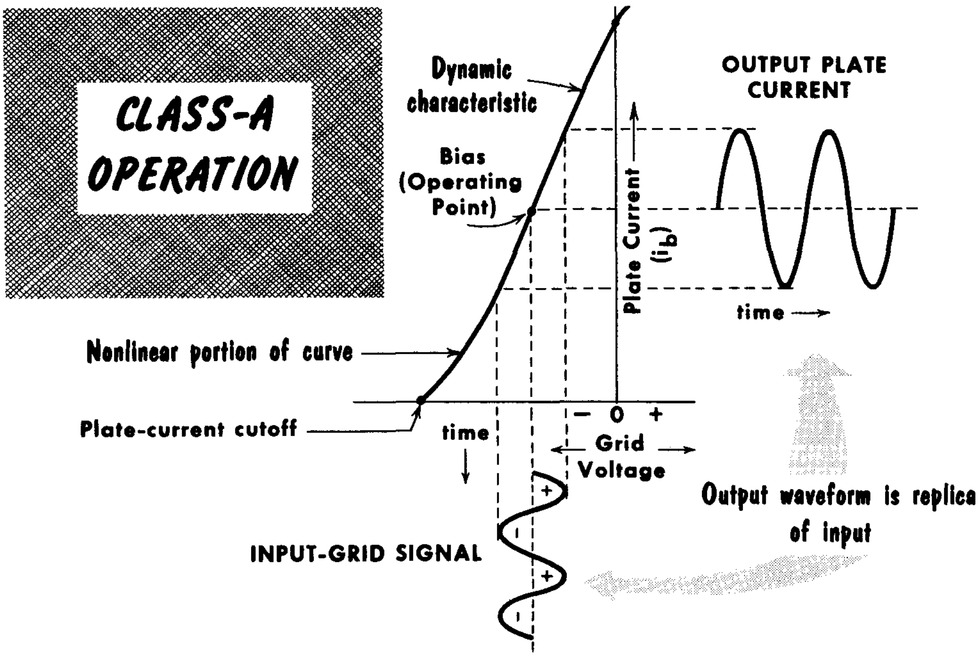
CAPACITANCE EXISTS BETWEEN the ELEMENTS of a TUBE



But at higher radio frequencies (from 100 kc up), when their reactance becomes low, they play an important role in influencing the tube's operation. The capacitance between grid and plate may have the very undesirable effect of coupling the output (plate) circuit back to the input (grid) circuit, resulting in the feedback of energy from plate to grid.

As we shall learn later in our study of oscillators, energy feedback is useful. However, the feedback of energy as a result of interaction between the grid and the plate circuits in an amplifier frequently presents serious problems in achieving circuit isolation. A reduction of interelectrode capacitances by the additional shielding of electrodes is achieved in multi-electrode tubes, such as the tetrode and pentode. Without this shielding, there would still be undesirable coupling and feedback effects between the plate and grid circuits.

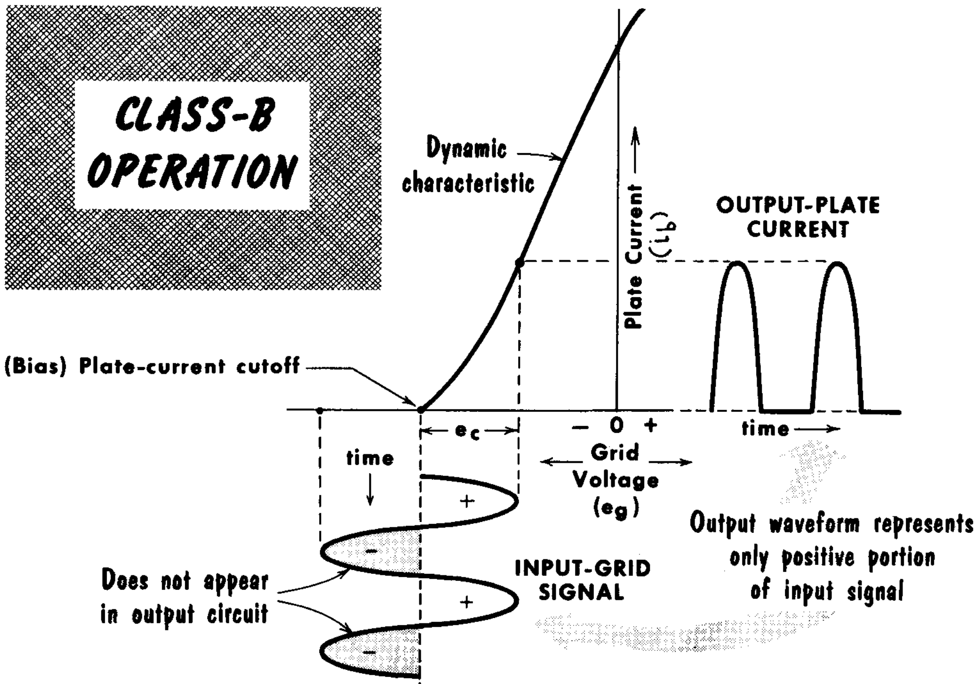
Class-A Operation



The position of the quiescent or operating point on the dynamic transfer characteristic, established by the d-c grid bias voltage, determines the different classes of operation of an amplifier. These are: class-A, -B, -AB, and -C. In radio receivers, the most commonly used class of operation is class-A. In a class-A amplifier, the grid bias is such that plate current flows in the output circuit during 360° of the input grid voltage cycle. In short, in a class-A amplifier, plate current flows continuously. Minimum distortion of the output waveform takes place because class-A amplifiers are generally small-signal low-power units in which the entire operation takes place over the linear portion of a tube's dynamic characteristic. At no time does the grid go positive under normal operating conditions, nor does it swing into the nonlinear portion of the curve on its negative cycle.

Should the control grid go positive on the positive half of the input cycle, part of the input grid signal would be lost or clipped, and the positive half of the output plate current waveform distorted. Similarly, should the negative half of the input grid signal swing beyond the plate current cutoff point, plate current would stop flowing and the negative half of the output plate current waveform would be clipped, with accompanying distortion. In most instances, the operating point of a class-A amplifier is in the center of the linear portion of the dynamic characteristic at about one-half plate current cutoff value. In terms of plate efficiency, which can be defined as the ratio of a-c power output developed across R_L to the d-c power supplied to the plate, class-A amplifiers are quite inefficient. They run as low as 20% or less, due to the high average value of plate current and, consequently, high plate power dissipation.

Class-B Operation



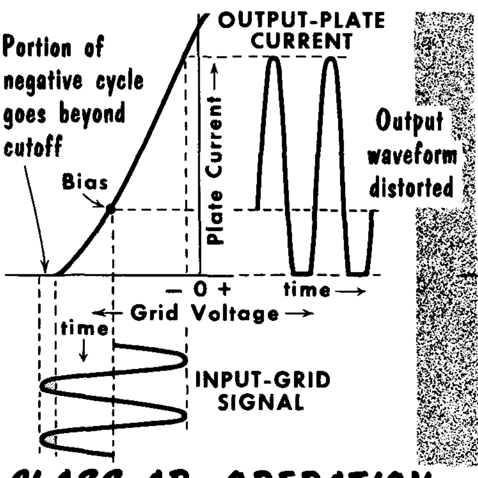
In a class-B amplifier, plate current flows during 180° of the input grid voltage cycle. That is, a class-B amplifier is biased at cutoff, so that there is no plate current flow when no signal is applied. Plate current flows only during the positive half of the input signal. Since the output plate current waveform represents only the positive half of the input signal, this class of operation cannot be used where an exact replica of the entire input must be reproduced in the output circuit. Single-ended (single-tube) class-B amplifiers are used in radio-frequency amplifier stages having a parallel-tuned circuit as the plate load. This tank circuit provides the second half of the output cycle by means of L-C charge and discharge action. In audio amplification, where the output waveform must be exactly like the input waveform for minimum distortion, two tubes must be used in "push-pull," an arrangement which we shall discuss later, where each tube supplies that half of the output waveform not supplied by the other.

Class-B amplifiers are characterized by medium power output, medium plate efficiency (approximately 50%), and moderate power amplification. Since the a-c component of plate current is proportional to the amplitude of the grid signal voltage, the output power is proportional to the square of this voltage. Being biased at plate current cutoff, the positive cycle must swing through the nonlinear portion of the dynamic characteristic, producing a certain amount of distortion in the output. When used in high-power amplifiers, class-B operation is often such that the positive cycle swings into the positive grid voltage area, and the grid draws current.

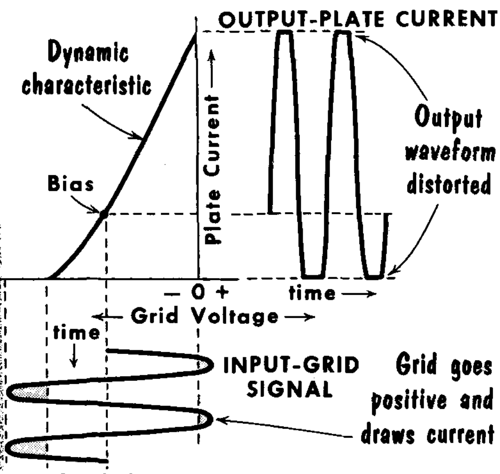
Class-AB and Class-C Operation

A class-AB amplifier operates in the region between class-A and class-B. That is, plate current in a class-AB amplifier flows for more than 180° of the input grid voltage cycle, but for less than 360°. In class-AB_1 operation, a limit is set on the input signal swing, so that the grid is not driven positive during any part of the input cycle. In class-AB_2 operation, the input signal exceeds the negative bias on the positive swing, and the grid goes positive, causing grid current to flow. Class-AB operation is essentially a compromise between the low distortion of the class-A amplifier and the high efficiency of the class-B amplifier. Single-tube class-AB operation cannot be used in audio circuits.

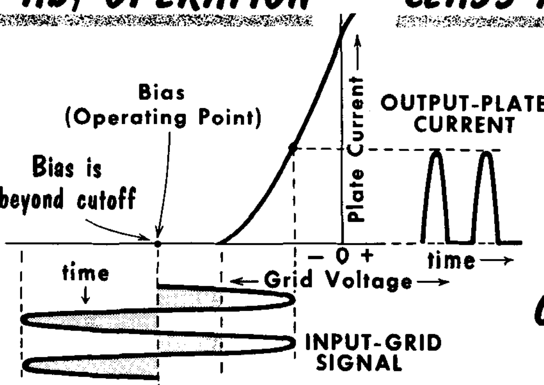
The class-C amplifier is used primarily as an r-f power amplifier in radio transmitters, and is discussed in Volume 6. The characteristic of a class-C amplifier is that plate current flows for less than 180° of the input grid voltage cycle. As a result, class-C amplifiers are noted for their extreme efficiency, upward to 80%. The high distortion of a class-C amplifier is overcome by the "flywheel" effect of tuned circuits.



CLASS AB_1 OPERATION



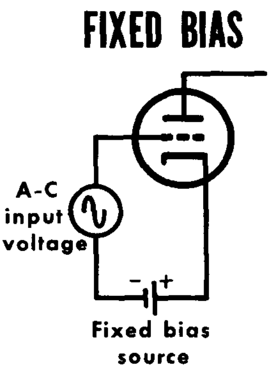
CLASS AB_2 OPERATION

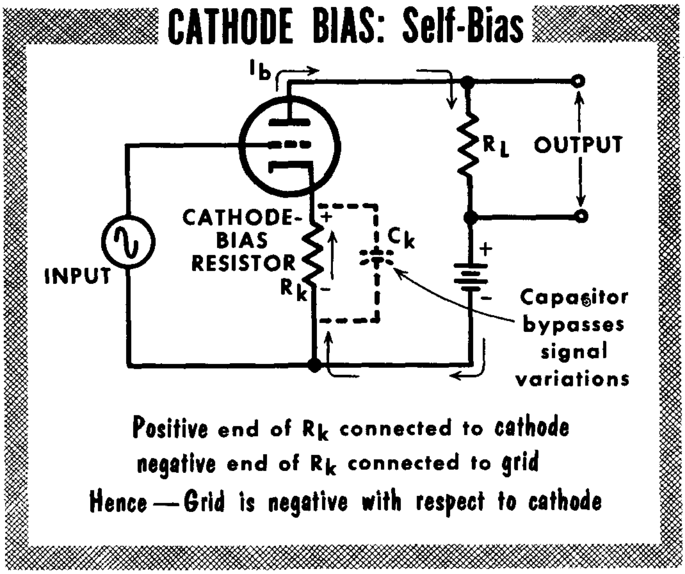


CLASS-C OPERATION

Fixed Bias and Self-Bias

In any amplifier, the location of the operating point on the dynamic characteristic curve (hence, the class of amplifier operation) depends on the d-c grid bias voltage. Basically, there are two types of bias – fixed bias and self-bias. Fixed bias is usually supplied from a separate voltage source, such as cells, or from a negative voltage tap in the power supply. The fixed bias is generally placed in series with the grid signal input. The d-c bias voltage is applied so that the grid is made negative with respect to the cathode.





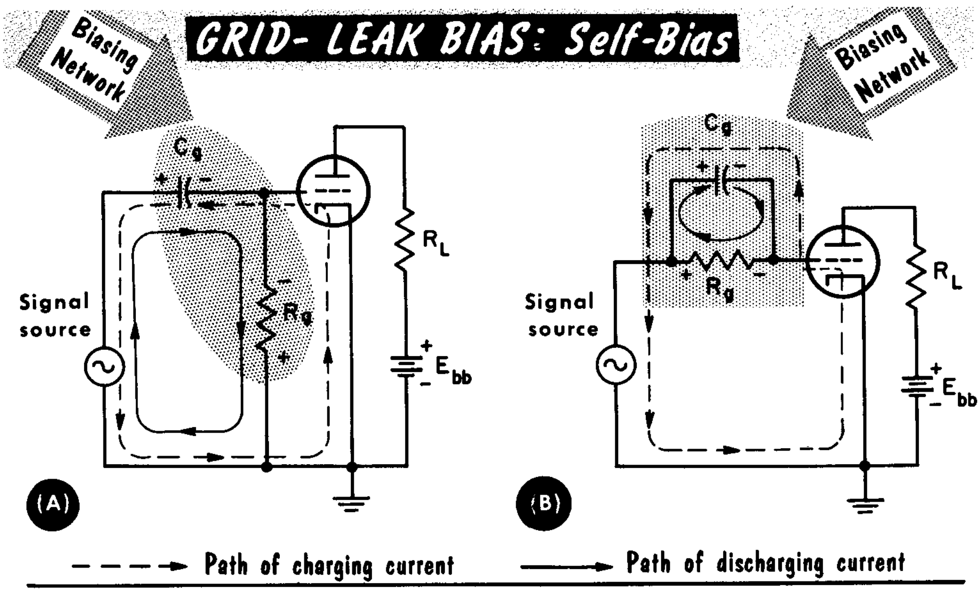
ode. The total voltage between cathode and grid is thus equal to the sum of the d-c bias voltage plus the instantaneous value of the signal voltage. We should remember, however, that the d-c bias voltage is completely independent of the signal voltage.

The most common form of self-biasing is cathode bias. This is accomplished by placing a resistor (R_k) in series with the cathode circuit so that all plate and screen current must flow through it. The voltage drop across R_k is then such as to make the grid negative with respect to the cathode. With zero signal input, the current through R_k establishes the fixed grid bias. An applied input signal results in plate current variations, and hence, variations in the voltage drop across R_k. To prevent these variations and maintain a steady d-c bias, a bypass capacitor (C_k) is placed in parallel with R_k. Basically, this is a simple R-C filter. As the voltage drop across R_k reaches maximum, C_k charges to the full wave. When the voltage drop across C_k falls off, C_k discharges across R_k to maintain the zero signal value. The overall effect, as in the case of power supply filters, is to maintain a relatively constant voltage drop across R_k. In our study of amplifier circuits, we shall discuss values of R_k and C_k.

Grid-Leak Bias

Grid-leak bias is another form of self-bias. In circuit A, the value of C_g is sufficiently large so that at the input signal frequency, its capacitive reactance is small compared to the resistance of R_g, whose value is large. No bias voltage exists for zero signal voltage. When the a-c signal voltage is applied, it appears across R_g, making the grid alternately positive and negative with respect to the cathode.

During the positive alternation of the applied signal, the grid becomes positive with respect to the cathode. It draws current that flows in the grid circuit, charging C_g to the maximum or peak value of the input voltage. The capacitor plate connected to the grid becomes negative. The charging current path is through the tube. During the negative alternation of the signal voltage, the grid becomes negative with respect to the cathode. C_g discharges slightly through resistor R_g, which has a high value. The top of R_g becomes negative with respect to the bottom. During the next cycle of the input voltage, C_g charges up again to full charge, then again slightly discharges, maintaining the voltage across R_g. Due to the action of R_gC_g, a d-c bias voltage is developed across R_g and applied to the grid of the tube.

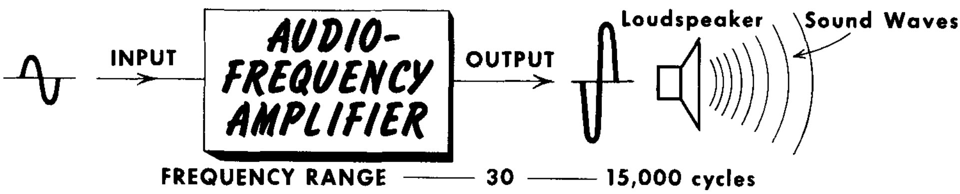


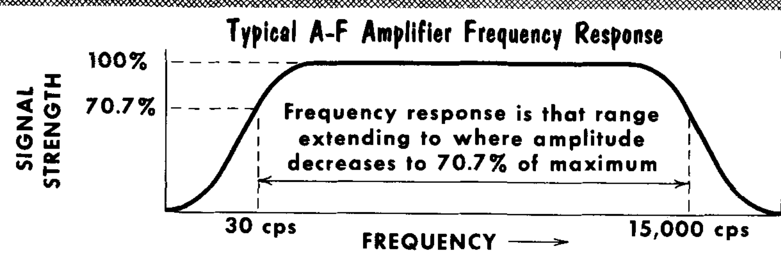
Another means of obtaining grid-leak bias voltage is to connect C_g in parallel with R_g. The operation of circuit B is similar to the one described above, but the discharge current path of C_g is through R_g alone. The charging current paths are identical. The average voltage drop across R_g is the d-c grid bias voltage. The total grid voltage is the sum of the d-c bias voltage plus the input signal voltage. In both grid-leak bias voltage circuits, the value of the grid bias voltage depends on the signal voltage and the grid resistor.

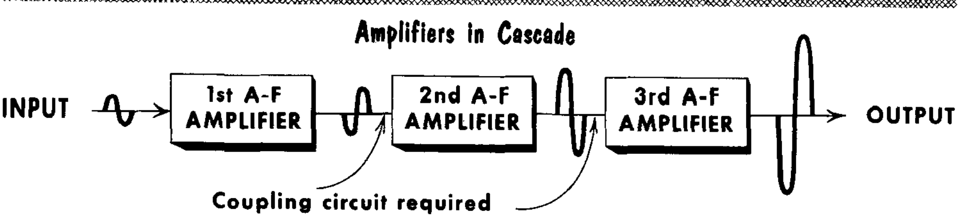
Audio-Frequency Amplifiers

Now that we have discussed the basic amplifier circuit, we can begin the study of amplifier characteristics. A common classification of audio amplifiers is audio-frequency (a-f) and radio-frequency (r-f) types. We shall take up first a-f amplifiers (sometimes called low-frequency amplifiers). Basically they are designed to amplify electrical signals of from about 30 to 15,000 cycles. These are called audio-frequencies because air moving back and forth at that rate can be "heard" by the human ear. While the range of hearing varies from person to person, 30 to 15,000 cycles represents an average hearing range. Certain animals, notably bats, have a hearing range that extends well beyond 15,000 cycles.

Two principal types of a-f amplifiers are voltage and power. Primarily, a voltage amplifier is designed to produce a large output voltage with respect to the input voltage. A power amplifier develops primarily a large signal current in the output circuit. Schematically, there is no way of distinguishing between voltage and power amplifiers except by their types of loads, a power amplifier in a radio receiver generally being used to drive a loudspeaker. In most instances, one stage of audio amplification is insufficient to accomplish most needs. Audio amplifiers must be cascaded; that is, the output of one feeds into the input of a second. The arrangement of transferring electrical energy from one stage to another is called coupling, and as we shall see, the type of coupling used greatly affects the amplifier frequency characteristics.

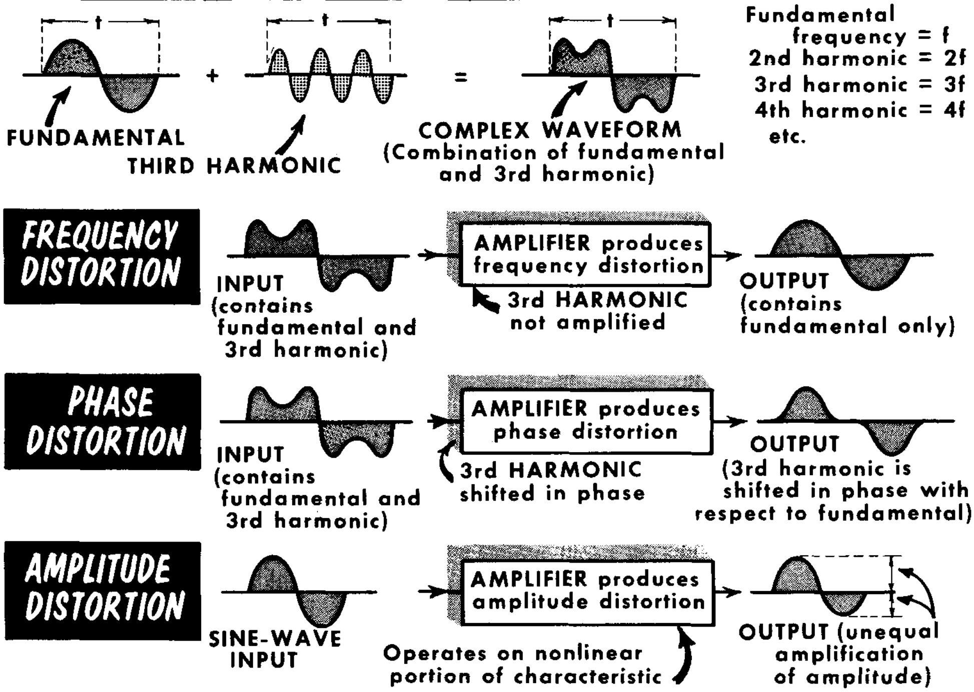






Amplifier Distortion

COMPLEX WAVES CONTAIN TWO or MORE FREQUENCY COMPONENTS



Before embarking on our study of coupling, let us give further thought to the nature of the signal being amplified. Regardless of whether the input signal voltage consists of a single sine waveform or a complex wave containing many frequencies, the function of an amplifier is to strengthen the signal without introducing any distortion in the process. Three types of distortion that may occur in amplifiers are: frequency, phase, and amplitude or non-linear distortion.

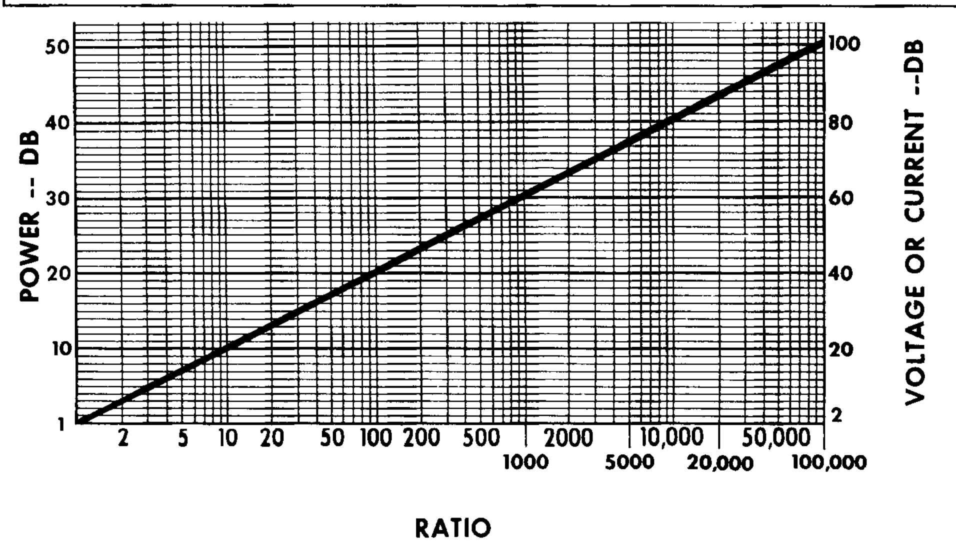
<u>Frequency distortion</u> occurs when some frequency components of a signal are amplified more than others. For example, a signal consisting of a fundamental and a third harmonic may pass through a two-stage amplifier which introduces frequency distortion by which only the fundamental has been amplified, and the third harmonic component does not appear in the output. In <u>phase distortion</u>, the output waveform is considerably different from the input because the phase of the third harmonic has been shifted with respect to the fundamental. If a signal is passed through a vacuum tube operating on any nonlinear part of its characteristic, <u>amplitude distortion</u> occurs. In this region, any change in grid voltage does not result in a change in plate current, which is directly proportional to the change in grid voltage.

The Decibel

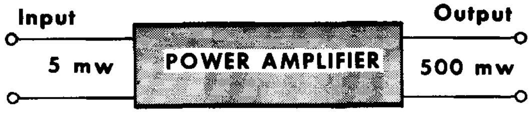
Various voltage, current, and power gains and losses in electronic equipment are often spoken of in terms of decibels. This is an outgrowth from the telephone field, where the decibel was used as a mathematical expression that represented the property of the human ear to respond to ratios of sound intensity. Mathematically, the decibel is a logarithmic ratio, and is used in electronics as a shorthand notation for power ratios. The decibel is a relative unit of measurement that originally was used to express changes in audio power, and the ability of the human ear to recognize these changes. For instance, if an amplifier produced an output power through its loudspeaker of 10 watts, it would then have to increase its output power 10 times to 100 watts for our ear to detect twice as much "loudness" (2 is the logarithm of 100).

Mathematically, we define the decibel as being equal to $10 \times \log_{10}$ of P2/P1, where P2 is always the larger power in watts, and P1 is always the smaller power in watts. The expression "$\log_{10}$" merely describes the type of logarithms used, generally called common logarithms. Thus, it should be understood that the decibel is only a unit of comparing two levels; it does not give an absolute value for either. If we compare two voltages, we say that the decibel gain or loss is equal to $20 \times \log_{10}$ of (E2/E1). This is because P has relationships of E^2/R and I^2R. Hence, currents would be written as $20 \times \log I2/I1$. This process is simplified by using the table below. For instance, let us assume that the output power of an amplifier is twice the input power. We look up the ratio of 2 in our table and find that it corresponds to 3 db. We then say the amplifier has a power gain of 3 db. The table is designed to read voltage, current, and power db projected from the ratio axis. When we have a loss, it is represented as -db.

DECIBELS AND POWER, VOLTAGE, OR CURRENT RATIOS



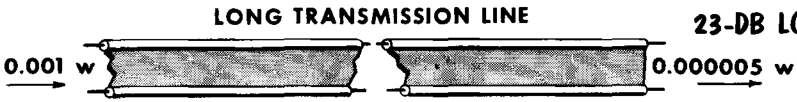
PRACTICAL EXAMPLES ILLUSTRATING the USE of DECIBELS



20-DB GAIN

$$\frac{P2}{P1} = \frac{500}{5} = \text{Ratio of } 100$$

POWER GAIN

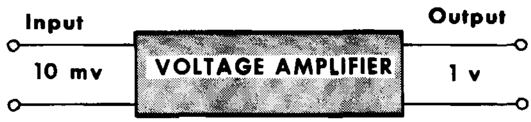


23-DB LOSS

$$\frac{P2}{P1} = \frac{0.001}{0.000005} = \text{Ratio of } 200$$

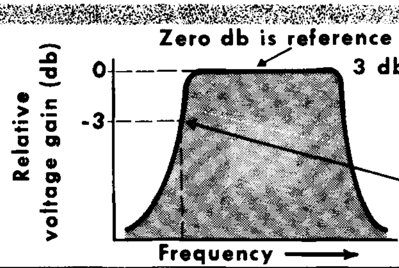
POWER LOSS

40-DB GAIN



$$\frac{V2}{V1} = \frac{1}{0.01} = \text{Ratio of } 100$$

VOLTAGE GAIN



3 db voltage loss = ratio of $\frac{1.41}{1}$

Loss of 3 db means curve has fallen off down to ratio of $\frac{1.41}{1}$ or to 71% of maximum

VOLTAGE LOSS

TABLE OF DB FOR P, E, AND I RATIOS

	RATIO			RATIO			RATIO	
DB	P	E or I*	DB	P	E or I*	DB	P	E or I*
1.0	1.26	1.12	4.4	2.75	1.66	7.8	6.03	2.46
1.2	1.32	1.15	4.6	2.88	1.70	8.0	6.31	2.51
1.4	1.38	1.17	4.8	3.02	1.74	8.2	6.61	2.57
1.6	1.44	1.20	5.0	3.16	1.78	8.4	6.92	2.63
1.8	1.51	1.23	5.2	3.31	1.82	8.6	7.24	2.69
2.0	1.58	1.26	5.4	3.47	1.86	8.8	7.59	2.75
2.2	1.66	1.29	5.6	3.63	1.91	9.0	7.94	2.81
2.4	1.74	1.32	5.8	3.80	1.95	9.2	8.32	2.88
2.6	1.82	1.35	6.0	3.98	1.99	9.4	8.71	2.95
2.8	1.91	1.38	6.2	4.17	2.04	9.6	9.12	3.02
3.0	1.99	1.41	6.4	4.36	2.09	9.8	9.55	3.09
3.2	2.09	1.44	6.6	4.57	2.14			
3.4	2.19	1.48	6.8	4.79	2.19	10.0	10.00	3.16
3.6	2.29	1.51	7.0	5.01	2.24	20.0	100.00	10.00
3.8	2.40	1.55	7.2	5.25	2.29	30.0	1,000.00	31.60
4.0	2.51	1.58	7.4	5.50	2.34	40.0	10,000.00	100.00
4.2	2.63	1.62	7.6	5.75	2.40	50.0	100,000.00	316.00

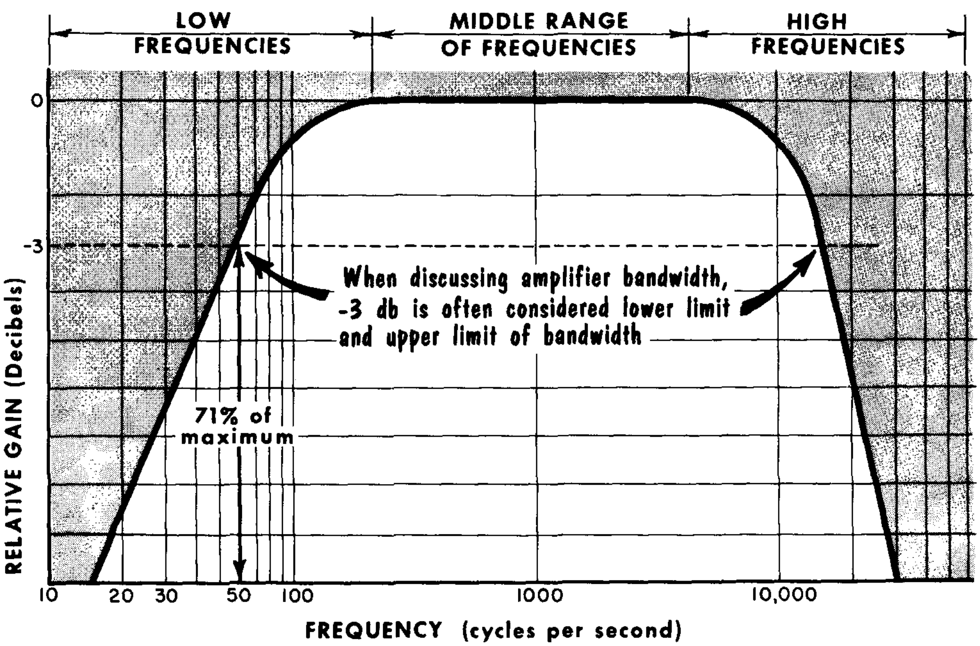
*May be used only when input and output impedances are equal.

Frequency Response

It is seldom that an amplifier is called upon to handle a single frequency. Much more often, it must handle a wide range of frequencies. Because of various factors, to be discussed later, most amplifiers have a characteristic of being able to amplify a certain middle range of frequencies relatively evenly, and then providing less amplification for both the lower and higher range of frequencies. If we draw a graph of this, using the vertical axis to represent relative gain, and the horizontal axis to represent frequency, we will get what is known as a frequency-response curve. It is really a "picture" of how an amplifier will amplify a wide range of frequencies. The vertical axis can be measured in terms of current, voltage, or power, or it could measure relative gain in terms of db.

The horizontal axis is usually logarithmically spaced rather than linearly spaced. That is, the distance from 10 cycles to 100 cycles is the same as the distance from 100 to 1000, or 1000 to 10,000. The reason for this type of scale is that the ratios of the frequency are important, and a very wide range of frequencies may thus be shown on the graph. When the logarithmic frequency scale is used, the frequency response for frequencies as low as 20 cycles can be shown, as well as for 20,000 cycles. This is not true of a linear frequency scale. Very often, the vertical axis is measured in terms of decibels, with the maximum amplification at the middle frequencies equal to 0 decibels. Thus, as the curve falls off at the high and low frequencies, the graph indicates how much db loss takes place at these frequencies.

TYPICAL FREQUENCY RESPONSE CURVE FOR AUDIO AMPLIFIER



The dynamic characteristics of a tube can be described as a graphic portrayal of tube behavior under load conditions.

When a signal source and a load are used, the following mathematical relationships hold true: $e_c = E_{cc} + e_g$; $e_b = E_{bb} - i_b R_L$; $e_{R_L} = i_b R_L$.

A load line displays the way in which the output of the plate supply voltage is distributed between the load and the internal resistance of the tube under different conditions of plate current.

The dynamic transfer characteristic curve correlates the plate current – grid voltage relationship with the load present in the circuit.

The total change in plate voltage is always equal to the change in voltage across the load.

Nonlinearity in electron tubes is a source of distortion and is highly undesirable.

If an amplifier is to operate without grid current, the amount of fixed negative bias must be at least equal to the peak value of the positive half of the input grid signal.

The effect of interelectrode capacitance can lead to feedback and oscillation.

Gain is the ratio of output to input. The greater the number of stages in an amplifier, the greater the gain.

An a-f amplifier amplifies frequencies in the audio range.

The amount of distortion in an amplifier depends partially on the linearity of its dynamic characteristics. The more linear the characteristic, the less the distortion.

A Class-A amplifier is one in which plate current flows all the time. The input grid signal operates along the most linear portion of the dynamic characteristic.

Plate efficiency is the ratio of a-c power output developed across the load to the d-c power supplied to the plate.

Grid-leak bias is obtained through the action of the input grid signal.

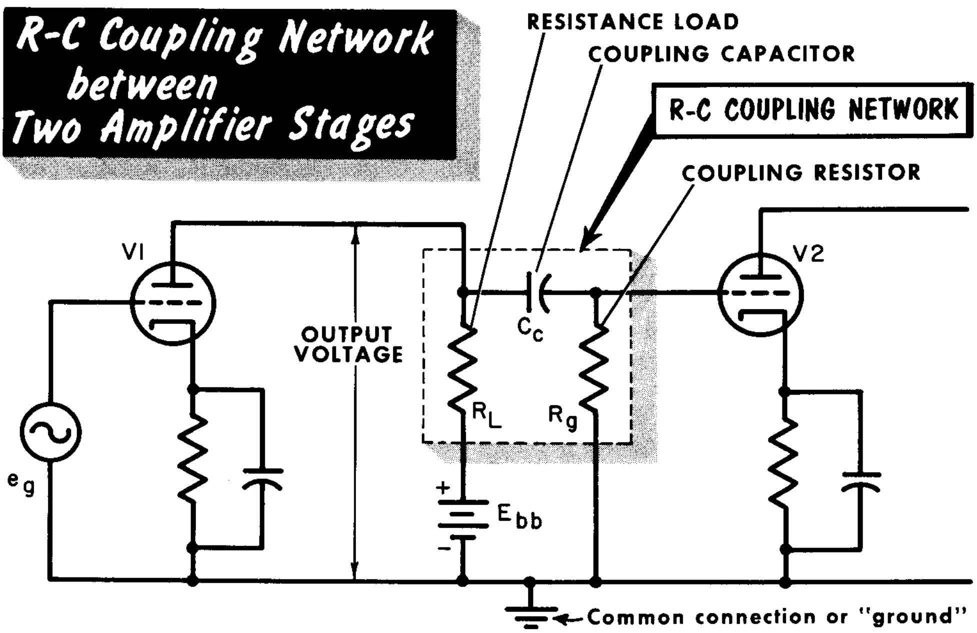
In an electron tube circuit, the a-c grid voltage, plate current, and voltage drop across R_L are all in phase, and 180° out of phase with plate voltage.

REVIEW QUESTIONS

1. What is a dynamic transfer characteristic curve?
2. How would you go about drawing a load line for a given set of grid-family curves?
3. What does a load line indicate?
4. Why does nonlinearity of characteristic curves cause distortion? How can it be minimized?
5. Is interelectrode capacitance more noticeable at higher or lower frequencies? Why?
6. What is meant by the operating point of a tube?
7. How is cathode bias developed?
8. How is grid-leak bias developed?
9. In a triode amplifier, what is the phase relationship between grid voltage, plate current, and plate voltage?
10. Explain Class-A operation.
11. What is meant by an amplifier's frequency response?
12. Compare the plate efficiencies of Class-A, -AB, -B, and -C amplifiers.

Resistance-Capacitance Coupling

The most common type of coupling network for transferring electrical energy from one circuit to another, is the R-C, or resistance-capacitance type. This is generally known as resistance coupling. Our diagram shows two triode amplifiers coupled by an R-C coupling circuit. When a varying signal voltage e_g is applied to the grid of V1, it causes the plate current to vary through the tube and through R_L. The changing current through R_L produces a varying voltage drop across it. The output signal of V1 is the varying voltage between its plate and ground. This output voltage is equal to the fixed plate supply voltage minus the varying voltage across R_L. It is desirable to make R_L as large as possible. As this resistance is increased (within limits), a larger signal voltage appears across it. As a result, the output voltage from V1 is increased, and the stage is said to have greater amplification. There is a limit, however, to the value of R_L; if it is made too large, it produces an excessive d-c voltage drop. This reduces the plate voltage on the tube and the resultant plate current, so that the V1 output is reduced. Typical values for

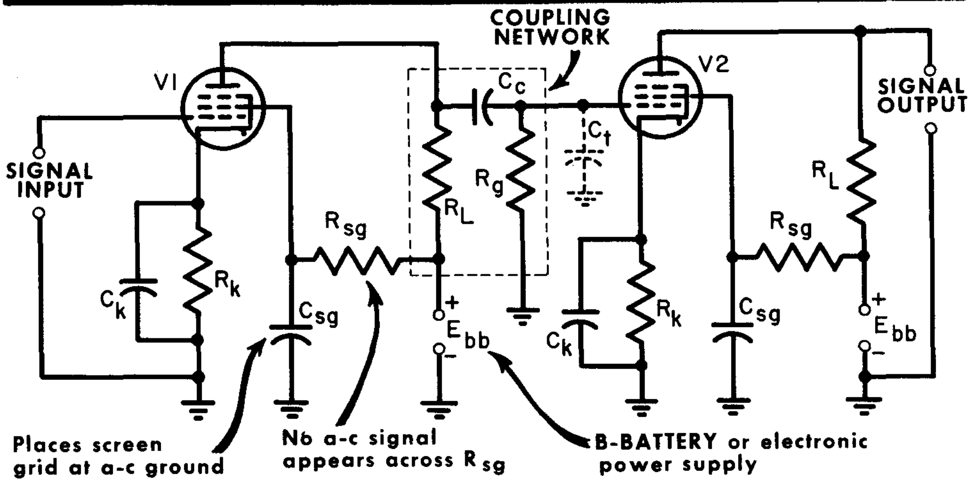


R_L range from about 25,000 to 500,000 ohms, with the larger resistances used in pentode circuits where the internal plate resistance of the tube is very high, requiring a very high R_L for proper voltage gain.

The output signal of V1 is coupled through coupling capacitor C_c. This capacitor blocks the high positive plate voltage of V1 from being applied to the grid of V2. Because C_c blocks or prevents the passage of dc, it sometimes is referred to as a blocking capacitor. Since this capacitor must pass the varying signal voltage easily, its reactance should be low. Typical values of C_c range from 0.001 to approximately 0.1 μf. Larger values normally are not used because of their excessive stray capacitance to ground.

Resistance-Capacitance Coupling (Cont'd)

R-C COUPLED AMPLIFIER USING PENTODES



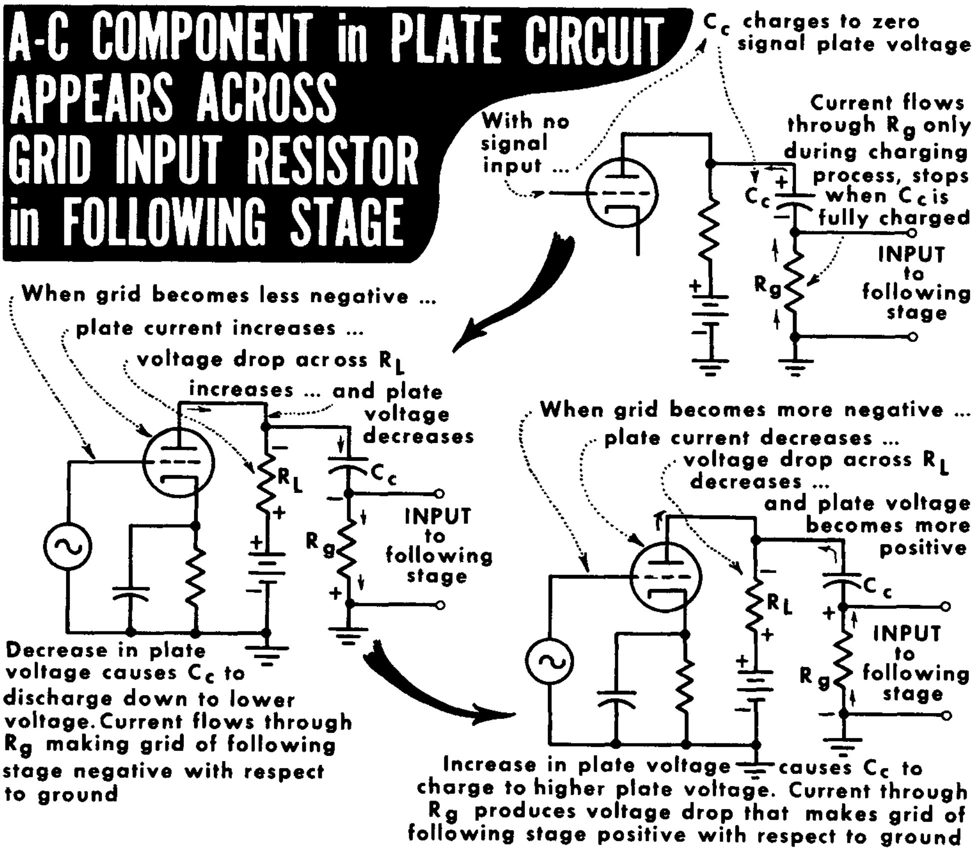
The a-c signal coupled through C_c is applied to grid resistor R_g of V2. The V2 input signal voltage drop across R_g is applied in series with the cathode bias voltage between grid and cathode of V2. R_g has other uses beside acting as a load across which the input signal to V2 is developed. R_g also provides a grid return for the grid of V2. That is, it connects the grid of the tube to this source of bias voltage – in this case, to one end of the cathode resistance. In this way the grid does not float. R_g also provides a discharge path for C_c, preventing an improper accumulation of electrical energy. Because R_g provides a path through which electrons can leak off C_c, it is often referred to as a grid-leak resistor. R_g can also be used as a source of grid-leak bias, and, as we shall learn, contact potential bias. Typical values for R_g range from 0. 5 megohm to several megohms.

C_c and R_g form an a-c voltage divider. The output of this is the voltage drop across R_g which becomes the actual signal input to V2. Because of the infinite reactance of C_c to dc, all of the d-c voltage drop appears across it, and no dc from the plate of V1 is applied to the grid of V2. The reactance of C_c at audio frequencies is made much smaller than the resistance of R_g. Thus, very little of the a-c signal is lost across C_c and most all of it appears across R_g. C_t represents the total shunt capacitance of the circuit. It is a stray capacitance made up of the interelectrode capacitance of the tube, and wiring capacitances.

R-C coupling is also commonly used for pentode circuits, but screen grids of pentodes obtain their d-c operating voltages from series-dropping resistors R_{sg}. Capacitor C_{sg} bypasses any a-c signal voltage that appears in the screen-grid circuit, thereby preventing it from causing a fluctuation in screen-grid voltage. C_{sg} provides a virtual short-circuit for a-c signals from screen to ground.

Action in an R-C Circuit

To fully comprehend coupling action in R-C coupling, we must think of C_c and R_g as being in parallel with R_L, so far as a-c is concerned. While we see R_L in series with the plate supply voltage, in actual practice this supply would be bypassed by electrolytic filter capacitors in the power supply filter. So effectively, with respect to a-c, the low end of R_L is at ground potential. Thus any variation in voltage across R_L appears across C_c and R_g in series. From our illustration, we can see that coupling is nothing more than a continual charge and discharge action of C_c as it tries to maintain the voltage that is on the plate of the preceding stage. In so doing, an a-c voltage is

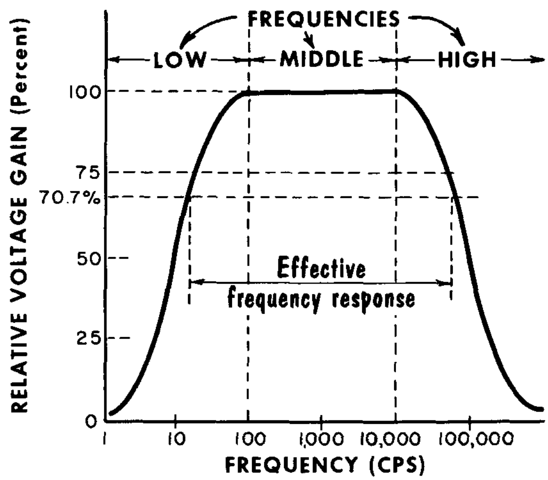


developed across R_g which becomes the input voltage to the following stage. In a-f amplifiers, the capacitance of C_c must be made sufficiently large so that its reactance will be extremely low when compared to the resistance of R_g. We want an a-c voltage drop only across R_g; this is useful. A voltage drop across the reactance of C_c is wasted. Hence, the reactance of C_c at the lowest audio frequency to be passed should be less than 10% of the value of R_g. As can be seen, the capacitance of C_c becomes a low-frequency limitation of a resistance-coupled amplifier.

Gain of R-C Coupled Amplifier

The gain of an amplifier, or its ability to amplify a signal, is generally greater at one frequency than another. By plotting the gain of an amplifier at various frequencies, we obtain a frequency response curve or response characteristic which indicates the gain of an amplifier over a wide range of frequencies. Frequency is plotted along the horizontal or X axis, while relative voltage gain is plotted along the vertical or Y axis. As we see, the response of an R-C coupled a-f amplifier is flat over the middle range of frequencies and then falls off at either end. The low-frequency tapering is caused by the increase in reactance of C_c. As the frequency to be amplified gets lower and lower, the reactance of C_c gets higher and higher. In the meantime, the resistance of R_g remains constant. Therefore, a greater and greater portion of the output signal is being dropped across the capacitive reactance of C_c and less signal appears across R_g, which represents the input source of the following stage. At the high-frequency end, the interelectrode and wiring capacitances begin to take on a shunting effect, and more and more of the output voltage is bypassed to ground; hence, amplifier gain decreases.

VOLTAGE GAIN CHARACTERISTIC of an R-C COUPLED A-F AMPLIFIER



GAIN OF A TRIODE R-C AMPLIFIER

A A-C plate current (i_p) $= \dfrac{\mu e_g}{r_p + R_L}$

B Output voltage (e_{out}) $= i_p \times R_L$

C Substituting for i_p from (A) $e_{out} = \dfrac{\mu e_g}{r_p + R_L} \times R_L$

D Voltage gain $A = \dfrac{e_{out}}{e_g}$

E Thus, substituting for e_{out} in (C) we get:

$$A = \dfrac{\dfrac{\mu e_g R_L}{r_p + R_L}}{e_g}$$

F Simplifying, we get: $A = \dfrac{\mu R_L}{r_p + R_L}$

GAIN OF A PENTODE R-C AMPLIFIER

Relationship: $\mu = g_m \times r_p$

Substituting for μ in (F) we get: $A = \dfrac{g_m \times r_p \times R_L}{r_p + R_L}$

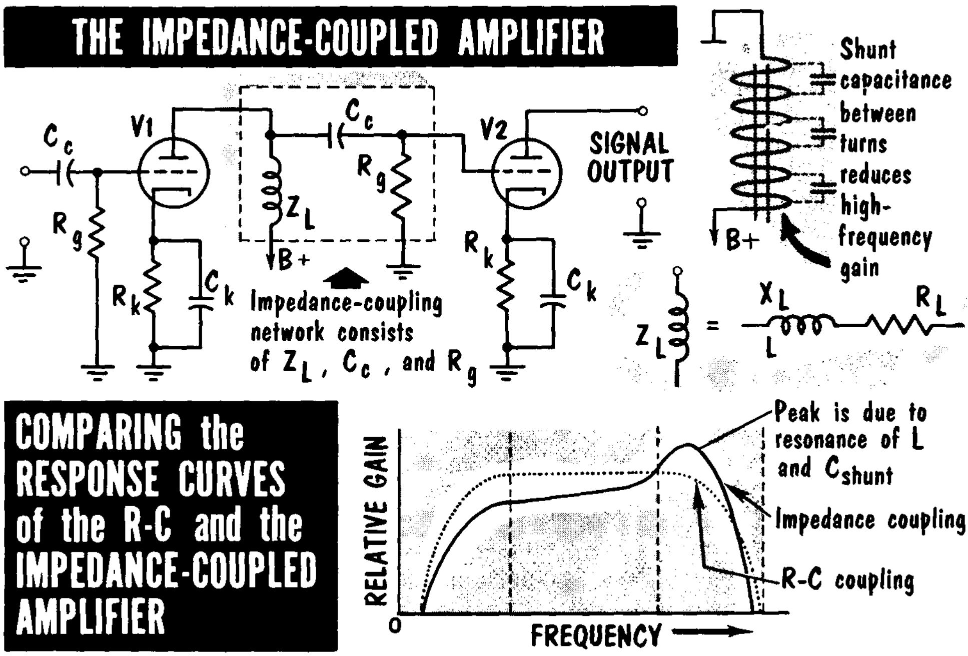
Since in a pentode is much larger than

We simplify the formula to: $A = \dfrac{g_m \times r_p \times R_L}{r_p}$

The r_p factor cancels out and we have: $A = g_m R_L$

for the middle range of frequencies

Impedance Coupling

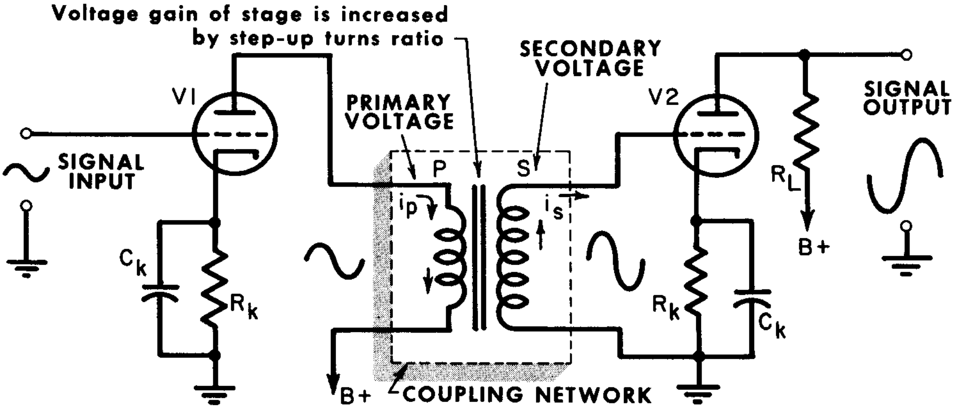


In the impedance-coupled amplifier, the plate load is an inductor; hence, instead of having a resistive load, we have an inductive load. Since all coils contain a certain amount of resistance, we refer to the load as an impedance, Z_L. To obtain as much amplification as possible, particularly at the lower frequencies, the inductance is made as large as is practical. A closed-shell type of inductor is generally used to avoid undesirable magnetic coupling. Because of the low resistance of an inductive load, the d-c voltage drop across it is small, and a greater amount of the supply voltage is available at the plate of the tube. The important characteristic of an inductive load is that its impedance changes with frequency. At low frequencies, its reactance is low; as the input frequency increases, the reactance (and hence, impedance) increases. As a result, impedance coupling is generally used in amplifiers where a relatively narrow range of frequencies are to be amplified.

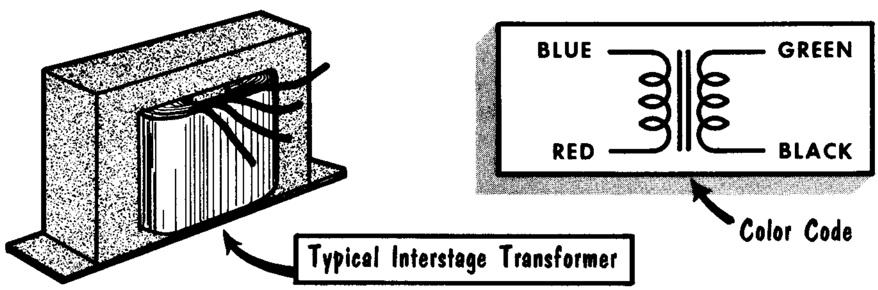
The gain of the impedance-coupled amplifier is the ratio of the signal voltage drop developed across Z_L to the input signal voltage. Hence, the low-frequency limitation of this type of coupling is in the decreasing reactance of Z_L and the increasing reactance of C_c. At the higher frequencies, the shunt capacitance between the turns of the inductor reduces the gain, as well as the shunt capacitance of the circuit itself. In some situations, shunt capacitance may be sufficient to resonate with the inductance to cause a high peak to occur in the frequency response curve. The outstanding feature of this type of amplifier is that almost all of the plate supply voltage appears on the plate with very little wasted as a voltage drop across the load impedance. The cost of an inductor places this type of coupling at a disadvantage to R-C coupling.

Transformer Coupling

A transformer-coupled stage has certain advantages over other types of coupling. For one thing, the voltage amplification of this stage may exceed the amplification of the tube, if the transformer has a step-up turns ratio. Another advantage is that the grid of the following tube is completely protected from the d-c plate supply voltage through the isolated primary and



TRANSFORMER-COUPLED AMPLIFIER

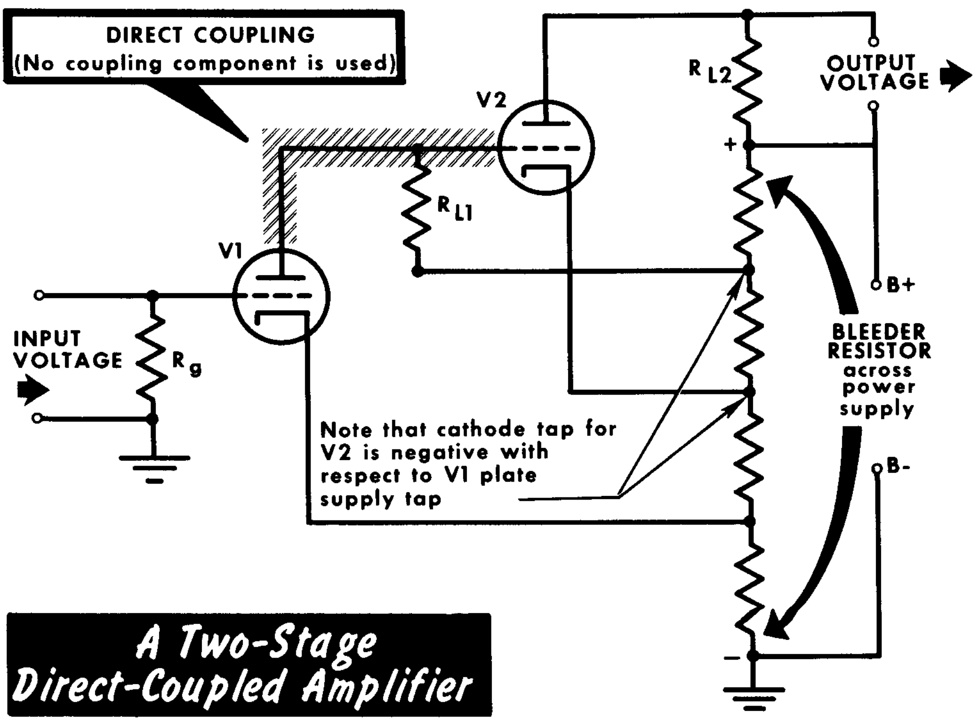


secondary windings. Transformer coupling provides much of the frequency characteristics of impedance coupling in that the primary is an inductor. The frequency response curve would show that a transformer-coupled voltage amplifier has a relatively high gain and uniform frequency response over the middle range of frequencies, but poor response at both the low and high audio frequencies. The low- and high-frequency limitations are much the same as those encountered in the impedance-coupled amplifier. Transformer coupling has the additional disadvantages of greater cost, greater space requirements, and the necessity for greater shielding. Transformer coupling is seldom used in voltage amplifiers because its frequency response is poor compared to the R-C coupled amplifier. The additional voltage gain available by a step-up turns ratio is generally not sufficient compensation for its poorer frequency response.

Direct Coupling

In the coupling circuits we have studied so far, the coupling device isolates the d-c voltage of the plate circuit from the d-c voltage of the grid circuit, allowing only a-c components of the output to pass through the coupling device. In the direct-coupled amplifier, the plate of one tube is connected directly to the grid of the next tube without any intervening capacitor, transformer, or other coupling device. Since the plate of the first tube must have a positive voltage with respect to its cathode, and the grid of the next tube must have a negative voltage with respect to its cathode, proper circuit operation demands the use of a special voltage divider. This is obtained by using a multitap bleeder resistor across the output of the power supply.

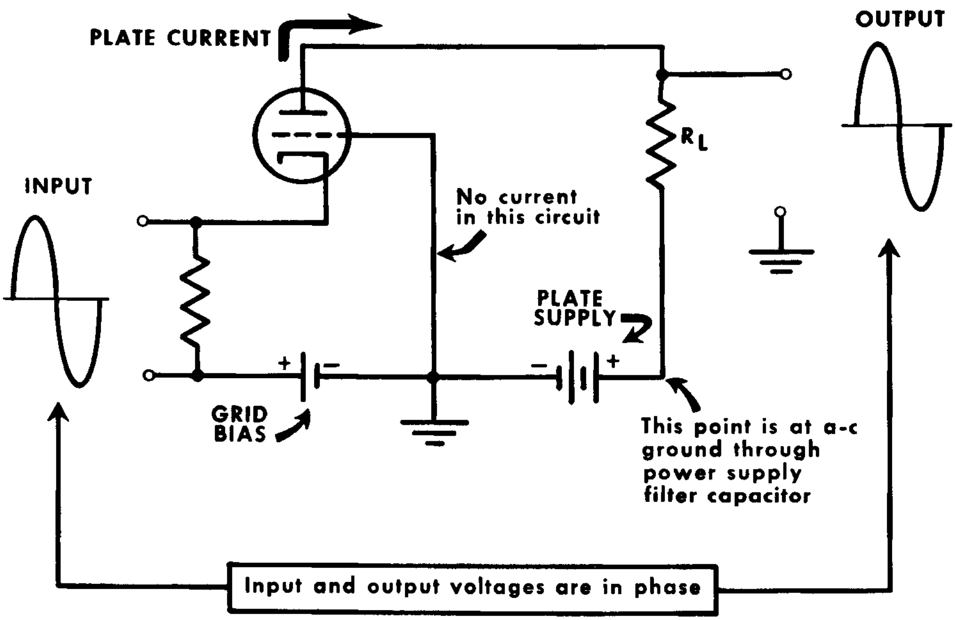
This type of circuit is particularly effective with low-frequency amplification because the impedance of the coupling element does not vary with frequency. Thus, a direct-coupled amplifier is capable of amplifying very low variations of signal input voltage. Note that the plate load resistor of the first stage acts as the grid resistor for the following stage. When the tube voltages are properly adjusted to give Class-A operation, the circuit serves as a distortionless amplifier whose response is uniform over a wide frequency range. The particular disadvantage of the direct-coupled amplifier is in the severe demands made on the power supply to ensure voltage stability. Any change in the various voltages produces drift. This is especially troublesome in high-gain amplifiers, since drift that takes place in the input stage is then amplified before appearing in the output circuit.



A Two-Stage Direct-Coupled Amplifier

The Grounded-Grid Amplifier

All amplifiers discussed so far have made use of a common or grounded cathode; that is, the input signal is delivered into the grid cathode circuit with the output taken from the plate cathode circuit. Very often used at higher frequencies is the grounded-grid amplifier. In this circuit, the input voltage is applied to the cathode with the grid being grounded, and the output is taken from the plate and is in phase with the input signal. If the input voltage is applied between cathode and grid, it is the same as placing an opposite voltage between grid and cathode. (Making the grid 5 volts negative with respect to the cathode is the same as making the cathode 5 volts positive with respect to the grid.) No current flows in the grid circuit because it is negative with respect to the cathode and repels all electron flow. Current flows in the cathode circuit and it is the same current as flows in the plate circuit. In the grounded-cathode arrangement, there is no current in the input circuit; hence, the a-c input resistance is extremely high. In the grounded-grid amplifier, there is current flow, and the input resistance is quite low, for example, 2000 ohms. The control grid acts as a grounded shield between the output circuit, plate to ground, and the input circuit, cathode to ground; thus, electrical energy transfer is largely avoided between the input and output circuits through the capacitances of the tube. In the grounded-grid amplifier, the input voltage is 180° out of phase with the plate current, and is thus in phase with the output or plate voltage signal which appears across the plate load resistor.

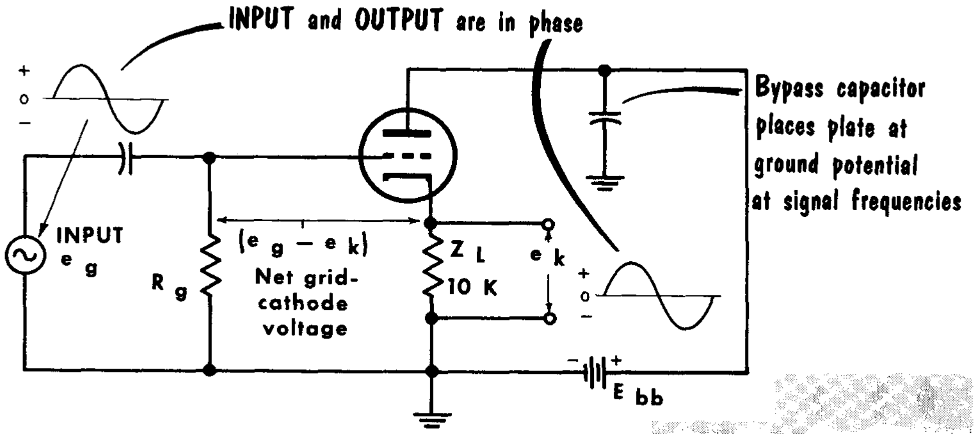


The GROUNDED-GRID AMPLIFIER

The Cathode Follower

CATHODE FOLLOWER (GROUNDED-PLATE) CIRCUIT

Provides no voltage gain



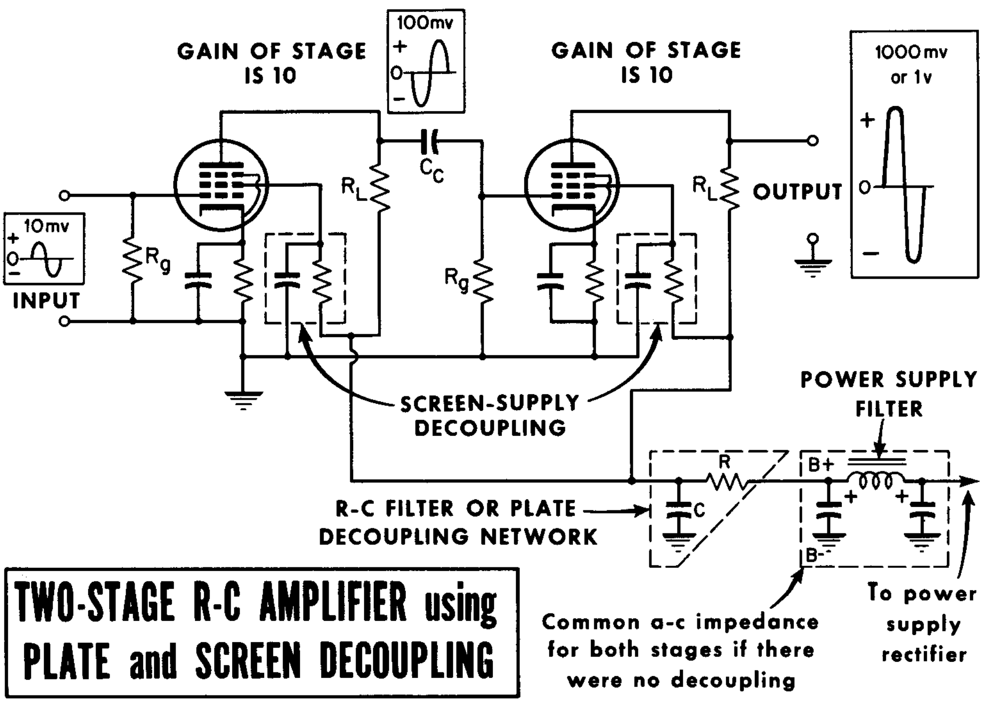
The cathode follower is essentially a grounded-plate amplifier in which the load impedance is located in the cathode circuit and the plate is at a-c ground through a bypass capacitor. The cathode follower is generally operated as a Class-A amplifier, the output of which appears across the unbypassed cathode resistor. This introduces degeneration (discussed later) and prevents the circuit from producing a voltage gain. The grid circuit, in drawing no grid current, presents a high input impedance. The load impedance, which may be anywhere from 50 to 20,000 ohms, presents a low output impedance. This circuit then becomes ideal for matching a high-impedance source to a low-impedance load.

The advantages obtained by the use of a cathode follower exist only at the price of a voltage gain which is less than unity or 1. However, a power gain is possible. The name cathode follower is derived from the output voltage which follows the input voltage; that is, the output voltage not only has the same waveform, but also the same instantaneous polarity (phase). Note that the amplifier output voltage e_k developed across Z_L is in series with the cathode and grid, and thus has a polarity that opposes the input signal e_g. Thus, the net grid-cathode voltage is $e_g - e_k$. Hence, the voltage e_k developed across the load impedance will always be somewhat less than the signal voltage e_g. However, voltage e_g can be developed across a high impedance, whereas E_k exists across a relatively small load impedance. In this way, the cathode follower acts as an impedance transformer in which power amplification can be obtained at the same time as the impedance level is reduced. The cathode follower has excellent response, especially at higher frequencies.

Decoupling and Bypassing

When amplifier stages are cascaded (that is, the output of one stage is fed into the input of the next stage), precautions have to be taken in circuits that are common to both stages. If the signals from one stage are not isolated from the signals of a previous stage, electrical energy may be fed back in such phase as to oppose the amplification of a signal. This undesirable feedback can be eliminated by isolating one stage from another in those areas where feedback can take place easily. The most common location for feedback problems is in the B+ lines - those lines feeding the screen grids and plates. Here, the signals from the various circuits are fed into the common power supply together with the d-c components. If signals from two successive stages produce voltage drops 180° out of phase across the common power supply impedance, undesirable feedback is produced. This can be eliminated by the use of decoupling networks.

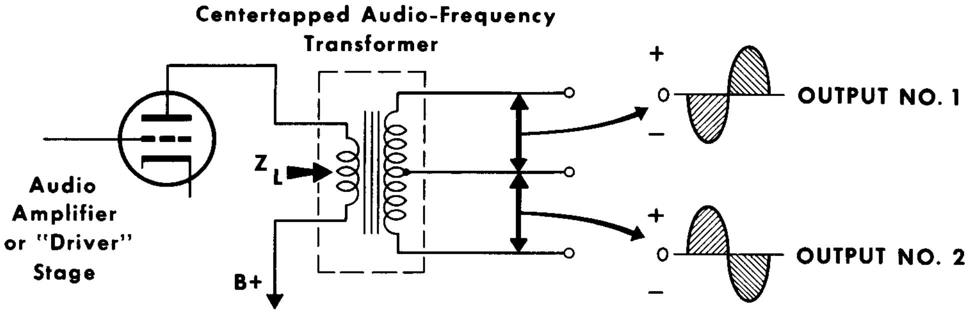
The most common form is a simple R-C filter circuit in series with the plate load of a stage, in which the a-c component is bypassed to ground. To be effective, the reactance of the decoupling capacitor should be no more than 10% of the decoupling resistance at the lowest frequency to be handled by that circuit. Should the reactance of the decoupling capacitor become too great, a significant portion of the a-c component would take the return path through the decoupling resistor and the power supply to ground, rather than through the capacitor. When this occurs, degeneration takes place across the power supply impedance. In some instances, energy may be fed back in phase to produce undesired oscillation.



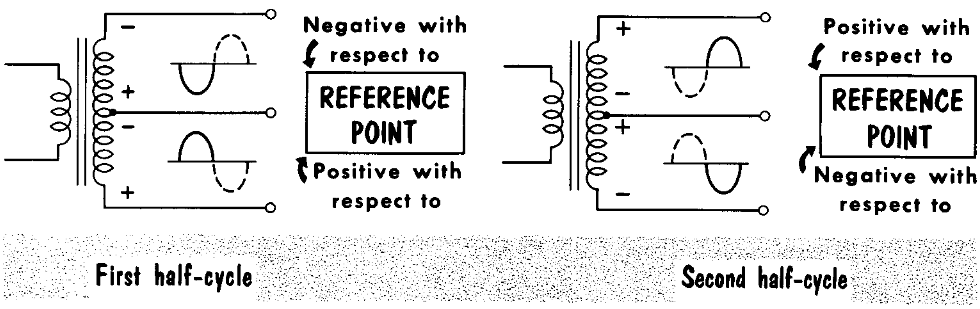
TWO-STAGE R-C AMPLIFIER using PLATE and SCREEN DECOUPLING

Transformer-Type Phase Inverter

CENTERTAPPED SECONDARY provides TWO EQUAL and OPPOSITE SIGNALS



Phase Relationships in Secondary

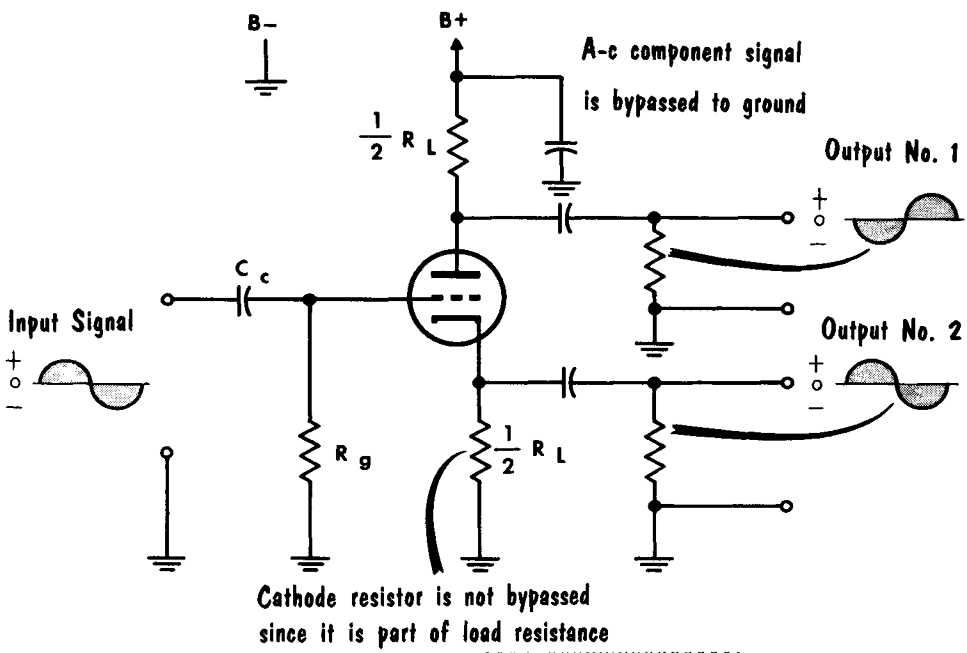


A very popular audio-frequency power amplifier is the push-pull type, which we shall discuss later. This circuit requires two input signals that are identical in every respect, except that one is 180° out of phase with the other. Various circuits are used to obtain these inputs. However, the simplest arrangement is through the use of an a-f interstage transformer in which the secondary is centertapped. The primary of this transformer, often called an input transformer, acts as the plate load impedance of the previous a-f voltage amplifier or "driver" stage. The secondary has more turns than the primary, and hence, provides a voltage step-up. The centertapped secondary assures that the voltages developed between the centertap and each end of the secondary are equal. Also, since the ends of the secondary are 180° out of phase, with the centertap as a reference point, the voltage developed between the centertap and the top half of the secondary will be 180° out of phase with the voltage developed between the centertap and the bottom half of the secondary. With the ends of the secondary connected to the two grids of the push-pull stage, the centertap must go to the negative side of the grid bias source; very often this is ground. The feature of this system is its simplicity. The secondary centertap provides two equal and opposite voltages.

The Split-Load Resistor

One simple way to make a phase inverter is to connect one half of the plate load resistance between B+ and plate, and the other half between cathode and ground. Since these resistances are equal and the same current flows through both, each produces the same d-c voltage drop and the same audio fluctuations. For example, if the input signal to the phase inverter goes

SPLIT-LOAD PHASE INVERTER



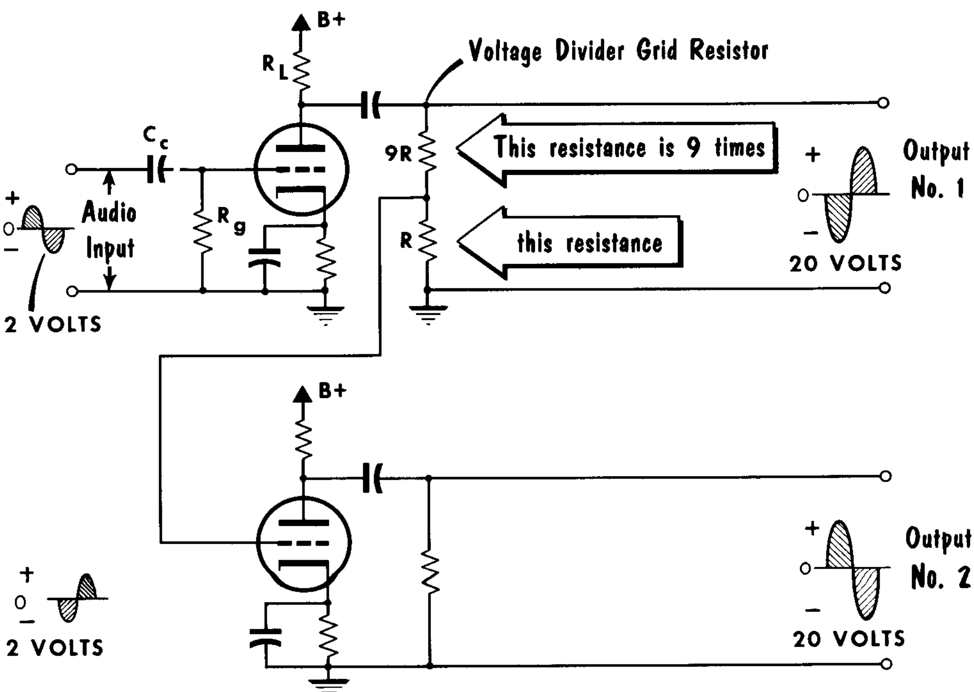
positive, it produces an increase in plate current. This causes an increase in voltage drop across the plate resistor, making the plate less positive with respect to ground. At the same time, this same plate current flowing through the cathode resistor produces the same increase in voltage drop and makes the cathode more positive with respect to ground. Taking our two outputs from the plate and cathode with respect to ground, we have equal and opposite signals. This circuit is simple, gives good balance, and provides good frequency response. Its disadvantages are that it provides no amplification and has a small signal-handling capacity.

The Paraphase Amplifier

Another phase-splitting arrangement is the so-called paraphase circuit. Basically, it uses two tubes. The output from the plate of one tube is fed by R-C coupling to one grid of the push-pull circuit. From this same point, a voltage-divider arrangement reduces the voltage and applies it to the grid of the second paraphase tube, which amplifies the voltage by as much as the resistance divider reduced it, producing an output voltage for driving the second push-pull grid. For instance, a positive fluctuation of 2 volts at the grid of the first paraphase tube may produce a negative fluctuation on its plate of, say, 20 volts. This 20 volts represents the output No. 1 of the paraphase amplifier. This 20-volt fluctuation is also divided to produce a negative fluctuation of 2 volts for the grid of the second paraphase tube which, in turn, becomes a positive fluctuation of 20 volts at the plate of the second tube. The positive 20-volt fluctuation represents output No. 2 from the circuit.

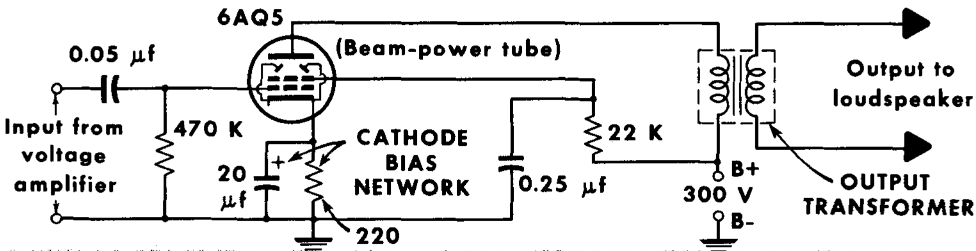
For this circuit to operate correctly, the voltage division produced by the resistor feeding the second paraphase tube must be in exactly the same ratio as the gain provided by the second paraphase tube. In the example given, the tube multiplies by 10 and the voltage divider divides by 10. This circuit has good handling capacity, gives some gain, and provides good frequency response. Its principal disadvantage is that the balance between the two outputs is sometimes difficult to maintain due to variations in circuit components.

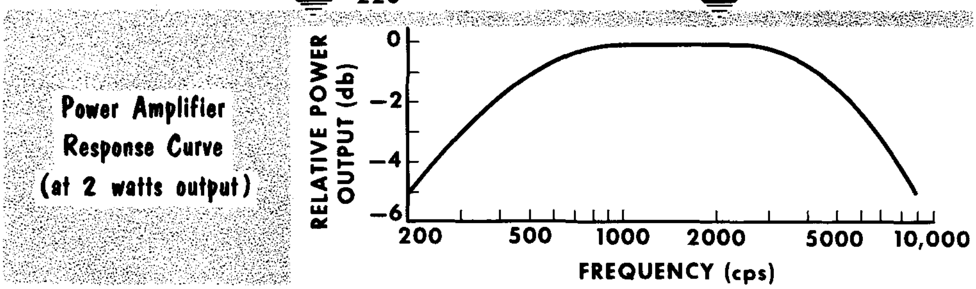
CIRCUITRY of the PARAPHASE AMPLIFIER



Power Amplifiers

TYPICAL CLASS-A AUDIO-FREQUENCY POWER AMPLIFIER used as output stage in radio receivers





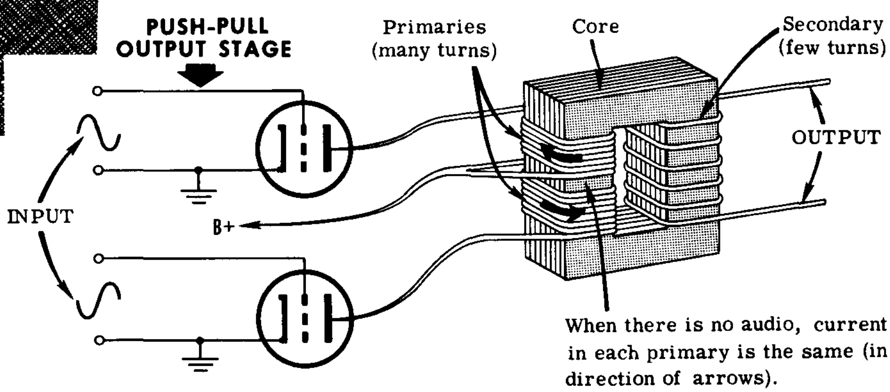
The primary function of the voltage amplifiers just discussed is to increase the voltage of a signal to a higher value without distorting the waveform. Normally, voltage amplifiers consume no power from the preceding stage nor do they supply any appreciable power to the following stage; they merely provide voltage gain. The primary function of a power amplifier, however, is to deliver power to a load, any accompanying increase in signal voltage being of secondary importance. Output power is proportional to the square of the grid voltage. Hence, the power amplifier must usually be preceded by one or more stages of voltage amplification, to raise the signal to the proper value for operating or driving the power stage. In radio receivers, the power amplifier is used as the audio-output stage to drive a loudspeaker, which is considered the load. When used as a single tube in a-f amplification, the power amplifier must be Class-A operated.

Tubes used in audio power amplification are generally of the beam power type, capable of high power sensitivity and high plate current. The typical a-f power amplifier, used as the output stage of a receiver, looks much the same as a voltage amplifier, except that the plate load impedance is the primary of the output transformer. Conventional cathode bias is generally used. Tubes such as the 6AQ5, 6L6, and 6V6, are capable of providing in excess of 5 watts output to the loudspeaker. Since the human ear is not particularly sensitive to distortion below about 5%, this much may be allowed in the output signal. The term "undistorted" output refers to distortion of less than 5%. Maximum undistorted power output is often achieved when the load impedance is approximately twice the plate resistance of the tube, and the plate current variations are at the maximum permissible value for Class-A operation.

Push-Pull Amplifiers

The next step in improving the power output capacity of an amplifier stage is to use two tubes in a connection known as push-pull. This arrangement uses transformer coupling, but there are two primaries (the primary winding has two halves) through which the current flows in opposite directions. B+ is connected to the center point of the primary, with the plate of one of the tubes connected to each end. The current therefore, flows from each plate outward through an equal number of turns to the center point. This means that the total magnetizing effect on the core of the transformer is neutralized as far as the dc is concerned. (The transformer core only has to carry the magnetization due to the audio fluctuation.) This simplifies the design and cost of the transformer, but the big advantage is in the tube operation.

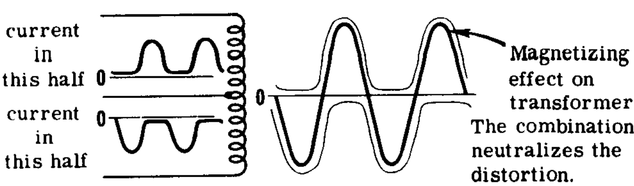
With a single tube, matching the output load to the tube plate resistance results in a poor output waveform, which is rounded at the bottom and sharpened at the top. When the tubes are worked in push-pull, the current flows in opposite directions around the transformer core, and consequently, what is the top of the current waveform in the upper part of the winding, becomes the bottom of the current in the lower half of the winding. Thus both halves of the current waveform have a sharpened portion added to a rounded portion, and the effect averages out, producing a much better waveform for the load value used. To achieve this, we must provide the correct audio voltages at the grids of the tubes. We shall consider this problem presently.



When there is no audio, current in each primary is the same (in direction of arrows).

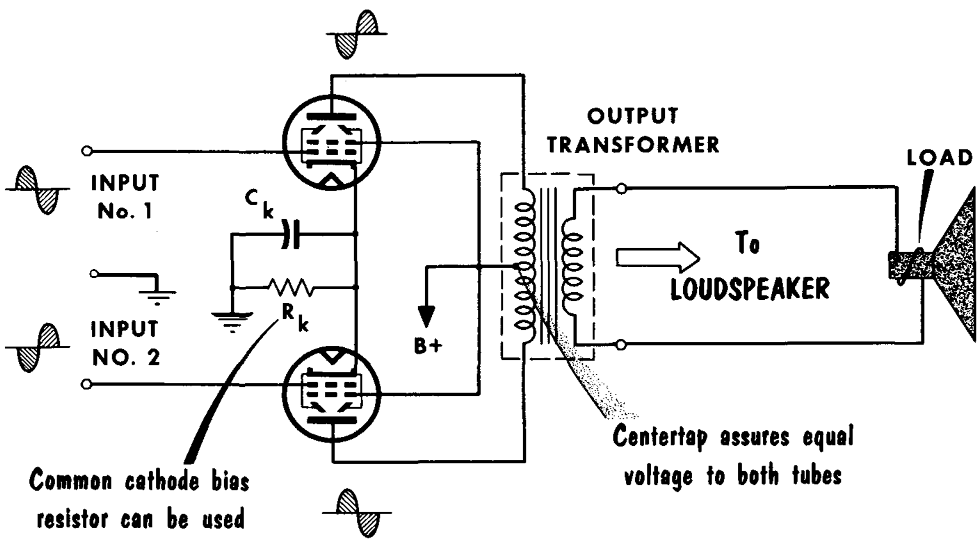
When audio signal is fed into output stage, current in one primary rises, while current in the other falls. On next half-wave, opposite primary currents rise and fall.

Push-Pull reduces distortion by neutralizing curvature



The Practical Push-Pull Circuit

BASIC PUSH-PULL AMPLIFIER

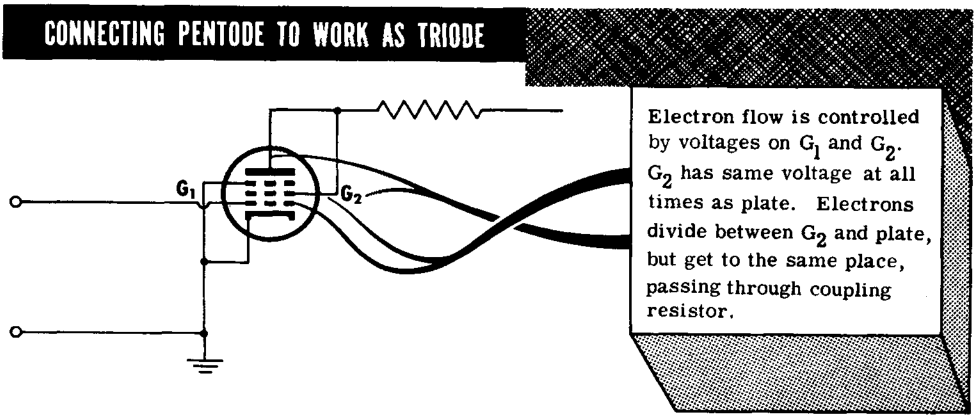


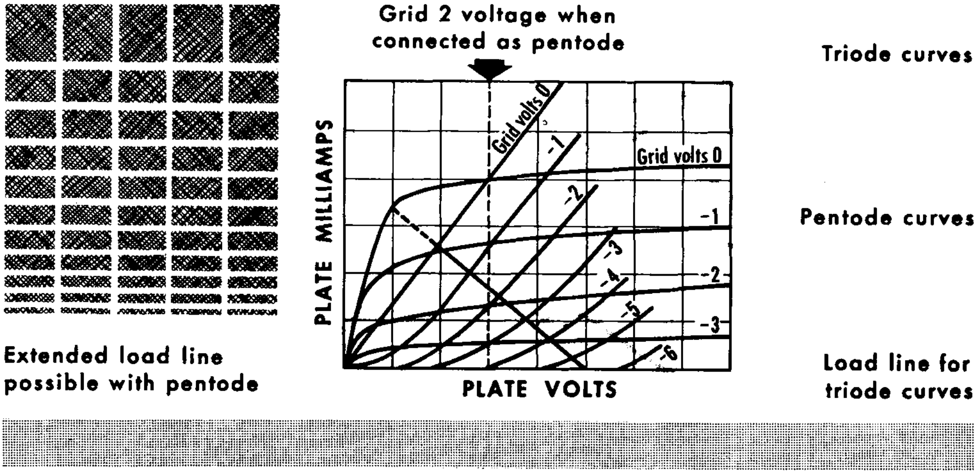
A number of advantages are to be gained by the use of a push-pull amplifier as the output stage of an a-f amplifier. Second harmonics, and all even-numbered harmonics, as well as even-order combinations of frequencies, will be effectively eliminated if the tubes are properly balanced, and if the frequencies are introduced within the output tubes themselves. Hum from the plate power supply, which may be present in the single-tube amplifier, is substantially reduced in the push-pull amplifier because ripple components in the two halves of the primary transformer are in phase, and tend to counteract each other in the output. Plate current flow through the two halves of the primary winding is equal, and in opposite directions. Therefore, there is no d-c core saturation, and the low-frequency response is improved. Regeneration is also eliminated because signal currents do not flow through the plate voltage supply when the circuit is operated as a Class-A amplifier.

The last voltage amplifier preceding the push-pull power amplifier stage may be either resistance- or transformer-coupled to the power stage. If the power amplifier is operated Class-A or Class-AB, the driver commonly employs resistance coupling because it affords a better frequency response. A phase-inverter tube, or section of a tube, must be used in connection with the resistance-coupled driver to provide the correct phase relation at the input of the push-pull stage. When the power tubes are operated Class-B, an input transformer employing a step-down turns ratio is commonly used. The transformer not only supplies the grid current necessary for Class-B operation, but at the same time permits an instantaneous signal voltage of the correct polarity to be applied to the grids of the two power tubes. Class-B power amplifiers draw practically no plate current when no signal is applied, and their plate efficiency is much higher than that of Class-A amplifiers.

Triode Connection of Pentodes

Any pentode-type tube can be made to work as a triode-type tube. This is done by connecting the second grid directly to the plate, so that both swing together at the same voltage. Because the screen grid, in combination with the control grid, is principally responsible for controlling the plate current, the presence of the suppressor grid between the screen grid and plate does not materially alter the tube's performance from that of a triode.



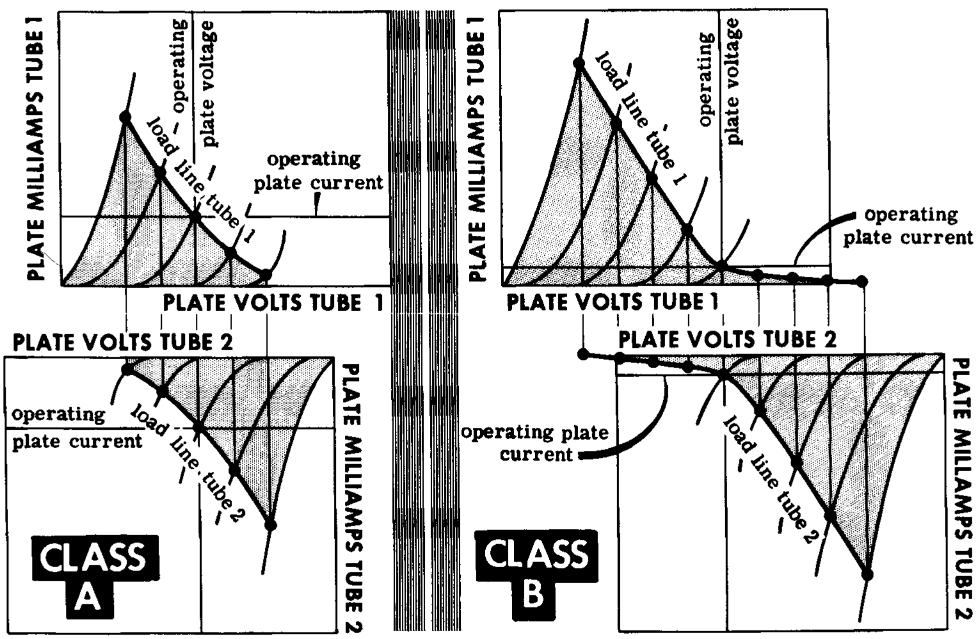


With the best possible load resistance with triode-connected tubes, the voltage fluctuation between B+ and plate reaches little more than half the B+ supply voltage. Changing the method of connection to pentode alters the curve so that the zero grid voltage curve is pushed out into a "knee." This extends, very considerably, both the voltage and current fluctuation available in the plate circuit, which, in turn, triples or quadruples the power that any pair of tubes will give.

The Lowering of Distortion

Using two tubes in push-pull helps the waveform problems, so that the distortion produced by one tube cancels that of the other. This can be understood better if we think of each tube as having a curved load line. The input voltages to the grids are equal but 180° out of phase. The plate voltages likewise are out of phase because of the coupling between the two primary windings of the output transformer. So the changes in plate current must adjust between the tubes to allow this condition, while the two of them supply the total current fluctuation to the load at all points. The ratio between total voltage and current fluctuations of both tubes is set by the load resistance matched to the secondary of the transformer, but each tube feeds a load resistance whose value is constantly changing, as represented by the curves.

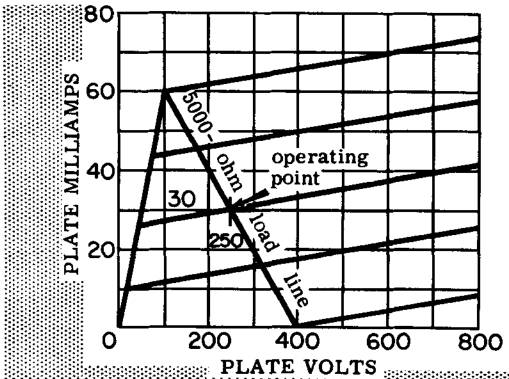
This effect can be extended further, to increase the efficiency of the output stage. Normally the steady plate current is about half the maximum plate



current (which occurs when the grid voltage fluctuation goes from the operating point up to zero). The current fluctuation in the tube at maximum power level swings between almost zero current and twice the steady current. This sets a considerable limitation on the power-handling capacity of the tube because the steady component is such a large proportion of the maximum current the tubes take. Using a greater negative bias on the grids of the tubes makes the audio fluctuations carry the current from almost zero up to a maximum in one direction, and cuts the tube off so that no current flows in the other direction. This makes possible a considerable increase in efficiency and available power output.

Power Output from a Class-A Stage

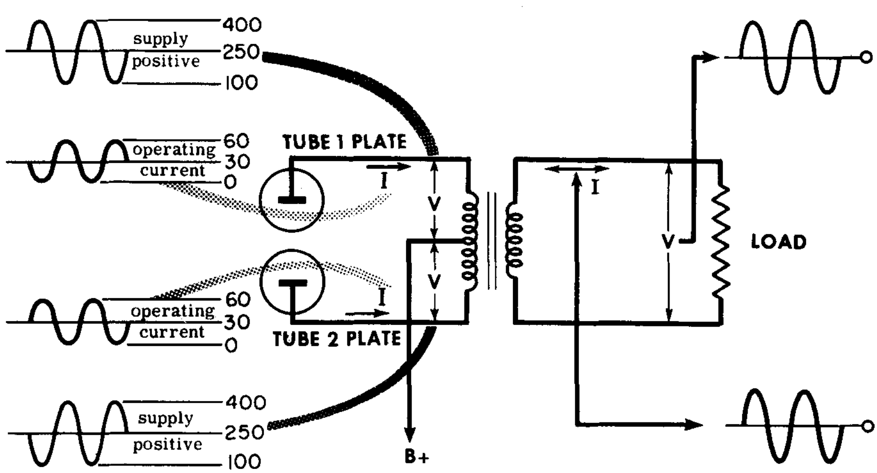
As an example, suppose that in ordinary push-pull (or Class-A with both tubes conducting current all the time), the operating point for each tube is 250 volts at 30 milliamperes, and that the load value presented to each tube is 5000 ohms with pentode operation. Disregarding the curves to make the calculation simpler (if approximate), the audio fluctuation should carry the plate between 100 volts at 60 milliamperes, and 400 volts at zero milliamperes.



Calculations for Class A push-pull

Load line is for each tube, but both work at once.

The characteristics for the other tube (not shown) would be inverted underneath.

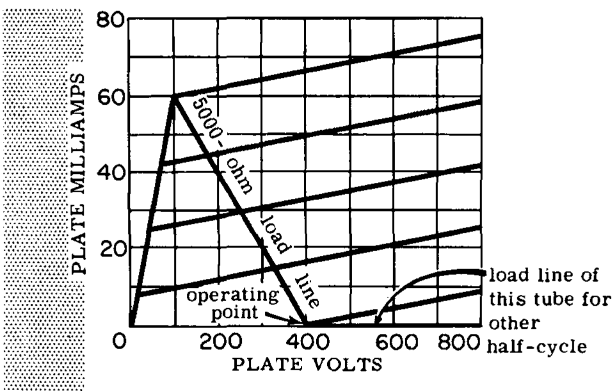


This represents a peak fluctuation from each tube of 150 volts and 30 ma in each direction, which is a peak power of (150×0.03) or 4.5 watts per tube, or 9 watts for the two tubes in push-pull. The average power, using a sine wave to drive the output, will be half this figure (4.5 watts for the two tubes).

Class-B Operation

If we wish to use an extreme economy measure, known as Class-B operation, we bias each tube for approximately zero current. (Actually it does not go quite to zero current because of the curvature of the characteristic, but it goes to where zero current would be if the tube characteristics were all straight. (This is called projected cutoff.) Using the same load line, the operating point for each tube would, theoretically, be 400 volts at zero milliamperes. This means that the plate potential voltage will swing from 100 volts to 700 volts. Plate current will swing, during one half-cycle from zero to 60 milliamperes and back, while in the other half-cycle, no current flows in that tube.

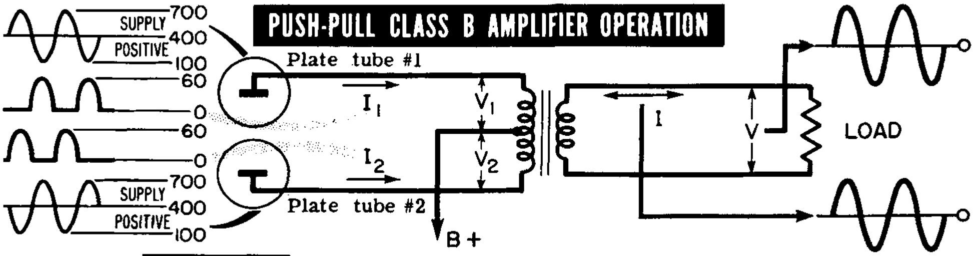
The permissible maximum voltage on the plate is considerably increased by this method of operation (from 400 volts to 700 volts). There is less danger of breakdown between the plate and some other electrode when no plate current is flowing. There is, however, a maximum rated voltage even under this condition, which sometimes restricts the amount by which this method of operation can improve efficiency.

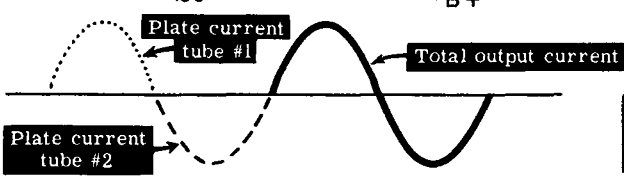


Calculations for Class B push-pull

Load line is for each tube, but only one works at a time on alternate halves of waves.

PUSH-PULL CLASS B AMPLIFIER OPERATION

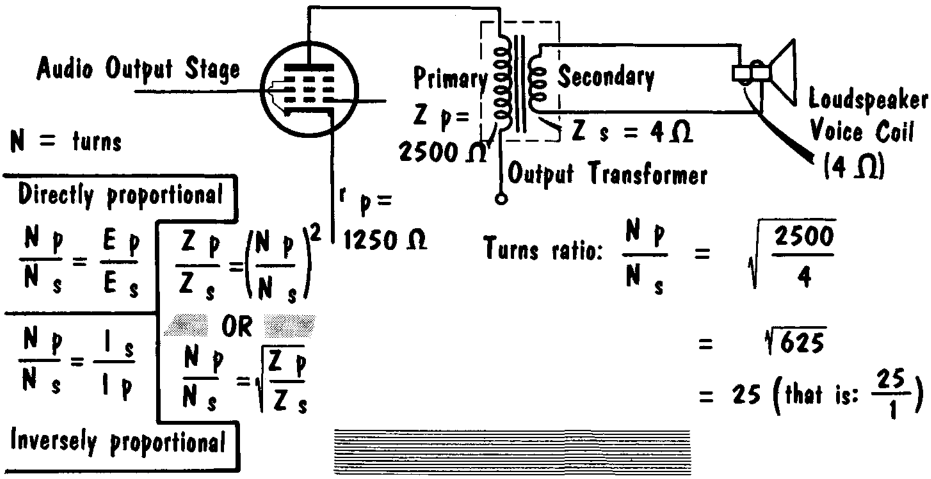




Current Waveforms for push-pull class B amplifier operation

Impedance Matching

OUTPUT TRANSFORMER PROVIDES CORRECT IMPEDANCE MATCH



The output transformer used with a-f power amplifiers serves as an impedance-matching device. Since the plate resistance of a power amplifier tube may range from perhaps 1000 ohms to more than 20,000 ohms, and since the impedance of the loudspeaker may range down to 4 ohms, the output transformer has a step-down turns ratio to provide the correct ratio of primary voltage and current to secondary voltage and current. The ratio of the two impedances that a transformer can match is equal to the turns ratio squared.

As a practical example, let us find the turns ratio needed for the transformer shown in our illustration. Since the plate resistance is 1250 ohms, the primary impedance is considered as twice this value, to permit maximum undistorted power output. The power fed to the 4-ohm voice coil, however, will be reduced unless the proper impedance is afforded by the transformer. The turns ratio between the primary and secondary that satisfies this condition is 25. The amount of power that can be handled by a transformer is determined by the current and voltage ratings of the windings. The primary frequency contains a d-c component that limits its inductance and frequency response. In a given transformer, the induced voltage is proportional to the frequency and the flux density. At low frequencies, the flux density is high and more distortion is introduced because of the saturation of the iron. The maximum allowable flux density is determined by the allowable distortion.

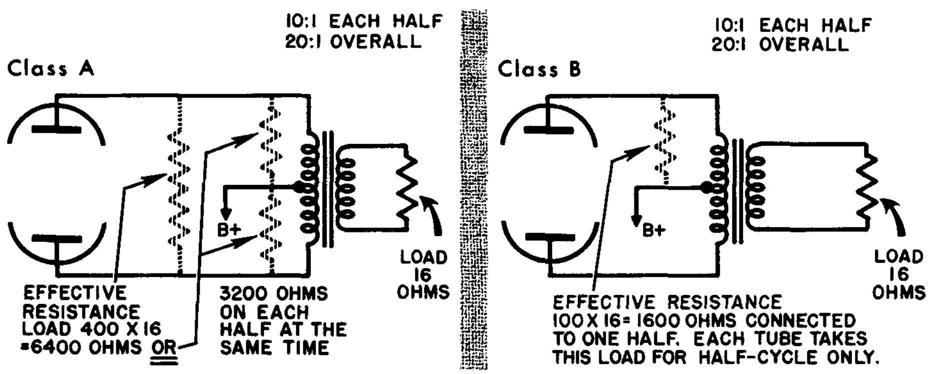
The output transformer causes a reduction in the output of a power amplifier at both the high and low frequencies. The reduced output at the low frequencies results from the shunting action of the transformer primary inductance on the load. The reduced output at the high frequencies results from the loss in voltage from the leakage reactances as a result of load current and capacitive current due to shunting capacitance.

The Output Transformer

The most common form of audio-frequency transformer is used in the output of an a-f power amplifier to match the circuit load impedance (usually a loudspeaker) to that required by the output tubes. The transformer here serves the additional purpose of avoiding both supply and audio losses, because the winding resistances are low compared to their respective impedances. The way impedance reflects in a push-pull transformer depends to some extent on the way tubes are operated. In Class-A, both tubes are delivering part of the power throughout the cycle, so the load is shared between them. If the ratio makes 16 ohms actual impedance equivalent to 6400 ohms at the primary, each tube has a load of 3200 ohms average. But in Class-B, only one half of the primary works at a time. The other is inactive for that half-cycle because its tube is cut off. The impedance transformation is based on the ratio to each half winding. If the whole ratio is 20:1, this is 10:1 each half. So 16 ohms connected to the secondary makes a load of 1600 ohms for each tube, but the tube takes the load for only half a cycle. A further advantage of push-pull operation in the transformer is that the magnet-

OUTPUT TRANSFORMER

MATCHES HIGH-IMPEDANCE VACUUM TUBE OUTPUT TO LOW-IMPEDANCE LOAD



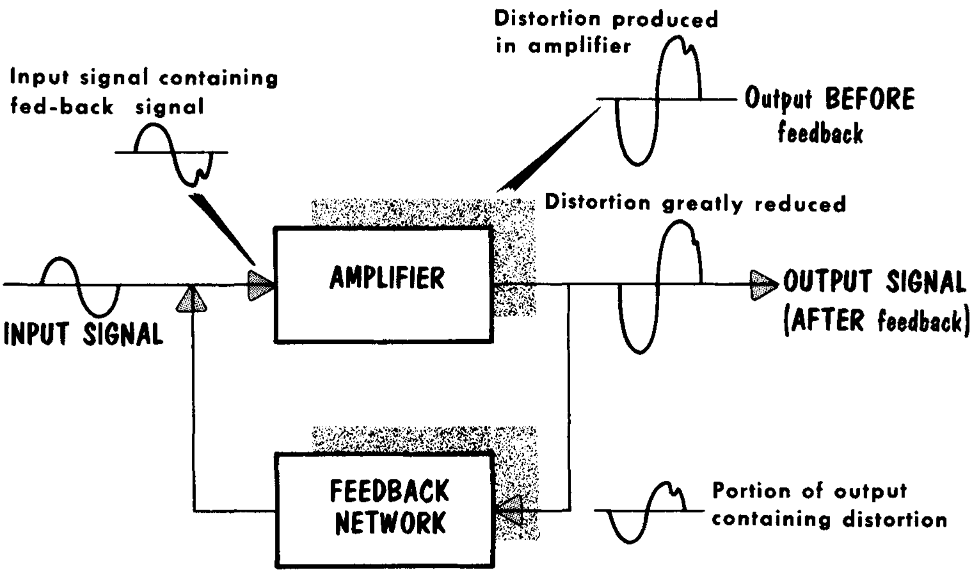
izing effect due to steady plate current cancels, whether the tubes are operated Class-A or Class-B. In turn, this allows a much smaller core to be used for providing an adequate primary inductance with the available turns in the primary winding.

The load reflects an impedance back into the primary. When current flows through the secondary winding, the resultant magnetic flux opposes the current in the primary winding and causes a new "impedance" to appear that was not present before secondary current began to flow. The effect of this inductive opposition is equivalent to adding an impedance in series with the primary winding. This impedance is known as reflected impedance. The reflected impedance becomes greater as the coefficient of coupling increases.

Negative and Positive Feedback

As the term implies, feedback involves the transfer of electrical energy from the output of an amplifier back to its input. If the signal is fed back in phase with the input signal, it is called positive or regenerative, because it adds to the input voltage. If the signal fed back to the input is 180° out of phase with the input signal, it is called negative, inverse, or degenerative, because it

Negative Feedback Reduces Amplifier Distortion



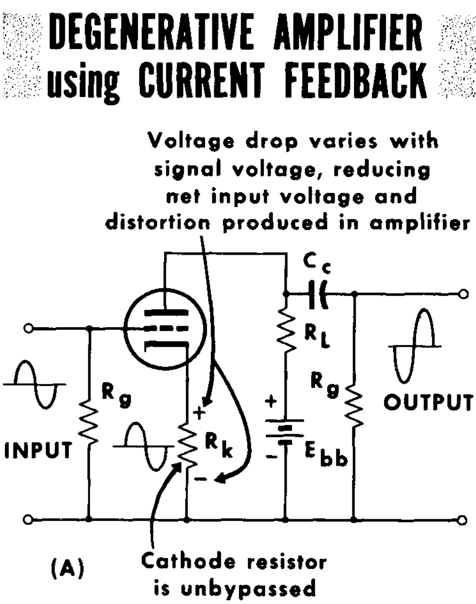
subtracts from the input voltage. In general, positive feedback is avoided in amplifiers since it produces oscillation; negative feedback is used for the reduction of signal distortion. Negative feedback may be used to reduce nonlinear distortion – that is, to make the output waveform more nearly similar to the input waveform by reducing nonlinearities that are introduced within the amplifier tube itself.

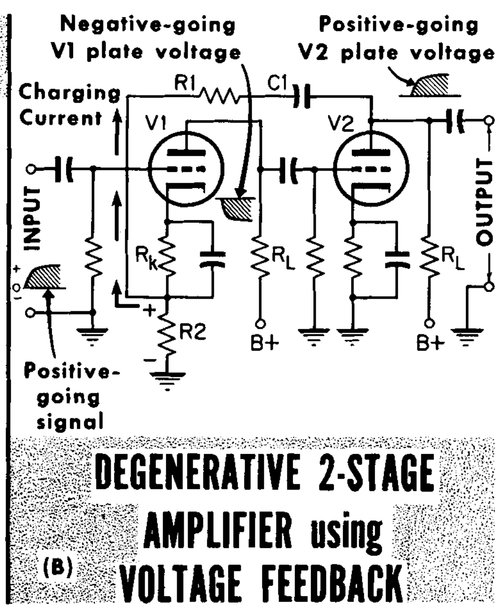
The input signal applied to the grid of an electron-tube is amplified by an amount determined by the μ of the tube, but any nonlinearities introduced within the tube are not amplified. If a portion of the output is fed back 180° out of phase with the input, the distortion component of this feedback signal will be amplified along with the input signal. The amplified distortion component will tend to cancel the distortion component introduced within the tube, and the output may be practically free of nonlinear distortion. It is necessary that the distortion occur in the plate circuit of the stage across which negative feedback is to be applied, to separate the distortion from the desired signal. However, the overall gain of the desired signal will also be reduced. This reduction may be compensated for by increasing the number of stages.

Feedback Circuits

A popular method of obtaining feedback is shown in (A). This method employs current feedback. The cathode bypass capacitor has been omitted, producing degenerative action. When an input signal makes the grid less negative, plate current increases as does the voltage drop across R_k. Since R_k is not bypassed, the plate circuit signal currents flowing through R_k add to the bias produced by the zero-signal component. The grid-to-cathode voltage on the positive-going half-cycle is thus equal to the difference in the input and the drop across R_k. Hence, the net grid cathode voltage is not as great as it would be without feedback, because the drop across R_k is increased. The fact that an output voltage in phase with the input voltage may be developed across an unbypassed cathode resistor has been used in cathode followers and phase inverters.

Negative feedback involving more than one stage may be used. Diagram (B) shows a negative-feedback two-stage amplifier employing voltage feedback. In this case, special attention must be paid to the phase relations in the circuit. Assume that at a given instant the input voltage is such as to make the grid of V1 less negative. Plate current then increases in V1 and plate voltage decreases, making the grid of V2 more negative. At the same time, the plate of V2 becomes more positive because of the reduction in plate current. This increase in V2 plate voltage increases the charge of C1. The operating current flows through ground, up through R2 and R1, to the left plate of C1, making the top end of R2 more positive with respect to ground. The voltage increase across R2 acts in series with the input and the bias across R_k to reduce the magnitude of the positive-going signal on the grid. In short, the grid input signal is reduced by the amount of the feedback voltage, because these two voltages act 180° out of phase.





The Loudspeaker

The conventional radio receiver may be considered as terminating in a loudspeaker. This is the device, often called a transducer, that converts electrical energy into mechanical energy, and finally, into acoustical energy. When the loudspeaker sets air in motion, the human ear should hear the same sounds that were emitted at the microphone of the transmitting station. The

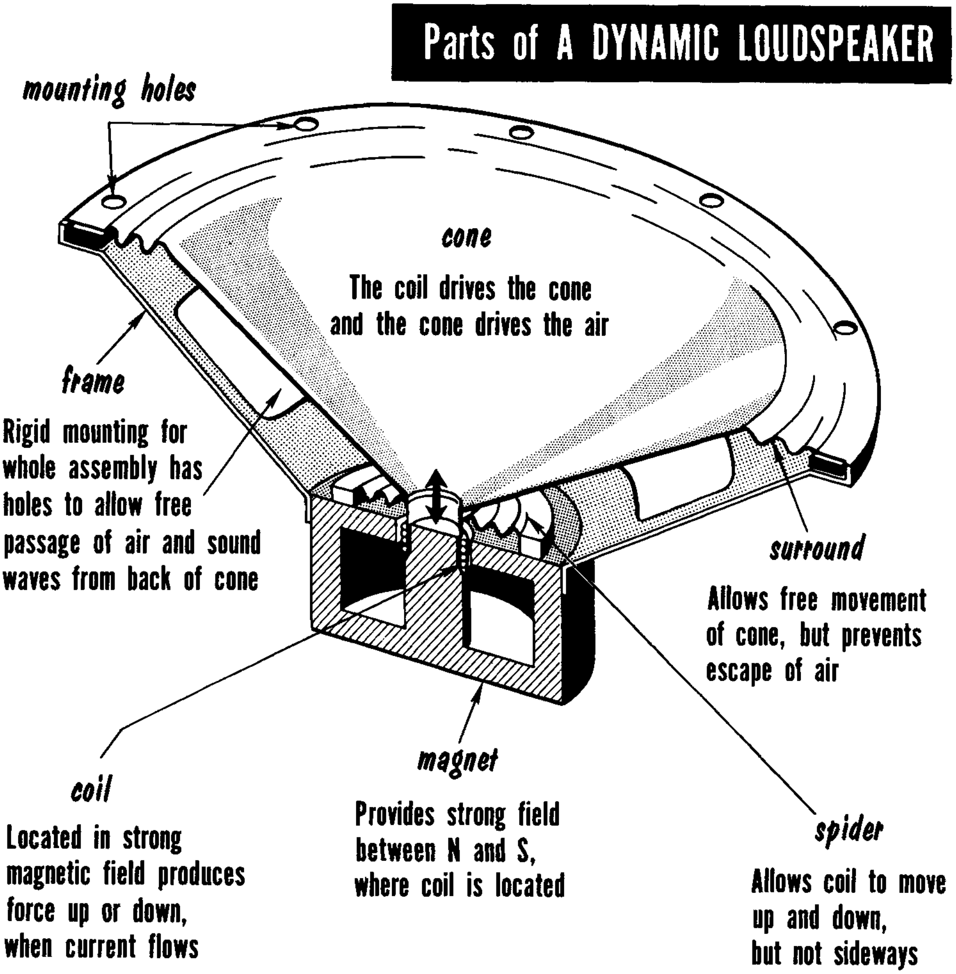
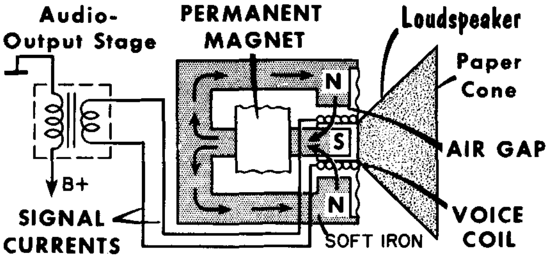


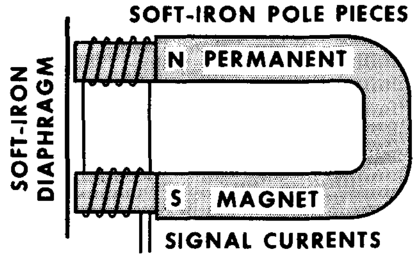
diagram shows a cross-sectional view of the conventional loudspeaker in use today, the dynamic or moving-coil speaker. The voice coil impedance is usually about 4 ohms, although 8-ohm and 16-ohm voice coils are not uncommon. The voice coil is small and light and is suspended by a light flexible material that allows free movement at all frequencies. The requirements for getting maximum force to drive the cone (or diaphragm) are a strong magnetic field and as great a length of wire as is possible in the field.

The Loudspeaker Circuit

In the permanent magnet dynamic type of loudspeaker, a strong field is established between the pole pieces by a powerful permanent magnet. The flux is concentrated in the air gap between a permeable soft-iron core and an external yoke. The voice coil is mounted in the air gap. When a-c signal currents flow in the coil, a force proportional to the strength of the current is applied to the coil, and the coil is moved axially in accordance with the a-c signal. The loudspeaker diaphragm is attached to the voice coil and moves in accordance with the signal currents, thus setting up sound waves in the air. The corrugated diaphragm to which the speaker cone is attached keeps



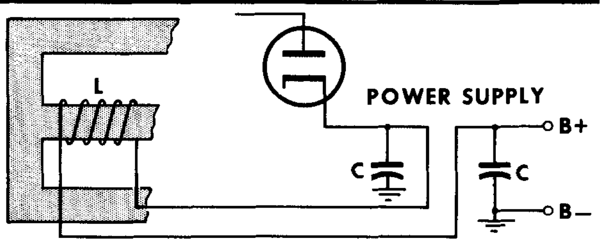
PERMANENT-MAGNET DYNAMIC SPEAKER



BASIC COMPONENTS of HEADPHONES

Many older loudspeakers used **FIELD COILS** in place of permanent magnets.

They acted as inductors in the power supply

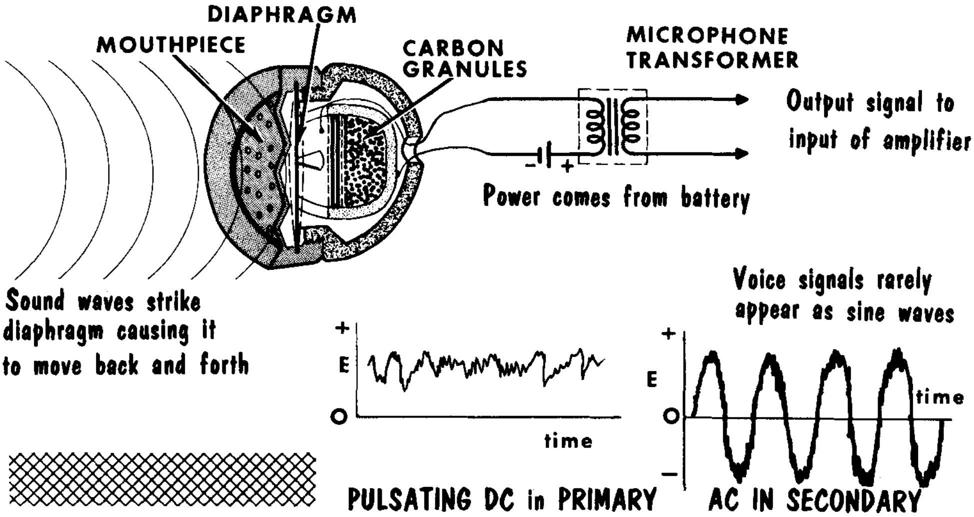


the cone in place and properly centered. An electromagnet may be used in place of the permanent magnet to form an electromagnetic dynamic speaker. However, in this instance sufficient d-c power must be available to energize the field.

Using the relatively sensitive headphone, signals can be heard directly from the output of the audio voltage amplifier. When no signal currents are present, the permanent magnet exerts a steady pull on the soft-iron diaphragm. Signal current flowing through the coils mounted on the soft-iron pole pieces develops a magnetomotive force that either adds to or subtracts from the field of the permanent magnet. The diaphragm thus moves in or out according to the resultant field. Sound waves then are reproduced that have amplitude and frequency (within the capability of the headphone) similar to the amplitude and frequency of the signal currents.

The Carbon Microphone

The CARBON MICROPHONE produces a "PULSATING" DIRECT CURRENT



Although the microphone is not primarily associated with radio receivers, it is helpful to consider the microphone in conjunction with the loudspeaker in that the microphone is often the origin of electrical signals, and the loudspeaker the conclusion. The microphone is also a transducer. It converts acoustical or mechanical energy into electrical energy. There are many types of microphones, but here we shall consider the three most common – the carbon, dynamic, and crystal types. We shall compare them in terms of frequency response, impedance, and sensitivity.

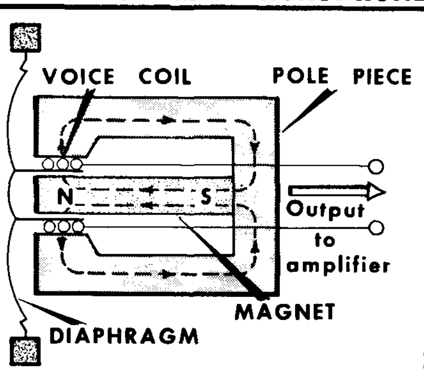
The carbon microphone is the most common. It operates on the principle that a change in sound pressure on a diaphragm coupled to a small volume of carbon granules will cause a corresponding change in the electrical resistance of the granules. In the carbon microphone a diaphragm is mounted against carbon granules contained in a cup. To produce an output voltage, the microphone is connected in a series circuit which contains a battery and the primary of a microphone transformer. The pressure of the sound waves on the diaphragm coupled to the carbon granules causes the resistance of the granules to vary. Thus, a varying or pulsating direct-current in the primary produces an alternating voltage in the secondary of the transformer. This voltage has essentially the same waveform as that of the sound waves striking the diaphragm. Commercial types of carbon microphones give good response from 60 to 6000 cycles, and have a relatively high output. Their low internal impedance requires the use of a transformer. The main disadvantages are the requirement of an external voltage source and a high degree of noise.

The Dynamic and Crystal Microphones

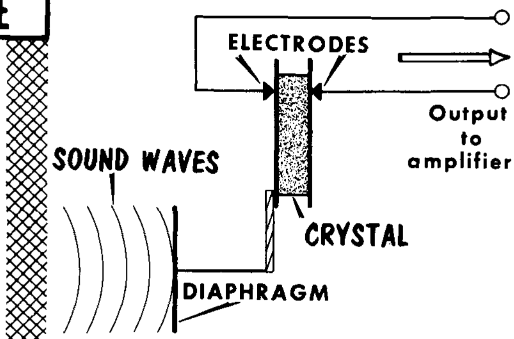
The dynamic or moving-coil microphone consists of a coil of wire attached to a diaphragm and is so constructed that the coil is suspended and free to move in a radial magnetic field. Note its similarity to the dynamic loudspeaker. Sound waves striking the diaphragm cause it to vibrate. This vibration moves the voice coil through the magnetic field so that the turns cut the lines of force in the field. This action generates a voltage in the coil that has the same waveform as the sound waves striking the diaphragm. The dynamic microphone requires no external voltage source and has good fidelity (approximately 20 to 9000 cycles). Its output, however, is extremely low as is its internal impedance (50 ohms or less). Its low impedance makes it desirable for connection to relatively long transmission lines without excessive attenuation of the high frequencies.

The crystal microphone utilizes the property of certain crystals such as quartz or Rochelle salts, known as the piezoelectric effect. The bending of the crystal resulting from the pressure of the sound waves produces a difference of potential across the faces of the crystal. This voltage is applied to

BASIC DYNAMIC MICROPHONE



Sound waves striking diaphragm induce voltage in voice coil



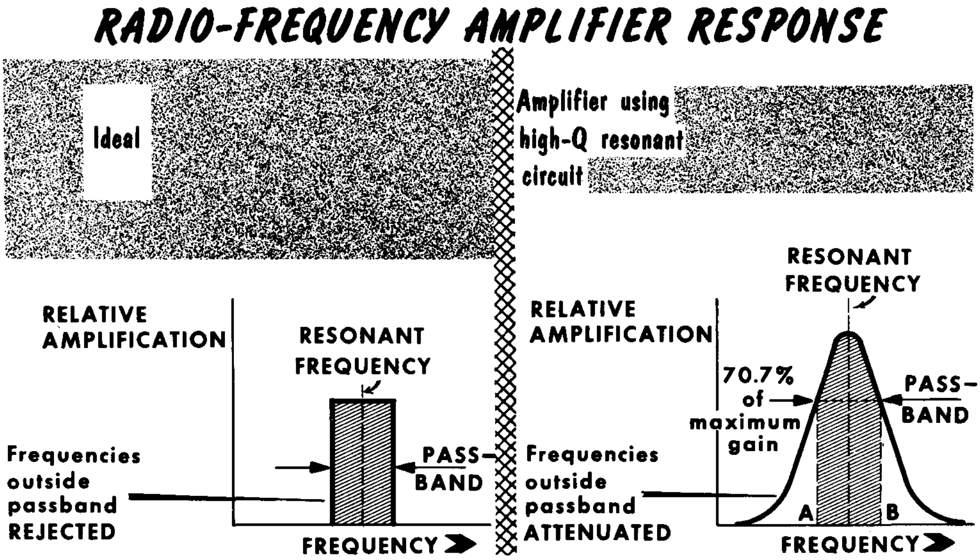
Pressure of sound waves distorts shape of crystal, producing emf across electrodes

BASIC CRYSTAL MICROPHONE

the input of an amplifier. The diaphragm may be cemented directly on one surface of the crystal or it may be cemented by mechanical coupling. A metal plate or electrode is attached to the other surface of the crystal. When sound waves strike the diaphragm, the vibrations of the diaphragm produce a varying pressure on the faces of the crystal, and therefore, a voltage is induced across the electrode. This voltage is essentially the same waveform as the sound wave striking the diaphragm. The crystal microphone has a relatively high output voltage as well as an extremely high impedance (several hundred thousand ohms). It is comparatively light, requires no battery, and has an excellent frequency response (up to 17,000 cycles). Its output is much higher than that of the dynamic microphone. It is widely used in that it can take rough handling while producing a high output voltage.

Radio-Frequency Amplification

Up to this point we have considered amplifiers designed primarily for the strengthening or amplification of a-f signals; that is, frequencies extending to about 16,000 cycles. In radio communications, receivers must also be able to handle frequencies well up into the megacycle range. For example, the standard radio broadcast band is from 535 kc up to 1605 kc. Above this there is the so-called "shortwave" bands which extend beyond 100 mc (100,000,000

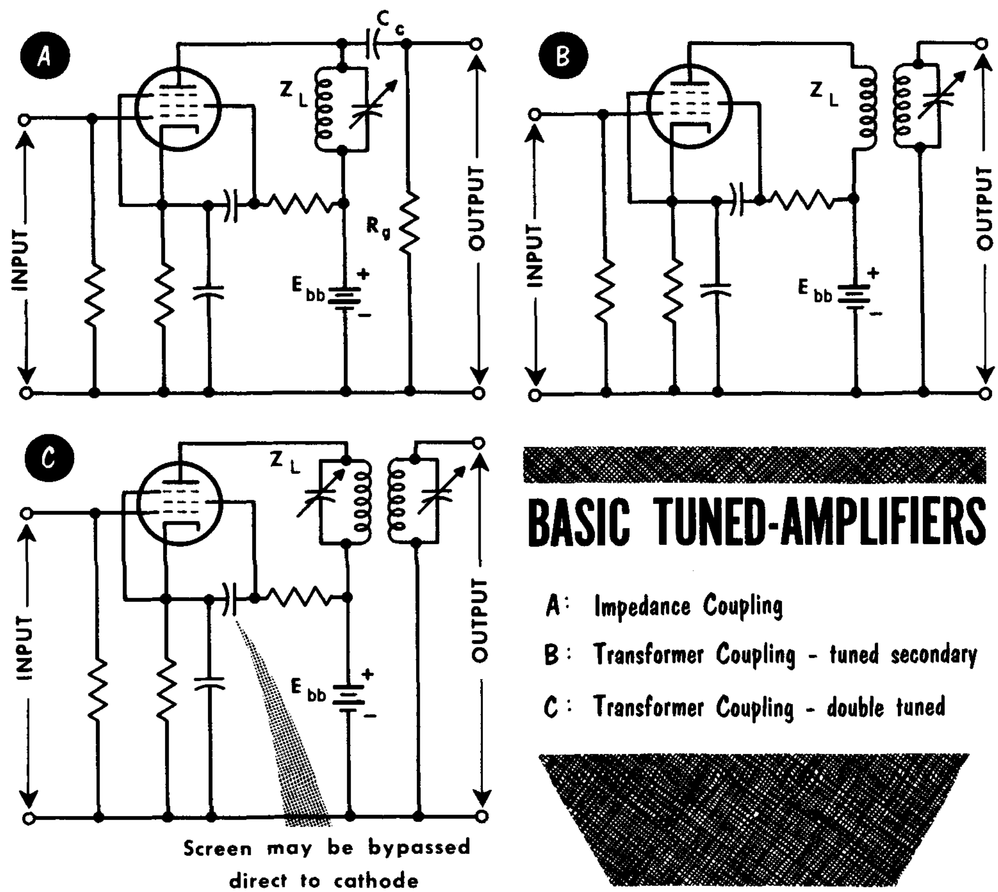


cycles per second). Amplification of these signal frequencies involve two basic changes: the tubes used are almost exclusively pentodes because of their low interelectrode capacitances (important at high frequencies) and methods of coupling. In general, r-f amplifiers utilize tuned or resonant circuits as grid loads, plate loads, or both.

An r-f amplifier is judged on effective gain at its tuned frequency, frequency response to signals lying at each side of the resonant frequency, and ability to discriminate against frequencies immediately adjacent to the passband. The band of frequencies to be passed by an r-f amplifier is generally a very small percentage of the resonant frequency. The tuned r-f circuits of a receiver operating at 1000 kc must pass a band of 10 kc – only 1%. An ideal r-f amplifier may be viewed as a bandpass filter having a response of that shown. Practical resonant circuits cannot be made to give this perfection. However, when the tuning components have a sufficiently high Q, this performance can be approached quite closely. The selectivity of a receiver is its ability to amplify a given band of desired frequencies and reject all others. The narrowest possible bandpass is not always desired. In some forms of communications, we purposely broaden the bandpass since the received signal covers a broad range; in others we make the response as narrow as possible. This will be covered in our discussion on receivers.

Tuned R-F Amplifiers

An important part of the coupling circuit of a tuned amplifier is the resonant circuit. This is used in plate circuits because it offers a high impedance at the desired frequency and a low impedance at other frequencies, thus permitting high amplification of a relatively narrow band of frequencies. In addition, the limitations imposed on untuned amplifiers by interelectrode and distributed capacitances are used to advantage because these capacitances become part of the tuned circuit. An r-f amplifier may be single- or double-tuned, depending on whether the plate circuit only, or both the plate and grid circuits, contain a tuned circuit. Our diagram shows the three common forms of tuned-amplifier circuits. In (A), the tuned circuit acts merely as a parallel L-C circuit, presenting a high plate impedance at resonance and a lower impedance above and below resonance. We can consider this circuit as being impedance-coupled. The single-tuned transformer-coupled amplifier of (B) is commonly used as a Class-A r-f voltage amplifier. Here magnetic coupling is used, with frequency selection taking place in the grid circuit of the second stage. Circuit (C) is double-tuned and transformer-coupled, with tuned circuits in the plate and grid circuits of coupled stages.



BASIC TUNED-AMPLIFIERS

A: Impedance Coupling

B: Transformer Coupling - tuned secondary

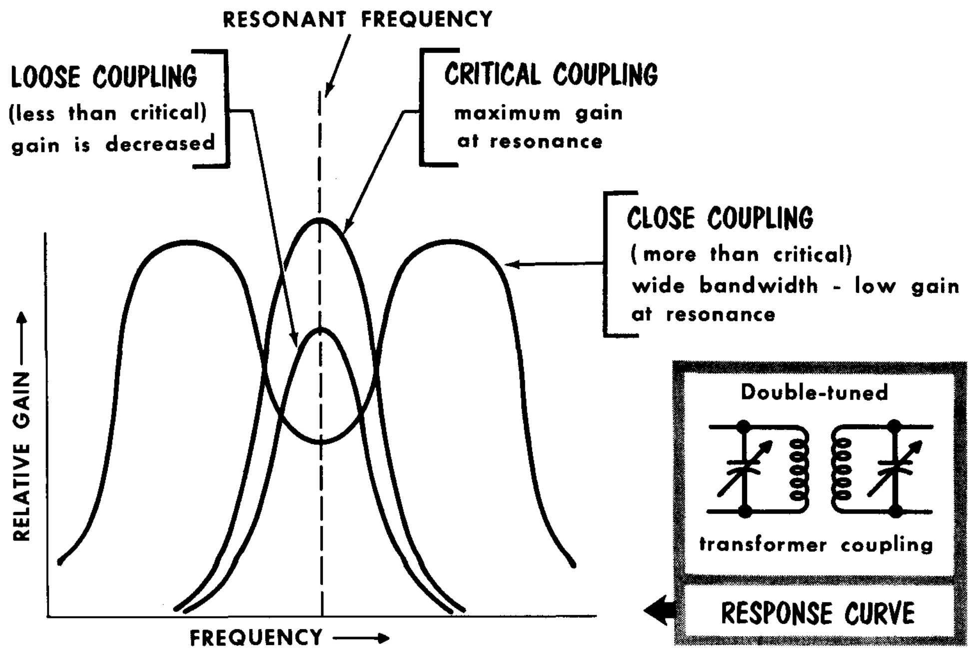
C: Transformer Coupling - double tuned

R-F Amplifier Coupling Characteristics

In tuning, the values of either L or C can be varied. C is varied by a variable capacitor; L can be varied by varying the powdered-iron core in the transformer. The powdered-iron core provides extremely high Q. Circuit Q is extremely important in that it determines the sharpness of the frequency response curve. A high-Q circuit gives a very sharp response, coupled with high gain, because of the extremely high impedance of the parallel resonant circuit. Using the double-tuned transformer, a relatively narrow band of frequencies may be amplified and frequencies outside this band are sharply reduced. These characteristics make this type of coupling highly desirable in intermediate amplifiers (which are discussed in Volume 4). The frequency bandpass is determined largely by the coupling between the primary and secondary of the r-f transformer. As can be seen from the curve, the degree of coupling may be varied to obtain a particular frequency response.

For low-power operation such as required in receivers, pentode r-f amplifiers, because of their low grid-to-plate capacitance, provide the highest gain with the least tendency to break into self-oscillation. Bypassing and decoupling become increasingly important at the higher frequencies. For maximum gain, pentode r-f amplifiers use high-Q circuits and tubes having a high mutual conductance.

DEGREE of COUPLING AFFECTS FREQUENCY RESPONSE



When the output of one stage is coupled to the input of another stage, a coupling network is used for transferring the energy.

One of the most common types of coupling networks is the R-C arrangement. The plate of one amplifier stage is connected to the grid of the following stage through a coupling capacitor. Resistors are used as the plate load and in the grid circuit.

A voltage gain or response characteristic curve indicates the amount of voltage gain an amplifier has over a wide range of frequencies.

The response characteristic of an R-C coupled amplifier drops off at the low frequencies because of the high reactance presented by the coupling capacitor.

The response characteristic of an R-C coupled amplifier drops off at the high frequencies because of the low reactance presented by the interelectrode, stray, and wiring capacitances.

In a transformer-coupled amplifier, the primary of the transformer is connected in the plate circuit of one tube and the secondary is connected in the grid circuit of the second tube.

If the feedback signal in an amplifier <u>aids</u> the original input signal, the feedback is called <u>regenerative</u> or <u>positive</u>. If the feedback signal <u>opposes</u> the original input signal, it is called <u>degenerative</u>, <u>negative</u>, or <u>inverse.</u>

A push-pull amplifier consists of two tubes arranged so that the plate current of one tube is 180° out of phase with the plate current of the other tube. The magnitudes of the currents are equal.

The input-grid signals for a push-pull amplifier are obtained from a transformer which has a centertapped secondary, or from a paraphase amplifier.

Compared to single-tube operation, less distortion is obtained in push-pull operation since its dynamic characteristic is more linear. Also, a greater grid-signal swing is permissible.

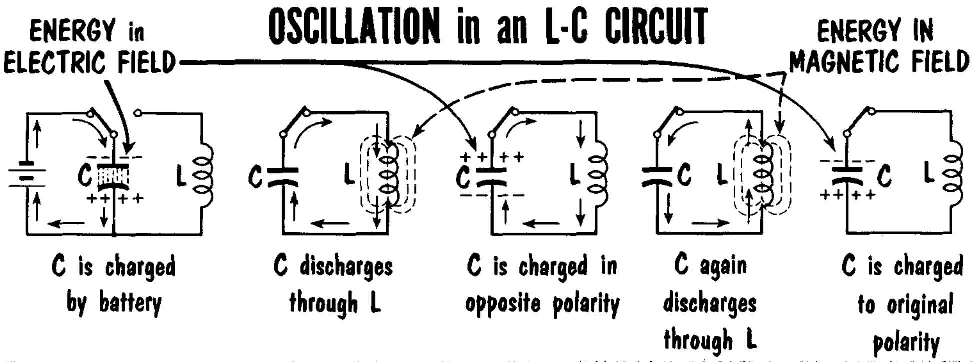
The dynamic loudspeaker operates on the basis of interacting magnetic fields from the voice-coil signals and the permanent magnet.

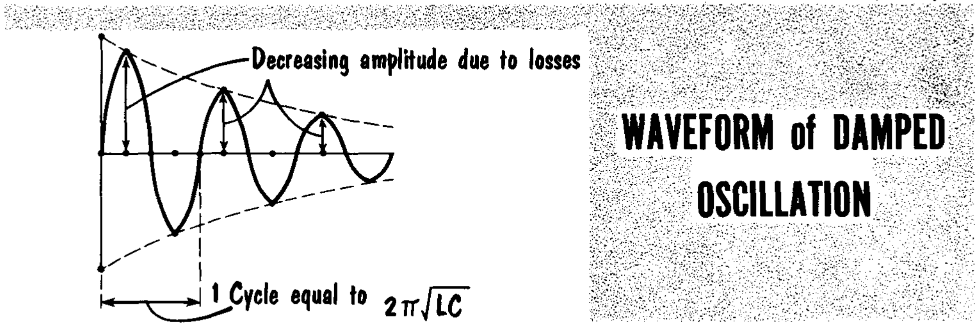
If the output waveform of an amplifier is identical with its input waveform, the amplifier is said to be distortionless. Types of distortion include amplitude distortion, frequency distortion, phase distortion, and harmonic distortion.

REVIEW QUESTIONS

1. What is the fundamental difference between voltage and power amplifiers?
2. What is the function of a coupling network?
3. What is the function of a d-c blocking capacitor?
4. Why does the voltage gain of an amplifier decrease at the low and high frequencies?
5. What are the advantages and disadvantages of transformer coupling?
6. Give an important application of the cathode follower.
7. What is a paraphase amplifier?
8. What advantages are realized from the push-pull amplifier?
9. What is the purpose of impedance matching?
10. Describe the operation of the dynamic loudspeaker.
11. Describe the operation of the carbon microphone.
12. Name some basic differences between r-f and a-f amplifiers.

Oscillation



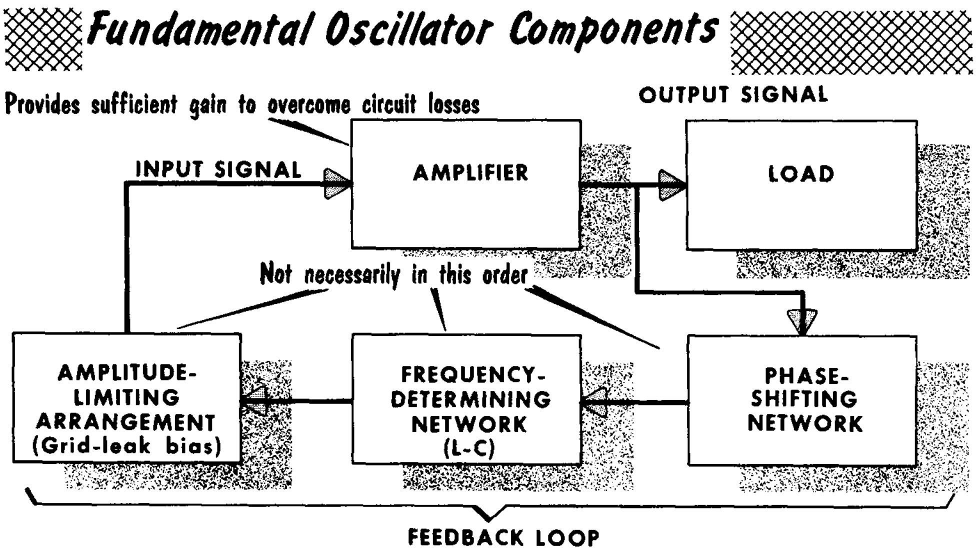


Before embarking on a study of electron tube oscillators, let us first investigate just what we mean by oscillation. If we take a 100-volt battery and place it across the plates of capacitor C, that capacitor would charge to a value of 100 volts. The actual energy would be in the form of the electrical lines of force existing in the dielectric between the negative and positive plates. On the assumption that C is of perfect quality, the charge would remain forever if the battery were removed. We now place inductor L across C. Since L represents a d-c path for the electrons, C begins to discharge through the inductor. The electron flow through L produces a magnetic field. Thus, in effect, the energy contained in the electric field of C is now transferred into the form of a magnetic field about L. When this magnetic field collapses, electrons flow into C until it is fully charged in the opposite direction, and energy is once more in the form of electric lines of force. When discharging, the reverse occurs and one complete cycle has taken place.

The back-and-forth flow of electrons between C and L is called oscillation. If there were no losses in L and C, these oscillations would continue forever since all the energy stored in the electric and magnetic fields is returned to the circuit. However, due to resistance losses, the energy in each alternation is successively lower. In actual practice, a train of "damped" oscillations occurs, until all the energy is consumed in the form of losses. In the electron tube oscillator, we make use of the L-C circuit characteristics; however, we provide enough additional energy to the circuit on each cycle to make up for the losses during that cycle, giving us continuous oscillations.

Oscillators

An oscillator is a device capable of converting dc into ac at a frequency determined by the values of the constants in the device. Regardless of its type, any oscillator may be divided into three basic elements: the frequency-determining network (often called the resonator), the amplifier, and the load or output circuit. From the block diagram, we see the fundamental oscillator components in greater detail. The output power of the amplifier (divided as



it is between the load and the feedback loop) must be sufficiently large to supply both. A second requirement for maintaining oscillation is that the feedback energy and the input energy of the amplifier be in phase. These two specifications can usually be met at only one frequency, since the resonator changes both the amplitude and phase of its output at frequencies other than the one to which it is tuned.

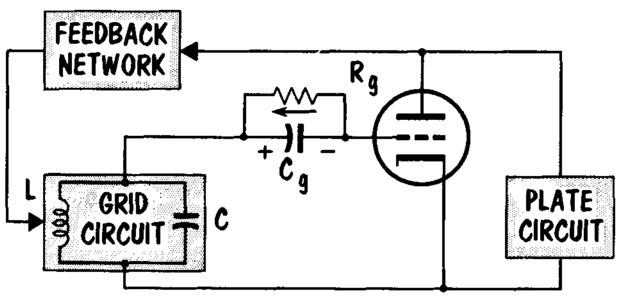
Starting with the amplifier, the assumption is that some circuit disturbance produces a small voltage change that constitutes input. The amplifier then acts upon this input signal to raise the output power to an adequately high level, so that both the feedback circuit and the load are properly excited. The useful power output remains, of course, in the load. The feedback loop contains the remaining three elements: the phase-shifting network satisfies the requirements of the resonator as far as feedback phase (as related to input phase) is concerned; the resonator comprises the L-C network, either lumped or distributed, and establishes the oscillator frequency. Finally the amplitude-limiting arrangement determines the amount of power to be circulated through the amplifier and feedback loop so that oscillation will neither cease because of insufficient feedback, nor become unstable as a result of excessive feedback. The amplitude-limiting function is usually performed by tube biasing.

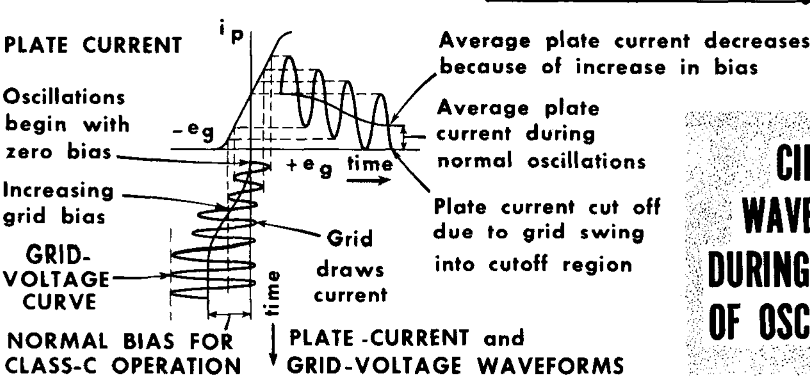
Grid-Leak Bias

The generation of oscillations starts with tiny chance disturbances such as non-uniform emission or changes in tube characteristics. The presence of a large Class-C bias at the outset would make it virtually impossible for oscillations to start spontaneously, since the grid signal must exceed the bias voltage. This situation is remedied by starting the cycle of events under conditions of low bias, and building the bias up to Class-C only after the input signal has become large enough. Class-C bias is used for most oscillators because of the high efficiency possible under these conditions. Grid-leak bias provides a natural situation for oscillation. A simple R-C network permits the build-up of oscillation.

In the diagram, the grid of the amplifier is supplied with signal voltage by way of the feedback network. At the beginning of the process, the grid bias is zero with respect to cathode, and the tube operates at a point on its characteristic curve at which the mutual conductance is reasonably high. These conditions encourage easy starting. As the amplitude of the oscillations grows, the alternating voltage appearing across L-C is applied to the grid in series with R_g and C_g. Because of the rectifying action of the grid, dc flows through R_g, causing a voltage having approximately the value of the peak signal voltage to appear across C_g. The development of this bias voltage continues to build up until Class-C operation is established. This is determined by the value of the R-C component in the grid circuit. During the intervals of zero grid current, C_g tends to discharge through R_g. With too short a time constant, no steady-state d-c bias appears, since the capacitor closely follows the variation of the voltage drop across R_g.

GRID-LEAK BIAS PERMITS EASY BUILD-UP OF OSCILLATIONS

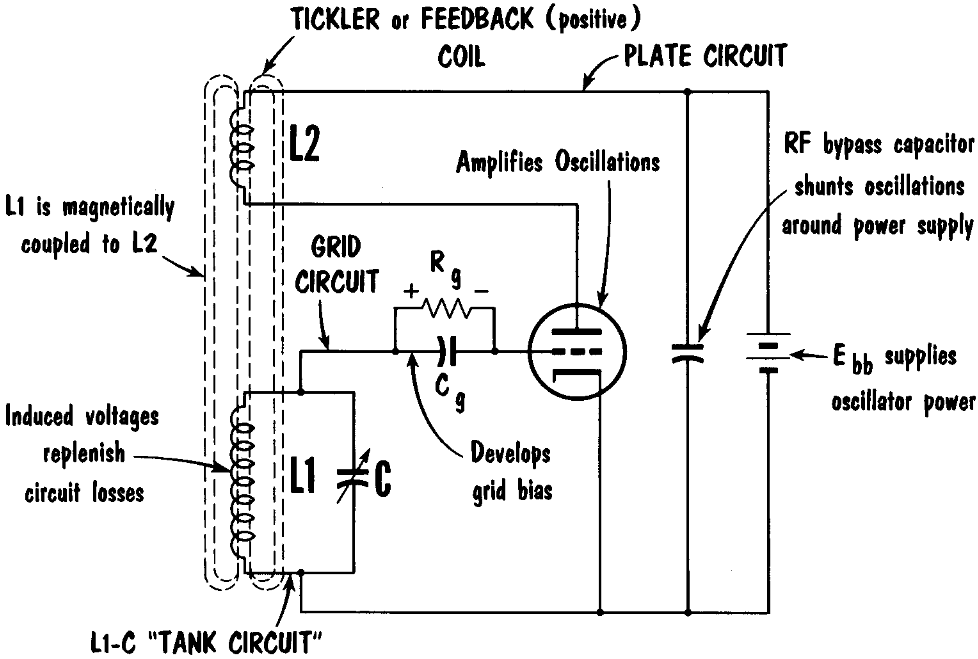




CIRCUIT WAVEFORMS DURING BUILD-UP OF OSCILLATIONS

The Tickler Feedback (Armstrong) Oscillator

TICKLER-FEEDBACK (ARMSTRONG) OSCILLATOR
can produce a-f and r-f oscillations



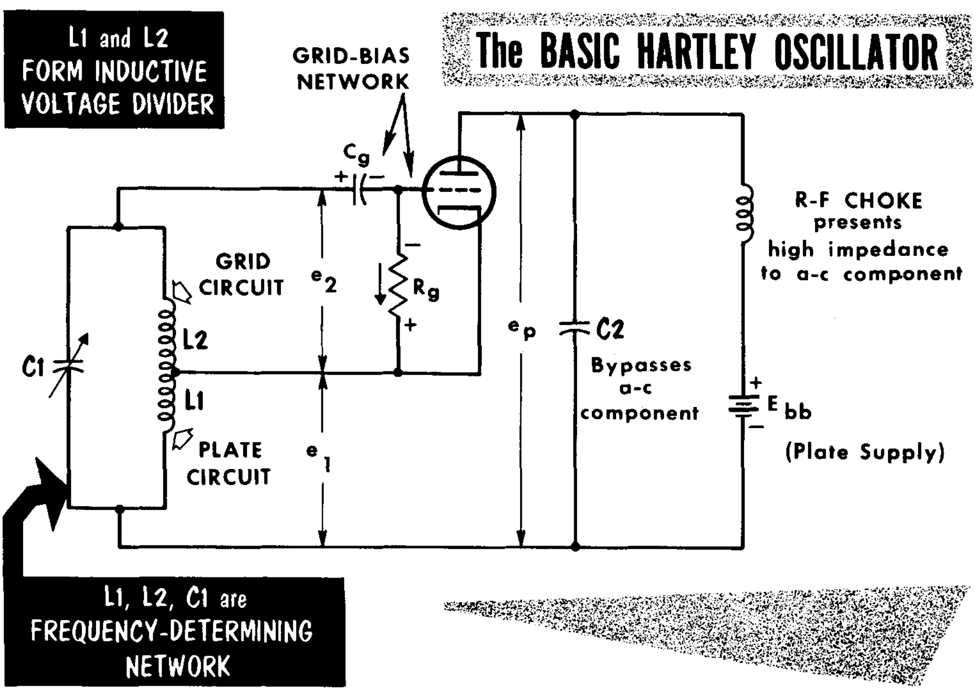
One of the simplest types of oscillator circuits is that employing tickler feedback. Feedback voltage of the proper phase from the plate circuit to the grid circuit is accomplished by mutual inductive coupling between the oscillator tank coil L1, and the tickler feedback coil, L2. The amount of feedback voltage is determined by the amount of flux from L2 that links L1, and can be varied by moving L2 with respect to L1. The frequency-determining part of the oscillator is the tank circuit L1-C. The coil and tuning capacitor interchange energy at the resonant frequency, and the excitation voltage developed across the tank circuit is applied to the grid in series with the grid-leak bias across R_gC_g.

The tickler coil is coupled to L1 in such a manner that there is regenerative feedback from the plate circuit to the grid circuit. This feedback is sufficient to overcome circuit losses and sustain oscillations in the L1-C circuit. The output or load is generally taken from the oscillating L-C circuit. This has the effect of loading down a circuit. For this reason, the load is considered to increase circuit losses. This can be compensated for by increasing the amount of regenerative feedback. Too great a load damps the oscillations in the L-C circuit, and can cause the oscillations to die out.

The Hartley Oscillator

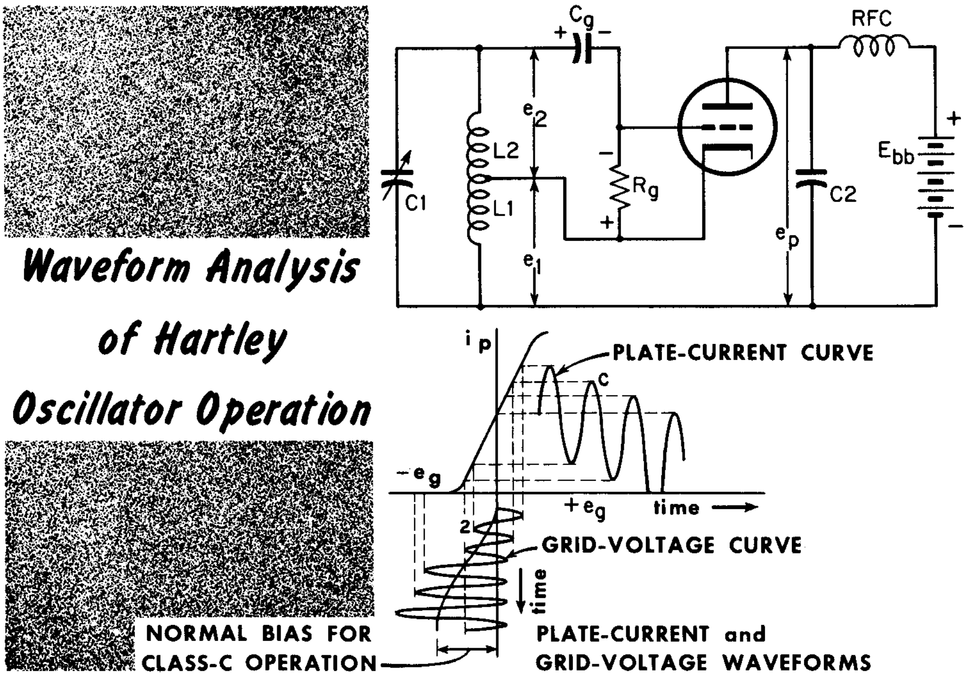
Along with the Armstrong oscillator, the Hartley oscillator is the most widely used type in radio. L1 is a part of the tuned circuit made up of L1, L2, and C1. It also is used to couple energy from the plate circuit back into the grid circuit by means of mutual inductance between L1 and L2. C_g blocks the d-c component of the grid circuit from L2, and together with R_g provides the necessary operating bias. C2 and the r-f choke keep the a-c component in the plate circuit out of the B supply. The B supply is returned to resonant tank coil L1. The tuned circuit therefore contains a d-c component of plate current in addition to the a-c signal component.

Now let us analyze the operation of this very important oscillator. When the tube warms up, plate current starts to flow, since B+ is applied. Because the grid is located in the electric field between the plate and cathode at a point positive with respect to the cathode, a small positive voltage exists on the grid. The increase in plate current through L1 is accompanied by an expanding magnetic field around L1 which induces voltage e2 in L2. The polarity of e2 makes the grid more positive with respect to the cathode, and plate current continues to increase until saturation. During this time, C1 is charging. Grid current flows as C_g acquires a small charge with the minus side facing the grid. The grid voltage during this time is e2 minus the voltage drop across C_g. Plate current stops increasing at saturation, and the field about L1 stops expanding. As a result, the induced voltage e2 falls to zero. The positive grid voltage (e_2 minus the drop across C_g) decreases, causing the plate current to decrease, and C1 begins to discharge.



The Hartley Oscillator (Cont'd.)

As the field about L1 collapses, it induces voltage e2 in L2 of opposite polarity to e2 when the field was expanding. Hence, the grid voltage goes negative with respect to the cathode, and plate current decreases further. The induced voltage e2 aids in the discharge of C1 and C_g. C1 discharges fully and begins to charge oppositely (its polarity reverses). However, C_g cannot dis-



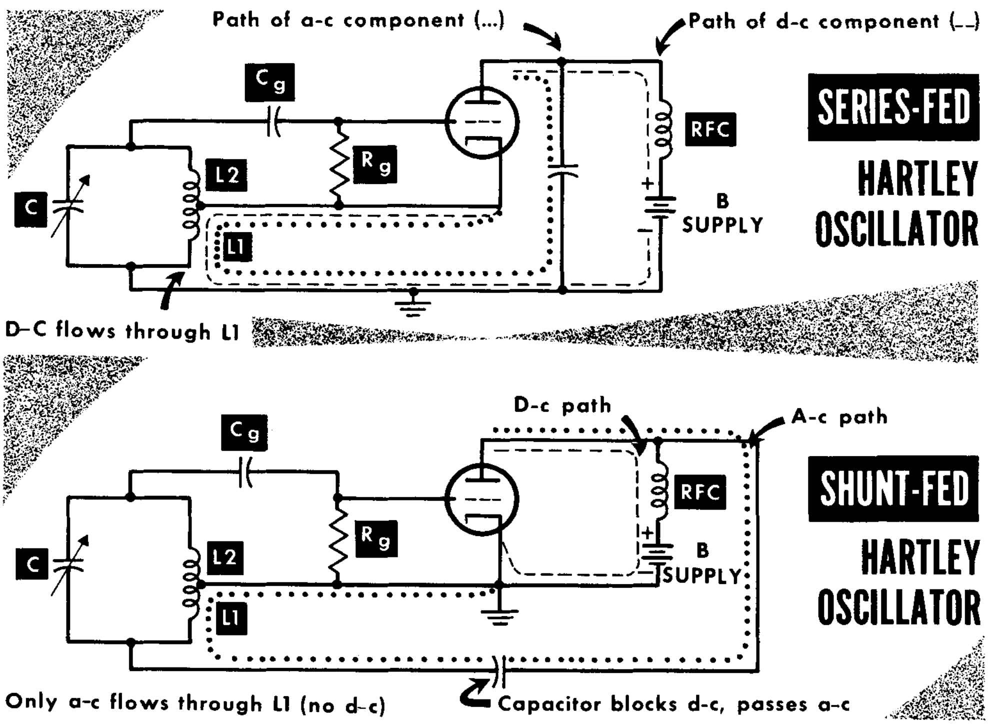
charge rapidly because of the long time constant R_gC_g. Grid voltage swings to a maximum negative condition (point 2 on grid curve), and C_g discharges slowly through R_g. Grid current does not flow during this part of the cycle, and the grid bias voltage is e2 plus the voltage drop across C_g. Plate current ceases to fall at this point. The field about L1 stops changing and e2 falls to zero. C1 begins to discharge.

The grid bias voltage swings in a positive direction again and plate current begins to rise. The expanding field about L1 again induces voltage e2 in L2, making the grid voltage more positive with respect to cathode. Current flows from cathode to grid into C_g, causing C_g to acquire a small additional charge, while plate current rises to maximum (point C on plate curve). From here, the cycle continues to repeat. On each subsequent cycle, bias voltage builds up across C_gR_g until it reaches a steady value. Normal bias indicates Class-C operation. The flywheel effect of the resonant-tank circuit maintains oscillations during the time that the plate current is zero and no energy is being supplied to the oscillator circuit.

Series and Shunt Feeding

There are two methods for applying plate voltage to the oscillator tube. The d-c plate voltage supply can be placed in series with the oscillating plate circuit, in which case the circuit is referred to as series-fed, or in parallel with the circuit, and the circuit is then referred to as shunt-fed. In either case, there must be a d-c return path from plate to cathode for the plate current. In the series-fed oscillator, the d-c plate current must pass through L1 before it can return to cathode. The disadvantage of this arrangement is that the plate supply is placed at a high a-c potential with respect to the cathode. Also, the supply has a large distributed capacitance to ground that is shunted across tank inductor L1.

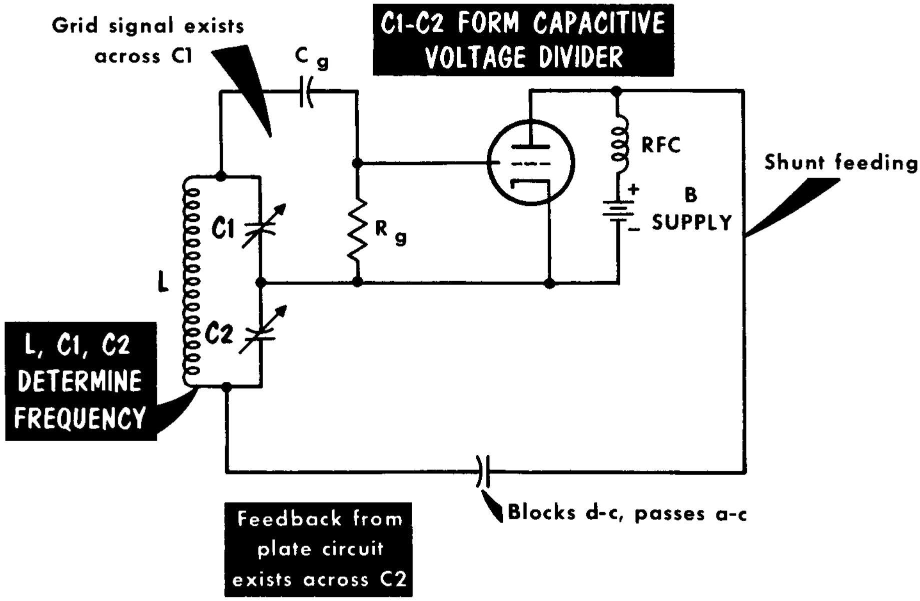
The disadvantage of the series-fed circuit can be overcome by keeping the d-c plate supply and the oscillating plate current separate. This is accomplished in the shunt-fed Hartley oscillator where the plate oscillations are



coupled to the split-inductance tank through a bypass capacitor. The capacitor prevents the d-c plate current from returning to the cathode through the tank. The plate current, therefore, can return only through the choke in series with the supply. This choke prevents any oscillations in the supply because its reactance is very high at the oscillation frequency. Thus, the principal advantage of shunt feed is that the high-voltage B supply is isolated from the tuned circuit.

The Colpitts Oscillator

THE COLPITTS (SPLIT-CAPACITANCE) OSCILLATOR



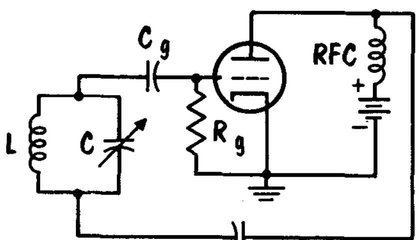
The Colpitts oscillator is similar to the shunt-fed Hartley type, with the exception that the Colpitts uses a split-tank capacitor as part of the feedback circuit instead of a split-tank inductor. Colpitts oscillators operate extremely well at high frequencies, and stable operation at several hundred megacycles is common. It is a relatively flexible oscillator because various circuit configurations are possible. Also, it may be made reasonably free of harmonics, and is easy to adjust. The frequency-determining network consists of L, C1, and C2, all connected in series. The plate-circuit signal-return path includes two parallel branches, one through C2 directly to the cathode, and the other through the series combination of L and C1. In this connection, the two tuning capacitors behave as a capacitor voltage divider, and the amount of plate-to-grid feedback depends upon the ratio of C1 to C2. To establish a particular frequency with a given inductance L, capacitors C1 and C2 together must total a specific capacitance; however, the smaller C1 is made and/or the larger C2 becomes, the greater the voltage coupled back will be. For this reason, both C1 and C2 are usually variable, enabling the operator to establish both the frequency and the amount of feedback.

When R_g is large enough (depending upon the bias required), it does not have much shunt effect upon the signal voltage applied on the grid. In certain frequencies, at which R_g must be small, an additional choke is necessary, and is inserted in series with R_g.

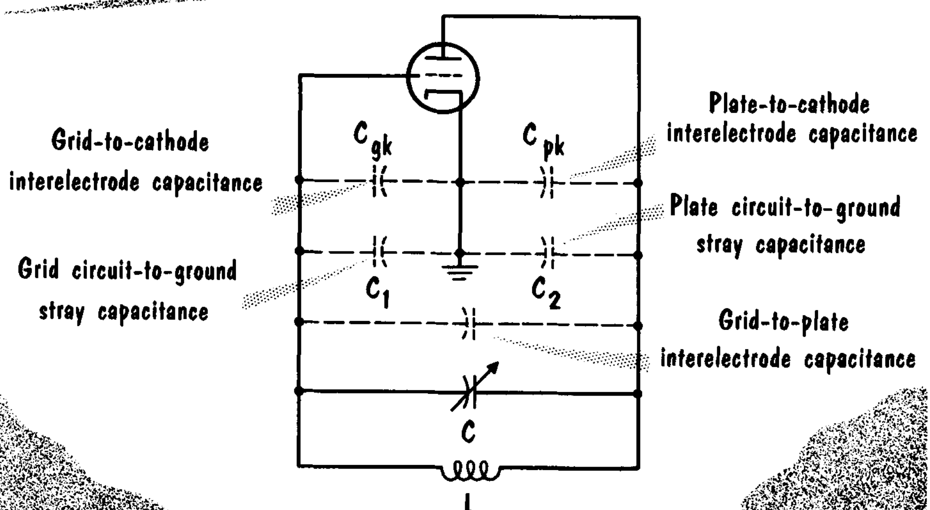
High-Frequency Oscillators

At very high frequencies, capacitances which are usually ignored, such as interelectrode and stray types, play an extremely important role. At audio and broadcast frequencies, the operating frequency is generally determined by the "lumped" constants of L and C in the circuit. We show a typical high-frequency oscillator often called the ultraudion. There would seem to be no proper feedback arrangement in the basic circuit. However, if d-c bias and plate supply voltages are ignored, and we observe the circuit directly from its a-c "equivalent," we see an arrangement similar to the Colpitts oscillator. The ratio of grid-cathode to plate-cathode alternating voltages is equal to $(C_{pk} + C2)/(C_{gk} + C1)$. The voltage drop across C_{gk} is appreciable at the operating frequency, and provides the grid excitation. The total tank circuit capacitance is equal to the sum of all the "branch" capacitances.

Thus it is obvious that care must be employed in working with high-frequency oscillators, inasmuch as moving wires can change stray capacitances and the substitution of one tube for another can change interelectrode capacitances.



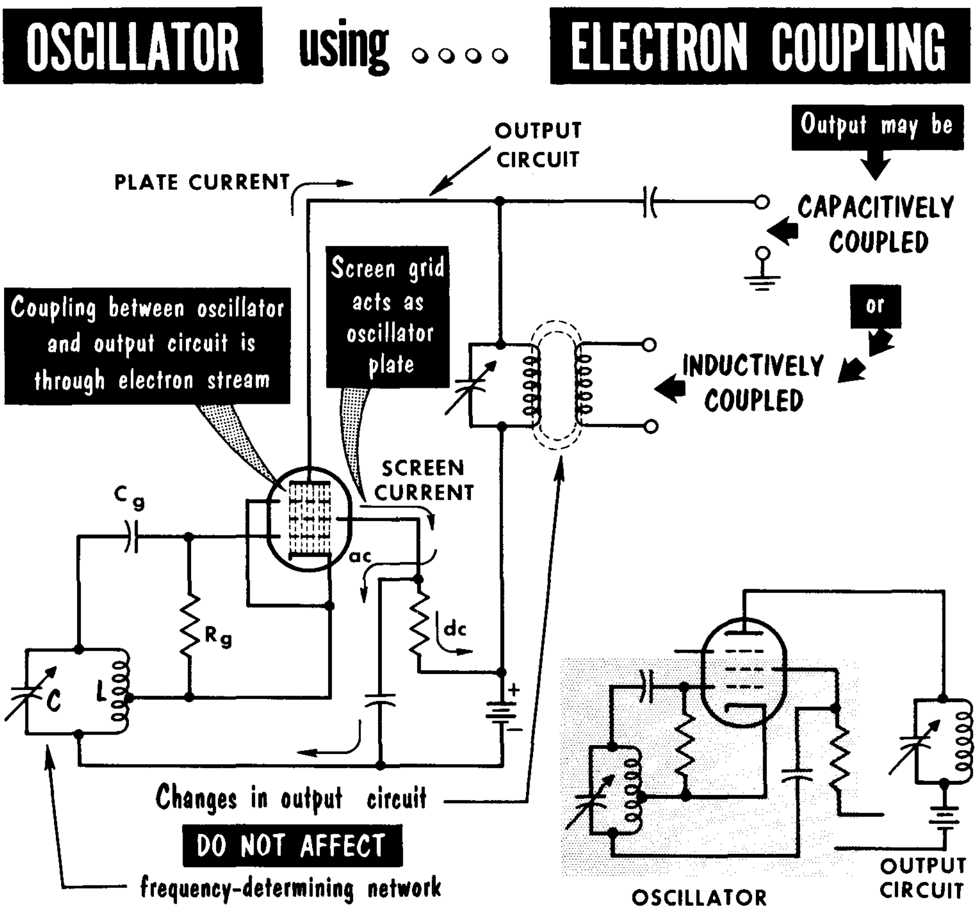
BASIC HIGH-FREQUENCY "ULTRAUDION" OSCILLATOR



Equivalent Circuit

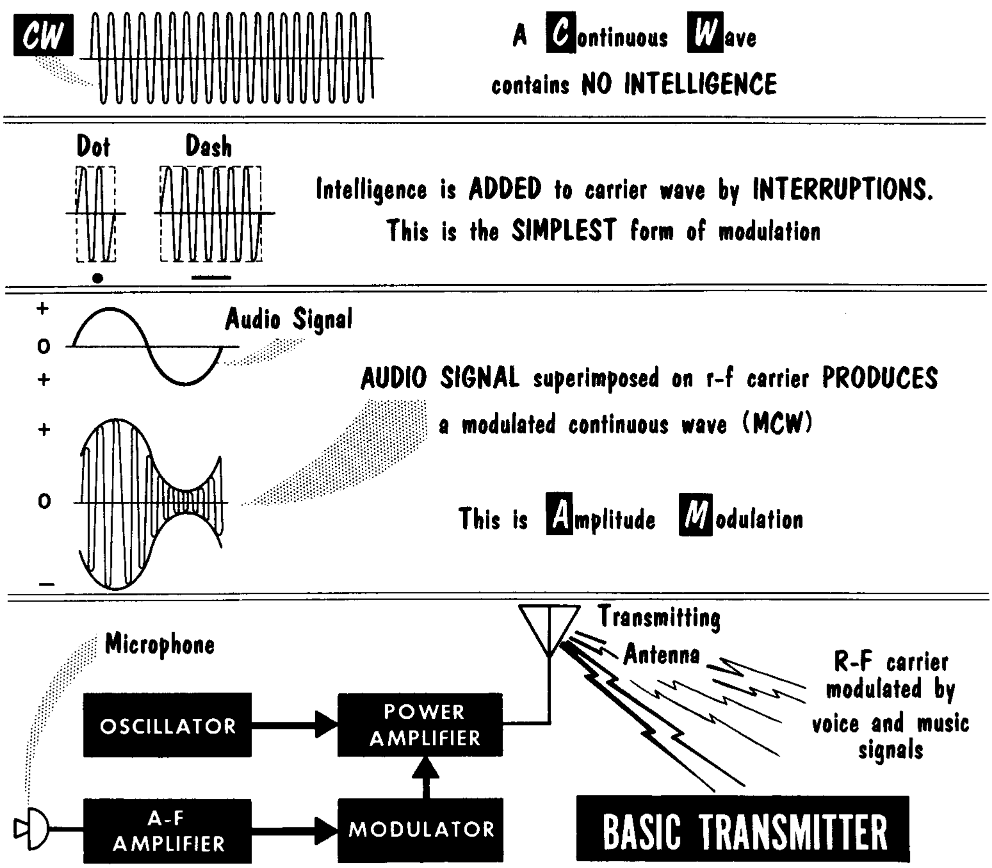
The Electron-Coupled Oscillator

As its name implies, the electron-coupled oscillator represents more a form of coupling than a basically new oscillator circuit. When a load is coupled to an oscillator tank circuit, the oscillator is subject to frequency variation with changes in the load. Any changes in the load cause changes in the oscillator circuit, and consequently a feedback phase shift that tends to make the frequency "drift." Electron coupling is used to isolate the frequency-determining tank circuit from possible variations due to loading. The use of a pentode tube, with the screen grid acting as a triode oscillator plate, keeps the plate circuit independent of the basic oscillator. The oscillations present at the control grid vary the electron stream flowing from cathode to plate. The screen grid, however, acts as the primary attracting force for the electrons. Because of the construction of the screen grid, the major portion of the electron stream goes past the screen and is attracted to the pentode plate where a voltage is developed across a tuned L-C circuit. The basic circuit used here is the Hartley; however, any of the oscillators discussed can be used with electron coupling.



Modulation

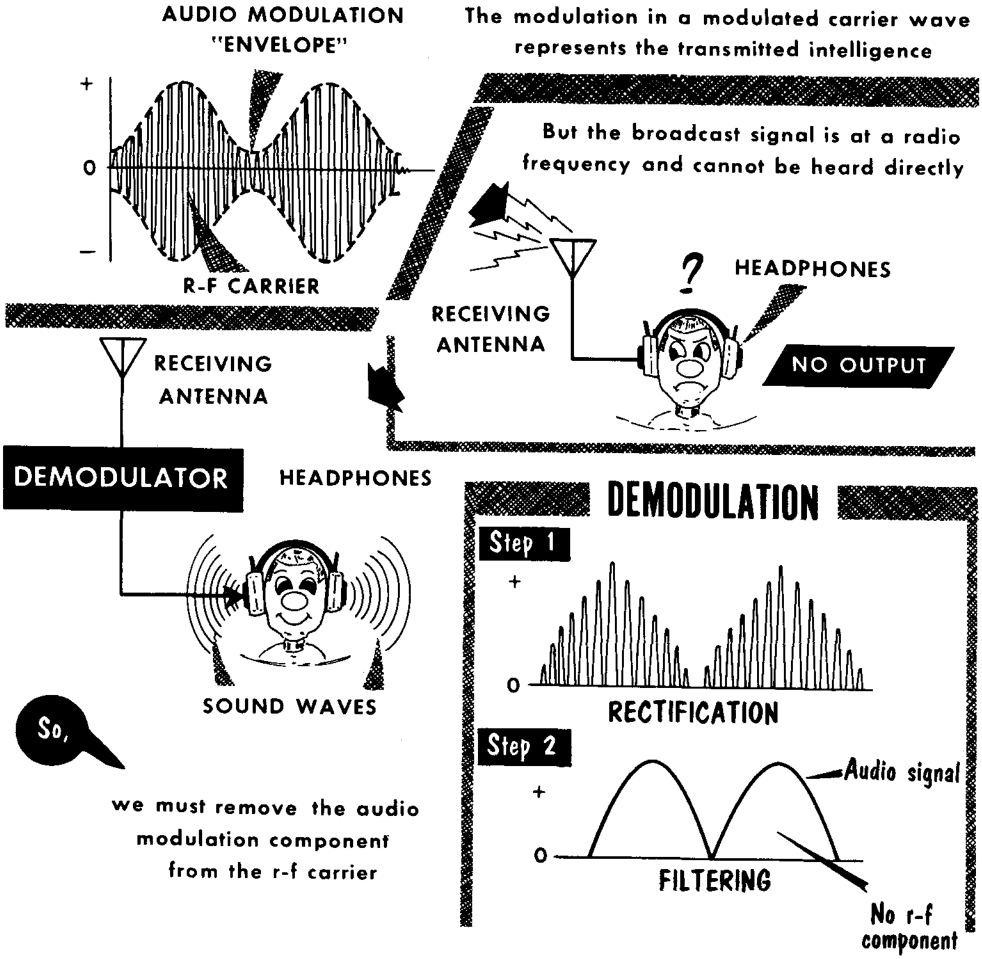
MODULATION is the PROCESS by which INTELLIGENCE is SUPERIMPOSED on a CARRIER WAVE



Radio signals are emitted from the transmitting antenna of a radio broadcast station. Basically, the broadcast station generates a powerful train of oscillations called a carrier wave. A circuit in the transmitter called the modulator causes the amplitude of this carrier wave to vary in accordance with the audio intelligence being broadcast. The process of making the amplitude of the carrier wave vary, is known as amplitude modulation (AM). Simple intelligence can be produced merely by a make-and-break process of the carrier wave to form dots and dashes; no real modulator is needed. The modulator is basically nothing more than a powerful audio amplifier. The carrier wave gets its name from actually "carrying" the audio modulation; it serves no other purpose than that. Modulation will be discussed in greater detail in Volume 6.

Demodulation

When the transmitted signal is intercepted by the receiving antenna of the radio receiver, its waveform is identical to that transmitted. It contains the high-frequency r-f carrier component and the audio component in the form of varying amplitude. It is the function of the modulator circuit in a receiver to

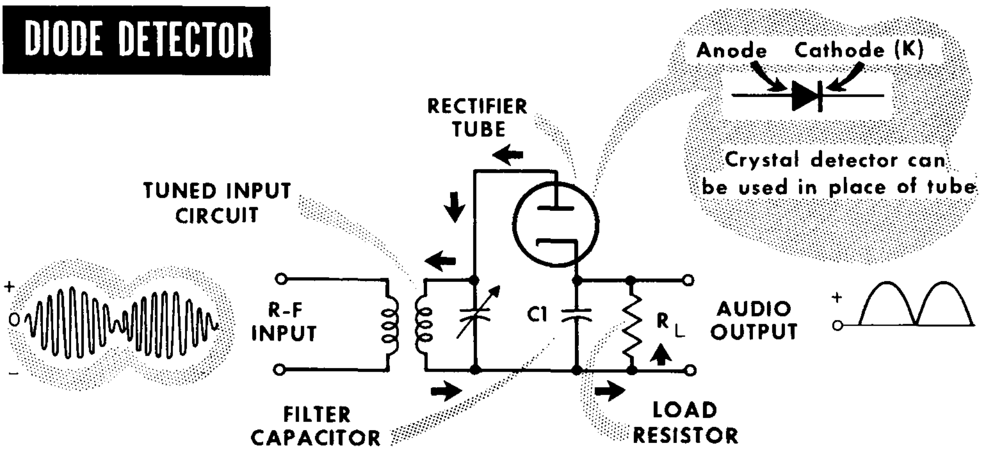


take this radio broadcast signal into its input and produce an audio signal at its output.

The modulator performs, basically, two tasks: one of rectification, in which the positive or negative alternations of the carrier are removed, and the second of filtering in which the r-f fluctuations are removed leaving only the a-f component or "modulation envelope." In the demodulator, the audio component is still not yet an a-c signal. It is a d-c signal that varies in amplitude at the audio rate.

The Diode Detector

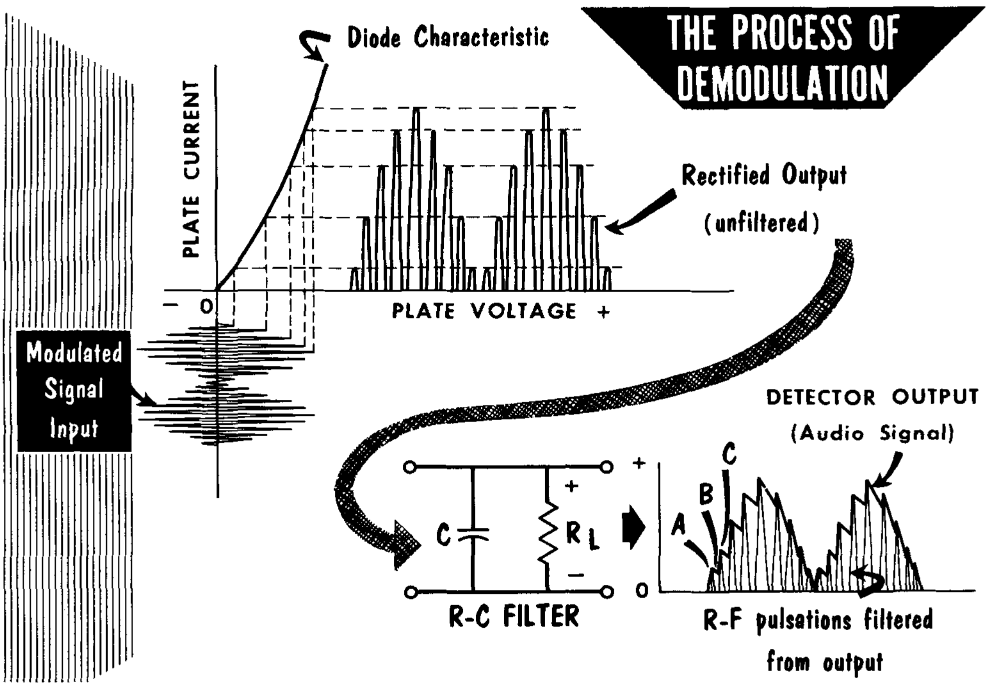
The simplest and most commonly used demodulator is the diode "detector" circuit. Its function is to convert the modulated r-f carrier to a direct current, varying at the a-f rate of the original modulated signal. The diode (tube or crystal) represents an ideal circuit device for signal rectification in that it permits current flow in one direction and not in the other. Because the diode characteristic is nearly straight, the diode detector is called a linear detector. However, with weak signals, the output of the detector follows the "square law", because weak signals force the operation to take place on the lower curved portion of the characteristic. In square-law detection, the output is proportional to the square of the input voltage. Detectors are generally rated in terms of their sensitivity – the ratio of the output signal to the input signal voltage; linearity – the ability to detect signals throughout the a-f range without distortion; and signal-handling capacity – the ease with which a detector circuit handles a signal without distortion.



The diode detector input is generally a tuned circuit. The other basic components are the diode load resistor and the filter capacitor. The diode detector can handle large signals without overloading, and it can provide an important avc voltage (discussed later) without extra tubes or special circuits. However, it has the disadvantage of drawing power from the input tuned circuit, because the diode and its load form a low-impedance shunt across the circuit. As a result, the circuit Q, the sensitivity, and the selectivity, are reduced. Because the diode detector distorts on weak signals, for optimum operation considerable amplification is needed before detection.

Let us now analyze the action of the diode detector. The incoming modulated r-f signal voltage is developed across the tuned circuit of the detector. Signal current flows through the diode only when the plate is positive with respect to the cathode (only on the positive half-cycles). The rectified signal flowing through the diode actually consists of a series of r-f pulses, and not a smooth outline or "envelope."

Action of the Diode Detector



On the first quarter-cycle of the applied r-f voltage, C1 charges almost up to peak value of the r-f voltage (point A). The small voltage drop in the tube prevents C1 from charging up completely. Then, as the applied r-f voltage falls below its applied value, some of the charge of C1 leaks through R_L, and the voltage drops only a small amount to B. When the r-f voltage applied to the plate on the next cycle exceeds the potential at which the capacitor holds the cathode (point B), diode current again flows, and the capacitor charges up to almost the peak value of the second positive cycle at C. Thus, the voltage across C closely follows the peak value of the applied r-f voltage, and reproduces the a-f modulation. The detector output after rectification and filtering is a d-c voltage that varies at an audio rate. The output voltage across C is shown somewhat jagged. Actually, the r-f component of this voltage is negligible, and after amplification, the speech or music originating at the transmitter is faithfully reproduced.

The choice of R_L and C1 in a diode detector is very important for maximum sensitivity and fidelity. R_L and the diode plate resistance act as a voltage divider to the received signal. Therefore, R_L should be high compared with the diode plate resistance for maximum output voltage. In addition, the value of C1 should be such that the R-C time constant is long, compared to the time of one r-f cycle. This is necessary because C must maintain a voltage across R_L during the time there is no plate current. Also, the R-C time constant must be short, compared with the time of one a-f cycle, so that the capacitor voltage can follow the modulation envelope.

The Grid-Leak Detector

The operation of the grid-leak detector is similar to that of the diode detector. The signal voltage applied to the grid of a triode is alternately positive and negative. Grid current flows during the half-cycle in which the grid is positive with respect to the cathode. As a result, pulsating dc flows through R_g. Filter C_g smooths the r-f pulses. A d-c voltage is produced across R_g which varies at an audio rate, just as in the diode detector. This audio voltage is used as the signal voltage input for the triode amplifier. As a result, an amplified audio signal appears in the plate circuit of the grid-leak detector. Capacitor C is an additional r-f filter.

The grid-leak detector is a square-law device, with the output varying as the square of the r-f input voltage. The development of higher-gain r-f amplifiers led to the replacement of the grid-leak detector by the diode detector. In comparison to the diode detector, the grid-leak detector has higher sensitivity, because of its amplification ability, and poorer linearity, because of its operation as a square-law detector. Selectivity is equally poor because it draws grid current through the tank circuit, lowering the Q, and it has a much lower signal-handling ability. The principal advantage of the grid-leak detector is that it provides a stage of audio amplification. The voltage applied to this circuit must not be so high that it causes the average grid voltage to exceed the plate current cutoff voltage for the tube. It is this characteristic that limits the power-handling capacity of the grid-leak detector.

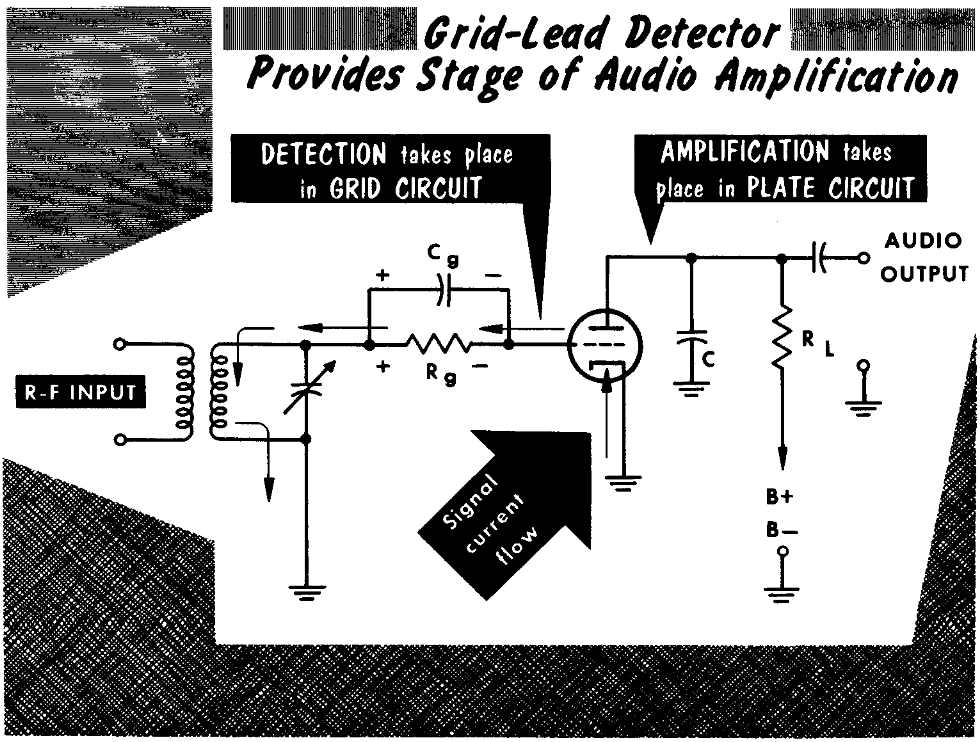
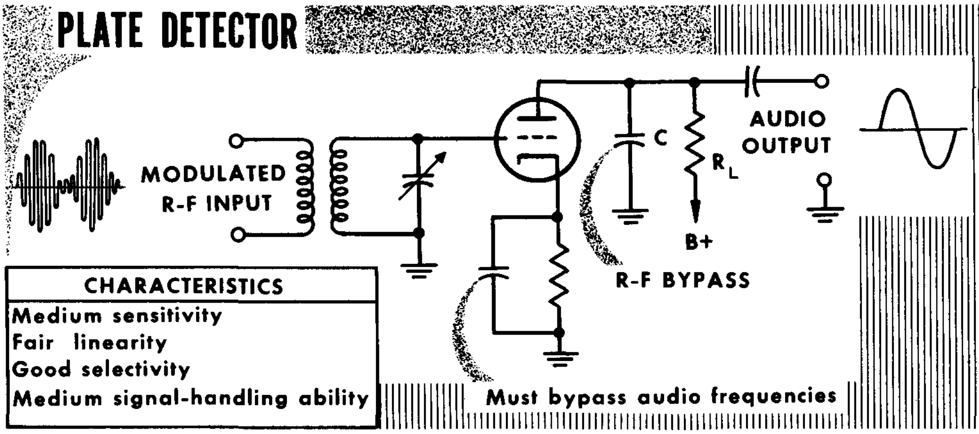
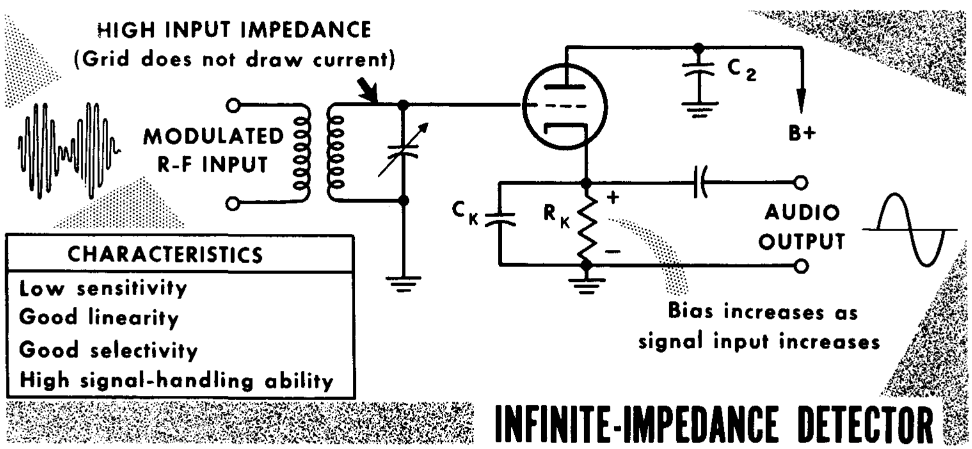


Plate Detector and Infinite Impedance Detector



The plate detector gets its name because detection occurs in the plate circuit. Operation is similar to that of a Class-B amplifier. Although cathode bias cannot produce plate current cutoff, operation at the lower end of the dynamic characteristic is possible. Normal plate current flows during the positive half-cycle of the input signal, with most of the negative half-cycle cut off. As a result, the average value of the plate current varies in accordance with the audio variations. Capacitor C acts as an r-f filter.

If the plate of the plate detector is connected directly to B+ and the output is taken across the cathode network, the result is an infinite impedance detector. Although there is no amplification of the signal in this circuit, which has the advantage of good reproduction, it has good signal-handling capacity for large inputs. The modulated r-f signal varies the d-c plate current through the tube. This current returning to the cathode network is filtered through C_k, and the current passing through R_k is dc, varying at an audio rate with negligible r-f ripple.

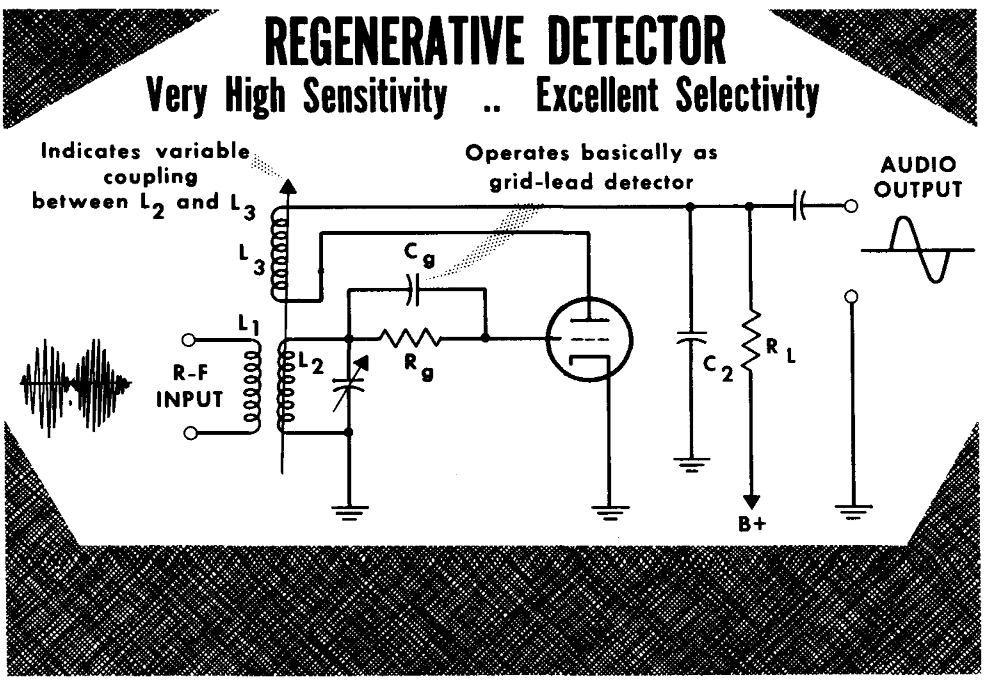


The Regenerative Detector

When high sensitivity and selectivity are the most important factors to be considered, a regenerative detector may be used. However, its linearity, as well as the ability to handle strong signals without overloading, is very poor. The process of feeding some of the output voltage of an electron-tube circuit back into the input circuit so that it adds to, or reinforces (is in phase with) the input voltage, is known as regeneration or positive feedback. The use of regeneration in a circuit greatly increases the amplification of the circuit because the output voltage fed back to the input circuit adds to the original input voltage, thus increasing the total voltage to be amplified by the tube.

A grid-leak detector may be modified to operate as a regenerative detector. Because an amplified r-f component is present in the plate circuit of the grid-leak detector, regeneration can be obtained by connecting coil L3 (known as a tickler coil) in series with the plate circuit, so that it is inductively coupled to grid coil L2. With an r-f signal across L2, an r-f component of plate current flows through L3. L3 is connected so that the voltage it induces in L2 is in phase with the incoming signal voltage applied to the grid. Thus, the voltage gain of the stage is increased.

It is important that the voltage fed back by the tickler coil be in phase with the incoming signal voltage. Otherwise the feedback will be degenerative, and amplification will be reduced. Furthermore, if the coupling between L2 and L3 is too great, oscillation occurs. The regenerative detector is the most sensitive triode detector circuit possible, when it is operated just below the point of oscillation.



To produce oscillations, an electron tube circuit must contain a tuned circuit having the proper amounts of inductance and capacitance to oscillate at the desired frequency; it must be capable of amplifying a signal at its control grid; it must have a means of providing the tuned circuit with sufficient regenerative energy to sustain oscillations.

The tuned grid oscillator obtains regenerative or positive feedback by coupling the plate circuit to the tuned grid circuit.

There are two basic types of split-tank oscillators – the Hartley and the Colpitts. The Hartley oscillator has a split-inductance tank divided between the grid and plate circuits; the Colpitts oscillator has a split-capacitance tank divided between the grid and plate circuits.

An oscillator is shunt-fed when its d-c plate supply is in parallel with the oscillating plate circuit.

The electron-coupled oscillator replaces the separate oscillator with an oscillator having "insulation" from loading effects. Coupling to the output circuit is through the electron stream in the tube.

Bias for r-f amplifiers is usually obtained from the action of the grid drawing current during part of the input cycle.

Oscillators generally use grid-leak bias.

Demodulation is the process by which a circuit separates the modulation component from the r-f component of a carrier wave.

The simple diode detector is the most commonly used demodulator.

Modulation is the process in which intelligence is superimposed on the r-f carrier wave at the broadcast station.

The grid-leak detector provides a high degree of sensitivity; however, its selectivity is poor.

The signal-handling ability of a detector is an indication of the amount of signal amplitude that can be handled without overloading the circuit. The sharpness with which a detector can be tuned determines its selectivity. The amount of distortion in the a-f signal as compared with the original sound is a measure of the linearity of a detector. A detector providing some amplification has greater sensitivity than one providing no amplification.

REVIEW QUESTIONS

1. Why does a damped oscillation occur in an L-C circuit?
2. Under what conditions can oscillations be maintained in an L-C circuit?
3. What is the function of the amplifier in an electron-tube oscillator?
4. Name the basic types of split-tank oscillators.
5. Explain the operation of the electron-coupled oscillator.
6. Explain how grid-leak bias is developed in an oscillator.
7. Explain the operation of the Armstrong oscillator.
8. Explain the fundamental characteristics of modulation and demodulation.
9. Explain how detection takes place in the diode detector.
10. Explain the operation of the grid-leak detector.
11. Explain the operation of the regenerative detector.
12. Compare the sensitivity of the diode, grid-leak, and plate detectors.

GLOSSARY

Amplification Factor: The ratio of a small change in plate voltage to a small change in grid voltage, with all other electrode voltages constant, required to produce tthe same change in plate current.

Amplifier: A device used to increase the signal voltage, current, or power, generally composed of a vacuum tube and associated circuit. It may contain several stages to obtain a desired gain.

Anode: A positive electrode. The plate of a vacuum tube.

Audio Frequency: A range of frequencies that can be detected as a sound by the human ear.

Beam-Power Tube: A vacuum tube in which the electron stream is directed in concentrated beams from the cathode to the plate.

Bias: The average d-c voltage maintained between the cathode and control grid of a vacuum tube.

Blocking Capacitor: A capacitor used to block the flow of dc while permitting the flow of ac.

Cathode: Negatively charged pole, electrode, conductor, or element from which current leaves. The primary source of electrons in a vacuum tube.

Coupling: The association of two circuits in such a way that energy may be transferred from one to the other.

Cutoff: The minimum value of negative grid bias which prevents the flow of plate current in a vacuum tube.

Detection: The process of separating the modulation component from the received signal.

Diode: A two-electrode vacuum tube containing a cathode and a plate.

Distortion: The production an an output waveform which is not a true reproduction of the input waveform.

Dynamic Characteristics: The relationship between the instantaneous plate voltage and plate current of a vacuum tube as the voltage applied to the grid is moved; thus, the characteristics of a vacuum tube during operation.

Feedback: A transfer of energy from the output circuit of a device back to its input.

Gain: The ratio of the output power, voltage, or current to the input power, voltage, or current respectively.

Grid: An electrode consisting of a wire mesh placed between cathode and plate in an electron tube, and used to control the electron flow through the tube.

Intermediate Frequency: The fixed frequency to which r-f carrier waves are converted in a superheterodyne receiver.

Local Oscillator: The oscillator used in a superheterodyne receiver, the output of which is mixed with the desired r-f carrier to form the intermediate frequency.

Modulation: The process of varying the amplitude (AM), the frequency (FM), or the phase (PM) of a carrier wave in accordance with other signals to convey intelligence.

Negative Feedback: The process whereby a part of the output signal of an amplifying device is returned to the input circuit in such a manner that it tends to cancel the input.

Oscillator: A circuit capable of converting dc into ac of a frequency determined by the constants of the circuit.

Paraphase Amplifier: An amplifier which converts a single input into a push-pull output.

Pentode: A five-electrode vacuum tube containing, a cathode, control grid, screen grid, suppressor grid, and plate.

Plate: The principal electrode in a tube to which the electron stream is attracted.

Plate-Load Impedance: The impedance in the plate circuit across which the output-signal voltage is developed by the alternating component of the plate current.

Plate Resistance: The internal resistance to the flow of ac between the cathode and plate of a tube. It is equal to a small change in plate voltage divided by the corresponding change in plate current, and is expressed in ohms.

Power Amplification: The process of amplifying a signal to produce a gain in power, as distinguished from voltage amplification.

Push-Pull Circuit: An amplifier circuit using two vacuum tubes in such a way that when one tube is operating on a positive alternation, the other operates on a negative alternation.

Radio Frequency: Any frequency of electrical energy capable of propagation into space.

Radio-Frequency Amplification: The amplification of a radio wave by a receiver before detection.

Rectifier: A device that changes alternating current into unidirectional current.

Self Bias: The bias of a tube created by the voltage drop developed across a resistor through which either its cathode or its grid current flows.

Shielding: A metallic covering to prevent magnetic or electrostatic coupling between adjacent circuits.

Space Current: The total current flowing between cathode and all the other electrodes in a tube. This includes the plate current, grid current, screen-grid current, and any other electrode current which may be present.

Static Characteristics: The characteristics of a tube with no output load and with d-c voltages applied to the grid and plate.

Superheterodyne: A receiver in which the incoming signal is mixed with a locally generated signal to produce a predetermined intermediate frequency.

Suppressor Grid: An electrode used in a vacuum tube to minimize the harmful effects of secondary emission from the plate.

Tetrode: A four-electrode vacuum tube containing a cathode, control grid, screen grid, and plate.

Thermionic Emission: Electron emission caused by heating an emitter.

Transconductance: The ratio of the change in plate current to the change in grid voltage producing this change in plate current, while all other electrode voltages remain constant.

Triode: A three-electrode vacuum tube, containing a cathode, control grid, and plate.

Vacuum Tube: Device consisting of an evacuated enclosure containing a number of electrodes that control the conduction of electrons through the vacuum.

Variable-mu tube: A vacuum tube in which the control grid is irregularly spaced, so that the grid exercises a different amount of control on the electron stream at different points within its operating range.

Voltage Regulation: A measure of the degree to which a power source maintains its output voltage stability under varying load conditions.

Voltage Amplification: The process of amplifying a signal to produce a gain in voltage. The voltage gain of an amplifier is the ratio of its alternating voltage output to its alternating voltage input.

INDEX TO VOL. III

(Note: A cumulative index covering all six volumes in this series is included at the end of Volume VI.)

basic radio

by MARVIN TEPPER

Electronic Services Division
Raytheon Company

Author of
FUNDAMENTALS OF RADIO TELEMETRY

VOL. 4

JOHN F. RIDER PUBLISHER, INC., NEW YORK

Copyright September 1961 by John F. Rider Publisher, Inc.

All rights reserved. This book or any parts thereof may not be reproduced in any form or in any language without permission of the publisher.

Library of Congress Catalog Card Number 61-11229

Printed in the United States of America

Fifth Printing, 1968

PREFACE

The purpose of this book is to fill a need for a text stating in plain, everyday language, what many people consider a complex technical subject. A technical subject need not be complex. Careful filling in of the background with essential information, and then leading step by step to the final explanation, provides a logical method of explaining the most difficult subject.

It would be impossible to cover in a single book or series of books, the immense scope implied in the word *electronics*. However, an understanding of radio circuits serves as a foundation for advanced study in all fields of electronics, such as television, radar, computers, etc. For teaching radio, the all-important basic tool of electronics, most available textbooks are woefully inadequate. One type contains information so brief as to acquaint rather than instruct. Another type is based on the premise that teaching a student to design a circuit is the best method of having him understand that circuit's operation.

Basic Radio represents the neglected middle ground. It is a course in radio communications, as distinct from a general course in electronics. The text deals with the circuitry and techniques used for the transmission and reception of intelligence via radio energy. Assuming no prior knowledge of electricity or electronics, the six volumes of this course "begin at the beginning" and carry the reader in logical steps through a study of electricity and electronics as required for a clear understanding of radio receivers and transmitters. Illustrations are used on every page to reinforce the highlights of that page. All examples given are based on actual or typical circuitry to make the course as practical and realistic as possible. Most important, the text provides a solid foundation upon which the reader can build his further, more advanced knowledge of electronics.

The sequence of *Basic Radio* first establishes a knowledge of d-c electricity. Upon this is built an understanding of the slightly more involved a-c electricity. Equipped with this information the reader is ready to study the operation of electron tubes and electron tube circuits, including power supplies, amplifiers, oscillators, etc. Having covered the components of electronic circuitry in Volumes 1 through 3, we assemble these components

in Volume 4, and develop the complete radio receiver, AM and FM. In Volume 5 we recognize the development of the transistor, and devote the entire volume to the theory and circuitry of transistor receivers and semiconductors. The last volume of the course, Volume 6, covers the long-neglected subject of transmitters, antennas, and transmission lines.

No prior knowledge of algebra, electricity, or any associated subject is required for the understanding of this series; it is self-contained. Embracing a vast amount of information, it cannot be read like a novel, skimming through for the high points. Each page contains a carefully selected thought or group of thoughts. Readers should take advantage of this, and study each individual page as a separate subject.

Whenever someone is presented with an award he gives thanks and acknowledgement to those "without whose help . . . " etc. It is no different here. The most patient, and long-suffering was my wife Celia, who typed, and typed, and typed. To her, the editorial staff of John F. Rider, and others in the "background", my heartfelt thanks and gratitude for their assistance and understanding patience.

MARVIN TEPPER

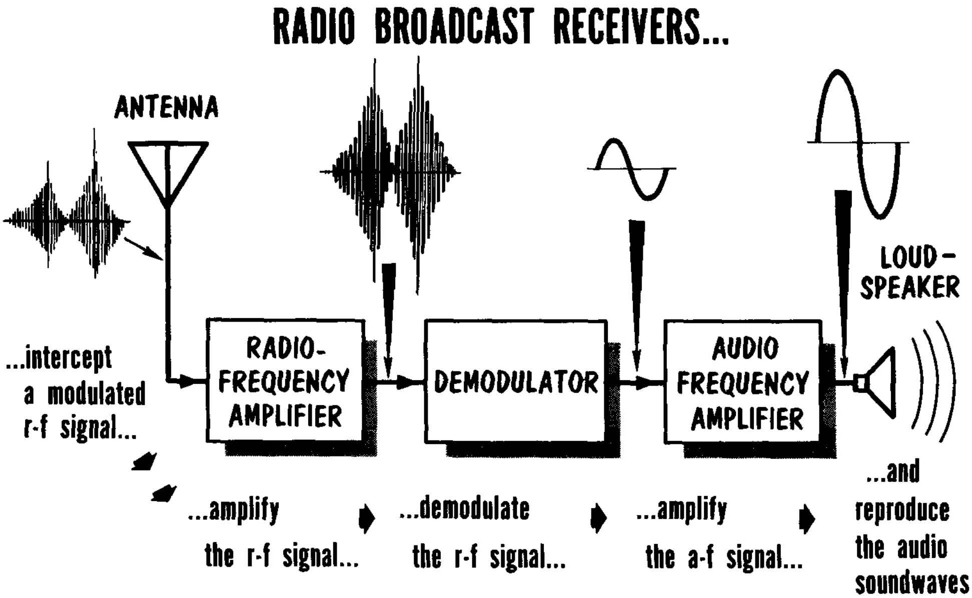
Malden, Mass.
September 1961

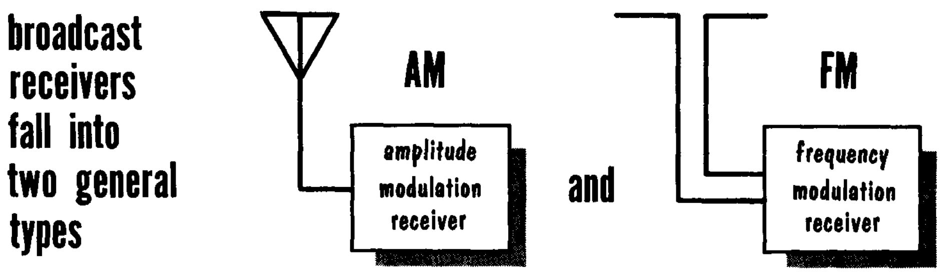
CONTENTS

VOL. IV — BASIC RADIO

The Radio Receiver

In Volume 3 we discussed electron tubes and electron tube circuits used in radio receivers. In this volume we will cover in detail the actual receivers and learn how the circuits fit together. We will see that the fundamental performance of all radio broadcast receivers is the same. All receivers intercept a radio signal at their antenna, amplify it, demodulate it, and then





reproduce the same audio-frequency signal that was emitted at the broadcast studio of the transmitting station. It is in the various techniques and processing of the signal that the receivers differ.

We have spoken of amplitude-modulated signals, but broadcast signals may also be frequency modulated, and special circuits are required for processing FM signals. There are many special circuits used only with complete radio receivers. These will be covered when we begin connecting circuits together to develop the complete modern radio broadcast receiver.

The Frequency Spectrum

Radio signals are radiated from transmitting antennas over a range known as the radio-frequency spectrum. The chart shows this very wide range. No one receiver is designed for the entire spectrum. Actually, most receivers are designed to receive radio signals over a small portion of one of the bands, although some receivers of the "multiband" type provide a range from 100 kc to 150 mc. At the higher frequencies many new circuit techniques and

THE RADIO-FREQUENCY SPECTRUM

frequency range	designation	wavelength
10-30 kc	Very Low Frequency (VLF)	30,000-10,000 meters
30-300 kc	Low Frequency (LF)	10,000-1000 meters
300-3000 kc	Medium Frequency (MF)	1000-100 meters
3-30 mc	High Frequency (HF)	100-10 meters
30-300 mc	Very High Frequency (VHF)	10-1 meter
300-3000 mc	Ultra High Frequency (UHF)	1—.1 meter
3000-30,000 mc	Super High Frequency (SHF)	1—.01 meter

FREQUENCY ALLOCATION (20kc — 300 mc)

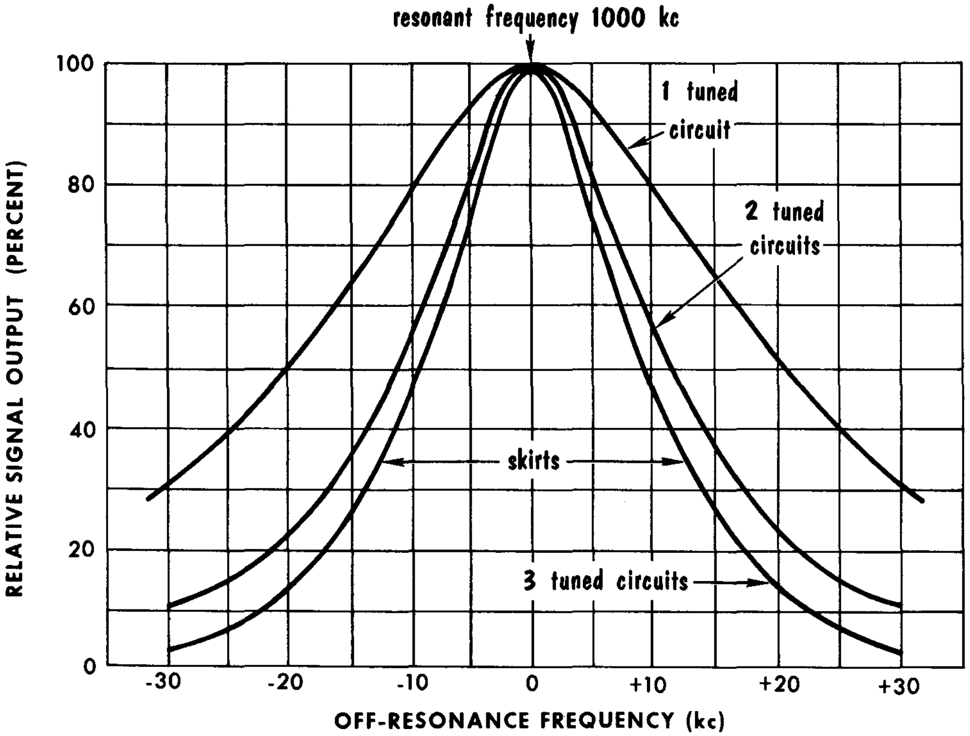
frequency band	uses
10 - 535 kc	Government, Commercial, Maritime, Aircraft, Ship-to-Shore, Transoceanic, High Power, Point-to-Point
535 - 1605 kc	Commercial Broadcast
1605 kc - 54 mc	Ship-to-shore, Amateur, Aircraft, Citizens Band, Police, Foreign, Experimental, Government
54 - 88 mc	Television (Channels 2, 3, 4, 5, 6)
88 - 108 mc	Frequency Modulation
108-300 mc	Television, Amateur, Aircraft, Government, Police

tubes come into play, and many of these will be covered here. However, we will concern ourselves primarily with the radio broadcast band of 535 kc to 1605 kc, and the shortwave bands up to 30 mc.

We have discussed operating frequencies. We can also think in terms of wavelength. Radio waves travel at the speed of 300,000,000 meters per second. Hence, frequency (in mc) is equal to $300/\lambda$ (wavelength, in meters), or wavelength (in meters) is equal to $300/f$ (frequency, in megacycles). Thus, a station operating at 30 mc may be said to be operating at 10 meters. The term wavelength is often used at frequencies above 3 mc.

Selectivity

The selectivity of a receiver is a measure of its ability to respond, or select a particular frequency or band of frequencies while rejecting all others. Higher selectivity does not necessarily make a better receiver. A radio broadcast signal consists of the carrier frequency and its sidebands. The sidebands are sum and difference frequencies produced at the transmitter by the modulating frequencies. For instance, a 5000-cycle audio tone might



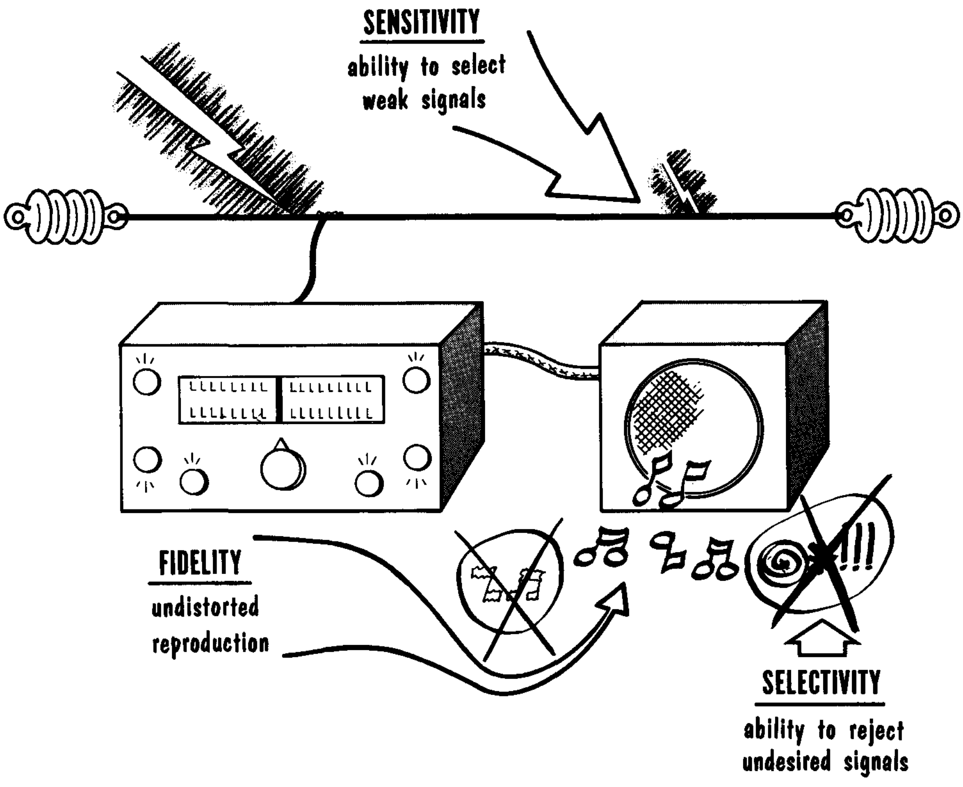
RECEIVER SELECTIVITY INCREASES WITH NUMBER OF TUNED CIRCUITS

be used to modulate an 800-kc carrier wave. This would produce frequencies of 800 kc, 800 kc plus 5 kc, or 805 kc, and 800 kc minus 5 kc, or 795 kc. Thus, for the receiver to properly reproduce the entire broadcast signal of from 795-805 kc, it would have to pass a band of frequencies of 10 kc. A receiver more selective than this would not pass both sidebands.

Not only must 10 kc be passed in this case, but all other r-f signals must be rejected. If the nearby sidebands of another station are received by the r-f circuits, they may "ride" through the receiver and be heard as background sounds. You may have heard more than one station at a time on receivers having low selectivity. Selectivity is determined by the tuned circuits in the r-f section of a receiver. The higher the Q of a circuit, the greater its selectivity.

Sensitivity and Fidelity

Sensitivity and fidelity are two other important characteristics of receivers. The sensitivity of a radio receiver is its ability to respond to weak signal voltages. It is measured in terms of the voltage that must be induced in the antenna by the signal to develop a standard output (arbitrarily chosen as 0.05 watt) from the power output amplifier. The sensitivity of broadcast receivers is seldom higher than a few microvolts in the most sensitive receivers, and

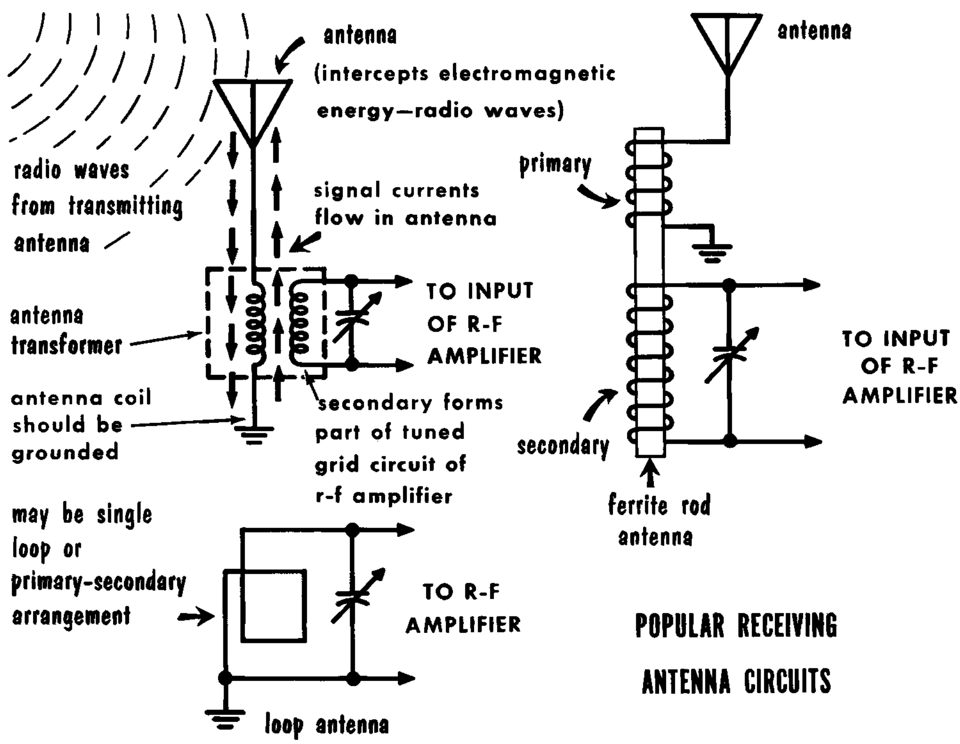


in the less-sensitive sets, is in the order of 100 microvolts. Under the most favorable conditions, such as in specially designed communications receivers, the highest usable sensitivity is in the order of less than 1 microvolt.

While sensitivity and selectivity are the measure of a receiver's ability to intercept a weak signal and to extract the intelligence from that signal to the exclusion of all others, fidelity is a measure of a receiver's ability to reproduce the intelligence of the signal. It must reproduce as faithfully as possible the signal or program which is received. That is, all the frequencies in the original input signal at the transmitter should be received and reproduced without distortion in the receiver output. With the growth of "high fidelity" audio equipment, it is becoming increasingly important that the r-f portions of a receiver do not introduce any distortion into the signal.

Antennas and Antenna Circuits

Radio signals are emitted from a transmitting antenna in the form of electromagnetic energy. As this energy radiates outward it will induce signal voltages in any conductor intercepting the electromagnetic waves. An antenna is a specially designed conductor whose function is to intercept this electromagnetic energy. As a result, signal currents flow in the antenna which eventually appear as r-f input voltages to the r-f amplifier. Actually, an antenna is itself a tuned circuit and provides some selectivity. The formula for resonance is $f = 1/2\pi\sqrt{LC}$. Since an antenna contains a certain amount of distributed inductance and capacitance, it will have a certain resonant frequency. In conventional broadcast receivers the antenna is not variable and



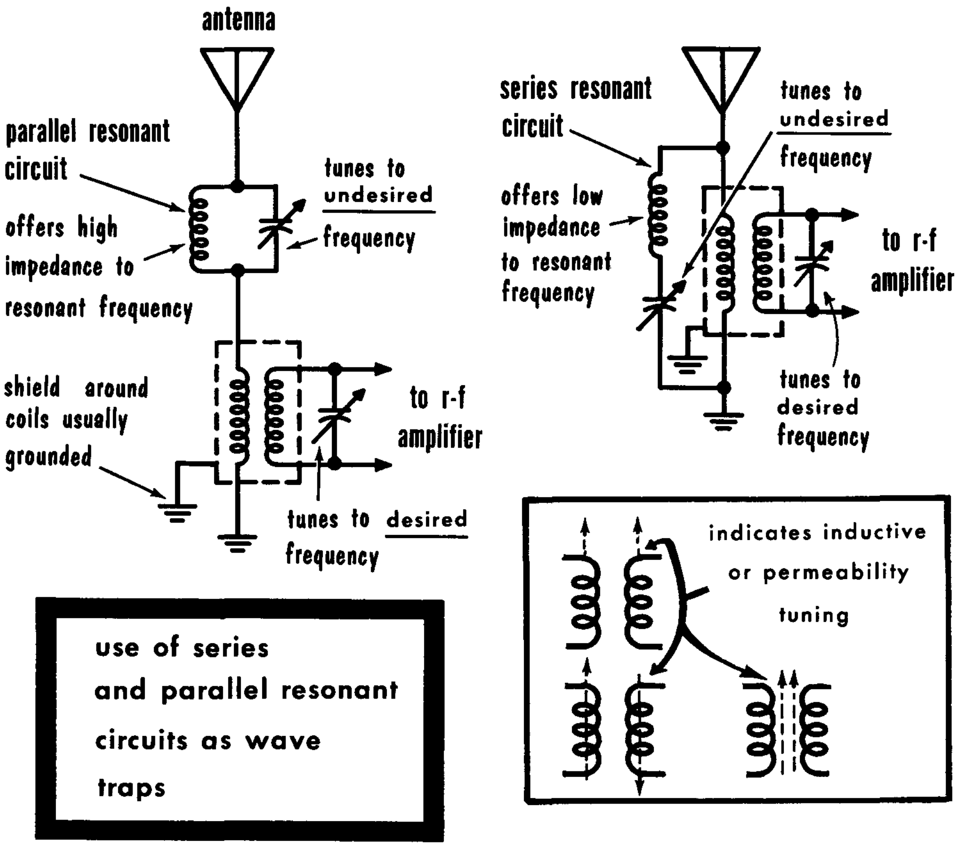
POPULAR RECEIVING ANTENNA CIRCUITS

is designed to resonate broadly through the broadcast band. In some more expensive communications receivers, provision is made for antenna tuning to obtain maximum selectivity. In general, tuned antennas are used at higher frequencies.

Many portable receivers use a loop antenna that is coupled through an antenna transformer to the input of the r-f amplifier. A recent innovation is the ferrite-rod antenna. This device has an extremely high permeability with excellent "pick up" characteristics. At low frequencies, such as the broadcast band, a long-wire antenna will produce optimum results.

Antennas and Antenna Circuits (Cont'd)

Antenna transformers are used to couple the antenna to the tuned circuit. Although a transformer, it is most often referred to as the antenna coil. The primary winding isolates the antenna from the tuned circuit and provides the proper impedance match to the antenna. Most often the antenna transformer uses an air core. However, at broadcast band frequencies, a powdered-iron



use of series and parallel resonant circuits as wave traps

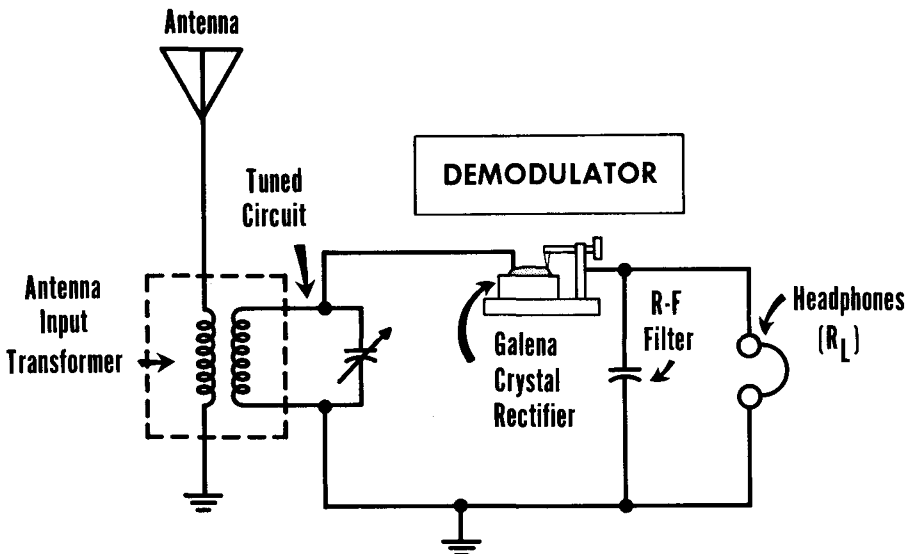
core may be used. Antenna transformer windings are often mounted in shielded containers with color codes used to identify the leads or lugs.

Wavetraps are often used in antenna circuits to improve the selectivity of the receiver. A wavetrap is an L-C device for preventing undesired signals from nearby transmitters or from extra powerful transmitters from drowning out desired signals from weaker or more distant transmitters. As shown, these wavetraps may be parallel- or series-resonant circuits. In each case the tuned trap is used to reject or attenuate the undesired signal. A separate tuned circuit in the r-f input selects the desired signal. In most instances the shield can in which the tuned circuit is mounted is grounded to the metal chassis of the receiver.

Basic Crystal Receiver

Among the first "receivers" used in the early days of radio was the crystal set. It was a receiver in that it did indeed intercept radio waves, demodulate them, and reproduce the audio signals. The heart of this receiver was a mineral crystal such as galena or carborundum. These possess the interesting characteristic of permitting current flow in only one direction--in short, they act as a diode. In most of the original crystal receivers galena was used, and many a radio enthusiast spent hours adjusting a "cat's whisker" to the surface of the galena to find the best spot for signal reception. The diagram shows the basic arrangement of the crystal detector.

The antenna usually consisted of an extremely long wire antenna for maximum signal pickup or interception. When the radio wave passes or cuts across the antenna, r-f voltages are induced in the antenna. By mutual in-



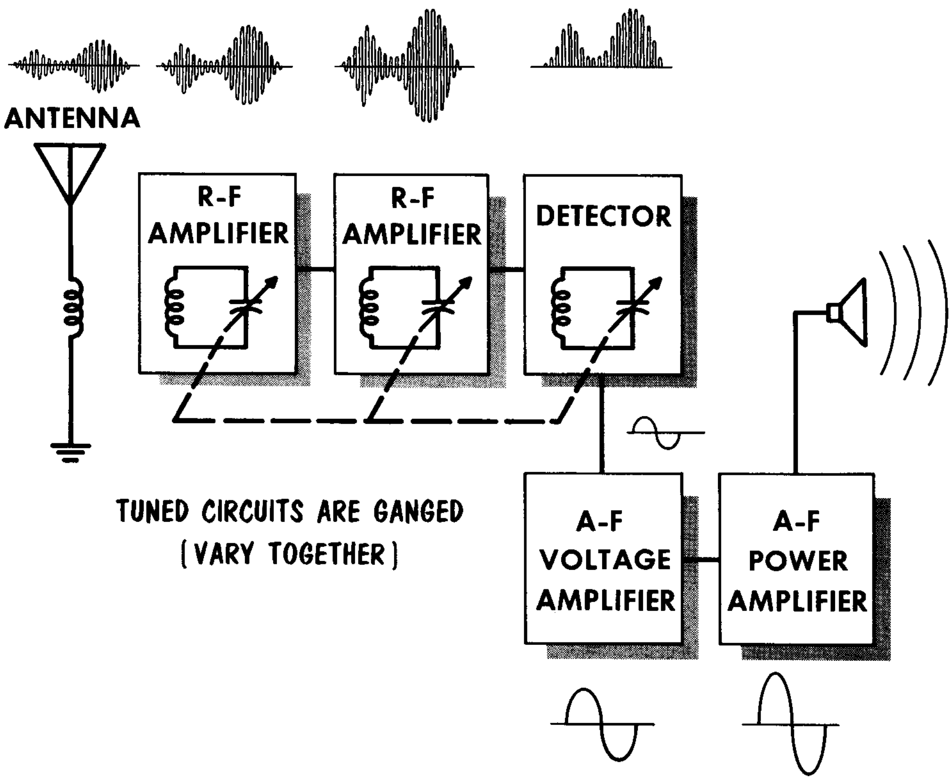
EARLY CRYSTAL DETECTOR

duction, current flow in the primary induces current flow in the secondary. The antenna transformer usually has an air core and step-up ratio, permitting a small amount of signal gain. The secondary is a series tuned circuit, with maximum circulating current at resonance. The crystal rectifies the signal, with the capacitor acting as the r-f filter in conjunction with the headphones which acts as the load. With one tuned circuit, this receiver has very little selectivity, and with no amplifiers, very little sensitivity.

The TRF Receiver

The first practical radio receiver in wide use was the tuned-radio-frequency type. This consisted of an antenna input system, two or more stages of r-f amplification, a stage of demodulation, and one or more stages of a-f amplification, followed by a loudspeaker. It had also, of course, a power supply. The amplitude of the signal at the input of the receiver is small because it has been attenuated in the space between the transmitter and the receiver. It consists of the carrier frequency and the modulation envelope. The r-f

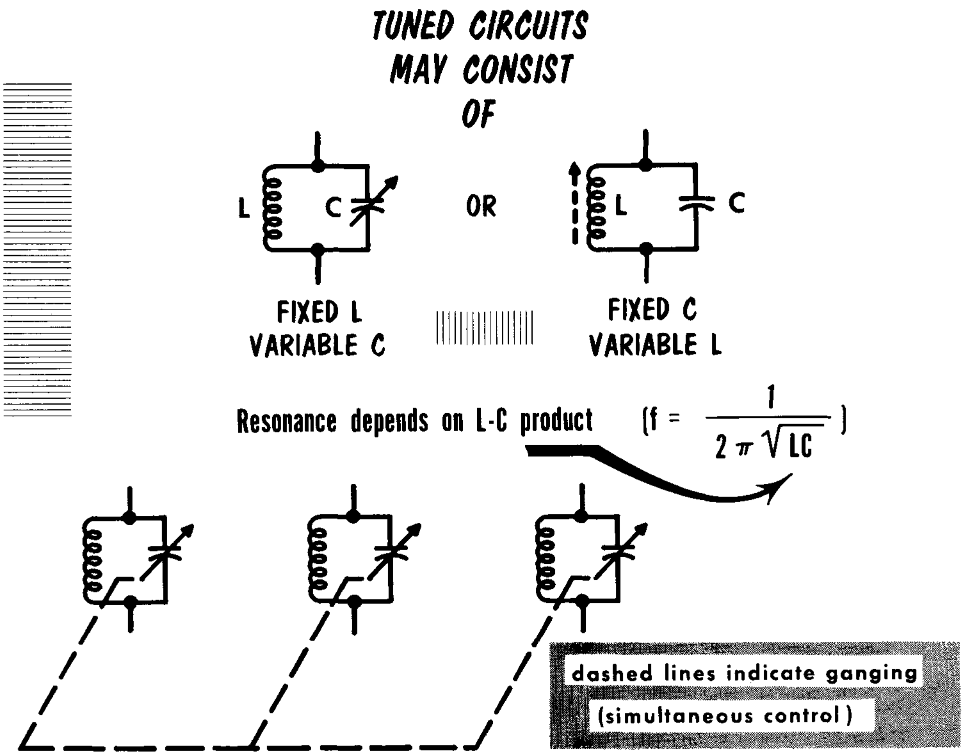
BASIC ORGANIZATION OF THE TRF RECEIVER



stages amplify the waveform, but they do not change its basic shape. The detector rectifies and removes the r-f component of the signal. The output of the detector is a weak signal made of the modulation envelope or component of the incoming signal. The a-f amplifier stages following the detector increase the amplitude of the audio signal to a value sufficient to drive a set of headphones or a loudspeaker. The TRF receiver was a practical device because it provided amplification, and its various tuned circuits provided good selectivity. While this receiver is largely replaced by the modern superheterodyne, the TRF receiver is still used today for certain applications. Let us now examine the circuitry of the TRF receiver.

Tuned Circuits

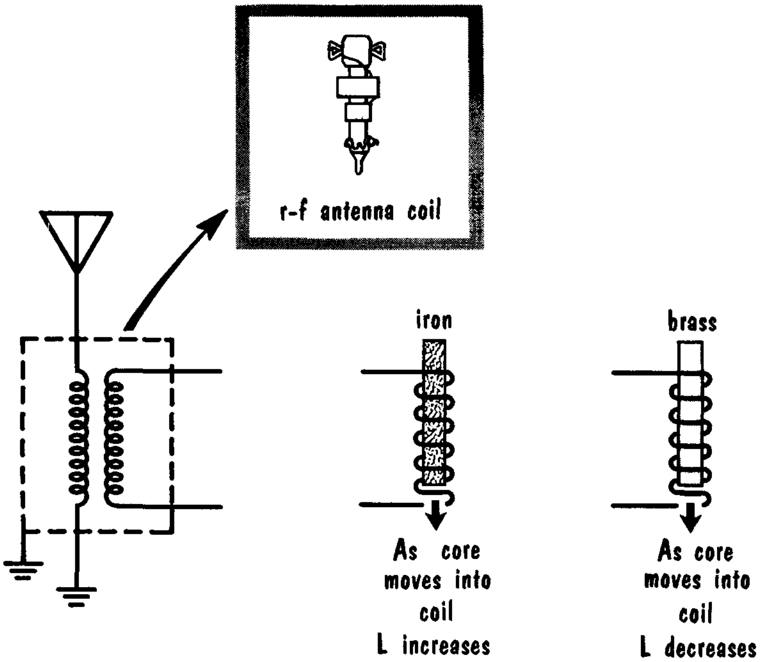
The TRF receiver introduces many tuned circuits in the r-f stages. All tuning is done by varying L or C or an LC circuit. Resonance depends on the value of L and C. Increasing the value of L or C decreases the resonant frequency; decreasing L or C increases the resonant frequency. Changing the value of both may change the resonant frequency--it depends on whether the



product of L and C is changed. If L and C are changed so that their product remains the same, the frequency remains the same. Thus for the same frequency L may be high and C low, or L low and C high. The L/C ratio is important to the Q of a tuned circuit. Since Q = Z/R, the ratio of impedance to resistance affects the Q. If the resistance remains the same, raising L and lowering C gives a higher ratio of Z to R, and a higher Q. A high Q has the advantage of providing sharp tuning and high gain, thus improving selectivity and sensitivity. If the Q is too high, tuning may be so sharp that it eliminates part of the sidebands and introduces distortion.

The tuned circuits in the TRF receiver are usually tuned by variable capacitors. The capacitors are ganged to be varied simultaneously by one control. Because the capacitors can be varied over a range of frequencies, the receiver is provided with its frequency coverage. Variable inductors also tune. They can be ganged like the variable capacitors.

R-F Coils



Coils used in r-f circuitry are designed for minimum loss. They are often made of woven multistrand "Litz" wire, (a fine, insulated wire). Its a-c resistance is low, therefore, the Q of the coil is high. Very light cardboard is used for the coil form. The form is varnished before and after winding to prevent absorption of moisture. The form contains both primary and secondary windings, with the primary wound near or directly over the secondary.

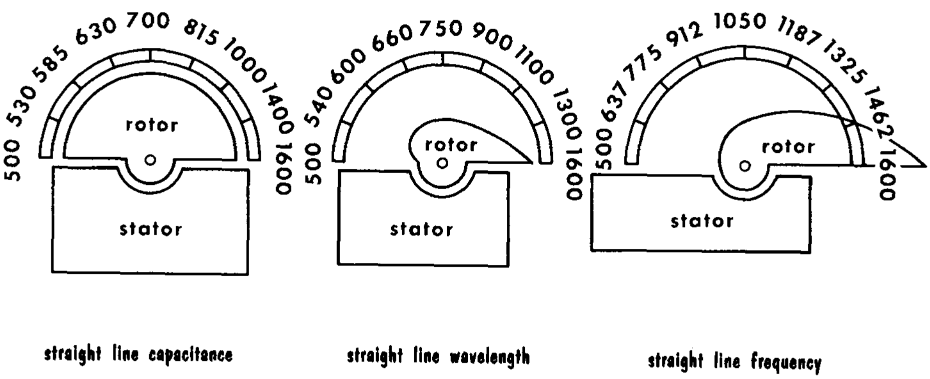
A shield covers the coil and shields it from electric and/or magnetic fields which might induce unwanted voltages. The shield must be a good conductor to shield against electric fields. Aluminum is usually used because it is light, cheap, and a good conductor. Shielding of coils results in some reduction of efficiency, due to eddy currents induced in the shield by the magnetic field of the coil. These eddy currents produce a magnetic field of their own which opposes the field of the coil. The shield should be well grounded.

R-f coils usually have air cores. A few, designed for low and medium frequencies have powdered-iron cores. The cores are sometimes adjustable. As the iron core moves into the coil, inductance increases and the frequency of the tuned circuit decreases. A brass core can be used for the opposite effect. As the brass core is moved into the coil, inductance decreases and frequency increases. In either case, the circuit can be tuned by moving the core. This is called permeability tuning.

Variable Capacitors

The plates of variable capacitors, both rotor and stator, are usually aluminum, but better receivers sometimes use silver-plated brass capacitors for improved r-f conductivity. The calibration pattern of a variable capacitor depends on the shape of its rotor, and falls into three classes--straight-line capacitance, straight-line wavelength, and straight-line frequency.

With the straight-line capacitance type, capacitance increases directly with the amount of rotation. Since frequency does not increase directly with the decrease in capacitance, calibration puts the upper half of a frequency band in about 1/8 of the dial rotation. The straight-line wavelength type wavelength



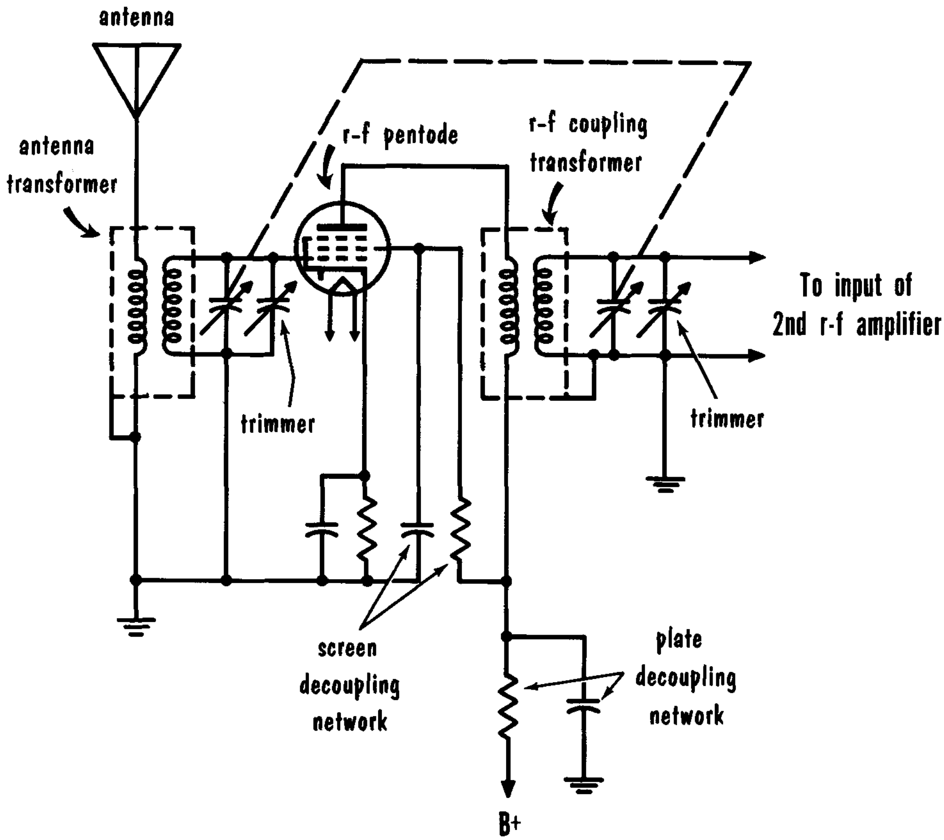
FREQUENCY CALIBRATION VARIES WITH CAPACITOR SHAPE

increases directly with the amount of rotation. The upper half of the band appears on 1/3 of the dial rotation. The straight-line frequency type frequency varies directly with the amount of rotation. This permits linear dial calibration. With this type capacitor a tuned circuit has the same sharpness of tuning over the whole band.

When variable capacitors are ganged it is almost impossible to manufacture them so that they "track". That is, to maintain equal capacitances at each setting of the rotor. To compensate for differences, each of the ganged capacitors is provided with a trimmer, an additional adjustable capacitor. It is connected in parallel with the main tuning capacitor. For further adjustment, the outer rotor plates on each section are slotted for ease in bending part of a plate and slightly changing the capacitance of one section.

The Radio-Frequency Amplifier

TYPICAL R-F AMPLIFIER STAGE



The electron tube used in r-f amplifier stages is almost always a pentode type. At the high radio frequencies it is very important that the interelectrode capacitances are kept as low as possible to prevent feedback from the plate circuit back to the grid circuit. Specially designed pentodes such as the 6AU6 provide excellent r-f amplification and isolation between input and output stages. R-f amplifiers in TRF receivers have tunable tanks in the grid circuits. Thus, the receiver may be tuned so that only one r-f signal within its tuning range is selected for amplification. When the tank is tuned to the desired frequency, it resonates and produces a relatively large circulating current. The grid of the r-f amplifier then receives a relatively large signal voltage at the resonant frequency, and minimum signals at other frequencies.

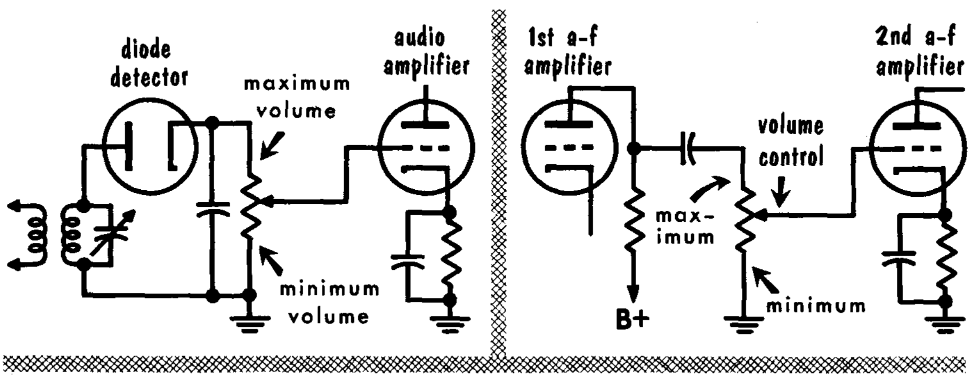
A typical r-f amplifier is shown in the diagram. The antenna circuit is inductively coupled to the grid circuit of the r-f amplifier. The plate circuit is also inductively coupled to the grid circuit of the following r-f stage.

Volume Control

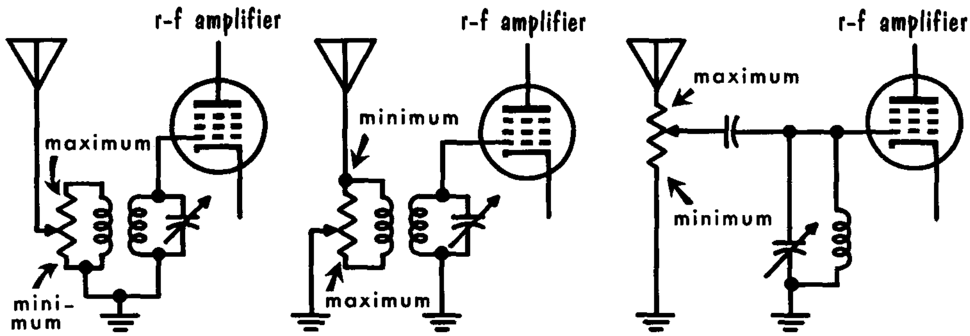
In audio amplifiers, we were concerned with the amount of amplification possible from a particular stage. However, in a receiver we are concerned with the amount of output. We want to be able to vary the amplitude of the output signal as some stations are received with a strong signal, others with a weak signal. The process of matching incoming signal strength to desired audio output involves the use of a volume control. One automatic and two manual methods of volume control are used.

The manual control of audio is accomplished by using a potentiometer as the detector-load resistor. The strength of the audio signal applied to the grid of the first audio amplifier can thus be varied by this means. The output could

MANUAL CONTROL OF AUDIO SIGNAL STRENGTH



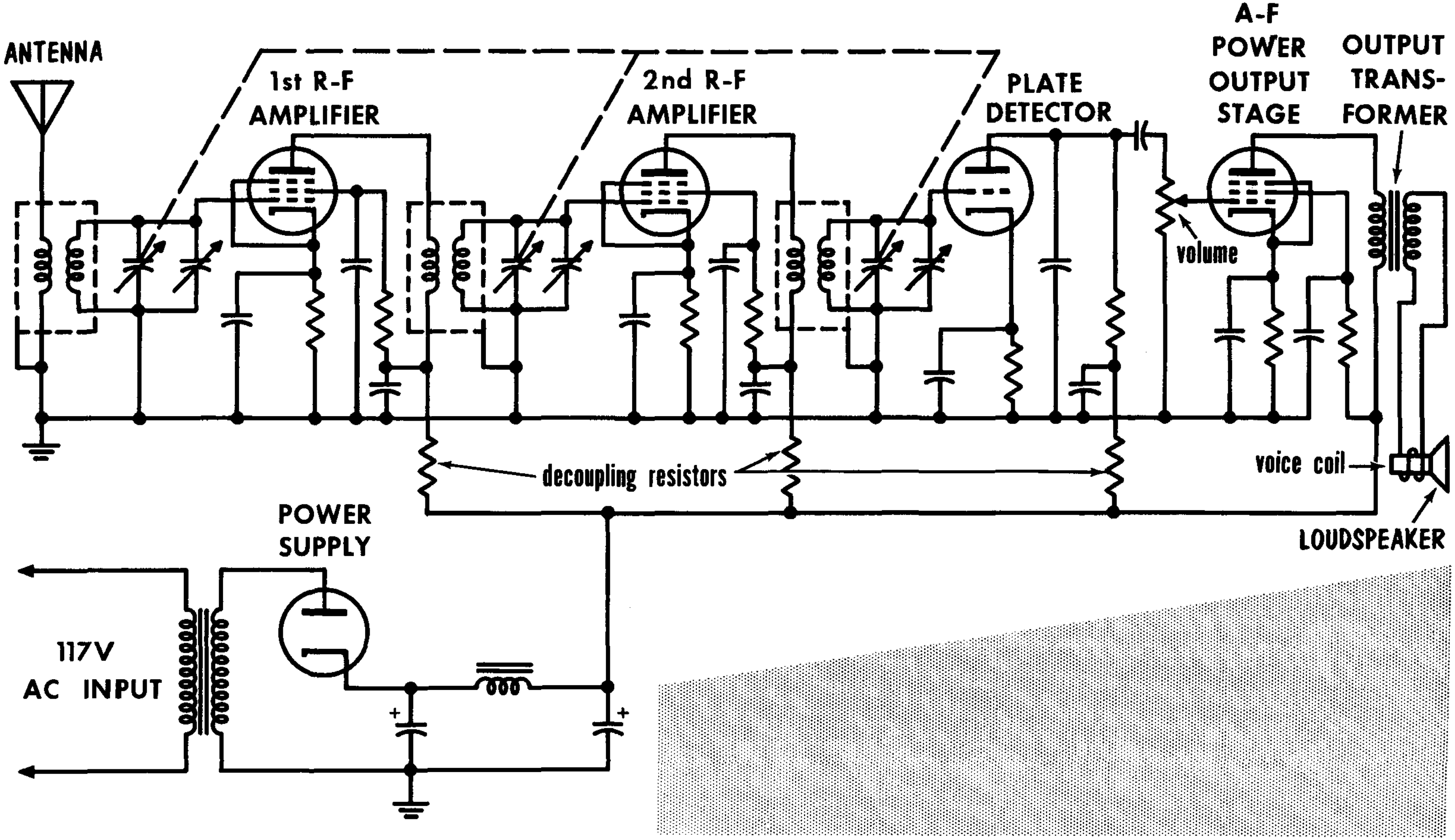
MANUAL CONTROL OF R-F SIGNAL STRENGTH



also be coupled to the potentiometer by capacitive means. The amount of audio voltage fed to the grid is that voltage existing across the potentiometer between ground and the tap connected to the grid.

R-f signal strength can be controlled in the antenna circuit. In each of the circuits shown, regulation of the potentiometer controls the amount of signal applied to the grid.

TYPICAL TUNED RADIO-FREQUENCY RECEIVER



A Basic TRF Receiver

We see on the opposite page the complete circuitry of a typical TRF receiver. The input signal is developed in the antenna circuit and coupled to the input of the first r-f amplifier. The incoming modulated r-f signal is amplified by the first r-f amplifier and fed, through transformer coupling to the input of the second r-f amplifier. Once again the signal is amplified and this time fed into the input circuit of a plate detector. The demodulated signal appears across the plate load resistance which feeds the input to the volume control, which represents the input to the final stage of audio power amplification. At this point an output transformer matches the plate impedance of the output tube to that of the voice coil, and the audio signal is reproduced by the loudspeaker. The power supply, a transformer-input half-wave circuit provides plate and screen supply for all stages.

There are three stages of signal selection, with all three tuned circuits ganged or connected mechanically. Each of the main tuning capacitors has its own trimmer capacitor connected in parallel for individual adjustments in total capacitance. Principal variations in this circuit would be three stages of r-f amplification, other types of demodulators, additional audio stages, and automatic volume control circuits.

The principal disadvantage of the TRF receiver is that its selectivity, or ability to separate signals, does not remain constant over its tuning range. As the set is tuned from the low-frequency end of its tuning range to the high-frequency end, selectivity decreases. Also, the amplification, or gain, of a TRF receiver is not constant over the tuning range of the receiver. The gain depends on r-f transformer gain, which increases with frequency. To improve the gain at the low-frequency end of the band, r-f transformers employing high-impedance (untuned) primaries are designed so that the primary inductance will resonate with the primary distributed capacitance at some frequency slightly below the low end of the tunable band. Thus, the gain is good at the low end of the band because of the resonant build-up of primary current.

The near-resonant condition of the primary at the low end more than offsets the effect of reduced transformer action. However, the shunting action of the primary distributed capacitance lowers the gain at the high-frequency end of the band. To make up for the resultant poor gain at the high end of the band, a small capacitor is connected between the plate and grid leads of adjacent r-f stages to supplement the transformer coupling. At the low end of the band the capacitive coupling is negligible. The <u>superheterodyne</u> receiver has been developed to overcome many of the disadvantages of the TRF receiver.

The popular superheterodyne receiver does not put the TRF in a position of antiquated equipment. In fact, there are certain applications in which the TRF has definite advantages. For one, the TRF does not contain a local oscillator which may often produce bothersome oscillations. A superheterodyne often radiates a portion of the oscillator signal from its antenna. Another advantage is that the TRF is inexpensive and easy to align. TRF receivers may contain as many as three radio-frequency amplifier stages, producing extremely high r-f gain for processing to the demodulator stage.

A receiver intercepts a small portion of the radio energy radiated by a transmitter and recovers the intelligence contained in it.

Basically, the receiving antenna intercepts a signal; tuned circuits select the desired r-f signal; tuned r-f amplifiers strengthen the r-f signal; the detector stage demodulates the signal; following detection, a-f amplifiers increase the strength of the audio signal; and a reproducer translates the audio variation into corresponding sound waves, thus reproducing the original intelligence.

A crystal receiver consists of an antenna, a tuned-input circuit, a crystal detector, and a reproducer.

A TRF receiver has one or more stages of r-f amplification, a detector, an audio amplifier, and a reproducer.

All the tuned circuits of a TRF receiver operate at the frequency of the incoming signal.

Selectivity is the ability of a receiver to differentiate between the desired signal frequency and all unwanted signal frequencies.

Fidelity is the characteristic of a receiver which permits it to amplify a band of frequencies containing the modulation without discrimination or distortion.

The sensitivity of a receiver is expressed in microvolts for a standard input.

Signal-to-noise ratio limits the usable sensitivity of a receiver; it is the ratio of signal strength to noise present at the receiver input.

Interstage coupling between r-f stages can be transformer coupling, impedance coupling, or resistance-capacitance coupling. Tuned-transformer coupling is most common, although tuned-impedance coupling can be used in conjunction with permeability tuning.

Trimmer capacitors are connected in parallel with tuned circuits to compensate for small irregularities, and permit tracking of all tuned circuits throughout the frequency range.

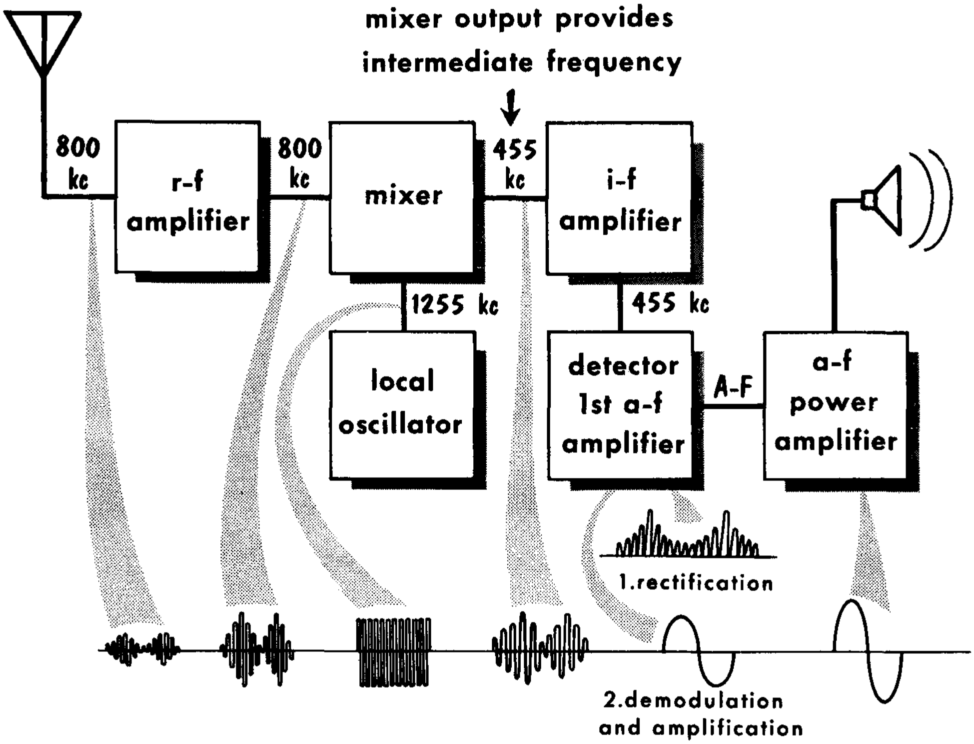
Pentodes in the r-f stages of a receiver have more gain than triodes.

Detection or demodulation is performed by rectifying the modulated carrier and filtering out the rf. The result is an audio frequency that corresponds to the modulation at the transmitter.

REVIEW QUESTIONS

1. Describe the basic functions of a receiver.
2. Describe the essential parts of a TRF receiver.
3. What are the advantages of pentodes as compared to triodes in TRF receivers?
4. Describe the function of the volume control.
5. Why is interelectrode capacitance an important consideration at high frequencies?
6. What is a variable capacitor?
7. What function does a shield serve in r-f amplifiers?
8. What is meant by permeability tuning?
9. What is a tuned circuit?
10. What is the function of an antenna?
11. What do we mean by the "sensitivity" of a receiver?
12. What do we mean by the selectivity of a receiver?

The Superheterodyne Receiver



BLOCK DIAGRAM AND WAVEFORMS OF BASIC SUPERHETERODYNE RECEIVER

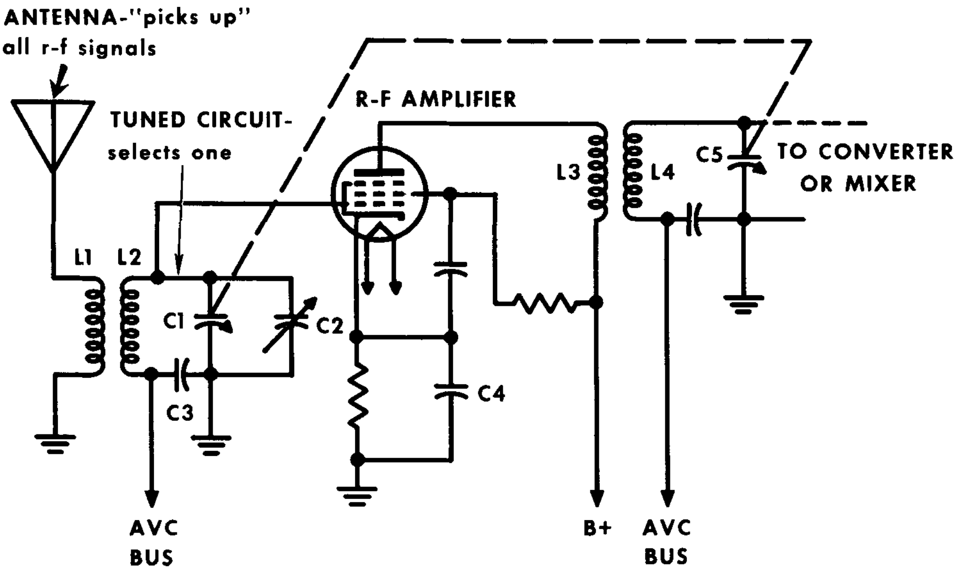
The superheterodyne receiver, often referred to as a "superhet", is used almost exclusively in modern radio receivers. We recall that in the TRF receiver all the tuned circuits were ganged and tuned throughout the entire range of desired broadcast frequencies. In the superheterodyne receiver only the r-f amplifier and mixer or converter stages tune throughout the broadcast band; the remaining high-frequency stages operate at a single frequency (the intermediate frequency) irrespective of what incoming r-f are being received.

The superheterodyne receiver uses a heterodyne or "signal beating" principle to convert the incoming r-f signals to a carrier of a fixed, lower frequency containing the same audio modulation as the original carrier. This lower frequency signal is fed to the intermediate amplifier (i-f) section which, because it is fixed-tuned, can be designed for optimum gain. Also, its band-pass characteristic can be tailored to any desired response for purposes of fidelity and selectivity. The incoming r-f signal is selected by a tuned r-f circuit, amplified as in the TRF receiver, and then fed to the mixer or converter stage. At this point, a locally generated oscillator signal of constant amplitude heterodynes or "beats" with the modulated r-f signal. A difference signal is thus produced in the plate circuit of the mixer stage, called the intermediate frequency.

The Preselector Stage

The first stage of a superheterodyne receiver consists of an r-f amplifier. Since it selects the desired r-f signal before the signal is fed to the mixer stage, the r-f amplifier is called the preselector. Communications type receivers may use two stages of r-f amplification. Besides amplifying the r-f signals, the r-f amplifier has other important functions. For example, it isolates the local oscillator from the antenna-ground system. If the antenna

The R-F Amplifier Is The First Stage Of Signal Selection



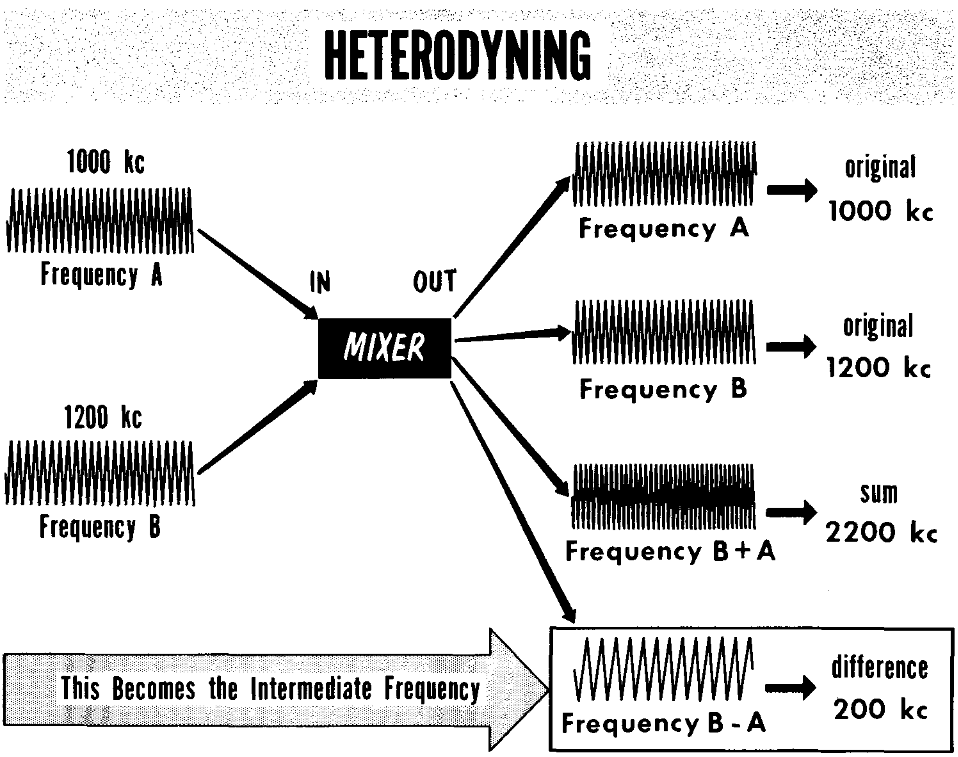
were connected directly to the mixer stage, a part of the local oscillator signal might be radiated into space causing interference with other receivers. As we will learn later, the preselector prevents image frequencies from being received.

A typical preselector stage is shown using cathode bias and a variable-mu pentode. This type tube has a curved characteristic that moves very gradually toward cutoff. As we will see, this makes it ideal for the application of automatic volume control (avc), a negative bias that becomes more negative as the incoming signal gets stronger. L1 is the antenna coil, L2 and C1 make up the tuned input circuit, and C2 is the trimmer used for alignment of the tuned circuit. C3 provides low-impedance coupling between the lower end of L2 and the grounded end of C1, thus bypassing the decoupling filters in the avc circuit. The r-f transformer in the output circuit consists of an untuned high-impedance primary L3, and a tuned secondary L4 which resonates with C5 at the station frequency.

The Phenomenon of Heterodyning

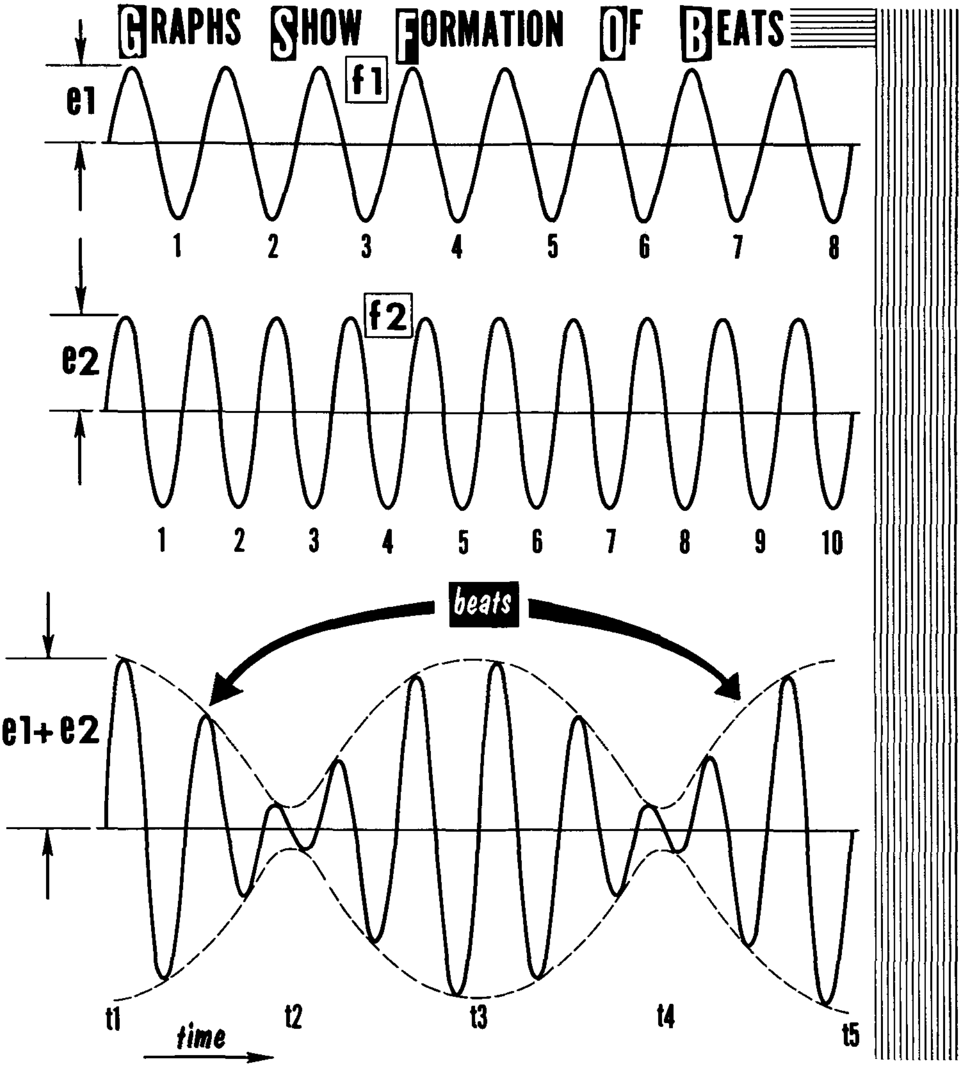
The output of the r-f amplifier or preselector is an amplified version of the incoming modulated r-f signal. We are now ready to introduce the signal to heterodyning action; let us see how this occurs. If two adjoining piano keys are struck simultaneously, a tone will be produced that rises and falls in intensity at regular intervals. This is a result of the rarefactions and condensations produced by the vibrating strings which gradually approach a condition in which they reinforce each other at regular intervals, while the sound intensity increases. At equal intervals the condensations and rarefactions gradually approach a condition where they oppose each other, and the intensity is periodically reduced.

This addition and subtraction of the intensities at regular intervals produces beat frequencies. The number of beat frequencies produced per second is equal to the sum and difference of the two frequencies. The production of



beats in a superheterodyne is somewhat analogous to the action of the piano, except that with the receiver the process is electrical and the frequencies are much higher. While both sum and difference frequencies are produced, we are interested only in the difference frequency. As we shall learn, the difference frequency produced when we mix the incoming r-f with the locally generated oscillator signal is called the intermediate frequency.

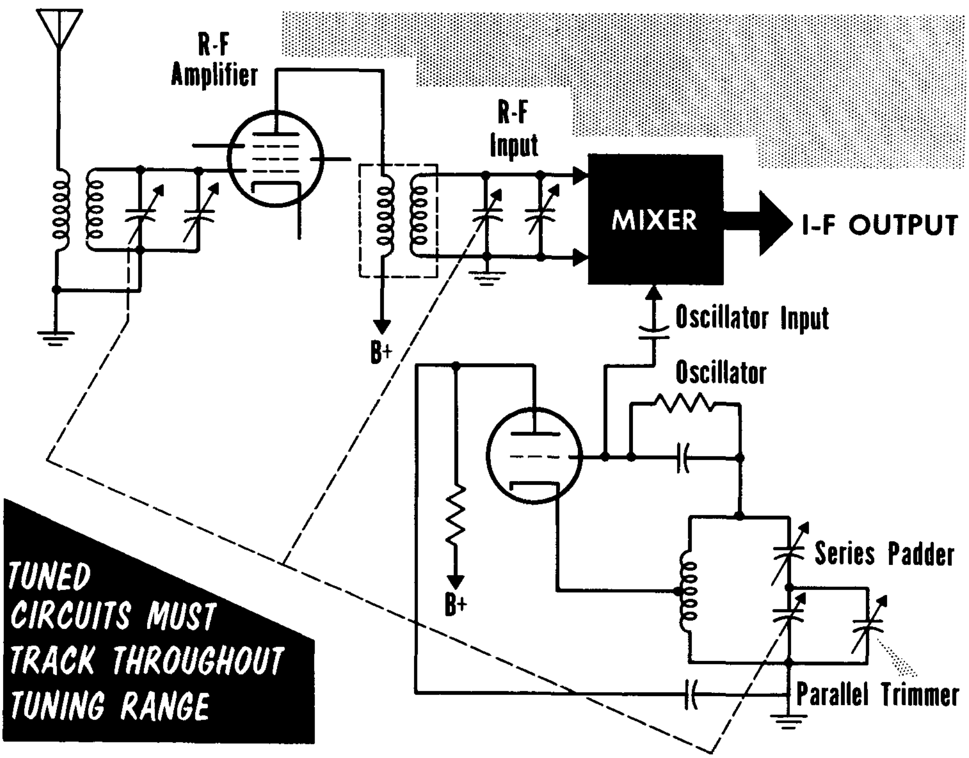
Demonstrating Heterodyning Action



The beat frequency is produced when signals of two different frequencies are combined in the mixer tube. The resultant envelope varies in amplitude at the difference frequencies indicated by the dashed lines. Frequency f1 is 8 cps; f2 is 10 cps. Initially, the amplitudes of f1 and f2 add at instant t1, but at t2 the relative phase of f2 has advanced enough to oppose f1, and the amplitude of the resultant envelope is reduced to a value dependent upon f1. At t3 the relative phase of f2 has advanced enough to permit the amplitudes to add again. Thus, 1 cycle of amplitude variation of the envelope takes place in the time interval that f2 needs to gain 1 cycle over f1. We see that f2 gains 2 cycles in the interval to t5. Therefore, the beat or difference frequency is 2 cycles per second.

The Local Oscillator

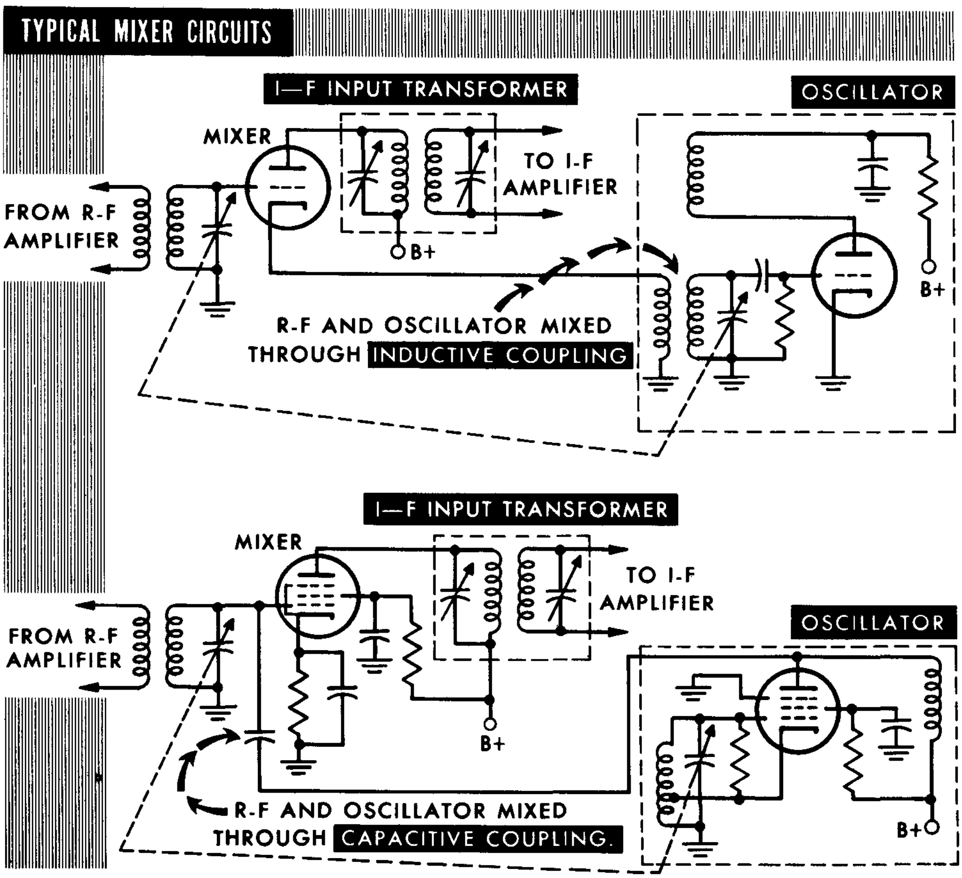
The oscillator used in the superheterodyne receiver is often referred to as the local oscillator since the oscillations are generated locally or within the receiver itself. The tuning capacitor of the oscillator is ganged to the tuning capacitor of the r-f amplifier in order to maintain a fixed frequency difference when the two signals are mixed together in the mixer stage. For instance, if we want the intermediate frequency to be 455 kc (a commonly used i-f), when the incoming r-f carrier is 500 kc, the oscillator frequency must be 955 kc; when the incoming r-f carrier is 1600 kc, the oscillator frequency must be 2055 kc. In each case the difference between the incoming



carrier frequency and the oscillator frequency is the intermediate frequency. Most superheterodyne receivers use variations of the Hartley and Armstrong circuits for the local oscillators. An important characteristic of a superhet receiver is its ability to "track". For example, if the difference frequency between the r-f and oscillator signals is 455 kc at one point of the broadcast band, it must be 455 kc at every point on the band. If this difference frequency should vary, the i-f would vary, and there would be a loss of amplification in the i-f amplifier. The trimmer capacitor in parallel with the variable capacitor is used to adjust the total capacitance for proper tracking. It is particularly effective at the high end of the band; a padder capacitor is sometimes connected in series; it has its greatest effect at the low-frequency end of the band.

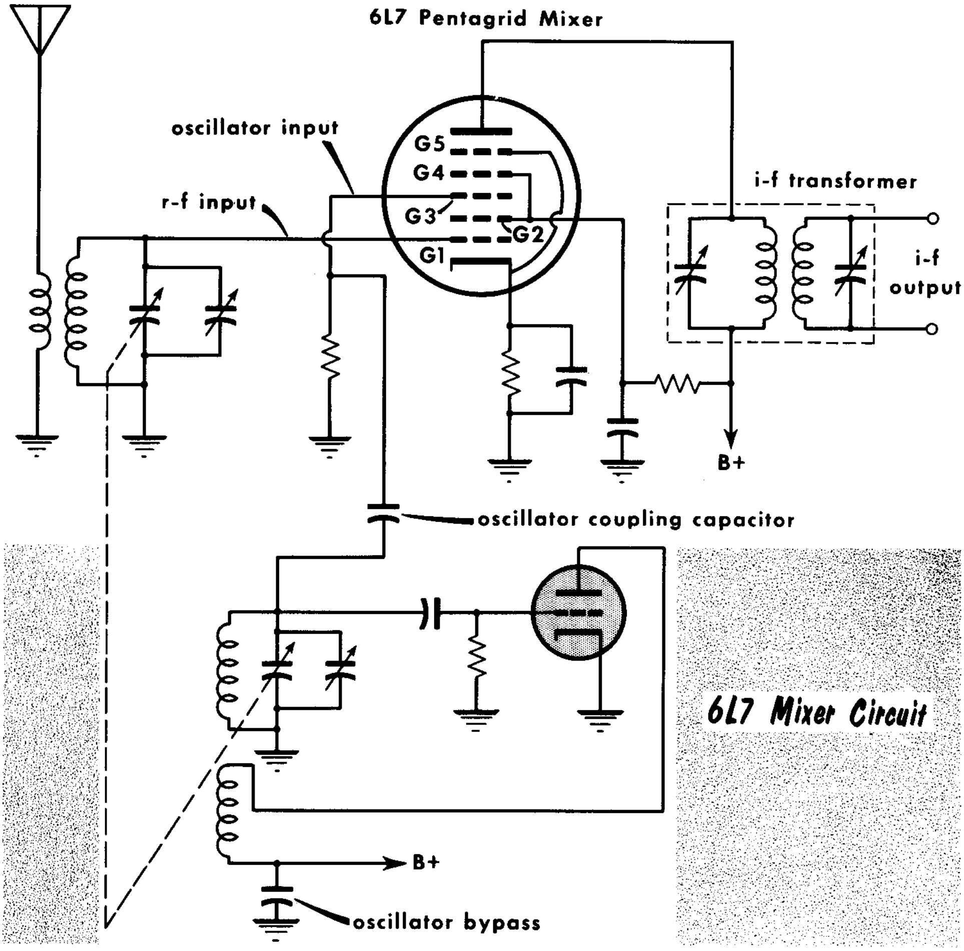
The Mixer

The mixer in a superheterodyne receiver mixes the incoming r-f signal with the locally generated oscillator signal to produce a difference, or intermediate frequency. Because the mixer tube operates on a nonlinear portion of its characteristic curve to produce the heterodyning, the circuit is often referred to as the first detector, since a new signal component is produced, the intermediate frequency. The simplest type of mixer circuit, a triode used as the mixer, is shown here. The modulated r-f signal is fed into the grid circuit, with the oscillator signal inductively coupled into the cathode circuit. Effec-



tively both signals are being applied in series and affecting the electron stream between cathode and plate. As a variation a Hartley type oscillator is capacitively coupled to the grid of a pentode mixer. One disadvantage is that the local oscillator is too closely related to the r-f amplifier and variations in the r-f amplifier can produce a frequency change or drift in the oscillator. This is undesirable because it changes the intermediate frequency. While the original signals plus the sum and difference signals appear in the plate circuit of the mixer, the tuned plate circuit selects only the difference signal.

Mixing Circuits



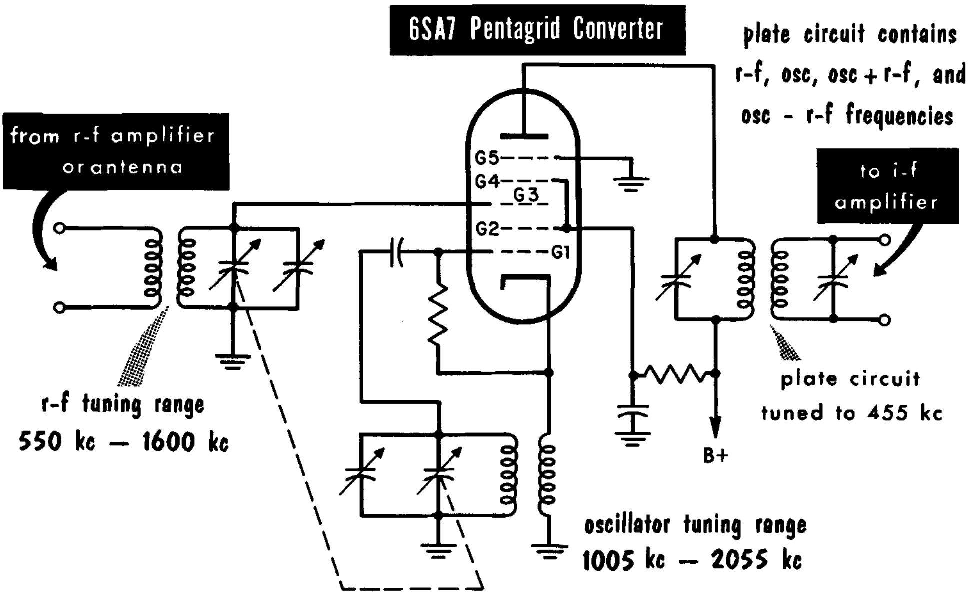
The problem of interference between the r-f amplifier and the oscillator caused special tubes to be designed to be used solely as mixers. One of the earliest was the type 6L7 tube. It is called a pentagrid mixer, because it has five grids and was designed as a mixer.

A circuit is shown using a type 6L7 pentagrid mixer. The r-f signal is applied to grid 1, which is closest to the cathode, and has the greatest control over plate current. The relatively weak r-f signal modulates the electron stream from the cathode. Together, grids 2 and 4 act as the screen grid, attracting the electrons from the cathode. Grid 3 has the local oscillator signal coupled to it. This further modulates the plate current by electron coupling, resulting in mixing of the two signals. Placing grid 2 between grids 1 and 3, makes it act as a shield, helping to keep the r-f and local oscillator signals from interfering with each other. Grid 5 is the suppressor grid, and is tied to the cathode.

The Converter

The mixer stage functions, for economy, as both a mixer and oscillator, using a multigrid tube, called a converter. While the oscillator can track 455 kc below the r-f signal, virtually all oscillators track above the r-f signals.

The 6SA7 is a pentagrid converter, with its five grids arranged as shown in the circuit. Grid 1 is the oscillator grid, and grid 2 is the oscillator plate, screen grid and shield. Grid 3 has the r-f signal applied to it. Grid 4 is physically connected to grid 2 and also is the screen grid. Grid 5 is the suppressor grid. The cathode, grid 1, and grid 2 act as a triode oscillator, with the majority of the modulated electron stream going by grid 2. Grid 3 has the r-f signal applied to it, further modulating the oscillator signal, and thus mixes the two signals by electron coupling. Grid 4 acts as a screen grid and is connected to grid 2. Grid 5 acts as a suppressor grid. In this circuit, the r-f signal and oscillator signal do not interfere with each other. The suppressor grid increases the efficiency of the tube. The major differences be-

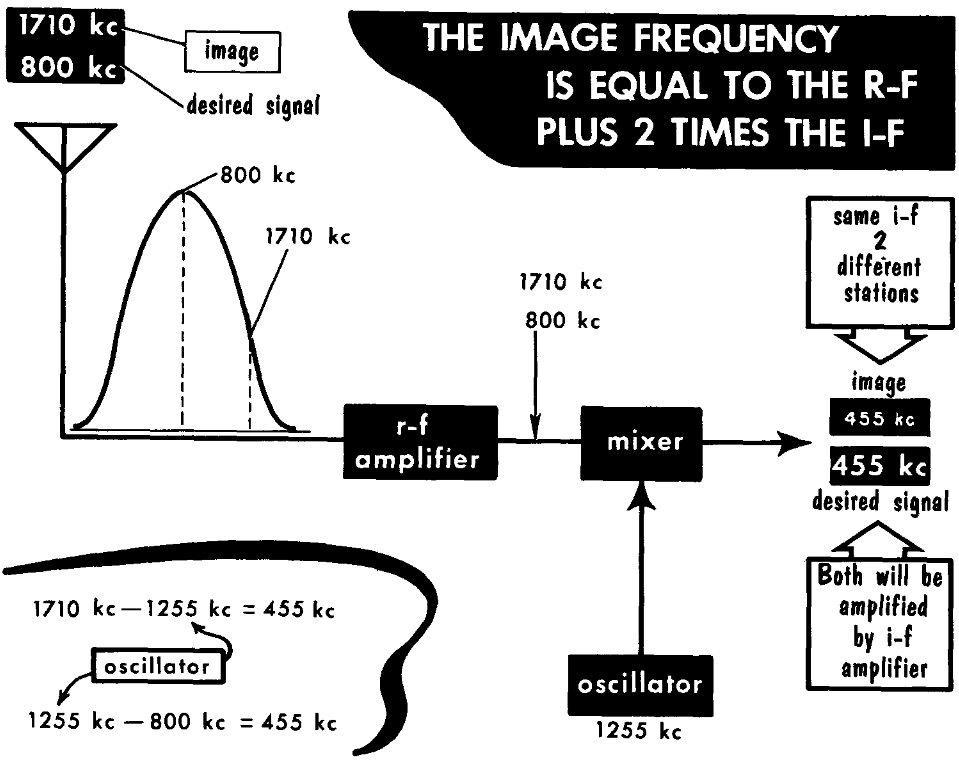


PENTAGRID CONVERTER SCHEMATIC

tween the types 6A8 and 6SA7 pentagrid converters are the removal of the grid used as the oscillator plate in the type 6A8 and the use of the screen grid as the oscillator plate in the type 6SA7. Considering the triode oscillator section of the type 6A8 as a virtual cathode, (since it is the source of electrons for the remainder of the tube), only a control grid, screen grid and plate are left. Thus, it is, in reality, a tetrode-type tube. The action of combining grid 2 of the 6SA7 as an oscillator plate, screen grid and shield permitted the addition of a suppressor grid, making this a pentode-type tube. Hence, superior performance is achieved.

Image Frequencies

There is one major disadvantage to superheterodyne operation. If a local oscillator frequency of 1255 kc can mix with an input frequency of 800 kc to produce a difference frequency of 455 kc, this same local oscillator frequency of 1255 kc can also mix with an input frequency of 1710 kc to produce a difference frequency of 455 kc. Thus, the mixer section might present to the

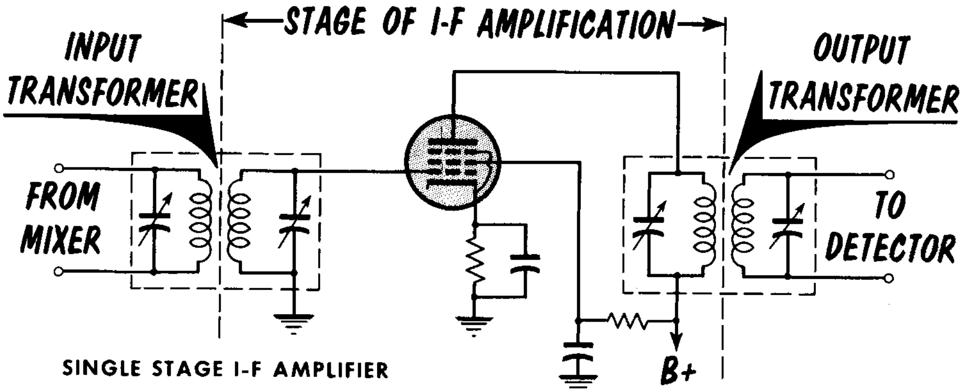


i-f amplifier the signal from two different stations at the same time, both converted to the same i-f. The i-f amplifier would accept and amplify both at the same time. The demodulator would detect the signal of both at the same time. The intelligence of both would be present in the speaker at the same time. Such a mixture of signals would be confusing, if not unintelligible.

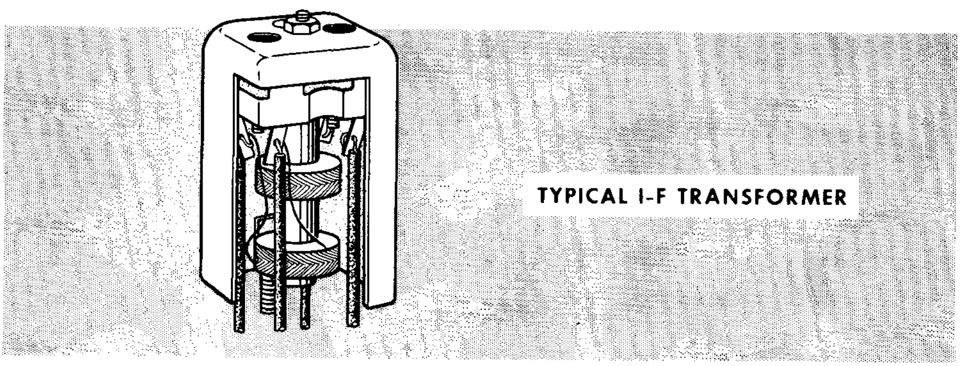
The second signal which might interfere with the desired signal is called the image frequency. Image frequencies can best be prevented by selective tuning of the r-f amplifier. Highly selective r-f amplifier tuned circuits, when tuned 455 kc below the oscillator frequency, will reject a frequency 455 kc above the oscillator. Hence, the r-f stage tuned to a frequency of 800 kc rejects the image frequency of 1710 kc. Early i-fs of 175 kc and 262 kc were increased to 455 kc and 465 kc to increase the separation between the desired and image signals, thus placing the image signal at a lower level of the input circuit selectivity curve.

The Intermediate Frequency Amplifier

An i-f amplifier is basically an r-f amplifier with a fixed-tuned input and output. It selects and amplifies the difference frequency signal produced in the plate circuit of the mixer. The tuned circuits act as bandpass filters, accepting the i-f, but rejecting other frequencies. The fixed-tuned circuits have a constant Q, and provide high selectivity and sensitivity. Pentode tubes



SINGLE STAGE I-F AMPLIFIER

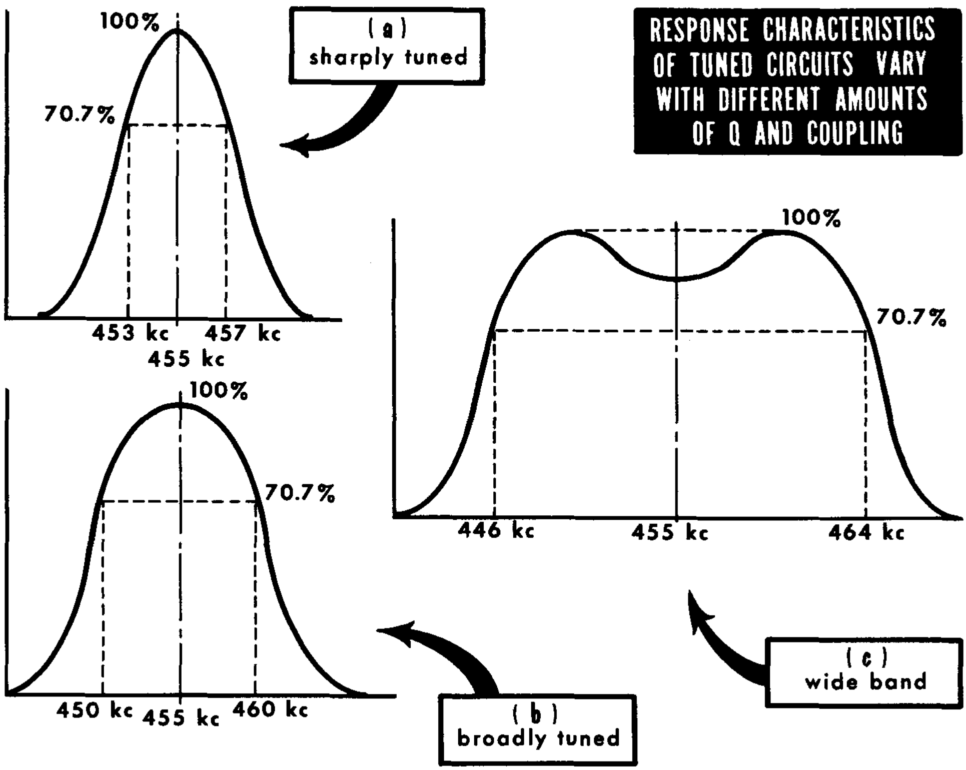


TYPICAL I-F TRANSFORMER

are used. Most receivers have one to three stages of i-f amplification, depending upon the amount of gain required. These stages operate as class-A voltage amplifiers.

The output transformer which couples the plate circuit to the grid circuit of the second i-f amplifier, is tuned by mica or air-trimmer capacitors. In some instances the capacitors are fixed, and the tuning is by movable powdered-iron core. This method is called permeability tuning. In special cases the secondary only is tuned. The coils and capacitors are mounted in small metal cans which serve as shields, and provision is made for tuning without removing the shield. The input i-f transformer has a lower coefficient of coupling than the output transformer in some receivers, to suppress noise from the mixer. A typical i-f stage in a broadcast receiver may have a gain of about 200.

I-F Amplifier Response

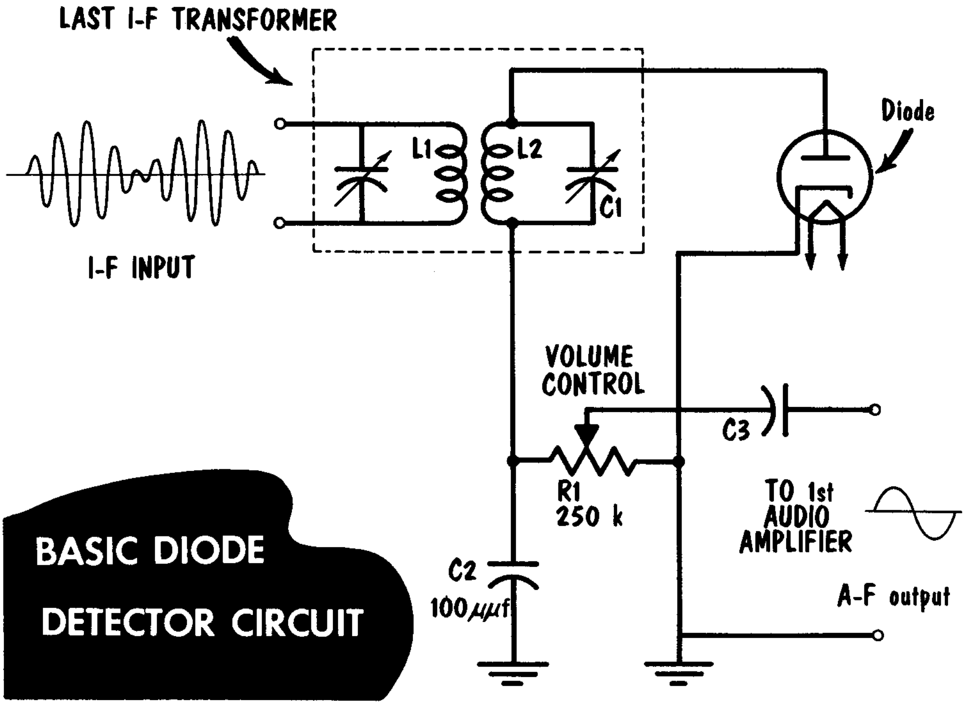


The i-f signal is identical in every respect to the incoming modulated r-f signal except for frequency. The heterodyning process in the mixer produces a fixed difference frequency, but the modulation components remain unchanged. Hence, the i-f amplifier must pass the carrier and its sidebands which contain the signal intelligence. The sharply tuned circuit of A presents a practical bandwidth of 4 kc; B shows the characteristic of a circuit with a 10-kc bandwidth. An overcoupled i-f bandpass characteristic found in high-fidelity tuners is shown in C. The bandpass is 18 kc wide for full sideband reception. A decrease in output voltage from a 100% level to a 70.7% level is equivalent to a 50% reduction in power, since power is a function of the square of the voltage ($P = E^2/R$). This is sometimes referred to as "3-db down".

Local AM broadcast stations in any given area have assigned frequencies spaced at least 40-kc apart. They normally have modulation frequency response up to at least 7.5 kc. However, to minimize interarea interference at night, control noise, and provide low distortion with economical design, the passband of most AM broadcast receivers is kept to about 10 kc. For specialized transmission such as ship-to-shore voice communication, high-gain narrow-bandpass i-f stages are used since 2 kc is the highest modulation frequency for voice intelligence.

The Second Detector

The name "second detector" is given to the demodulator of a superheterodyne receiver because the mixer or converter stage is often thought of as the first detector. The term "detector" usually means the audio demodulator. Most superhets use a diode detector. This type of detector is practical because of the high gain as well as the high selectivity of the i-f stages. The diode de-



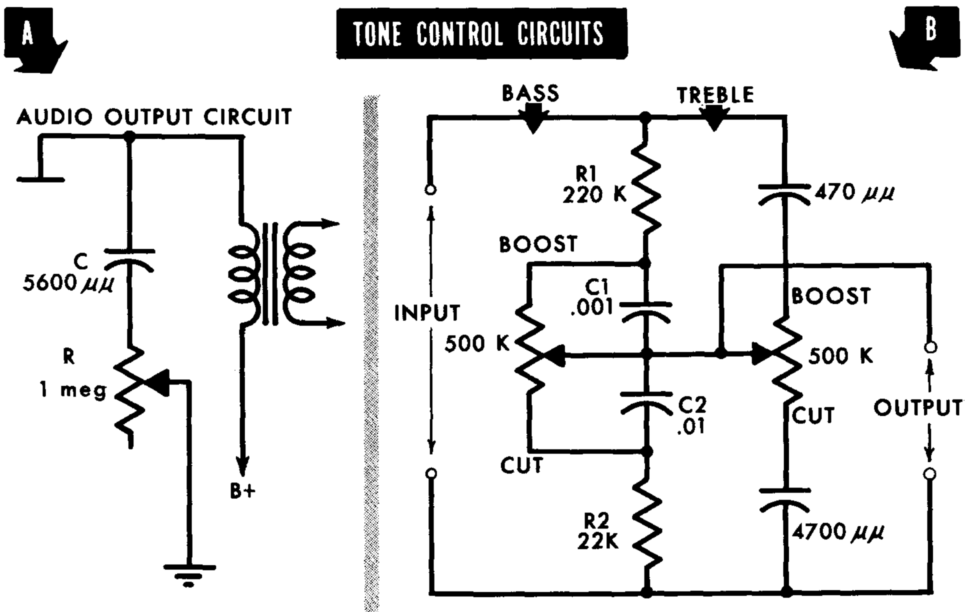
BASIC DIODE DETECTOR CIRCUIT

tector has good linearity and can handle large signals without overloading. To save space and money, the diode detector and first audio amplifier are often included in the same envelope in modern superhets.

A simple diode detector is shown with the rectified voltage appearing across R1, which also serves as the volume-control potentiometer. C2 bypasses the r-f component to ground, the C3 couples the output of the detector to the first audio amplifier stage. Tuned circuit L2C1 is the secondary of the last i-f transformer. The time constant of R1C2 is long compared to the time for one a-f cycle. If the i-f is 456 kc, the time for one i-f cycle in microseconds is 1/0.456, or 2.19 μsec. If R1 is 250,000 ohms, and C2 is 100 μμf, the time constant in microseconds is 0.25 x 100, or 25 μsec. C2 discharges through R1 in one-half the time for one a-f cycle (1/2f). The time required to completely discharge C2 is 5R1C2 seconds. Thus, 1/2f = 5R1C2, or f = 1/10 R1C2. This works out to 4000 cps, as the high audio frequency C2 is capable of following with distortion. To increase the response of the diode detector, the time constant of R1C2 must be reduced.

Tone Controls

Tone controls are networks that continually adjust frequency response, by increasing or decreasing the high- and the low-frequency output. The tone control usually acts to boost or reduce the frequencies toward one end of the range amplified. The treble control acts on the high-frequency end, the bass control on the low-frequency end. A tone control is effectively a variable filter. A simple tone control (A) consists of a series R-C circuit. This



simple tone control

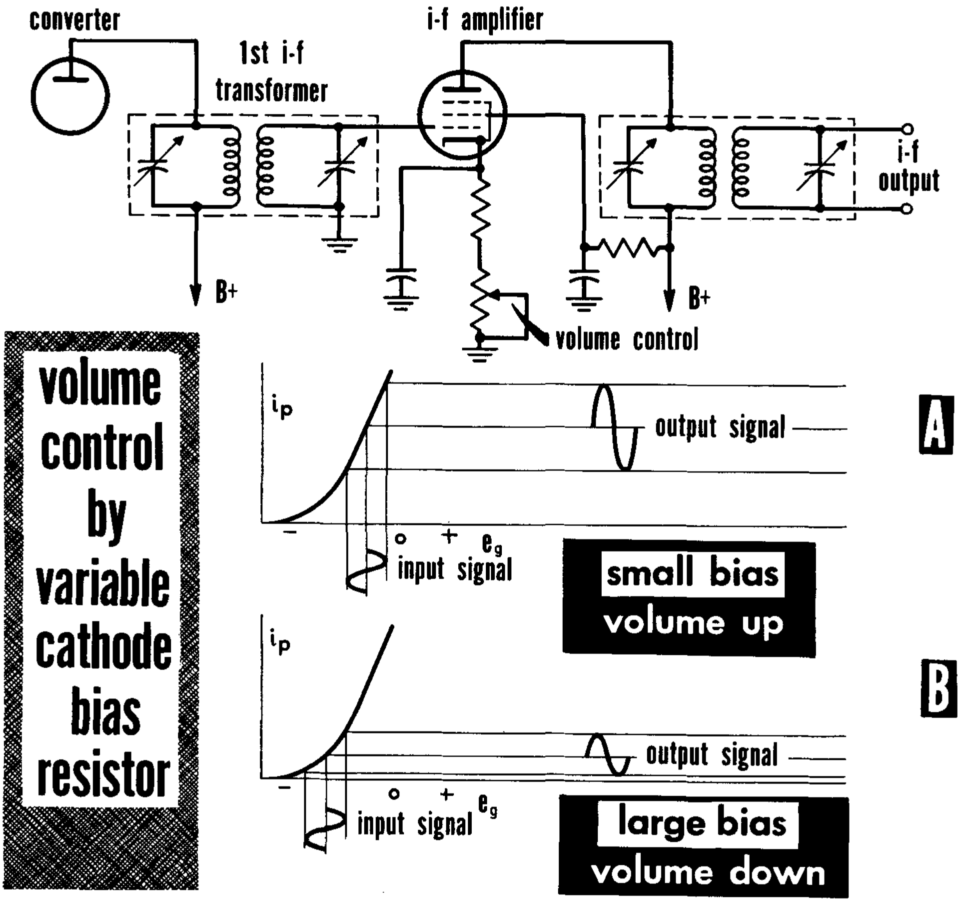
composite tone control

combination may be connected between plate and ground or grid and ground in any audio stage in the receiver. The value of C will bypass the high-frequency components to ground. The setting of the variable resistor determines the amount of high-frequency energy removed by the tone control circuit. When R is low, the frequencies are attenuated, when it is high, they appear in the output.

A more complete tone control (B) is also shown with capacitors in series with R1 and R2. Lower capacitor C2 develops a considerable audio voltage, particularly at the low frequencies, resulting in a larger voltage output at the very low frequencies than over the rest of the audio range. Putting a capacitor in series with R1 develops the greater part of the voltage at the low frequencies and reduces the amount developed across the lower resistor. This produces an attenuation or loss of the extremely low frequencies. Combining the arrangement and using a potentiometer across the two capacitors, provides a continuous adjustment going from bass boost to bass cut. Many modern tone control circuits combine the two arrangements with two controls, one for the treble boost and cut, the other for bass boost and cut.

Volume Control

We have discussed volume control by tapping off a voltage across the detector load resistor. Variable bias control used in the cathode circuit of an amplifier is another means of volume control. An i-f amplifier with a fixed cathode bias resistor and a variable control in series is shown. To operate correctly, this circuit must use a tube with remote cutoff or variable-mu characteristics, such as the type 6BA6 tube. When the arm is set at the top



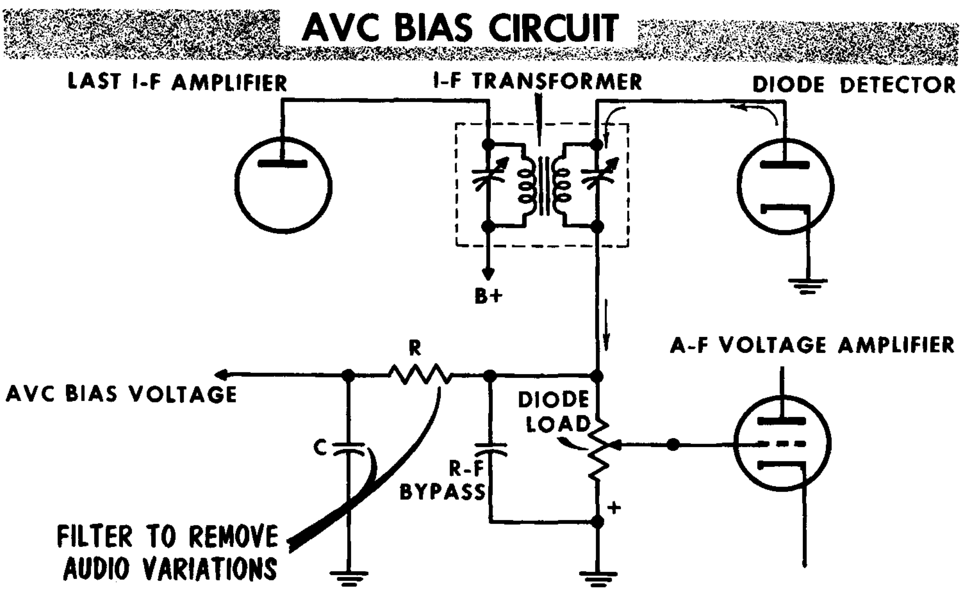
of the control, the minimum resistance of the fixed resistor is in the cathode circuit to develop bias. The small value of bias developed causes the grid voltage to go in a positive direction. The result is a higher plate current flow through the inductive plate load resulting in increased output. When the arm is set at the bottom of the control, the maximum resistance of both the fixed resistor and the bias control are used to develop bias. The large value of bias developed causes the grid voltage to go in a negative direction. The lower plate current flow through the inductive plate load results in decreased output.

Automatic Volume Control

Manual volume control is not the perfect answer. When tuning from a weak to a strong station, or vice versa, the increased or decreased signal strength may cause an annoying change in volume. To prevent this, an additional circuit is added to the receiver called automatic volume control (avc). It does not replace the manual volume control, but limits it to its main function of adjusting the loudspeaker volume. The avc circuit prevents changes in the volume as set by the manual volume control.

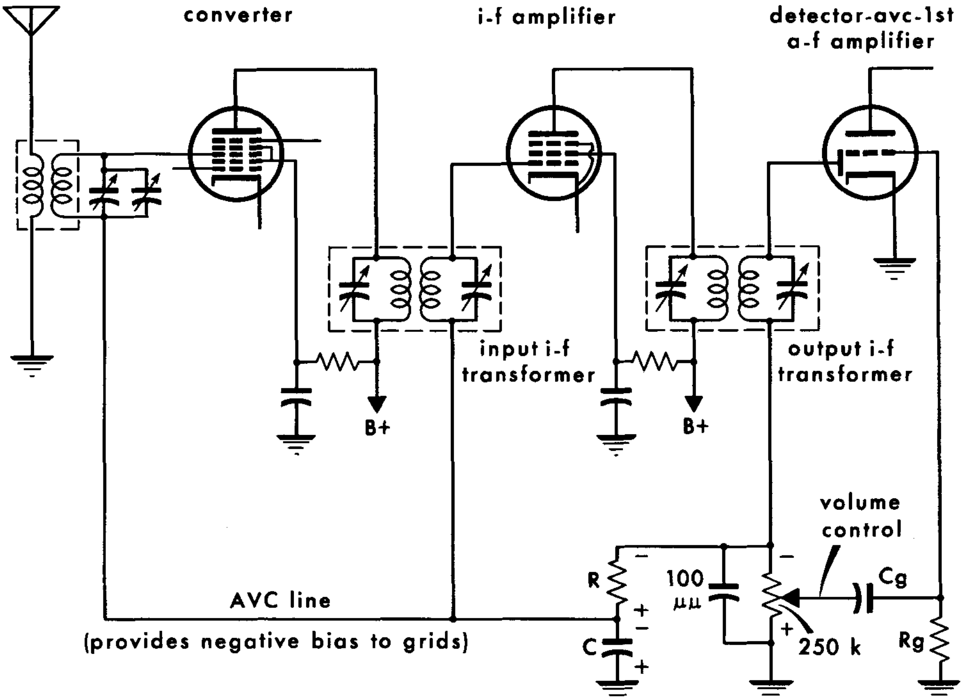
Avc action relies on the use of remote-cutoff characteristics of a tube such as the type 6BA6. A variable-bias control acting as a volume control is the basis of avc. The bias control is varied so that a weak signal would decrease the negative grid bias, thereby increasing the gain of the amplifier stage. When a strong signal is tuned in, the grid bias becomes more negative, decreasing the gain of the amplifier stage.

To make the circuit automatic, a voltage is needed which varies directly with the average value of the incoming signal. The simplest method is to apply a negative voltage to the grid circuit, where no current is drawn; only a varying voltage is necessary. The value of the voltage necessary to vary the grid



from higher to lower gain is approximately 3 to 5 volts for a pentode. All these requirements are available at the output of the second-detector circuit. The voltage polarity developed across the detector load resistor has the output negative with respect to ground. The a-f voltage developed across the load resistor varies directly with the strength of the signal received. To use the available voltage to provide a negative grid bias, we must remove the variations of the audio signal and filter it to get a steady d-c bias voltage.

Automatic Volume Control (Cont'd)



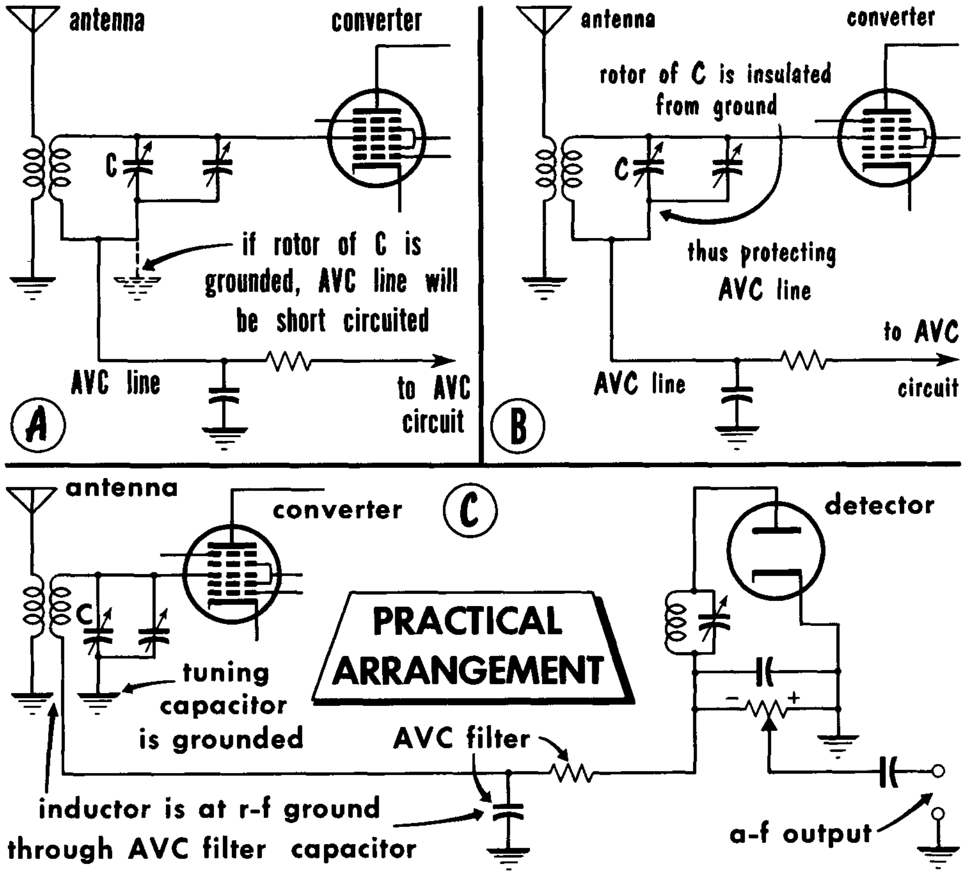
COMPLETE AVC CIRCUIT DIAGRAM

The resistor and capacitor of the avc filter are in parallel with the detector load. The voltage developed across the load will also be across R and C of the filter. The division of voltage is dependent upon the values of R and X_C. The avc bias voltage is the voltage developed across C. The charge on C remains constant despite the a-f fluctuations. C cannot discharge quickly, since the value of R is high. The values of R and C are important. The low-frequency audio signals must not be so long in time as to charge the capacitor to too high a value. The capacitor must still be able to discharge quickly enough to change the bias value to compensate for fading signals. A typical C value is between .01 and .05 μf; R ranges from 1 to 3 megohms. The voltage across C is a filtered d-c bias voltage, having a polarity negative with respect to ground, and varies directly with the value of signal strength received. This voltage is often applied to the grid of the i-f amplifier and the r-f grid of the converter.

Applying the avc bias voltage to the grid of the i-f amplifiers presents no problems. The "low" end of the secondary of the input i-f transformer is raised, and the avc voltage is applied directly to the grid in series with the signal voltage. The voltage to the r-f grid of the converter tube is also applied to the low end of the r-f tuning circuit.

Automatic Volume Control (Cont'd)

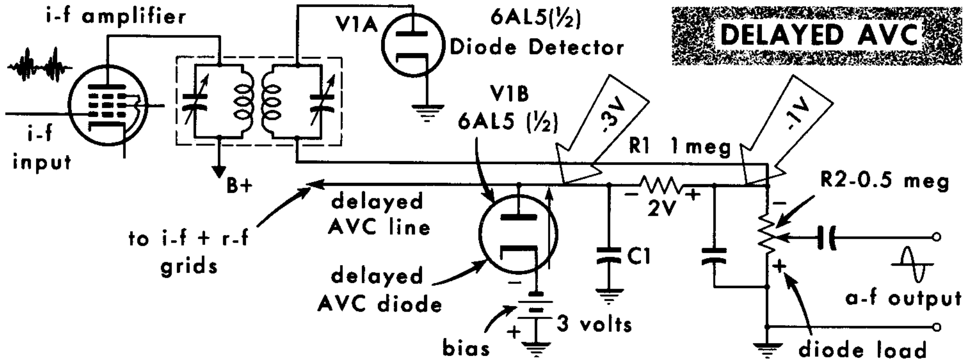
Applying avc to the r-f grid of the converter through the tuned circuit presents a problem. The rotor of the r-f tuning capacitor and the rotor of the oscillator tuning capacitor are on a common shaft and connected electrically. The rotor at ground potential (A), grounds out (shorts) the avc bias applied



to the r-f grid. To remove this obstacle, the tuned capacitor can be placed above ground (B), mounted on rubber mountings to insulate it from the metal chassis. The tuned circuit can then be returned to the avc line with the avc filter capacitor acting as an r-f return.

A more typical avc arrangement (C) uses a separate connection for the inductive portion of the tuned circuit. The variable capacitors are grounded and the inductors are connected directly to the avc line. The r-f return is through the avc filter capacitor. In this way, as far as r-f is concerned, the inductors are connected directly in parallel with the tuning capacitors, since the reactance of the avc filter capacitor is negligible at radio frequencies.

Delayed AVC



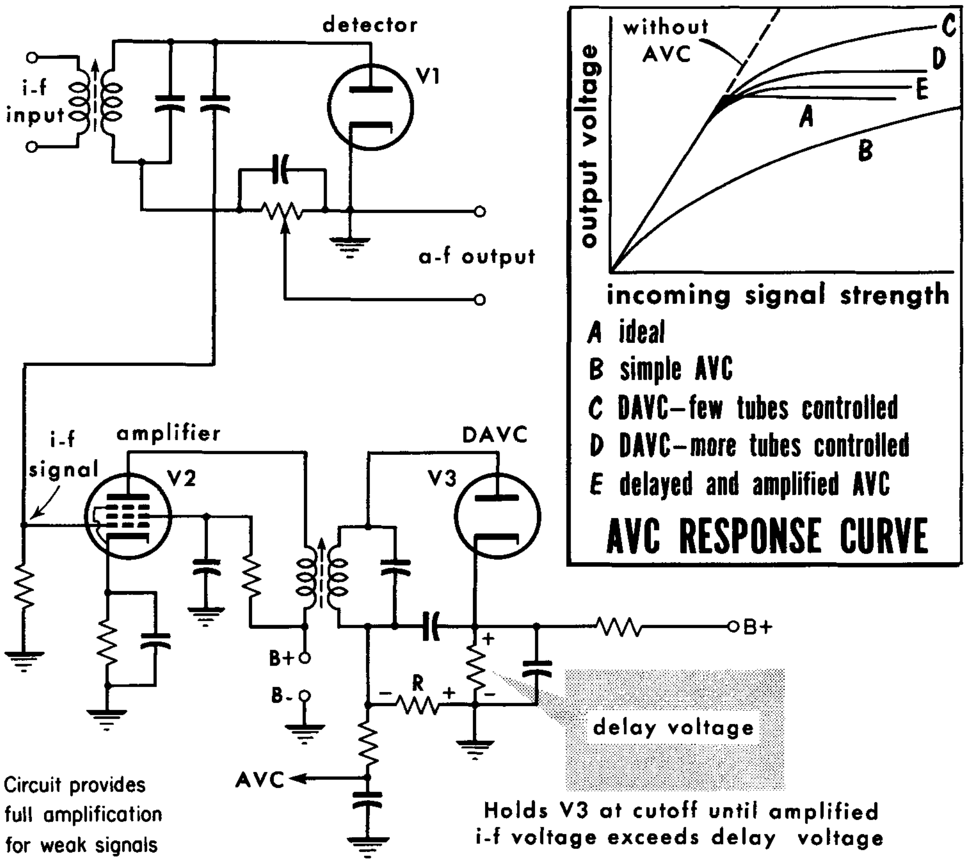
Weak signals developing avc voltage create a problem. The avc bias voltage developed cuts down the gain of the amplifier circuits when it is still needed for weak signals. A circuit is needed to hold back the avc bias voltage until the signal strength reaches the minimum level. When the incoming signal develops an avc bias voltage higher than the predetermined minimum level, avc action takes over and reduces the gain of the amplifier circuits. This permits weak signals to have full gain, while still maintaining avc action on strong signals. This is called delayed avc.

A typical delayed-avc circuit is shown, using a duo-diode, such as the type 6AL5. The rectified output of V1A is the regular audio output signal fed to the first a-f amplifier. The filter circuit for the avc voltage consists of R_1 and C_1. With no signal received, the path of electron flow from the 3-volt battery in the cathode of V1B is important. Assuming no drop across the diode's plate resistance, the flow of electrons will be through V1B, through R_1, and through R_2 to ground, which is at +3 volts potential. The voltage drops will divide, with 2 volts across the 1-megohm filter resistor, and 1 volt across the .5-megohm detector load. This places the plate of V1B at -3 volts with respect to ground. The bias is fixed at a -3 volt level, which is proper for the amplifier stages. When a strong signal is received, a typical IR drop across the detector load could be a value of 5 volts. This places the top of the detector load at -5 volts potential with respect to ground. This -5 volts will also be present at the plate of V1B. There is no loss across R_1, since there is no current flowing. The value of -5 volts at the plate of V1B puts the plate 2-volts negative with respect to the cathode; thus, V1B stops conducting. The value of bias voltage will then be that of the avc voltage of -5 volts. This increased negative signal applied to the grids of the amplifier stages will reduce their gain for the strong signal received; the output will then be normal, despite the strong signal.

The result is that all signals which develop 3 volts or less across the detector load will not vary the bias of the amplifier stages. All signals strong enough to develop 3 volts or more across the detector load will develop an avc bias voltage which will maintain the proper gain in the amplifier stages. In actual practice, the negative voltage required to bias diode V1B may be taken from a source such as the power supply.

Delayed and Amplified AVC

An ideal davc should operate so that weak signals get full amplification while strong signals are held to a satisfactory, uniform level. An ideal avc response is shown at A of the graph. B shows the response of simple avc. C shows the response of davc when only a few stages are controlled. D shows davc with more tubes controlled. E shows the response of a special kind of davc called delayed and amplified avc. Its response comes close to the ideal. To see how this is achieved, study the circuit for delayed and amplified avc. V1 is the detector. V3 is a separate davc tube. The i-f signal is applied to V1 for detection. It is also applied to V2, an i-f amplifier used to provide a separate channel for the davc circuit. The output of this ampli-

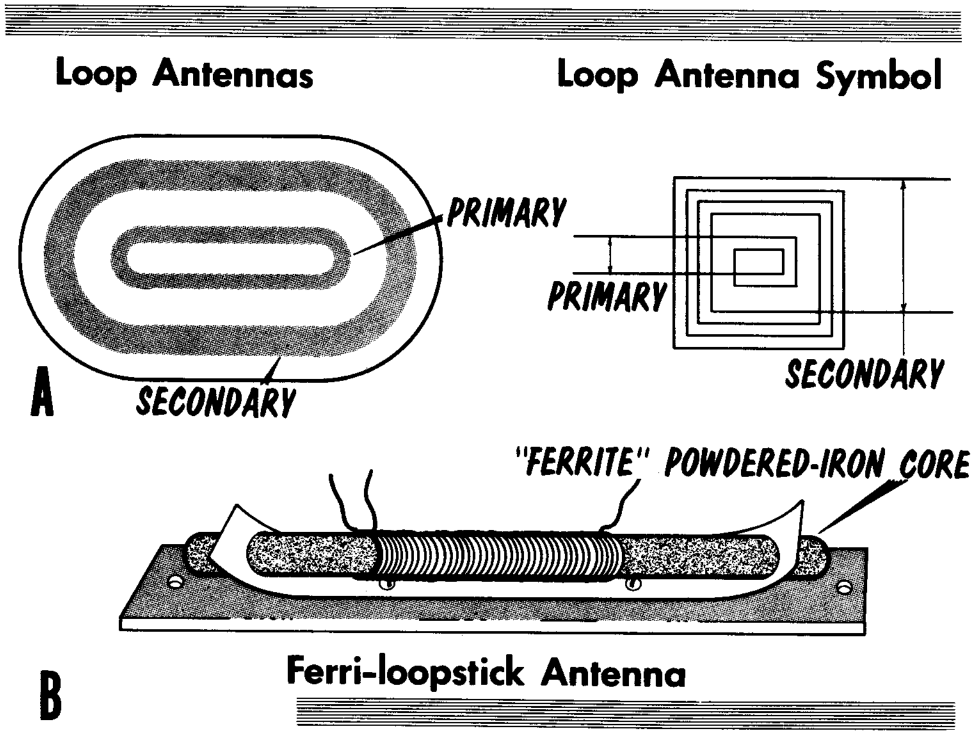


fier is applied to the davc diode. Because of the separate channel i-f amplifier, the i-f signal applied to the davc diode is stronger than the signal applied to the detector diode. As a result, the avc voltage taken across R increases more rapidly than the signal strength applied to the detector. Thus, with proper gain in V2, the response to strong signals is almost constant.

Superheterodyne Circuit Refinements – Loop Antennas

Prior to high-gain tubes and powerful broadcasting stations, the radio antenna occupied the present home of the TV antenna. Long antennas strung across the rooftops were required to pick up the maximum value of induced signal. Today's superheterodyne receivers, with their high sensitivity and strong broadcast signals, have permitted the use of small loop antennas.

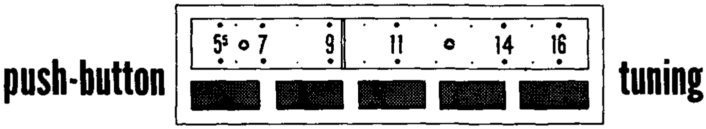
A common type of loop antenna is shown in (A), with its symbol. The primary of the loop is wrapped around the secondary. The secondary is the larger winding, and is tuned by the r-f section of the variable capacitor. The signal pickup by the secondary is usually sufficient to provide good re-

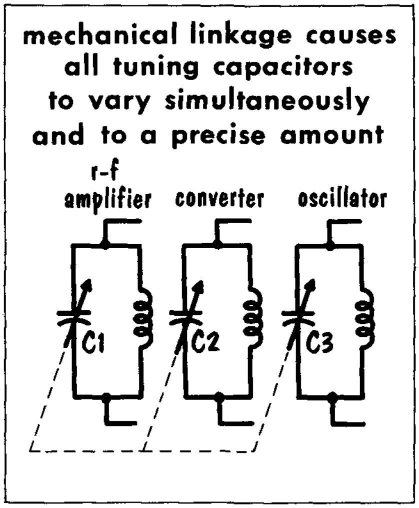


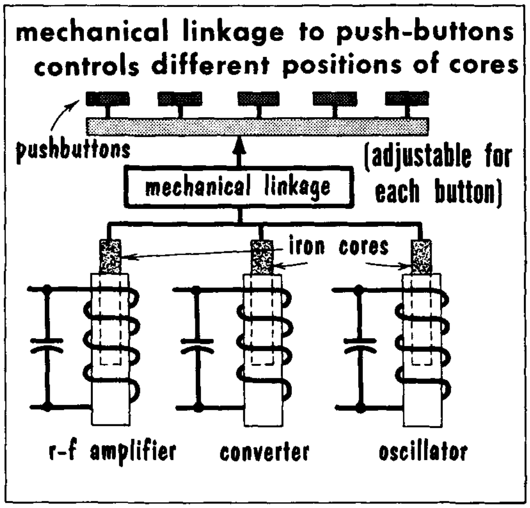
ception in strong signal areas. When the receiver is used in weak signal areas, an outdoor antenna is connected to the primary winding and the loop performs as an ordinary antenna transformer. The loop is small enough to fit inside the receiver cabinet, leaving no unsightly wires dangling behind. In addition, since the loop antenna is part of the receiver, a table-model can be carried around.

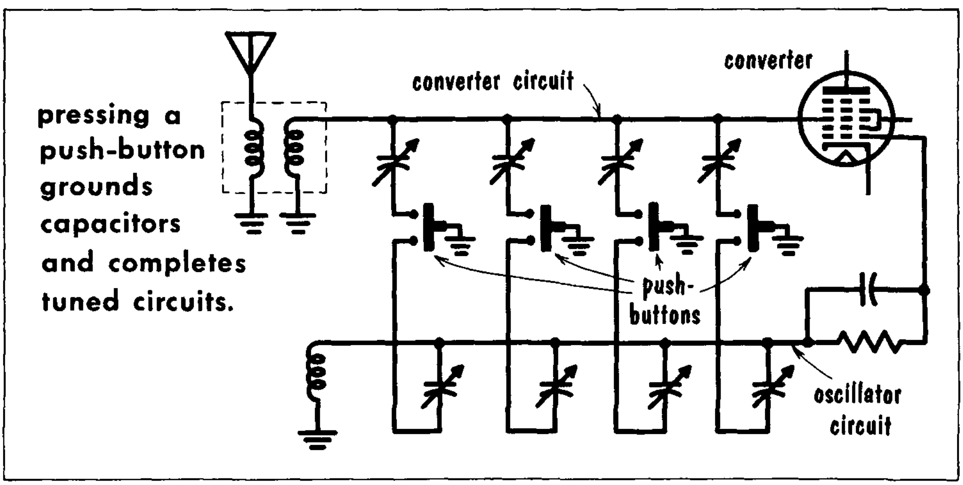
An improved version of the loop antenna, called a ferri-loopstick (B), uses a small, compact coil with a special powdered-iron core, producing a very high Q. This permits weak signals to develop a reasonably-high signal voltage in the voltage gain of the series-resonant, tuned-rf circuit.

Pushbutton Tuning



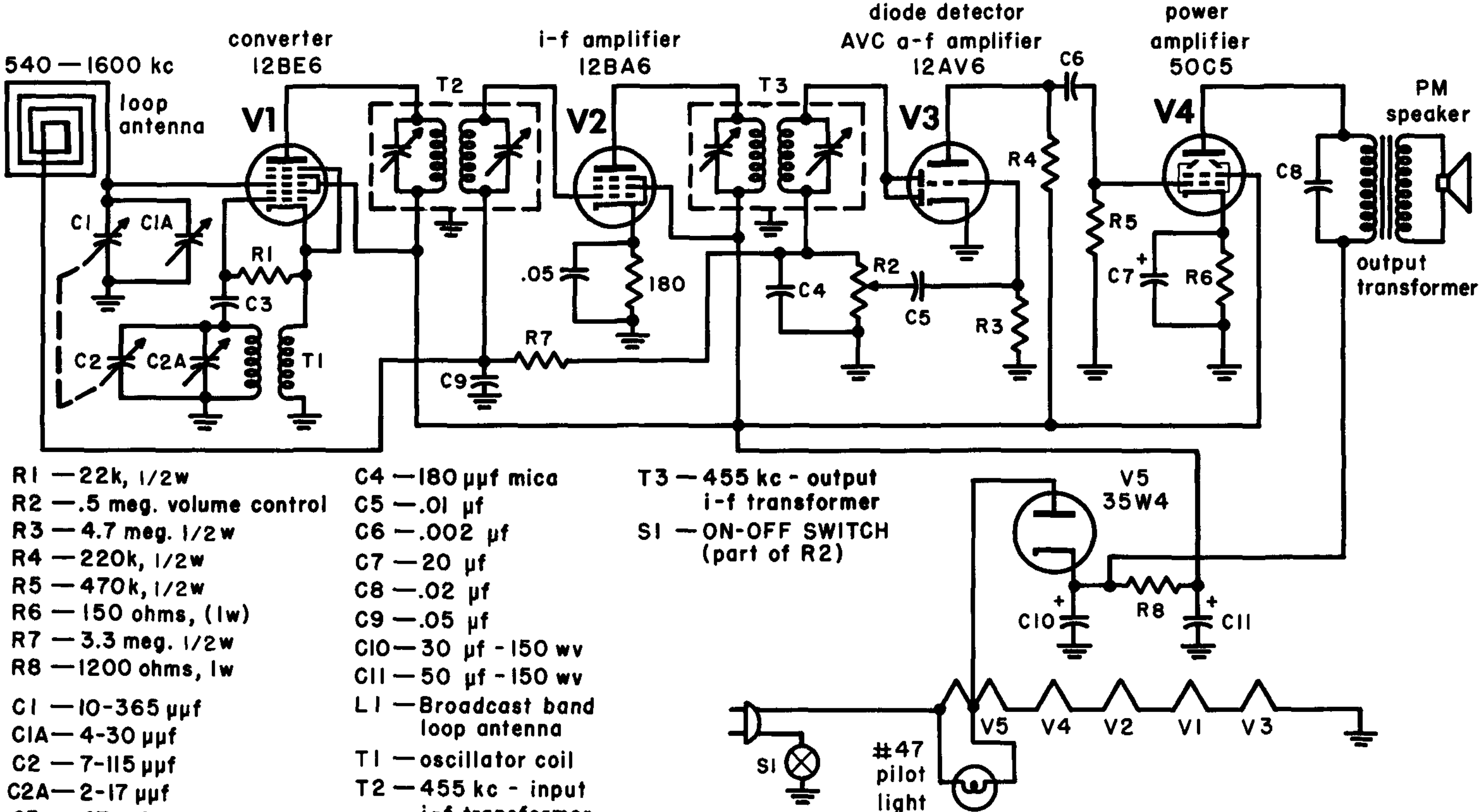






In most areas, there are only a few popular stations. Pushbuttons are used to locate them instantly, without turning the manual tuning control. Pushbuttons are popular in automobile receivers, but have almost disappeared in home radios. Pushbutton tuning either causes the inductive or capacitive component of a tuned circuit to vary in value, or places different values of C in parallel with one fixed L (or vice versa) for the station. The mechanical arrangements in pushbutton tuning are complex. They produce a specific L-C circuit variation or change when the button is depressed.

TYPICAL AC-DC SUPERHETERODYNE RECEIVER



R1 — 22k, 1/2w
R2 — .5 meg. volume control
R3 — 4.7 meg. 1/2w
R4 — 220k, 1/2w
R5 — 470k, 1/2w
R6 — 150 ohms, (1w)
R7 — 3.3 meg. 1/2w
R8 — 1200 ohms, 1w

C1 — 10-365 μμf
C1A — 4-30 μμf
C2 — 7-115 μμf
C2A — 2-17 μμf
C3 — 47 μμf mica
C4 — 180 μμf mica
C5 — .01 μf
C6 — .002 μf
C7 — 20 μf
C8 — .02 μf
C9 — .05 μf
C10 — 30 μf - 150 wv
C11 — 50 μf - 150 wv
L1 — Broadcast band loop antenna
T1 — oscillator coil
T2 — 455 kc - input i-f transformer
T3 — 455 kc - output i-f transformer
S1 — ON-OFF SWITCH (part of R2)

AC-DC Superheterodyne Receiver

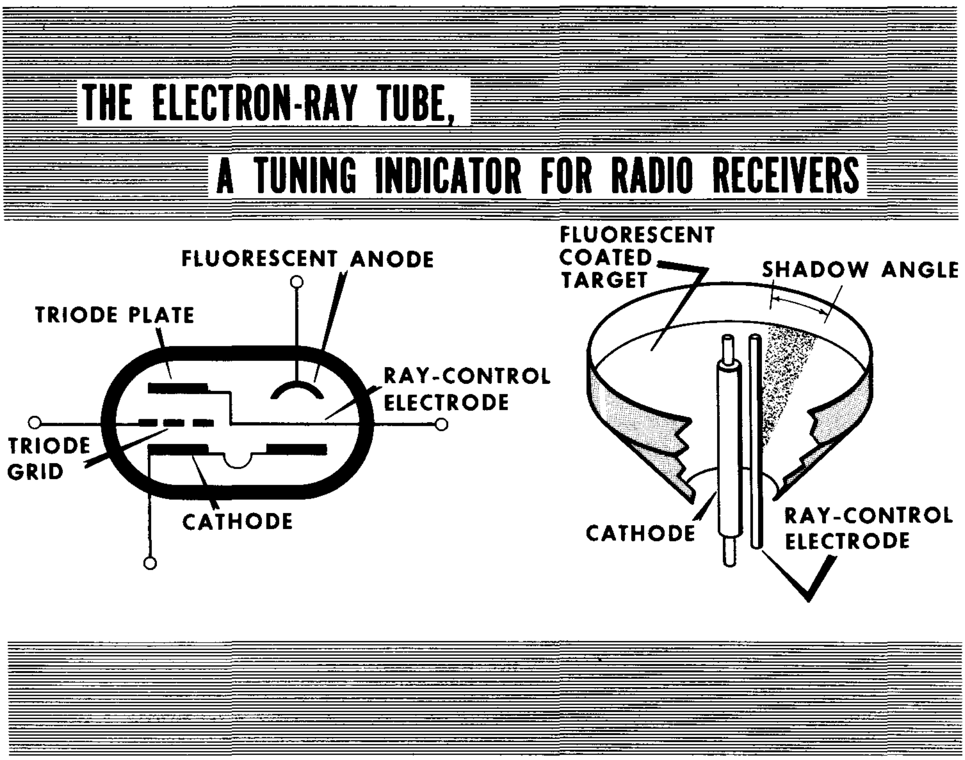
Having discussed the component parts of the modern superheterodyne receiver, we can examine the entire unit. On the opposite page is a five-tube AC-DC superheterodyne receiver commonly found in table model radios. Four circuits handle the signal; one, the 35W4 halfwave rectifier, provides the operating plate and screen voltages. Signal processing starts at the antenna input circuit, which is a loop antenna acting as an inductance in parallel with C1. This antenna input circuit tunes to the desired frequency in the broadcast band providing the necessary input voltage to the r-f grid of the 12BE6 converter. Grid 1 of the converter acts as the oscillator grid, grids 2 and 4, tied together, act as both a screen and oscillator plate. The feedback type oscillator's tank circuit is magnetically coupled to the winding in the cathode circuit. The screen grid of this tube reaches B+ by being tied directly to its line. The tubes in this receiver may have the screen operating at approximately the same voltage as the plate which makes this arrangement economical. The plate circuit of the 12BE6 is tuned to the i-f, 455 kc. The signal is fed by magnetic coupling to the input of the 12BA6 i-f amplifier circuit. Here the signal receives extreme amplification. Through magnetic coupling of the output i-f transformer, the i-f signal is fed to the input of the diode section of the 12AV6. Note: the diode plates are tied together. The circuit does not use delayed avc, so most manufacturers tie these plates together. Sometimes the lower plate is grounded or tied to the cathode. The detector load is also used as the volume control potentiometer, with the audio voltage taken from there and fed to the triode section of the 12AV6. The high side of the diode load represents the takeoff point for the avc line, which is fed through the R-C avc filter to the control grids of the i-f amplifier and converter. The output of the audio voltage amplifier is fed to the 50C5 power amplifier which drives the permanent magnet loudspeaker. The output transformer provides the necessary impedance match between the higher impedance plate circuit of the 50C5 and the lower impedance voice coil of the loudspeaker.

The 35W4 halfwave rectifier provides the necessary plate and screen voltages. Maximum B+, taken from the cathode of the 35W4, is fed to the plate of the power amplifier. The balance of the B voltages are taken from the output of the R-C filter. Note that the heaters are in a series-string arrangement, with the entire line voltage dropped across all the heaters. The 35W4 has provision across part of its tapped heater for connection of the pilot light. Most ac-dc receivers contain additional components, such as decoupling networks and voltage dropping resistors. This circuit represents the utmost in simplicity for a five-tube ac-dc superheterodyne receiver.

We might now ask how do more elaborate superheterodyne receivers differ from this basic circuit. For one thing, more elaborate receivers will use a transformer-type full-wave rectifier power supply which permits higher plate and screen voltages and greater voltage regulation and filtering action. In addition, a better receiver would have one or perhaps two stages of r-f amplification preceding the converter. It might also have a separate oscillator-mixer circuit which would provide greater frequency stability of the oscillator. Furthermore, it might have two or even three stages of i-f amplification, for much higher gain and selectivity.

Electron-Ray Tube

The electron-ray tube, or "magic eye", contains two sets of elements; one of which is a triode amplifier, the other a cathode-ray indicator. The plate of the triode section is internally connected to the ray-control electrode (A) so that the voltage on the ray-control electrode varies as plate voltage varies with the applied signal. The ray-control electrode is a flat, metal strip placed relative to the cathode to deflect some of the electrons emitted from



the cathode. The electrons that strike the anode, or target, cause it to fluoresce, or give off light. The electron deflection caused by the ray-control electrode prevents electrons from striking part of the target; thus, a wedge-shaped shadow is produced on the target. The size of this shadow is determined by the voltage on the ray-control electrode. When this electrode is at approximately the same potential as the fluorescent anode, the shadow disappears.

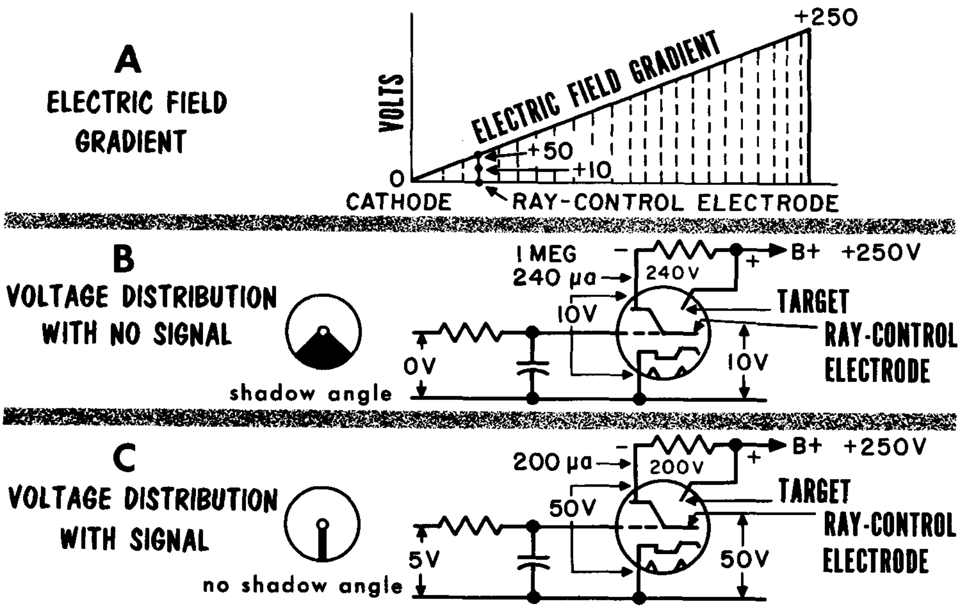
If the ray-control electrode is less positive than the anode, a shadow appears, the width of which is dependent upon the voltage on the ray-control electrode with respect to the anode. If the tube is calibrated, it may be used as a voltmeter, when rough measurements will suffice. The magic-eye tube is used principally as a tuning indicator in receiving sets and as a balance indicator in bridge circuits.

The Electron-Ray Circuit

Shadow angle width depends on the voltage between the ray-control electrode and ground, compared to the voltage between an electric field gradient point and ground near the ray-control electrode, (A).

When no signal is applied to the grid of the triode section, the plate current is 240 μa (B). The voltage on the ray-control electrode equals the plate supply voltage less the drop through the 1-megohm resistor, or 250-240 = 10 volts. The electric field gradient is assumed to vary as a straight line, starting at the cathode with zero potential and ending at the anode with a potential of +250 volts with respect to cathode. A point on the electric field gradient in the vicinity of the ray-control electrode has a potential of +50 volts with respect to ground. Thus, the ray-control electrode is negative with respect to the field at this point by an amount equal to -(50-10), or -40 volts. The negative charge repels electrons and the shadow angle is established.

In (C), a 5-volt signal is developed between grid and ground of the triode section of the magic-eye tube. The plate current is reduced to 200 μa, and the potential of the ray-control electrode is equal to 250-200, or +50 volts

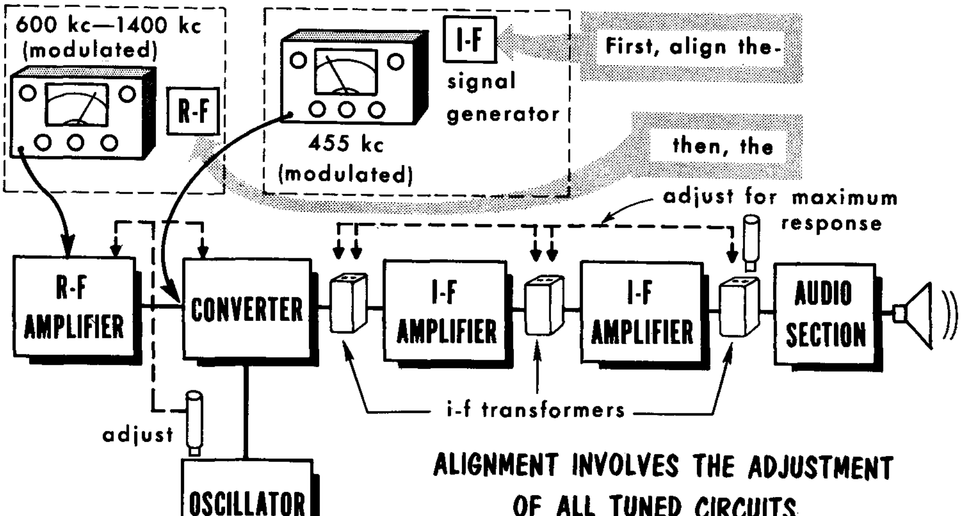


with respect to ground. Since the potential of a point on the electric field gradient in the immediate vicinity of the ray-control electrode is also +50 volts with respect to ground, there is no difference in potential between the control electrode and the field. Thus, the control electrode does not repel electrons, and the shadow angle closes, indicating the signal voltage applied to the triode grid.

Superheterodyne Alignment

The superheterodyne receiver contains several tuned circuits operating at different frequencies. For any given station, there is an incoming r-f signal, an oscillator signal, and a fixed i-f signal. For maximum gain and minimum distortion, it is important that all these tuned circuits be properly adjusted. When they are, the receiver is said to be aligned. Alignment is necessary in two situations. One, where the set is initially put together by the manufacturer; in this case, alignment is done directly at the factory. Second, after a set has been in use for some time, due to vibrations from the loudspeaker and changes in temperature and humidity, component values may change slightly and detune a circuit.

Alignment generally begins with the adjustment of the i-f amplifier. Most receivers give the i-f directly on the cabinet or chassis. Should the i-f be 455 kc, a signal generator tuned to that frequency and containing an audio



ALIGNMENT INVOLVES THE ADJUSTMENT OF ALL TUNED CIRCUITS FOR OPTIMUM SENSITIVITY AND SELECTIVITY.

modulation is connected to the r-f grid of the converter. The various tuned circuits of the i-f amplifier are then adjusted, starting from the secondary of the output transformer to the primary of the input i-f transformer. Adjustment is made for maximum output. It is usually desirable to keep the volume low so that changes in volume are easily noticed. At maximum response from i-f alignment, the signal generator is set at the high end of the broadcast band, usually about 1400 kc, and connected to the antenna terminals. The tuning dial is turned to 1400 kc and the tuning capacitors or inductors aligned for maximum output in the r-f amplifier, converter, and oscillator stages. The signal generator and tuning dial are set at the low end of the band, 600 kc, and adjustment for maximum response is made with the low-frequency padder capacitor. Manufacturers give precise alignment procedures for their receivers that must be followed carefully.

The superheterodyne receiver differs essentially from the TRF receiver in that it changes the frequency of the received signal to a lower, fixed value, at which the tuned amplifying circuits can be designed to operate with maximum stability, selectivity, and sensitivity.

When a modulated radio signal is heterodyned with a locally generated signal, the envelope of the resulting beat frequency (i-f), contains the modulation of the original radio signal.

Two r-f signals of different frequency interact only if they are combined in a mixer with a nonlinear characteristic. Of the many frequencies produced in the mixer stage, only the difference frequency (i-f) of the original signals is selected by a tuned circuit in the mixer-output stage.

The principle of frequency conversion inherently provides higher selectivity than a TRF receiver.

The basic components of a superheterodyne receiver are the antenna, the r-f amplifier, the frequency converter consisting of the mixer and local oscillator, the i-f amplifier, the detector, the a-f amplifier, and the reproducer.

The selectivity, fidelity, and gain of the superheterodyne receiver is controlled largely in the intermediate amplifier stages.

The i-f transformer usually consists of two coupled resonant circuits, whose bandpass characteristics depend on the degree of coupling. If the coupling is greater than the critical value, a double-peaked resonance curve results, which provides substantially uniform response between peaks.

Most superheterodyne receivers use a diode detector, because its linear characteristics give it high signal-handling ability with low distortion. A diode detector, however, has poor selectivity and low sensitivity.

An avc circuit derives a negative bias voltage proportional to the carrier amplitude by rectifying the carrier with a diode detector.

Push-button tuning permits quick selection of any one of a number of fixed-tuned stations; it can work mechanically or electrically.

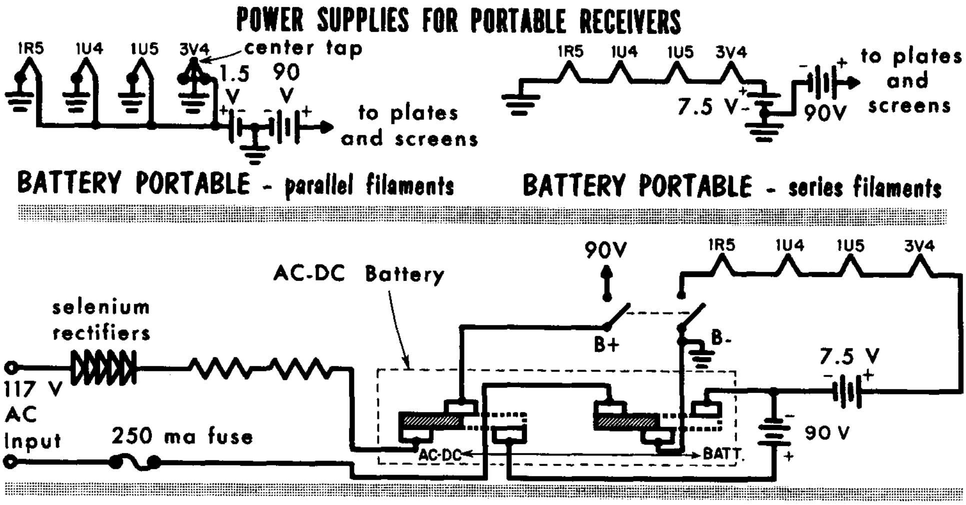
Manual sensitivity control is often incorporated into the r-f stage in the form of a variable cathode resistor, providing variable bias to the grids of remote cutoff amplifier tubes.

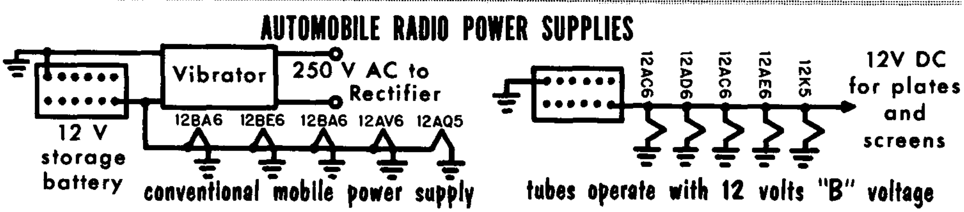
REVIEW QUESTIONS

1. State an important advantage of a superheterodyne over a TRF receiver.
2. State briefly the steps involved in the reception of an AM signal by a superheterodyne receiver.
3. Describe the process of frequency conversion in a mixer stage.
4. Give two methods by which the local oscillator voltage can be injected into the mixer stage.
5. Define critical coupling. Explain what happens when coupling is greater.
6. Explain the operation and advantage of the diode detector in a superheterodyne receiver.
7. What is the purpose of avc and how is it used?
8. What is an image frequency?
9. Describe briefly the alignment of a superheterodyne receiver.
10. Explain the operation of an electron-ray circuit.
11. What is the advantage of a mixer-oscillator circuit over a converter?
12. What is the function of a trimmer capacitor?

Portable and Automobile Receivers

Particular consideration must always be given to the power supply of portable receivers. Special circuitry and tubes are used with the usual 117-volt 60-cycle supply. Some portable receivers use tubes having 1.5-volt parallel-connected filaments, and are driven from a 1.5-volt battery combined with a





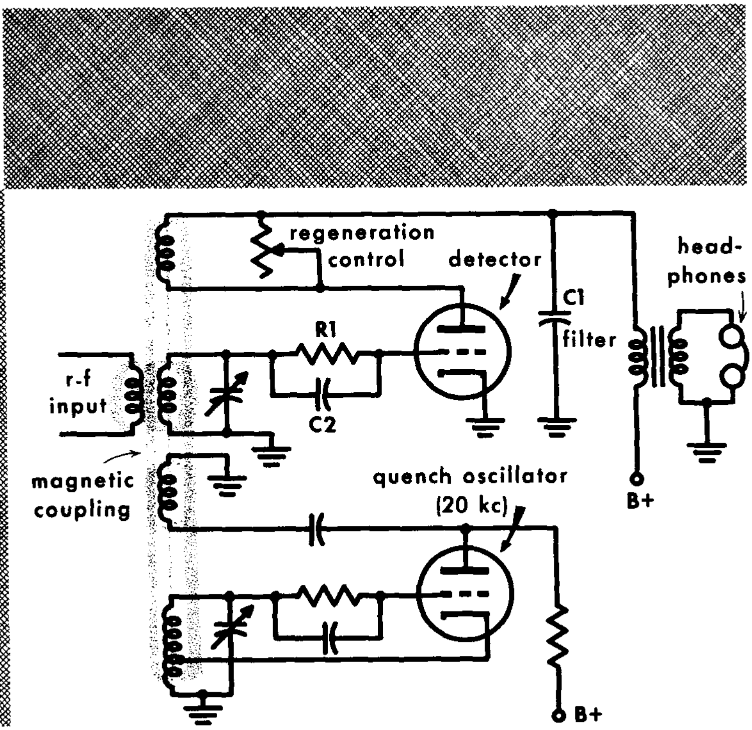
67.5- or 90-volt battery for plate and screen voltages. Some portables have filaments connected in series, driven by a 7.5- or 9-volt battery. Others of the ac-dc-battery type use a conventional power supply with provision for switching to battery operation.

Automobile receivers require special consideration. Because they are situated close to ignition noise, they must be well shielded. The antenna is relatively small, and they often operate in poor reception areas, as between cities, which calls for high sensitivity. Because of these requirements, "auto" radios have many stages of amplification, and very often use a preselector or r-f amplifier stage. The power supply for these units is, of course, the car's 6- or 12-volt battery (actually 6.6 or 13.2 volts; 2.2 volts to a fully charged cell). Originally, all auto radios used a vibrator-type power supply that converted the dc from the battery to ac that could be fed to transformers. However, with the recent development of tubes having 12-volt heaters and requiring only 12 volts on the plate and screen, no power supply other than the battery is needed. Transistors are also used (Vol. 5) that work directly from the storage battery.

The Superregenerative Receiver

An effective simple receiver of high-frequency CW (continuous wave) signals is the superregenerative detector. When the detector oscillates, the amplitude of oscillation is controlled by the amplitude of the r-f input signals. The time that the detector oscillates is controlled by the quench oscillator, which operates at about 20 kc and applies an additional signal to the detector

BASIC "SUPERREGEN" an excellent high-frequency detector

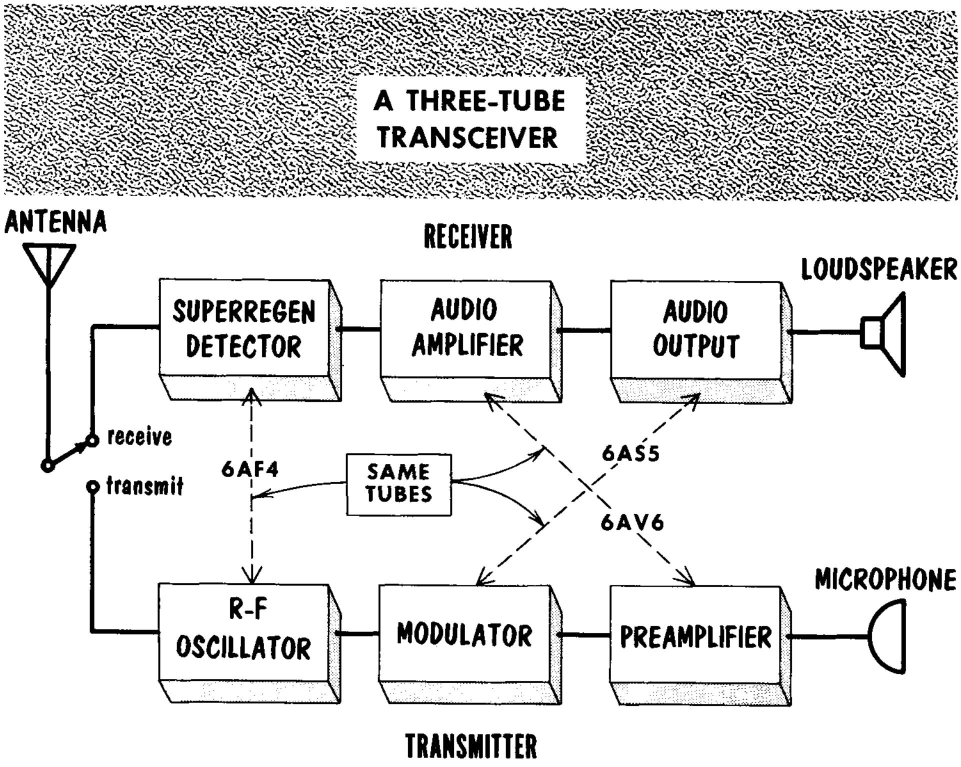


grid. When this signal is positive, the detector oscillates; when negative, the detector is cut off. Each time the detector oscillates, a voltage pulse is developed in the plate circuit of the detector. The amplitude of this pulse is controlled by the amplitude of the incoming r-f signal during the time of the pulse. Successive pulses, therefore, vary in amplitude according to the modulation envelope. C1 filters these pulses so that only the audio voltage is applied to the primary of the output transformer.

The quench oscillator must operate above the audio range to prevent the quench frequency from being heard in the output. The ratio of the r-f input frequency to the quench frequency must be at least 100 to 1, to prevent a large amount of noise from developing. This means that the minimum r-f at which this type receiver will operate satisfactorily is 2000 kc. Superregenerative detectors have high sensitivity, but poor linearity. They have poor selectivity because the grid current loads the tuned circuit. Their signal-handling ability is very good.

The Transceiver

The transceiver is a piece of electronic equipment that operates as both a receiver and a transmitter. In this type unit many of the tubes are used for both receiving and transmitting. Transceivers are often used as portable equipment, such as mobile radio for amateur use, and for "citizens band" radio, when the equipment should be as light as possible. The transceiver uses three general common circuits: A common antenna, for transmitting and receiving; a common power supply, for the transmitter and for the receiver; and many of the tubes serve both a transmitting and receiving function. The transceiver has a transmit-receive switch which places the neces-

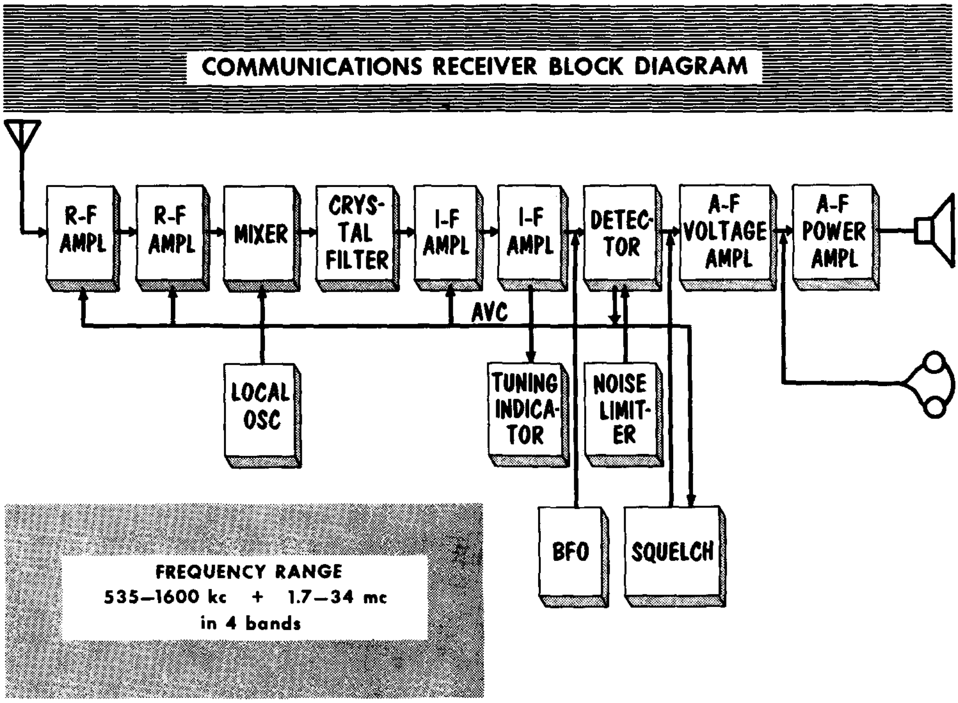


sary circuits in their proper position. Since a circuit performs two functions, its design must represent a compromise. A transceiver will rarely give the quality of performance of a separate transmitter-receiver combination.

In a typical transceiver the receiver's audio amplifier system serves as the modulation system for the transmitter; the r-f amplifier and mixer of the receiver can serve as the r-f amplifiers of the transmitter; the receiver's local oscillator can serve as the generator of carrier waves in the transmitter; or, a receiver's regenerative detector can serve as the transmitter's oscillator.

Communications Receivers

Fundamentally, a communications receiver is a basic superheterodyne receiver with special circuits. Because it most often works with very weak signals (less than 1 microvolt) it requires an extremely high order of sensitivity. To add sensitivity, additional stages of r-f and i-f amplification are

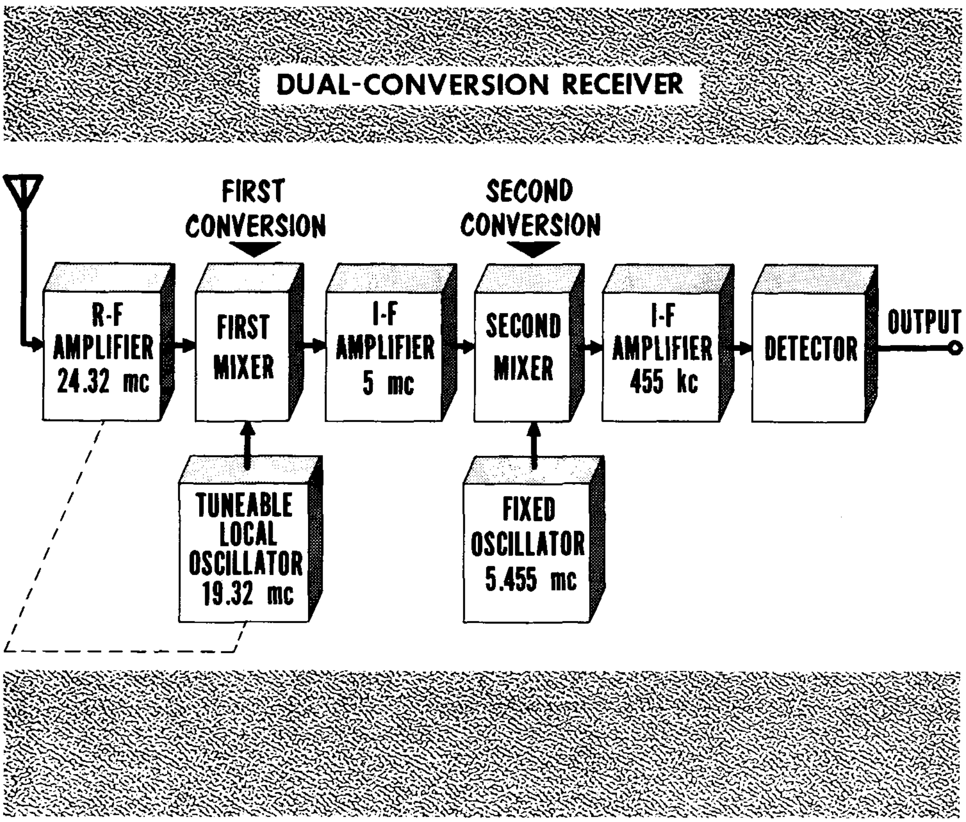


used. A communications receiver is rarely used for reception of music; it is used mainly for reception of speech and continuous wave (CW) code signals. Different type signals require different bandwidth response curves in the i-f stage tuned circuits, necessitating some form of bandwidth control.

A tuning indicator is used primarily to aid in selecting the desired signal from among many others. The nature of the tuning indicator lends itself to doubling as a signal-strength indicator. For detection of CW and single side-band signals, an additional oscillator, called the beat frequency oscillator (BFO), is required. A disturbing factor is the high susceptibility to noise due to high receiver sensitivity. Noise is an amplitude-modulated signal, riding in with the desired signal; it cannot be removed, but it can be limited in amplitude preventing it from drowning out the desired signal. To ease periods of extended listening, squelch circuits are used to quiet the audio output of the receiver when no signal is being received. Upon transmission the received signal automatically releases the squelch circuit, allowing the detected signal to pass through the audio circuits. The audio output can be applied to either a loudspeaker or earphones.

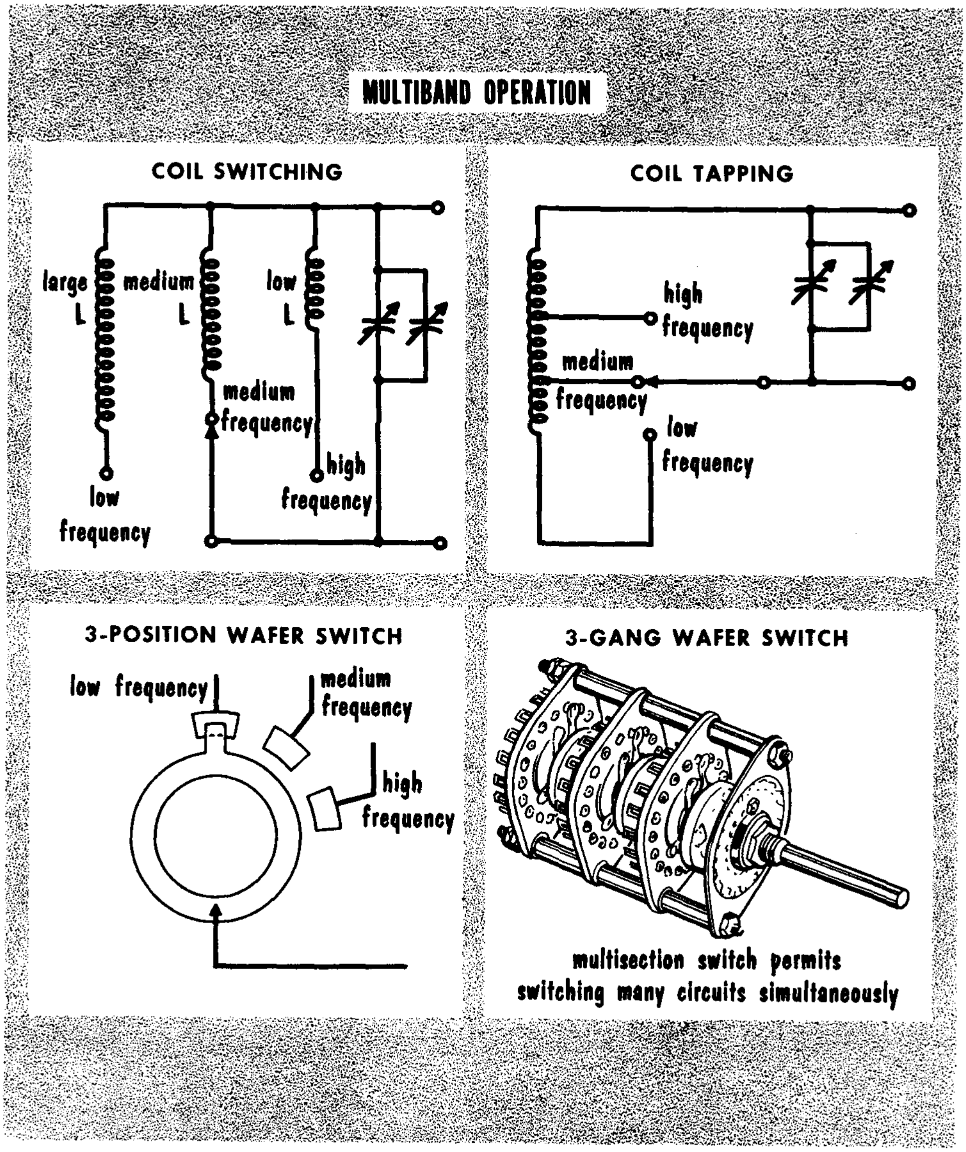
Dual Conversion

The use of a 455 kc i-f becomes a problem when a communications receiver is operating at high frequencies. The problem is trying to keep the local oscillator tuned exactly to the difference frequency above or below a typical high value of frequency such as 24.320 mc. In this example the local oscillator, if tuned above the incoming signal, must be set to 24.775 mc. At this frequency the slightest oscillator drift would result in a wrong value difference frequency. An upwards shift of 1/2 of 1% in the local oscillator frequency is equal to a shift of 123.875 kc. The new local oscillator frequency of approximately 24.899 mc would provide an i-f of 579 kc, completely out of the i-f range. To overcome this problem dual conversion is used.



The block diagram shows the high frequency r-f signal mixing with the local oscillator to produce a high value i-f of 5 mc. The wide bandpass of the 5-mc i-f covers any normal drift of the local oscillator. The output is then mixed with a fixed oscillator signal at 5.455 mc to produce a second and final i-f of 455 kc. Dual conversion permits increased gain and relaxes the need for special oscillator frequency control. It also aids image rejection, which can be troublesome at high frequencies.

Multiband Operation



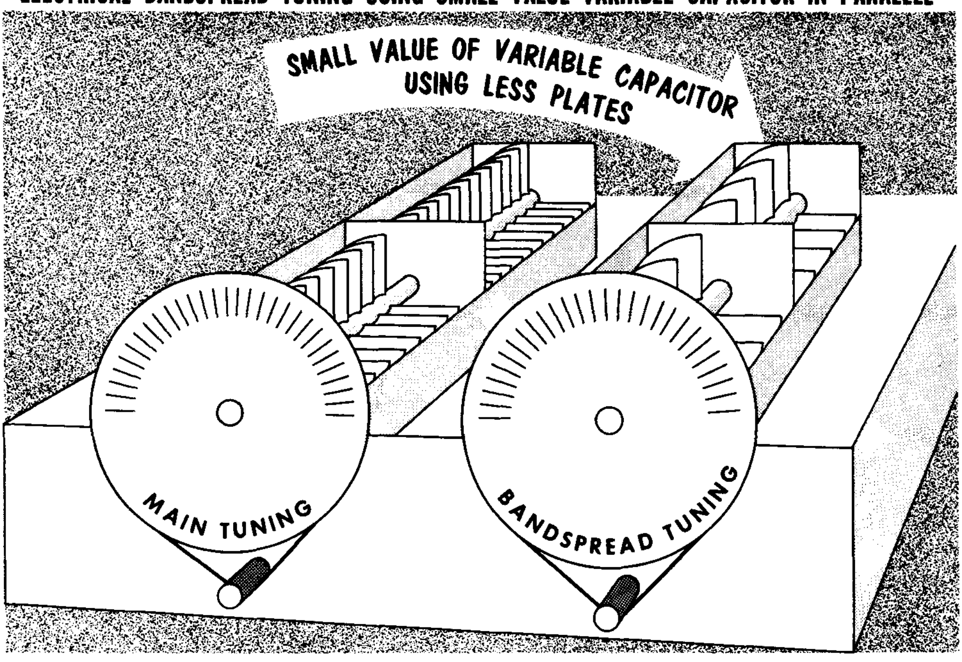
Communication receivers require extensive coverage. The average communication receiver operates from just above the broadcast band to 30 mc. Some receivers include the broadcast band, others range to as high as 50 mc. To tune in these higher frequencies, various methods are used to lower the value of the tuned-circuit components. It is difficult to switch in variable tuning capacitors for each band. Different coils are switched in, or, for economic reasons, one coil is tapped for the lower values of inductance necessary for high-frequency tuned circuits.

Bandspread Tuning

Tuning in a station in the broadcast band results in no apparent crowding. Since each station is at least 10 kc apart, they are relatively easy to tune. The broadcast band starts at 535 kc and ends at 1605 kc; thus the entire band is 1070 kc wide. Switching to a typical short wave band, the dial may read 6 mc at one end and 18 mc at the other. This results in a bandwidth of 12,000 kc on the same length of dial used for tuning the broadcast band. It becomes difficult to select stations, as each slight turn of the dial may pass the desired signal unnoted.

To overcome this difficulty, bandspread tuning is used. Mechanical and electrical bandspread tuning are used. Mechanical-bandspread tuning uses a special dial with vernier tuning (fine tuning). The dial may take up to 100

ELECTRICAL BANDSPREAD TUNING USING SMALL VALUE VARIABLE CAPACITOR IN PARALLEL



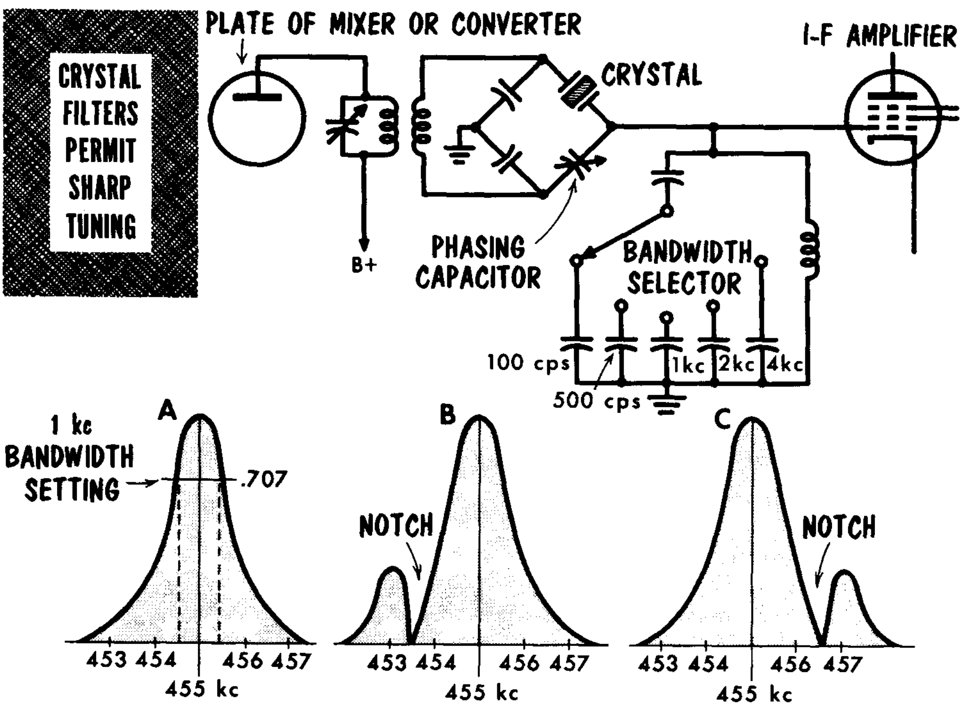
turns of the tuning knob to one turn of the variable-tuning capacitor. This very fine tuning permits reception throughout the band, removing the possibility of passing the desired station.

Electrical-bandspread tuning uses a small value variable-tuning capacitor wired in parallel with the main variable-tuning capacitor. Using the main variable-tuning capacitor, the dial is set to approximately the desired frequency. Varying the small value variable-tuning capacitor (called the bandspread capacitor) will, in effect, be making very small changes in the value of the main variable-tuning capacitor. This permits very fine tuning about the setting of the dial.

Adjustable Bandwidth, Crystal Filters

Tuning CW signals requires tuning over a narrow (approximately 200 cycles) band of frequencies since the CW signal is an interrupted carrier frequency. For such sharp tuning, and to remove close-by interfering signals the i-f stage will often have variable bandwidths, and a special filter. The bandwidth is varied mechanically by moving the primary and secondary windings of the i-f transformer closer or farther apart; or electrically by loading the i-f circuits with a resistor across the tuned circuit to lower the Q. Varying the bandwidth is helpful when nominal interference is present. A crystal filter circuit is used to vary the bandwidth and provide extremely sharp tuning.

A crystal filter makes use of the very high Q of a quartz crystal at its resonant frequency. A crystal cut to resonate at 455 kc and placed in a bridge circuit will provide the response curves shown. When the phasing capacitor is set to equal the value of the capacitance of the crystal holder, the circuit is balanced and the resonant curve is a symmetrical, high-Q, very narrow bandpass curve (A). Adding or subtracting some capacitance in the circuit

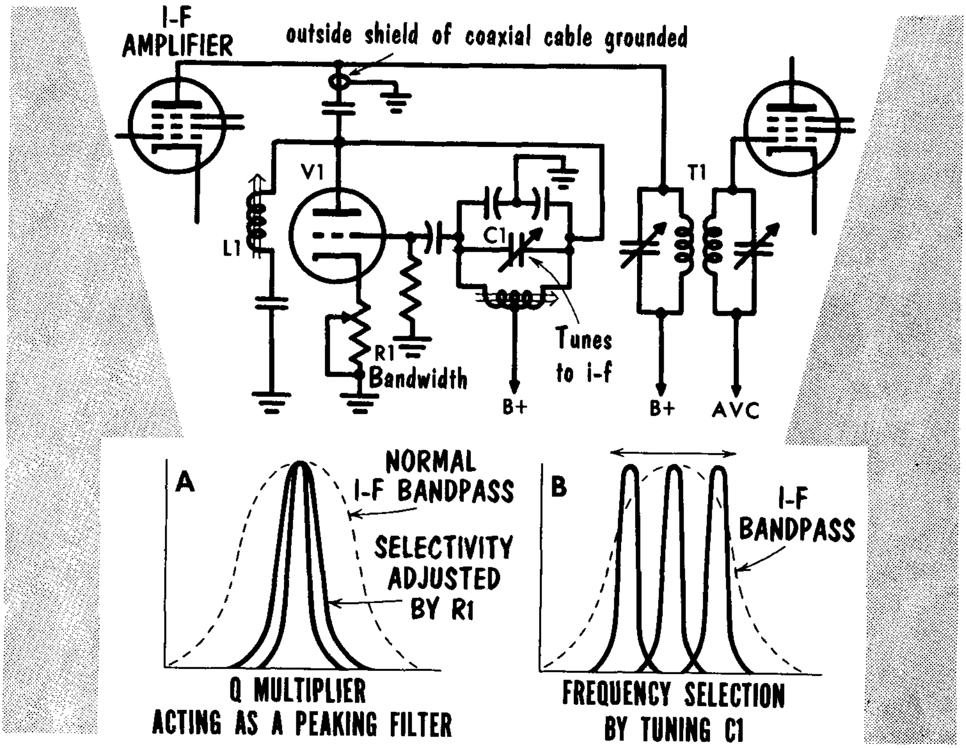


by varying the phasing capacitor will introduce a notch on either side of the resonance curve (B and C). By careful tuning an interfering signal can be placed in the notch for rejection while at the same time providing high gain to a narrow band of desired frequencies. Varying the value of the capacitor in the grid tank circuit varies the overall bandpass of the filter.

Q Multiplier

Crystal filters can be difficult to tune. Adjusting the phasing control also has an effect on the bandwidth. The Q multiplier circuit takes advantage of controlled regeneration to increase the Q of a tank circuit, peaking a signal, or it can use the same tank circuit as a trap to remove a signal.

Positive feedback from plate to grid of V1 is controlled to increase circuit gain, just short of breaking into oscillation. Much of the resistive loss of



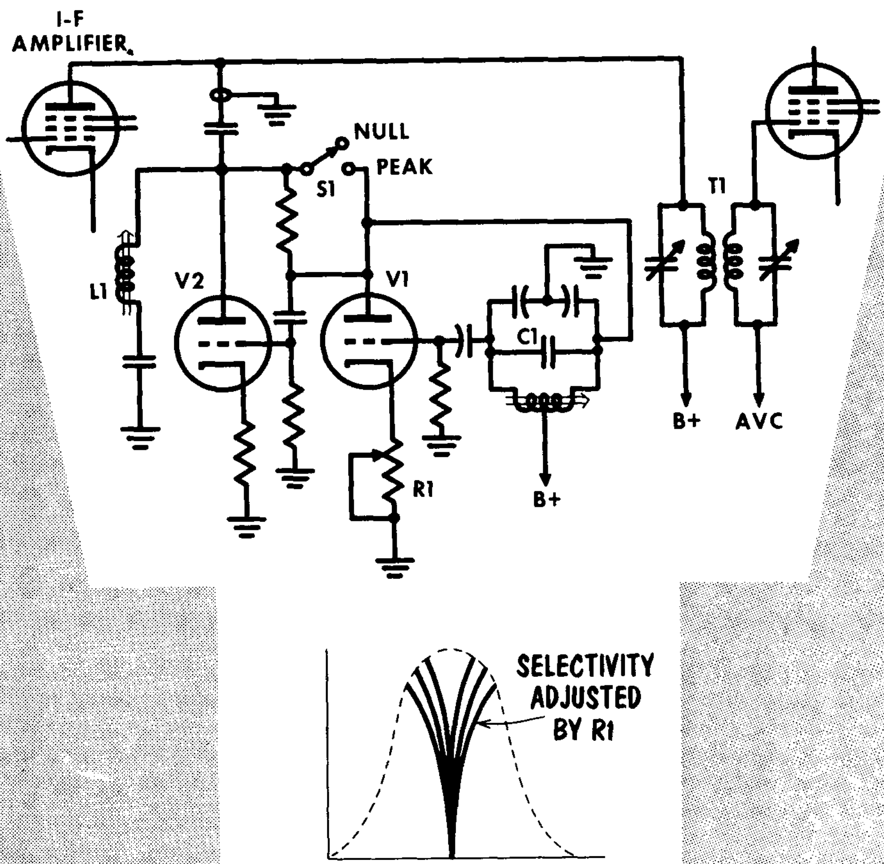
the tank circuit is made up by the gain of the tube. Since $Q = X_L/R$, and R is very low in this circuit, the result is a very high Q circuit equivalent to that found in crystal filters.

Connecting the Q multiplier to the plate of the i-f amplifier places a high-Q, high-impedance tank circuit in parallel with the primary of i-f transformer T1. This places an extremely narrow bandpass filter ahead of T1. To vary the bandwidth of the filter, the gain of V1 is varied by the cathode bias adjustment R1 (A). C1 tunes the tank circuit to the i-f, or, if desired, anywhere within the i-f bandpass (B). This allows selection of a desired signal within the i-f bandpass without re-tuning the receiver. To overcome losses in the interconnecting cable, inductor L1 tunes the capacitance of the connecting coaxial cable and associated capacitor to parallel resonance at the i-f.

Q Multiplier (Cont'd)–Null Circuit

Insertion of amplifier V2 inverts the action of the Q multiplier. In place of positive feedback, the combined circuit now offers negative feedback over the same narrow band of frequencies. The combined circuit becomes a series resonant circuit parallel to the primary of T1. At resonance a series resonant circuit is effectively a short circuit to ground.

At the same frequencies in which the high-Q circuit previously peaked the narrow band of frequencies, it now acts as a short circuit. The narrow band

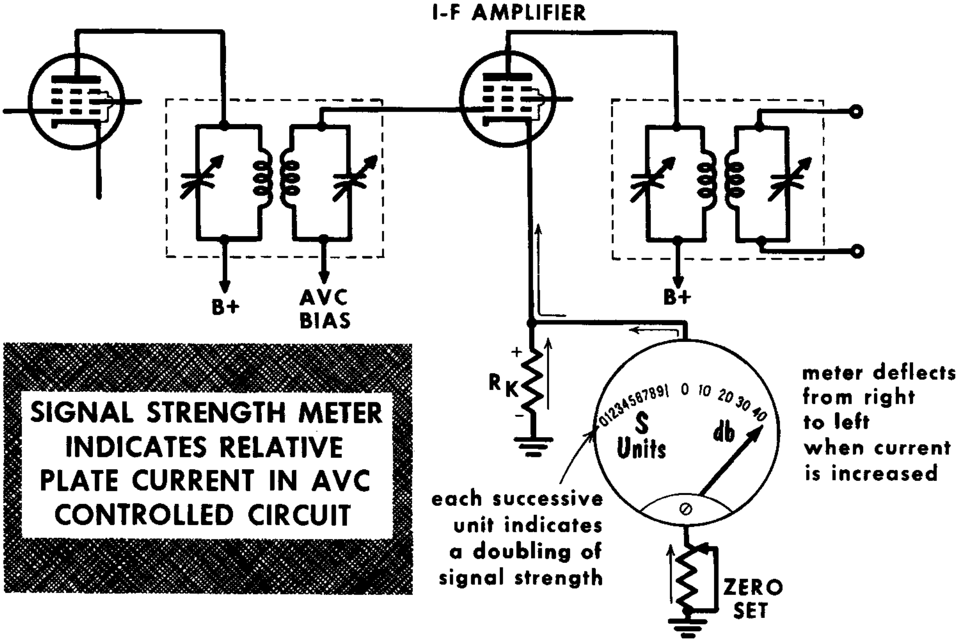


Q MULTIPLIER
ACTING AS A NULL FILTER

of frequencies are notched or nulled out of the i-f bandpass. As before, tuning C1 moves the notch across the bandpass. In this manner an interfering signal within the i-f bandpass can be nulled, making it easier to copy the desired signal. To resume action as a peaking circuit S1 is closed, bypassing inverter V2.

S Meters

The original function of an indicating meter was to aid in tuning. However, its position in the receiver circuits allows it to provide an indication of the signal strength of the received signal. To gauge the signal strength the meters are divided into S units. Most often the circuit is designed so that a



SIGNAL STRENGTH METER INDICATES RELATIVE PLATE CURRENT IN AVC CONTROLLED CIRCUIT

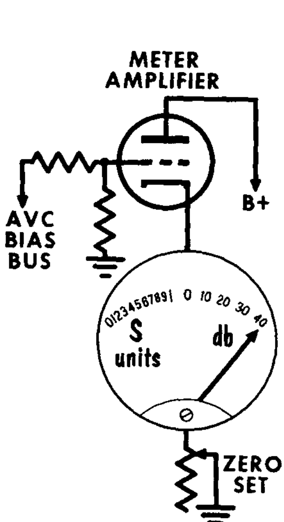
signal of 50 microvolts at the antenna input will provide an S9 reading on the meter; stronger signals are read in db above S9.

Reception and strength of a signal is indicated by using the change in plate current in an i-f amplifier tube to which avc bias voltage is being applied. In the circuit shown, a backwards-reading meter is placed in parallel with cathode bias resistor R_k. The avc bias voltage developed when a receiver is tuned to a signal reduces plate current, causing a smaller voltage drop across R_k, and a reduced reading in the parallel placed meter.

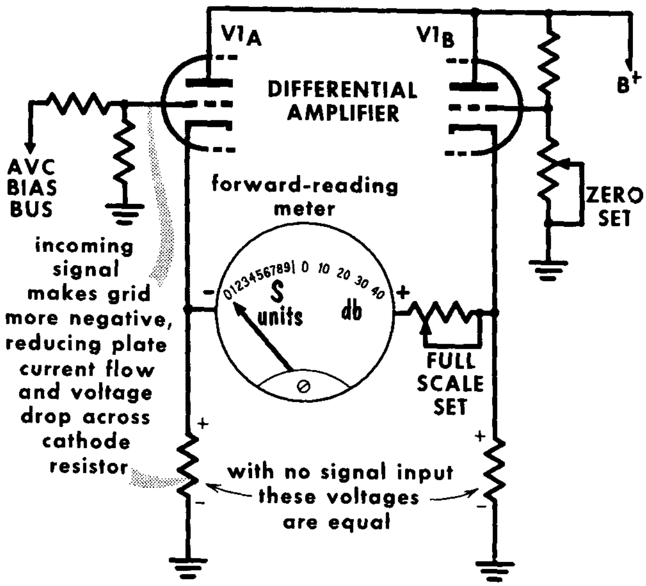
A reduced reading on a normal meter causes the needle to swing to the left. To indicate reception of a signal an S meter needle should swing to the right. A backward or "right zero" reading meter is constructed so that the zero-current setting, (and the pointer), is at the right side of the scale when no current is passing through the meter. The scale is printed normally, with 0 on the left. With no signal (no avc bias) the resulting high plate current causes the needle to swing full scale (zero scale reading). As a signal is tuned in the avc, bias results in reduced plate current, causing the needle to swing toward the right. A variable resistor placed in series with the meter permits calibrating for zero scale reading with no signal applied.

S Meters (Cont'd)

MEASURING SIGNAL STRENGTH DIRECTLY FROM AVC BUS



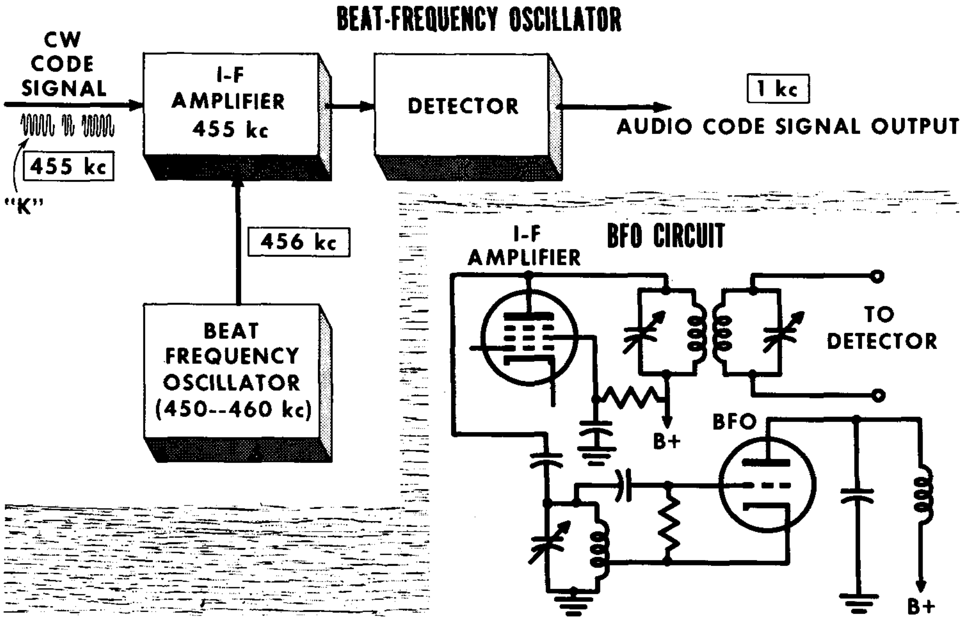
BRIDGE TYPE CIRCUIT MEASURES SIGNAL STRENGTH



Since the avc bias bus varies directly with signal strength, it would seem an ideal spot to place a meter. The avc bias voltage however is derived from a low-current, high-impedance source. Placing a meter across this source would load or place a near short circuit across the source. A meter amplifier circuit is used to isolate the meter from the bus, preventing the AVC bus from being loaded. The meter used is a backward-reading meter with the zero-scale reading on the left. The variable resistor in series with the meter is used to set it to read zero on the scale with maximum plate current, as set by no signal reception. Reception of a signal causes an increased avc bias on the amplifier grid, reducing plate current. The reduced plate current flow through the meter causes the needle to swing to the right, indicating an increase in signal strength.

A differential amplifier can be used as a sensitive S meter circuit; it also provides a good measure of linearity throughout the entire signal range. It consists of two triodes, one with a fixed bias (and resulting fixed-value plate current), the other with variable bias. With both triodes conducting equally there is no difference in potential between the two cathodes; the meter will read zero on the scale. A signal causing an increase in the avc bias applied to the grid of V1A causes its cathode to be less positive than the cathode of V1B. The difference in potential between the cathodes is indicated by the swing of the meter pointer. A variable resistor is placed in series with the meter to limit current flow to the maximum meter value. To set the meter to read zero with no signal, the bias of V1B is set to provide no difference in potential between the cathodes. An advantage of this circuit is its use of a normal forward-reading meter.

Beat-Frequency Oscillator



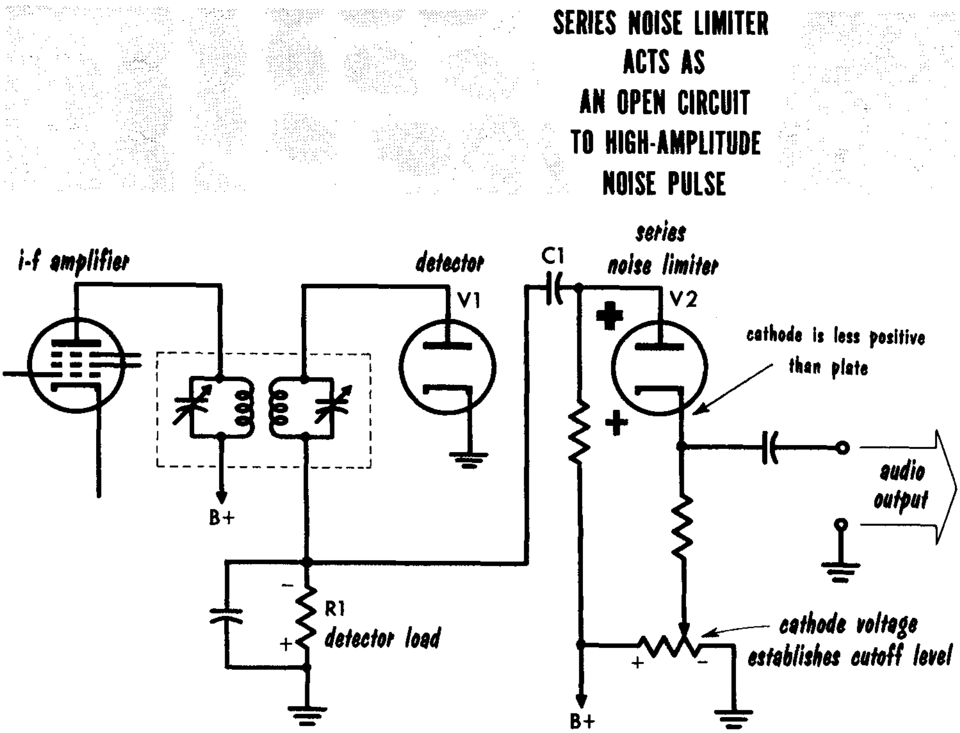
In voice communication, the r-f carrier is modulated to convey the intelligence. In code communication, as the International Morse code, the r-f carrier has a constant amplitude, but the carrier is interrupted in accordance with the code. When a code signal is received, amplified, and fed to the second detector, the r-f bypass capacitor filters the interrupted r-f signal, resulting in no output at the loudspeaker. To enable code reception, often referred to as CW (continuous wave), we must beat the incoming carrier against another signal to develop an audible beat frequency.

A beat frequency oscillator (bfo) is used. This oscillator tunes from approximately 5 kc below, to 5 kc above the i-f of the receiver. The output of the bfo is coupled either inductively or capacitively to the last i-f amplifier, or the second detector. When the code signal is received, and converted to the i-f of 455 kc, it beats with the output of the bfo, producing a difference frequency or beat note at an audio frequency at the detector. This is then fed to the audio amplifier. The frequency of the bfo can be varied to change the tone of the beat frequency, permitting the operator to select the tone most pleasant for listening, or a tone that will be most prominent among interfering signals. Assuming an input i-f signal of exactly 455 kc, with the bfo set for 450 or 460 kc, the audible beat note will be 5 kc. As the bfo frequency is brought closer to 455 kc the beat-note frequency becomes lower and lower. At exactly 455 kc there is no beat note; this is zero beat.

When using the bfo, the avc is usually removed by a switch that grounds the avc line, permitting the r-f and i-f amplifiers to operate at full gain. It prevents the output of the bfo from developing a bias voltage that would be fed back to the r-f and i-f amplifiers, reducing the receiver gain.

Noise Limiters

The high sensitivity of communications receivers creates a high background noise level. Most annoying is static and man-made noise from motors, automobile ignition systems, oil-burner ignition systems, etc. Since the noise frequencies are mixed with signal frequencies it is difficult to separate them.



A common method is to limit the amplitude of the detected signal applied to the audio amplifiers. This in turn limits the amplitude of a noise signal, preventing it from drowning out the desired signal.

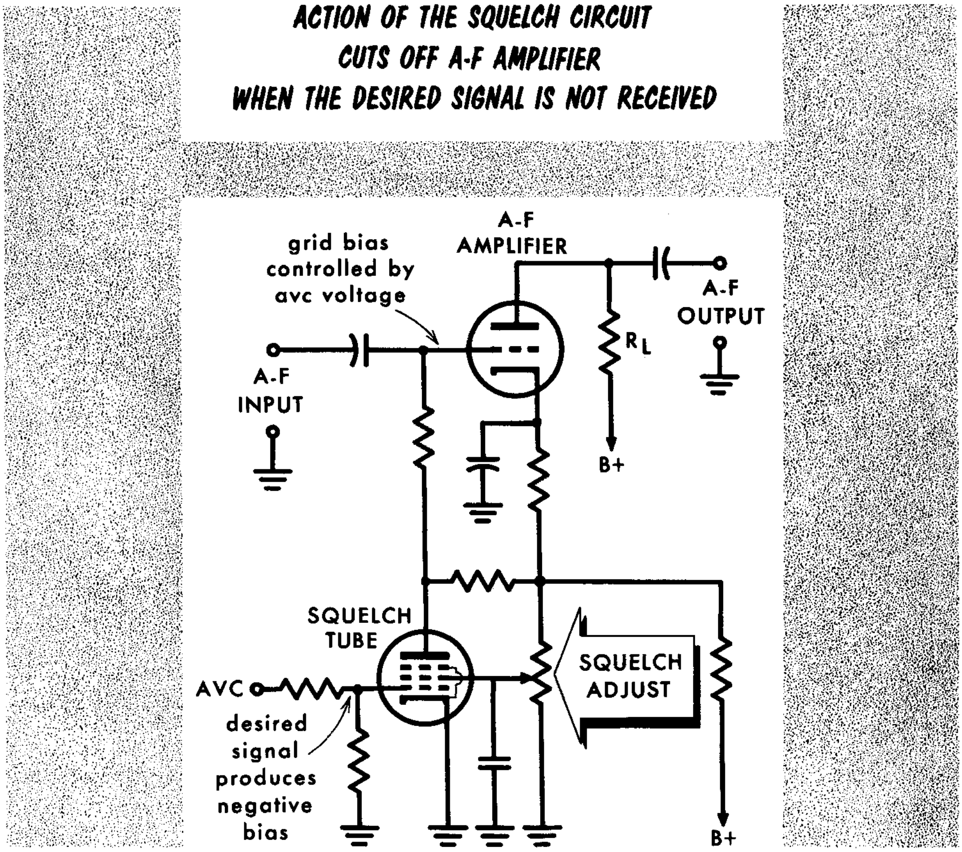
The circuit shown places the noise limiter tube in series with the detector output, and the input to the audio amplifiers. Because of this it is called a series noise limiter.

The cathode of V2 is set negative in respect to the plate by the value desired to act as a limit on the output signal. In this condition V2 is essentially a closed switch, allowing the detected signal to pass to the audio amplifiers. Reception of a noise pulse develops a large voltage across detector load R1. Capacitor C1 couples the increased signal voltage to the plate of V2. The high-value negative pulse places the plate of V2 negative in respect to the cathode. V2 acts as an open circuit and there is no output to the audio amplifiers. When the signal returns to its original level diode V2 again conducts to pass the audio signal.

Squelch Circuit

When operating a receiver on a fixed frequency over an extended period of time the constant background of noise and undesired signals can be very annoying. A squelch circuit automatically cuts off the undesired audio output. When the desired signal is received, the squelch circuit allows the signal to be amplified and passed through the audio circuits.

With no signal being received, there is no avc voltage on the grid of the squelch tube. The squelch tube conducts and its plate voltage drops. The



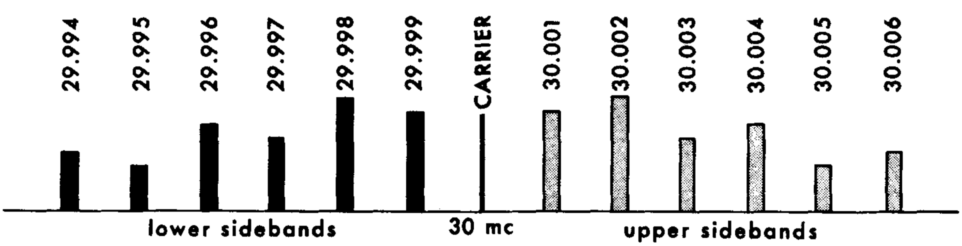
low plate voltage, through the connecting resistor to the grid of the audio amplifier cuts off the audio amplifier, and there is no audio output.

With a signal being received, the negative avc voltage developed is applied to the grid of the squelch tube, reducing plate current flow and increasing the plate voltage. The increased plate voltage places a more positive bias on the audio amplifier allowing it to act as a normal amplifier to provide an audio output. Varying the screen voltage of the squelch tube varies the point at which it cuts off the audio amplifier.

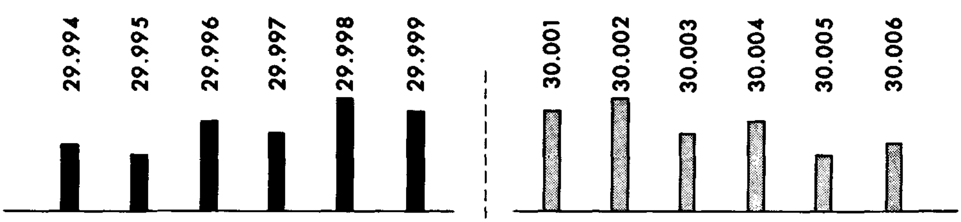
Single Sideband Signal

In previous discussions of AM we learned that sidebands were generated on either side of the carrier frequency. These sidebands contain the audio modulating intelligence. It was also learned that each sideband was a mirror image of the other. One set of sidebands is discarded at the receiver by the action of demodulation, and the other set filtered to recover the audio intelligence. Detection by a nonlinear device (diode) is similar to that of a mixer; the carrier beats against one of the sidebands to recover the audio intelligence. The carrier contains no information. Reviewing the three components of an AM signal; the lower sideband, carrier, and upper sideband, we find that all the intelligence is available in either sideband. The other sideband (it makes no difference, upper or lower) and carrier can be discarded with no loss of information. The result is a single sideband (SSB) signal.

CONVENTIONAL AND SINGLE SIDEBAND SIGNALS



CONVENTIONAL 30 mc AM SIGNAL WITH SIDEBANDS 1 kc APART



SINGLE SIDEBAND SIGNAL USING LOWER SIDEBANDS

SINGLE SIDEBAND SIGNAL USING UPPER SIDEBANDS

Power is conserved in a SSB transmitter; there is no power wasted on the useless carrier and duplicate sideband. A SSB transmitter can operate at peak power, whereas AM transmitters must operate at average power. Stated another way, a conventional 100 watt AM transmitter, converted to use its power on only one sideband, can provide up to 400 watts of peak power. In addition to conserving power, a SSB signal also conserves spectrum space. The SSB signal requires one-half the spectrum space taken by an AM signal. High stability is required of SSB transmitters and receivers. In a SSB receiver a substitute carrier has to be provided for detection. If the transmitted signal frequency shifts it will be quickly noticed in the received signal. The receiver, too, must have a high order of stability to prevent distorting the detected signal.

Balanced Modulation

A basic understanding of the generation of a SSB signal will aid in understanding the problems of SSB reception. The main problems are removal or suppression of the carrier and one sideband. The carrier is most often suppressed by a balanced modulator circuit. The balanced modulator has two inputs and one output. There is no output unless both input signals are applied together. Applying only the r-f carrier signal applies the signal in equal value and polarity to the grids of each tube. (This causes equal amounts of

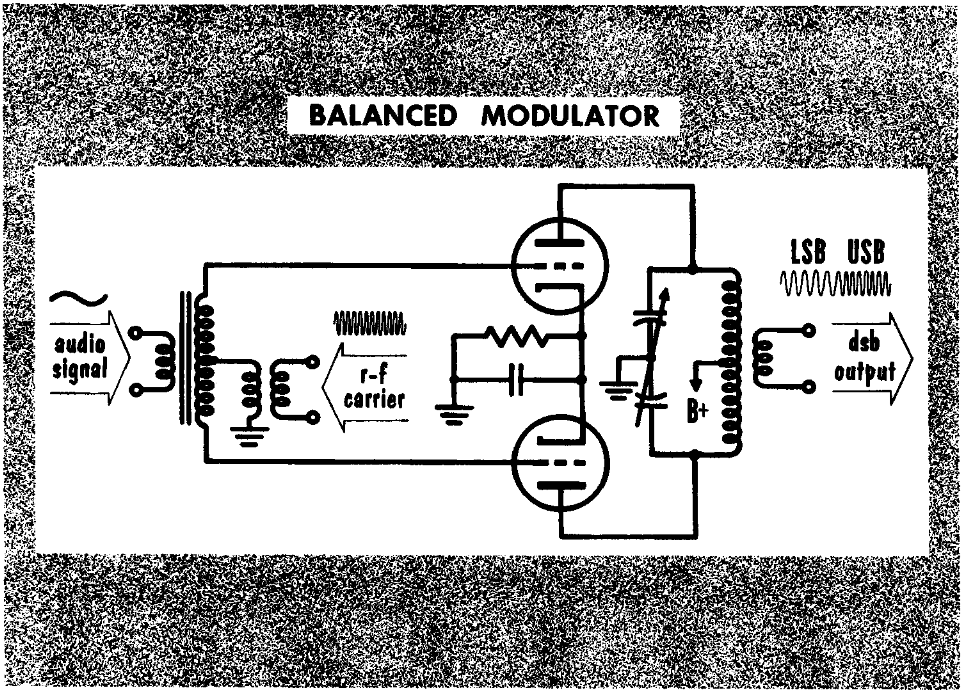
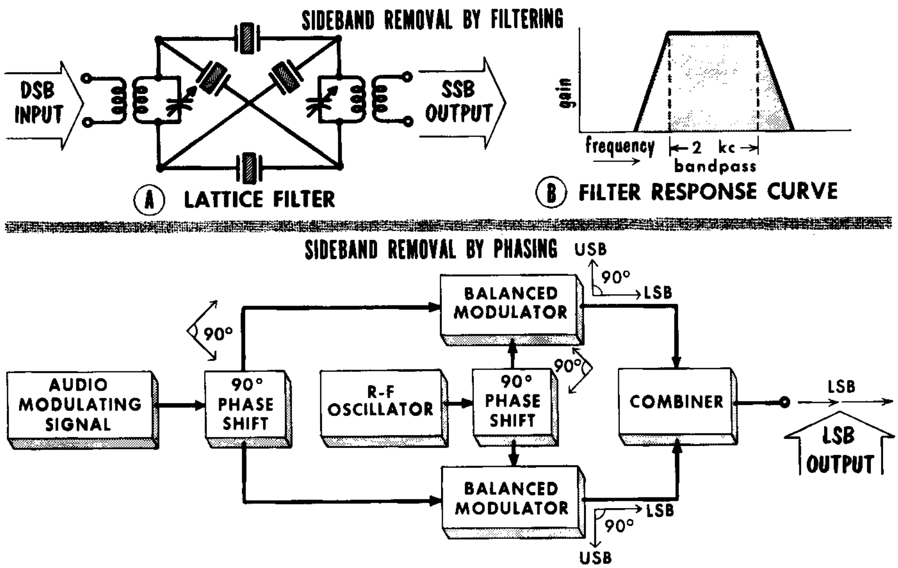


plate current to flow in opposite directions, cancelling each other, resulting in no r-f carrier output signal.) Applying the audio signal causes the circuit to act as a conventional push-pull amplifier. However, the plate circuit is tuned to the r-f carrier. With no appreciable plate load, there is no audio output signal.

Applying both the r-f carrier and audio signals results in the audio signal providing equal but opposite polarity signals at the grid of each tube, placing one grid positive and one grid negative. With the r-f carrier also applied, it places the same polarity signal at both grids. When the carrier places a negative signal on the grid of the tube that has a positive audio signal the result is cancellation of both signals. The negative carrier signal on the grid of the tube with a negative audio signal results in an additive signal. Mixing both the a-f and r-f signals generates an output signal composed of the carrier frequency plus or minus the audio frequency. There will be no output at either the audio or carrier frequency, only the two sideband frequencies. This type of signal is called a double sideband (DSB) signal.

Sideband Removal

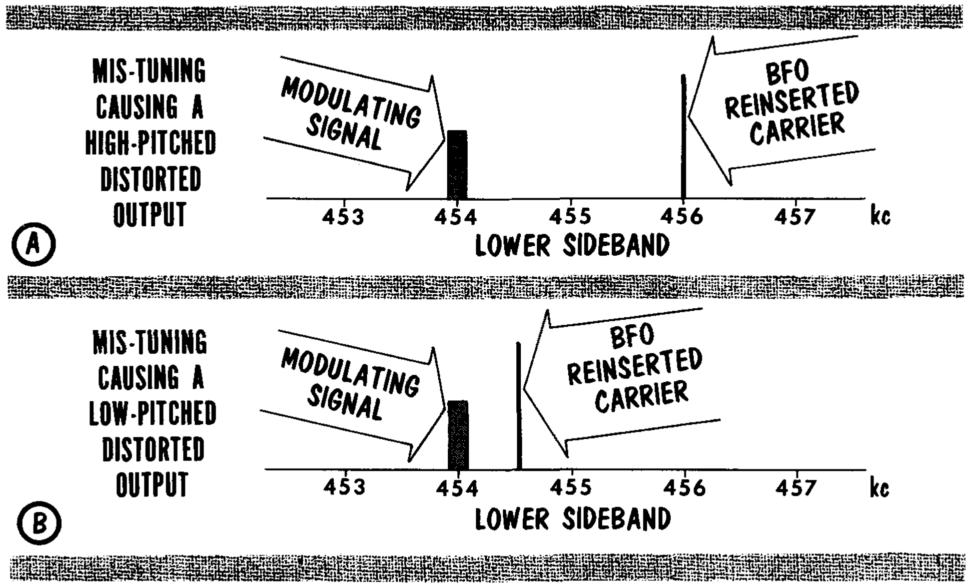
It is necessary to suppress one of the sidebands to obtain a SSB signal from the output of the balanced modulator. The two most popular methods are filtering and phasing. A crystal filter circuit of the type shown is called a lattice filter. Two pairs of crystals are used, each pair tuned 1 kc apart to provide a 2-kc bandpass. The filter bandpass is selected to match the frequencies to be passed. Most speech frequencies are centered within a 0–2000-cycle range. A suppressed carrier frequency of 8 mc, with the lower



sideband suppressed, would have upper sideband frequencies 8 – 8.002 mc. A crystal filter with the center frequency tuned to 8.001 mc will pass these frequencies. The frequencies within the filter's resonant bandpass receive very little attenuation. At off-resonant frequencies the high value of X_L or X_C provides high signal attenuation. The steep-slope narrow-bandpass characteristics of a crystal filter makes for an excellent filter for attenuation and isolation of undesired sidebands.

Removing a sideband by phasing is done by applying out-of-phase signals to mixing circuits. As shown in the block diagram the modulating audio signal is applied to a phase-shifting network supplying two equal-amplitude, 90°-out-of-phase signals. An r-f signal is also applied to a phase-shifting network to supply two equal-amplitude, 90°-out-of-phase signals. The two out-of-phase r-f and audio signals are applied to two balanced modulators. The outputs of the modulators provide two DSB signals, with the sidebands of each output 90°-out-of-phase. The out-of-phase signals are combined, the two upper sideband (USB) signals being 180° out of phase cancel each other. The two lower sideband (LSB) signals are in phase to add and produce a SSB signal. To reverse the output to an upper sideband signal the audio signals to the balanced modulators are reversed.

Single Sideband Reception



Communications receivers with stable oscillators and bandspread tuning can be used to receive SSB signals. The prime consideration is reinsertion of a carrier to beat with the incoming SSB signal for detection. This can be done by using the receiver bfo as a substitute carrier. The output level of the bfo is usually fixed at a preset level for detection of CW signals. Reception of SSB signals requires a higher level output than the average bfo can supply. To overcome this the level of the SSB signal at the detector is reduced by lowering the gain of the receiver. The incoming signal can be tuned to place the injected bfo frequency as close as possible to the carrier frequency. For example, assume the i-f signal to consist of the lower sideband; but the bfo calibration is defective. At the center position of the bfo tuning knob the output frequency is 456 kc instead of 455 kc (A). With a 1-kc modulating note as the incoming signal, it would produce a lower sideband frequency at 454 kc. Beating the 454-kc signal with the 456 kc bfo signal would provide a detected output signal of 2 kc. The speech frequencies of a gruff-voiced person sound high, providing a false high-pitched tone.

This situation can be corrected in two ways: retuning the local oscillator to shift the incoming signal frequency to a different value in the i-f stage; or retuning the bfo to place the reinserted carrier at the exact frequency. In retuning the local oscillator the need for good bandspread tuning becomes apparent when trying to tune to exact frequencies at the higher frequency ranges of the receiver. The slightest turn of the dial may tune over thousands of cycles. Receivers designed solely for SSB reception use special detector circuits. Typical is the balanced demodulator, similar in design and operation to the balanced modulator. A reinserted carrier is applied as one input, and the i-f signal as the other input, the difference signal is the detected output. In addition these receivers may use afc to control oscillator drift.

Portable radios are designed to operate from a battery power supply.

Many auto radios use a vibrator type power supply so that the 6- or 12-volt battery voltage can be stepped-up to higher plate and screen voltages.

The superregenerative detector has high sensitivity, but poor selectivity.

A transceiver is an electronic device that combines the functions of a receiver and a transmitter, with many tubes serving each function.

A communication receiver is generally a highly sensitive unit that can receive signals over many different frequency bands. It can usually receive both AM and CW signals through the use of a beat-frequency oscillator.

Dual conversion involves double heterodyning. It is used for high-frequency signals where a single heterodyning action is not sufficiently stable.

In multiband operation the basic circuitry remains the same for each band; however, the tuned circuits of each band are changed.

Crystal filters permit extremely sharp tuning, and removal of close-by interfering signals.

The Q multiplier uses controlled regeneration to increase circuit Q.

S meters are used to give an indication of signal strength.

The beat-frequency oscillator (bfo) is used in communication receivers for the reception of CW signals. It is a local oscillator that "beats" with the incoming CW to produce an audible signal.

Noise limiters are used to remove sudden bursts of noise from the signal.

Squelch circuits automatically cut off undesired audio output.

Single sideband transmission involves the partial or complete elimination of all components on one sideband of a carrier of an amplitude-modulated wave.

A balanced modulator has two inputs and one output. There is no output unless both input signals are applied together.

The two most popular methods of suppressing a sideband are through filtering and phasing.

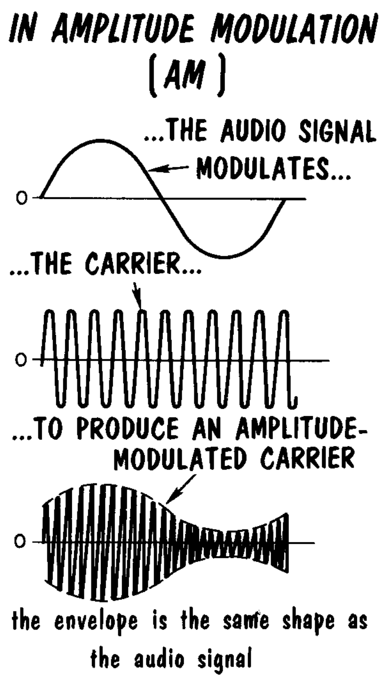
Bandspreading is a means of spreading stations in a single band of frequencies over an entire tuning dial.

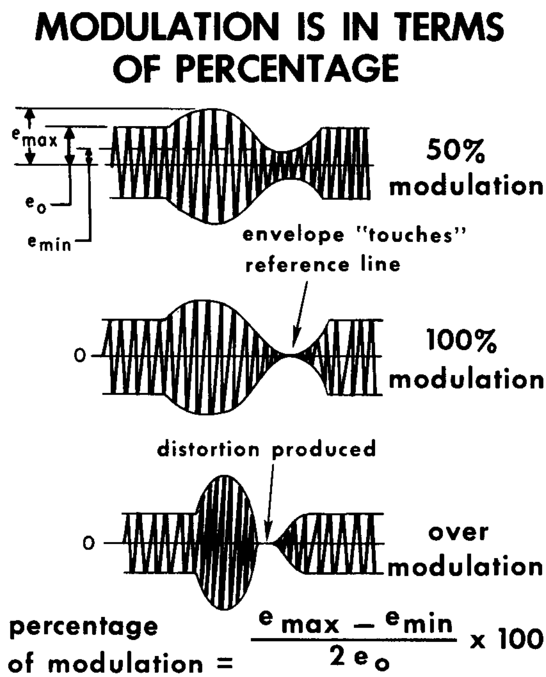
REVIEW QUESTIONS

1. Describe the operation of a superregenerative detector.
2. What is a transceiver, and what is its advantage?
3. Give some important characteristics of a communications receiver.
4. Explain the operation of an S meter.
5. How does a beat-frequency oscillator permit the reception of a CW signal?
6. Explain the operation of a noise limiter.
7. How does a squelch circuit automatically remove undesired signals?
8. Give two advantages of single-sideband reception.
9. Explain the operation of the balanced modulator.
10. How are sidebands removed in single-sideband operation?
11. What is meant by the expression "5 db above S9"?
12. What advantage is obtained by using a Q multiplier?

Review of Amplitude Modulation

Before beginning our study of frequency modulation (FM) let us review what we have learned about amplitude modulation (AM) so that it can serve as a basis of comparison. Modulation is the means by which intelligence is superimposed on the CW carrier wave. In amplitude modulation the modulating signal is used to vary the amplitude of the carrier wave. If we observe the envelope of the carrier, we see that it is an exact replica of the audio or



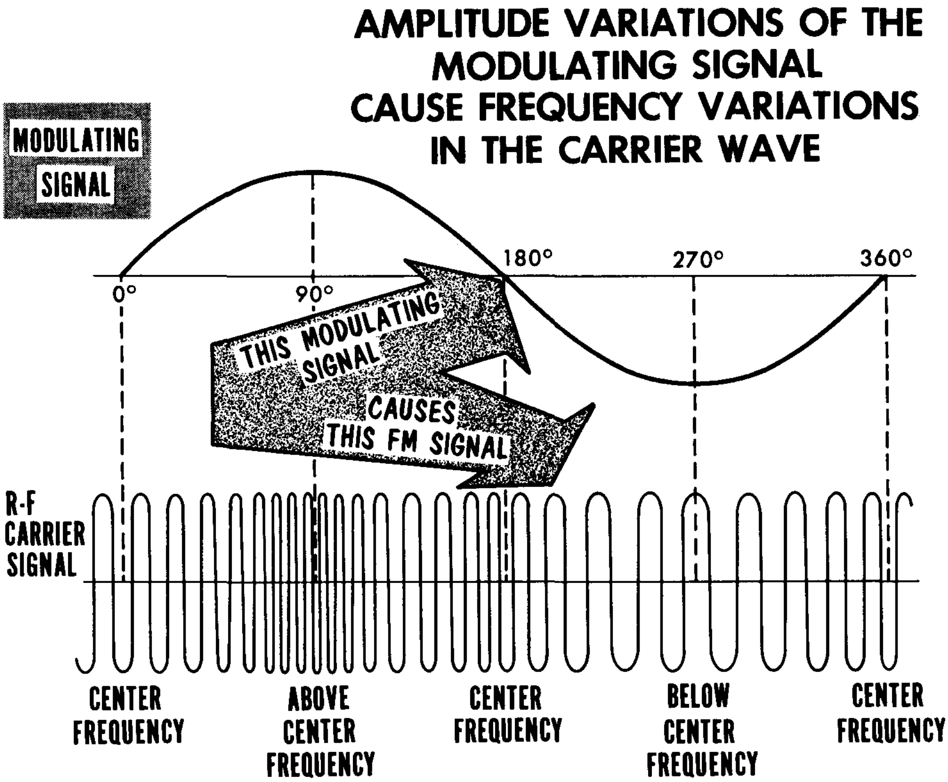


modulating frequency. The larger the audio signal, the greater will be the amplitude fluctuations in the carrier wave; the higher the frequency of the audio signal, the greater will be the rate of variation of the amplitude of the carrier wave.

When the carrier wave is modulated, sidebands are produced. If a 1000 kc carrier is modulated with a 5-kc audio signal, sidebands of 995 kc and 1005 kc are produced. Thus, the intelligence is actually contained in the sidebands of the radiated radio signal. The standard broadcast band is 10-kc wide, permitting modulating frequencies up to 5 kc. In certain instances of "clear channels", the FCC permits modulation of up to 7.5 kc. Modulation is measured in terms of "percentage", from 0 to 100%. Modulation in excess of 100% produces severe distortion.

A severe disadvantage of AM is its susceptibility to noise such as static. Static amplitude modulates a wave, and rides through as interference superimposed on the desired audio signal.

Frequency Modulation



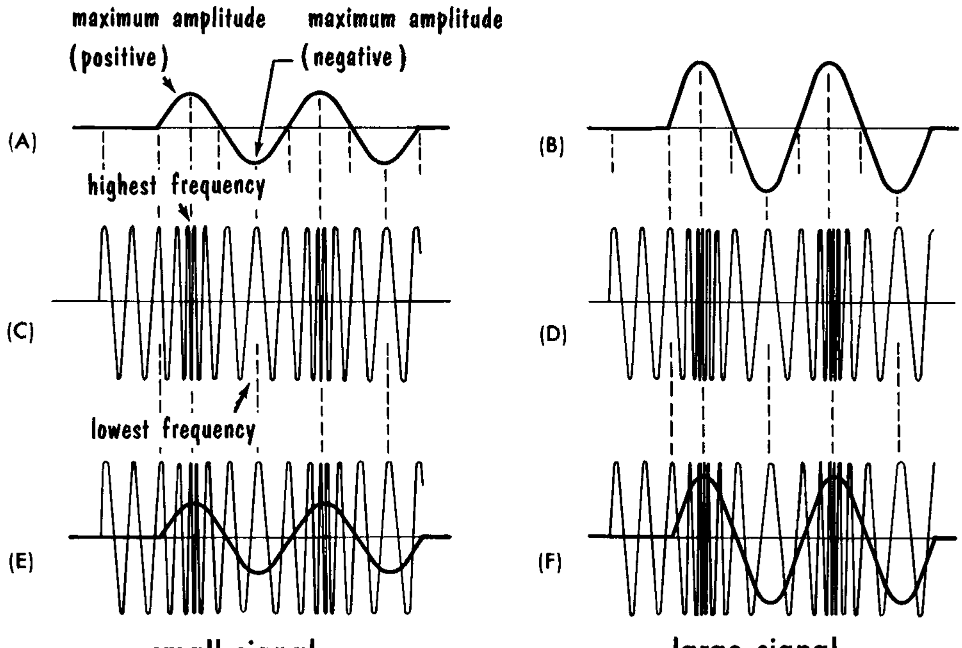
In the process of frequency modulation (FM) intelligence is superimposed on the carrier by varying the frequency of the carrier instead of the amplitude. A frequency-modulated carrier contains no variations in amplitude. All information is represented by changes in the carrier frequency. We can see an immediate advantage in FM – since noise or static amplitude modulates a signal, noise modulation will not affect the intelligence on the carrier. As we will study later, FM demodulators respond only to frequency variations, and not to amplitude variations. Thus, noise will not affect the fidelity of the FM signal.

Let us examine the FM signal. As the modulating signal increases from zero to its maximum positive value at 90°, the carrier increases from its center or resting frequency to maximum. As the value of the modulating signal is reduced to zero at 180°, the carrier frequency returns to its center or resting value. When the modulating signal varies from zero to its negative peak at 270°, the carrier frequency changes to its lowest value. Finally, when the modulating cycle is complete at 360°, the carrier is back to its center frequency. We can thus see the following basic condition: with no modulating signal, the carrier wave is at a fixed center frequency. As modulation is applied, the carrier frequency will increase and decrease in accordance with positive and negative alternations in the a-f modulation frequency.

Frequency Modulation – Effect of Audio Amplitude

We have just learned that the frequency modulated wave is at its highest and lowest frequency when the audio modulating signal is at its maximum positive and negative points. We see then, the first important characteristic of an FM wave – its frequency varies in accordance with the amplitude of the modulating signal. The larger the modulating signal, the greater will be the frequency shift or deviation from the center or resting frequency. For instance, if the center frequency of an FM signal is 90 mc, a weak audio signal may cause a frequency swing of plus and minus 10 kc, or a range of from 90.01 mc down to 89.99 mc. A strong audio signal may cause a frequency swing of

FREQUENCY SWING IS IN PROPORTION TO AUDIO AMPLITUDE



—small signal—
(small frequency swing)

—large signal—
(large frequency swing)

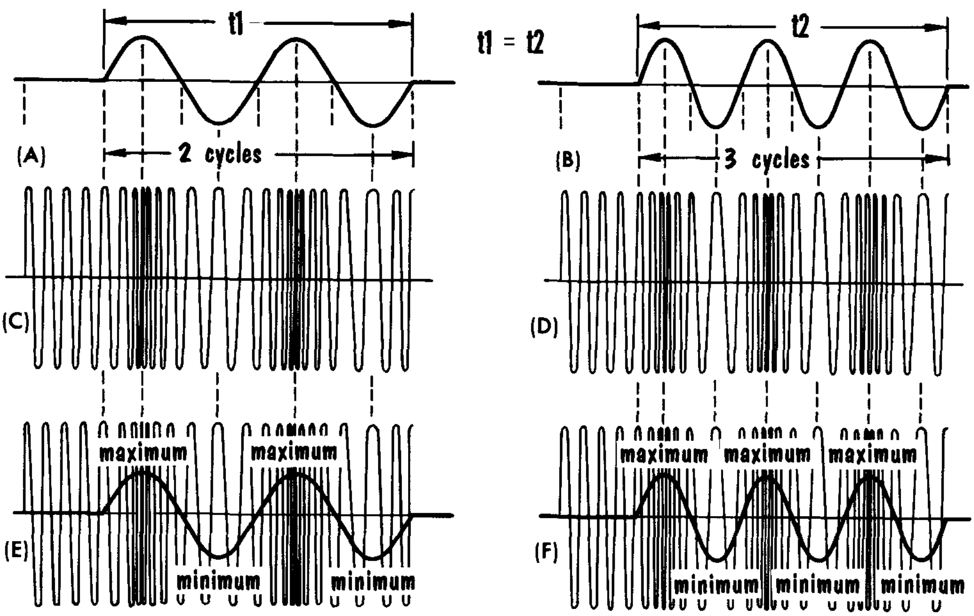
plus and minus 50 kc, or a range of from 90.05 mc down to 89.95 mc. Note that the amplitude of the audio signal has no effect on the amplitude of the carrier; it affects only the amount of frequency shift away from the center or carrier frequency.

The modulating signals (A) and (B) are the same frequency, but the amplitude (A) is less than (B). The louder signal (B) causes a greater frequency change in the modulated signal (D), shown by increased bunching and spreading, than does the smaller amplitude (A) on the modulated signal at (C). In (E) and (F) the modulating waves are superimposed on their respective modulated waves.

Frequency Modulation – Effect of Audio Frequency

The amount of frequency deviation during modulation is determined by the amplitude (loudness) of the audio signal. The audio signal frequency establishes still another characteristic of the FM signal – the rate at which the frequency deviation takes place. For example, if a 1-kc signal is used to modulate a carrier wave, the carrier will swing to its maximum upper and lower limits 1000 times per second. We have thus established two important

RATE OF FREQUENCY SWING IS DETERMINED BY MODULATING FREQUENCY



low frequency
(low rate of deviation)

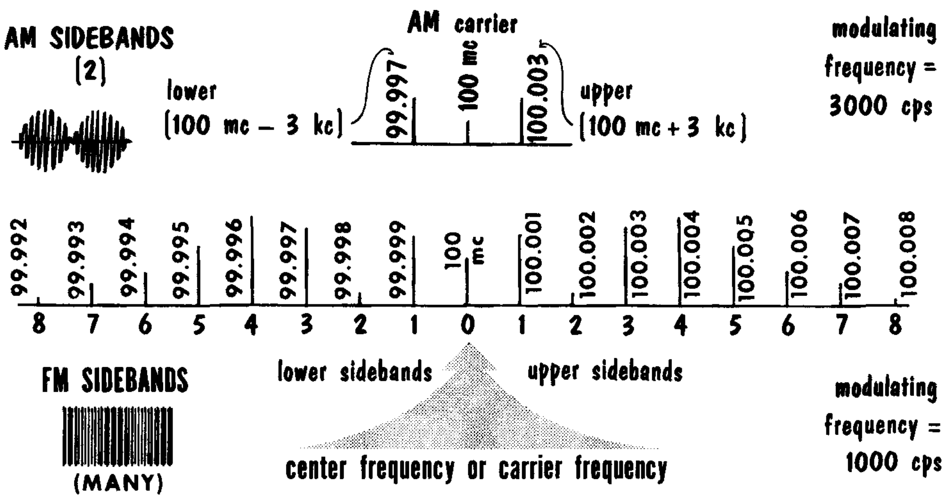
high frequency
(high rate of deviation)

points in the nature of the FM wave: the amount of frequency deviation indicates the amplitude of the audio signal, and the rate of the frequency deviation indicates the frequency of the audio signal.

In the illustration, the frequency of the modulating wave at A is less than that of B, the time intervals t1 and t2 being equal. Note the two groups of frequency changes in the modulated wave at C and the three in the wave at D, which occur in the same period of time. The modulating waves are superimposed on their respective modulated carriers in E and F.

FM Sidebands

In AM broadcasting, the limit of modulation is restricted by the value of 100% modulation. In FM, there is no limit. A hypothetical case might be that of an FM carrier of 100 mc. A modulating signal of sufficient strength could cause the carrier to shift 10 mc either side of the center frequency. In practice this would not be done, due to the possibility of signal interference. The FCC has laid down rules regarding the carrier shift. The maximum

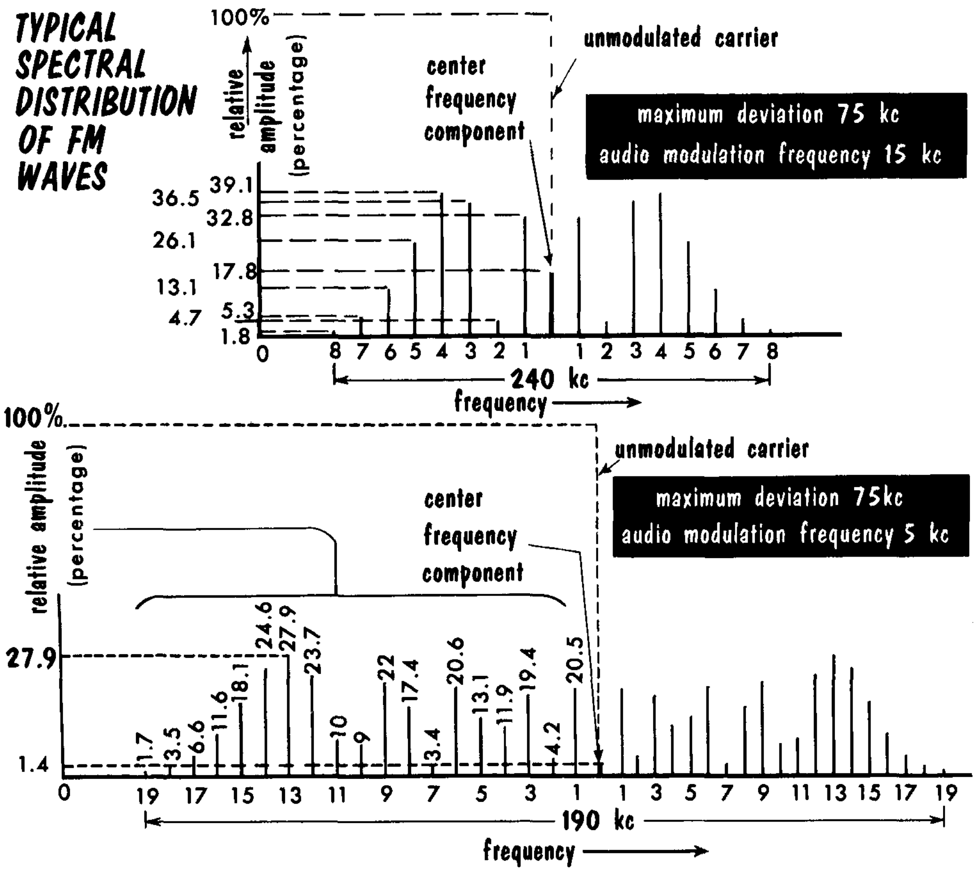


carrier shift should not exceed ± 75 kc from the center or assigned frequency. In addition, the FCC placed a 25-kc guardband on either side of the carrier's center frequency. Thus, 75 kc either side of the normal frequency equals 150 kc, plus a 25-kc guardband either side, allows a total of 200 kc for one FM channel.

In FM broadcasting, a modulating signal causes the carrier frequency to shift above and below its center frequency. While the carrier is shifting, sidebands are formed. These sidebands are unlimited in number and stretch out on either side of the carrier's center frequency. Note that some of the sideband components have larger amplitudes than the carrier frequency and other sideband components, indicating that most of the power of an FM signal is in the sidebands. Although the sideband components are unlimited in number, only 8 on either side of the carrier frequency are shown. This is due to the fact that in this case all the sidebands from the 9th on are so weak as to be totally ignored. All these sidebands are generated from one value of modulating frequency. The distance between each sideband is equal to the frequency of the modulating signal. An example would be a 1000-cycle modulating frequency. Assuming a center carrier frequency of 100 mc, there would be sidebands 1 kc apart above and below the center carrier frequency. These sideband frequencies would be at 100.001 mc, 100.002 mc, etc., up to 100.008 mc. In addition to the sidebands above, there are sidebands below the normal carrier frequency at 99.999 mc, 99.998 mc, etc., down to 99.992 mc.

FM Sidebands (Cont'd)

In FM, when a carrier is modulated, a number of sidebands form. The number of sidebands depends upon both the amplitude and the frequency of the modulating signal. The number of sideband pairs increases as the amplitude of the modulating signal increases. The number also increases as the frequency of the modulating signal decreases. At first glance, the amplitudes of successive sideband pairs contained in an FM wave seem to vary at random. Actually, these amplitudes are explained by the Bessel functions of higher mathematics. While the amplitude of many sidebands close to the carrier



may be less than that of sidebands farther away from the carrier, at a point sufficiently distant from the carrier the amplitude of sidebands diminishes and effectively disappears. The effective edge of the bandwidth of an FM signal is marked by the last sideband pair, the amplitude of which exceeds 1% of the amplitude of the unmodulated carrier. A detailed sideband distribution is shown where the modulating frequency is 15 kc. With 8 pairs of sidebands the total bandwidth is 240 kc. Using a lower modulating frequency of 5 kc, we have 19 pairs of sidebands with a total bandwidth of 190 kc.

Modulation Index

FM MODULATION INDEX - SIDEBAND PAIRS-AND BANDWIDTH

MODULATION INDEX	NUMBER OF SIDEBANDS EITHER SIDE OF CARRIER	BANDWIDTH (F-FREQUENCY OF MODULATING SIGNAL)
.5	2	4×F
1.0	3	6×F
2.0	4	8×F
3.0	6	12×F
4.0	7	14×F
5.0	8	16×F
6.0	9	18×F
7.0	11	22×F
8.0	12	24×F
9.0	13	26×F
10.0	14	28×F
11.0	15	30×F
12.0	16	32×F
13.0	17	34×F
14.0	18	36×F
15.0	19	38×F

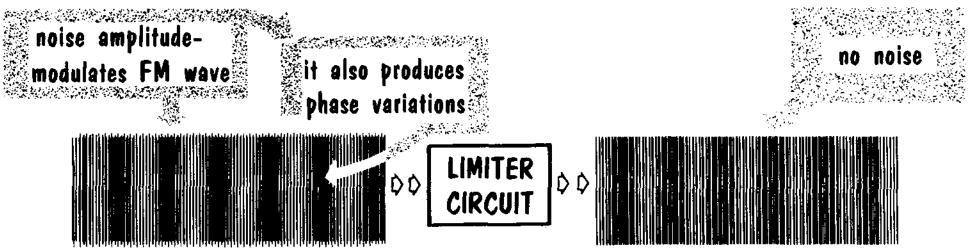
Modulation in FM is generally expressed in terms of the modulation index. The modulation index is the ratio of the frequency deviation to the modulation frequency. Thus, if the deviation is 10 kc when modulated by a 5-kc signal, the modulation index is 10/5 or 2. To see how the modulation index is related to the number of effective sidebands, look at the chart. Note that the number of effective sideband pairs increases as the modulation index increases. Since the modulation index is the ratio of deviation to modulation frequency, it can be increased either by increasing the deviation or by decreasing the modulation frequency. If the deviation is increased, the bandwidth increases and the number of effective sideband pairs increases. If the modulation frequency is decreased, the bandwidth decreases, while the number of effective sideband pairs increases.

Since a limit is placed on the amount of deviation and maximum modulation frequency, there is always a corresponding maximum value of the modulation index for the maximum modulation frequency. This is called the deviation ratio. For example, in commercial broadcasting, the maximum deviation of 75 kc divided by the maximum modulation frequency of 15 kc results in a corresponding modulation index (deviation ratio) of 5. However, a lower amplitude of a 15-kc modulation signal will produce a deviation less than 75 kc and a correspondingly smaller modulation index, while a 75-kc deviation produced by a modulation whose frequency is less than 15 kc results in a correspondingly larger modulation index.

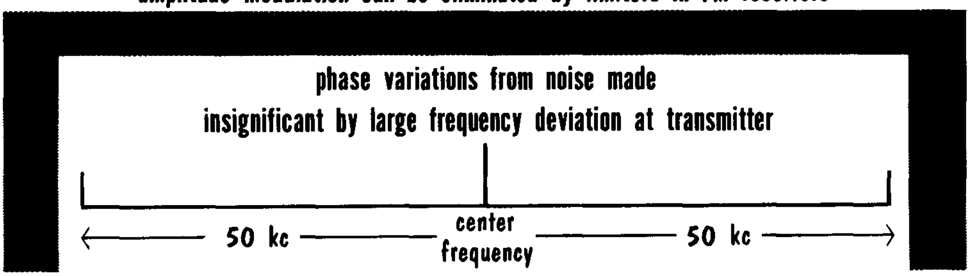
FM Noise and Interference

One of the features of frequency modulation is its ability to provide comparatively noise-free communication. While some noise can be avoided by operating at higher radio frequencies, this is not an inherent benefit of FM alone. Of greater importance is the characteristic of noise – it amplitude modulates the FM signal. This in itself is no problem since a special FM receiver circuit called the limiter removes amplitude variations from the FM signal. In addition, demodulator circuits such as the ratio detector are relatively insensitive to amplitude variations and would not pass them on, thus eliminating noise at the loudspeaker. Unfortunately, however, noise also produces phase variations in an FM signal which effectively represents changes in modulation.

These phase variations can be made ineffective by using a large frequency deviation at the transmitter (the modulation index of the signal should be high at full modulation). If the signal voltage is larger than the noise voltage, the modulation index produced by the noise voltage will be quite small, perhaps less than 1. Thus it is much smaller than the modulation index of the signal



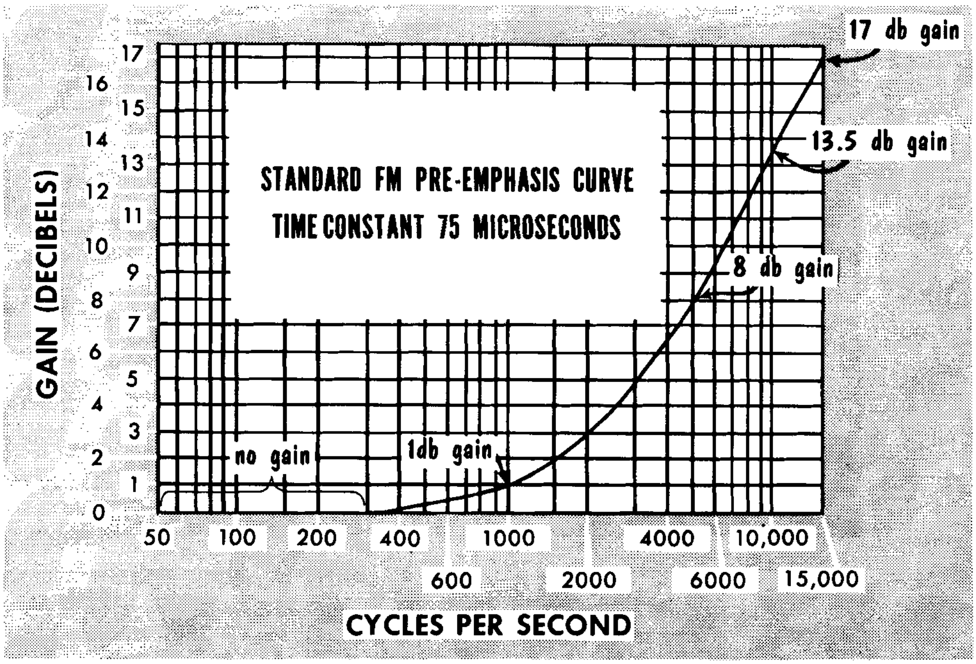
amplitude modulation can be eliminated by limiters in FM receivers



when the frequency deviation is large. For example, if a 500-cycle modulating signal produces a 50-kc swing, the signal modulation index is 50,000/500, or 100. A noise voltage as much as half as large as the signal would produce a phase deviation only .005 times as great as the swing produced by the signal voltage. Thus the noise voltage which is only slightly weaker than the signal voltage will be almost completely suppressed. The larger the frequency deviation of the FM carrier, the greater will the noise be suppressed. We see also that as long as the signal voltage is greater than the noise voltage, the signal-to-noise ratio of a wideband FM system is superior to that of a narrow-band system.

Pre-Emphasis

In ordinary speech and music, the higher audio frequencies are relatively weak and thus produce small frequency deviations. This in turn further reduces the modulation index of the desired signal at higher modulation frequencies. To overcome this, the amplitude of the higher modulation frequencies are increased before the FM transmitter is modulated. This process is called pre-emphasis. Basically, pre-emphasis involves increasing the



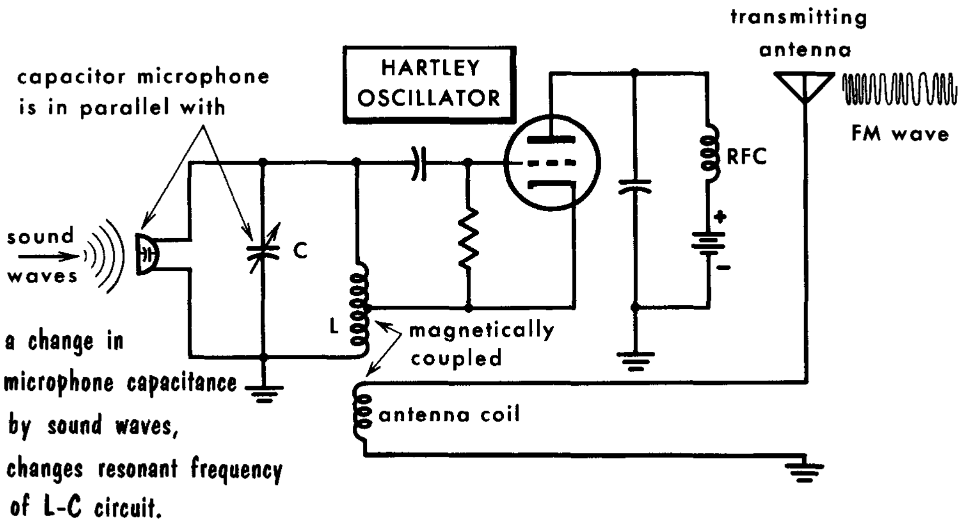
relative strength of the high-frequency components of the audio signal before it is passed into the modulator. In this way, the undesirable relationship between the high-frequency program material and the high-frequency noise is changed, because while the audio is increased, the noise remains the same. Thus, the high-frequency signal-to-noise ratio is increased. Of course, in the process, a defect is introduced—the natural balance between the high- and low-frequency tones in the program material is upset.

This defect is compensated for at the receiver by the de-emphasis network at the input to the audio amplifier. The de-emphasis circuit reduces the high-frequency audio signal exactly as the pre-emphasis increased it. However, it operates on both the high-frequency program material and the high-frequency noise. Thus, it does not change the improved high-frequency audio signal-to-noise ratio obtained by pre-emphasis. It does, however, re-establish the tonal balance of the program material lost in pre-emphasis. In FM broadcasting, the higher frequencies are emphasized in accordance with an R-L circuit having a time constant of 75 μsec.

The Basic FM Transmitter

Before beginning our study of FM receivers, we can get a better concept of frequency modulation by observing a basic FM transmitter. The simplest form of frequency modulation is that of a capacitor microphone, which shunts the oscillator-tank circuit LC, as shown. The capacitor microphone is equivalent to an air-dielectric capacitor, one plate of which forms the diaphragm of the microphone. Sound waves striking the diaphragm compress

A FREQUENCY MODULATED WAVE CAN BE PRODUCED BY VARYING THE RESONANT FREQUENCY OF AN OSCILLATOR TANK CIRCUIT.

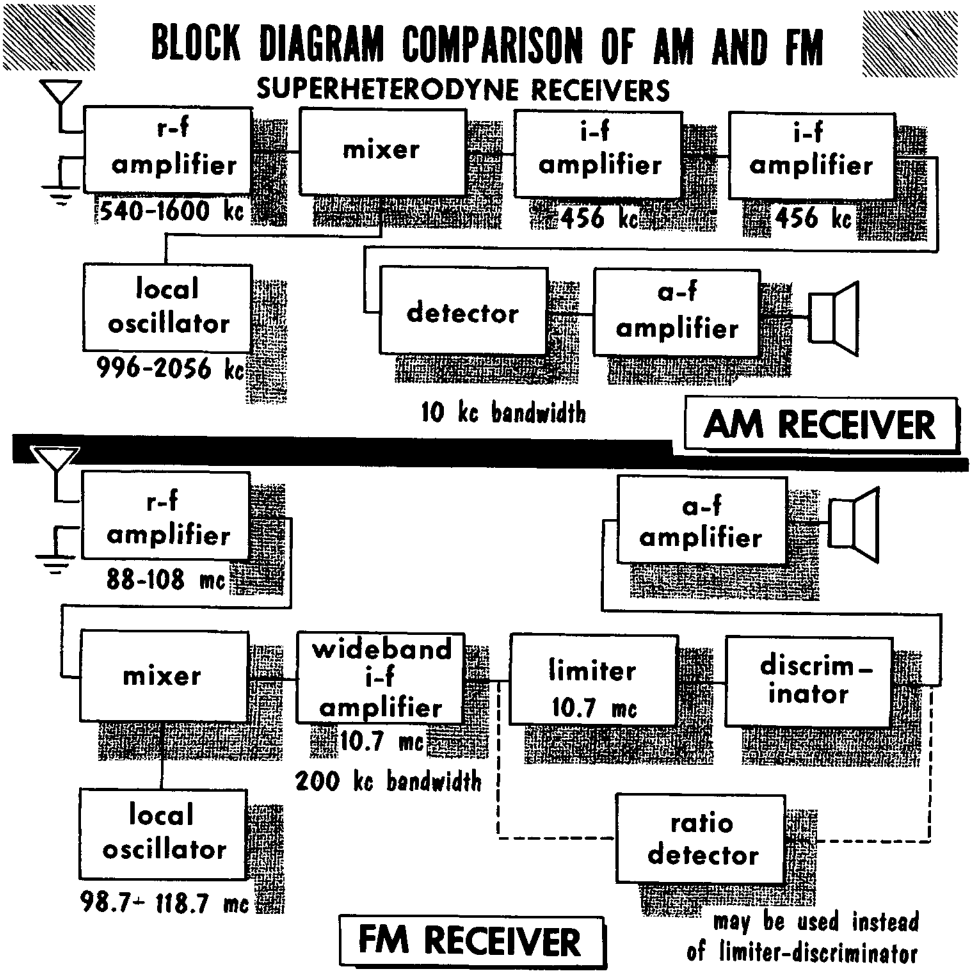


and release it, thus causing the capacitance to vary in accordance with the spacing between the plates. This type of transmitter is not practicable (among other reasons, the frequency deviation is very limited), but it is useful in explaining the principles of frequency modulation. The oscillator frequency depends on the inductance and capacitance of the tank circuit LC and, therefore, varies in accordance with the changing capacitance of the capacitor microphone.

If the sound waves vibrate the microphone diaphragm at a low frequency, the oscillator frequency is changed only a few times per second. If the sound frequency is higher, the oscillator frequency is changed more times per second. When the sound waves have low amplitude, the extent of the oscillator frequency change from the no-signal, or resting, frequency is small. A loud a-f signal changes the capacitance a greater amount and, therefore, deviates the oscillator frequency to a greater degree. Thus, the deviation frequency of the oscillator tank depends upon the amplitude of the modulating signal.

FM Receivers

The FM receiver is basically the same as the AM receiver we have studied. It is a superheterodyne type and operates commercially on a band of 88 – 108 mc. Because of these relatively high radio frequencies, there are some differences in the antenna, r-f amplifier, and mixer circuits. These are characteristic of high radio frequencies rather than the frequency modulated

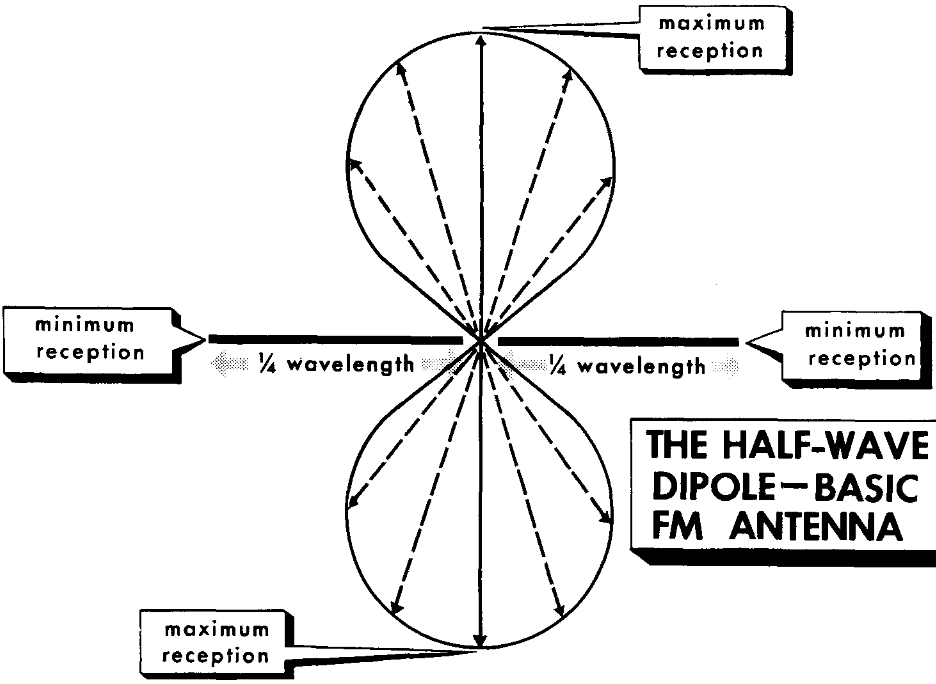


signal. R-f and i-f circuitry is designed for a broad bandpass compared to the relatively narrow bandpass used by AM receivers. The principal difference lies in the demodulator. Because of the nature of the modulation, special circuits are required to demodulate the FM signal. In general, these circuits are more involved than AM demodulators. The FM audio amplifier is like the AM except that it is more elaborate and designed to pass a wider range of audio frequencies. The reason for this is the higher modulating frequencies transmitted by FM stations.

FM Receiving Antennas

The AM broadcast band operates at frequencies of 535 – 1605 kc. In this range of frequencies, one wavelength is extremely long. For instance, at 1 mc or 1000 kc, one wavelength is equal to 300/1, or 300 meters (1 meter = 39.37 inches). At the FM band of 88 – 108 mc, the wavelength is much shorter. For example, at 100 mc, one wavelength is equal to 300/100, or 3 meters. This short wavelength permits us to use high-gain "tuned" antennas. Thus, the different antennas used for FM are not a result of the type of modulation used, but are types used for those high radio frequencies.

The type of antennas used for broadcast AM reception are not critical. AM stations usually produce a strong signal that can be received with most ungrounded conductors, including a short length of wire connected to the receiver's antenna terminal, or even a bedspring. FM stations usually operate

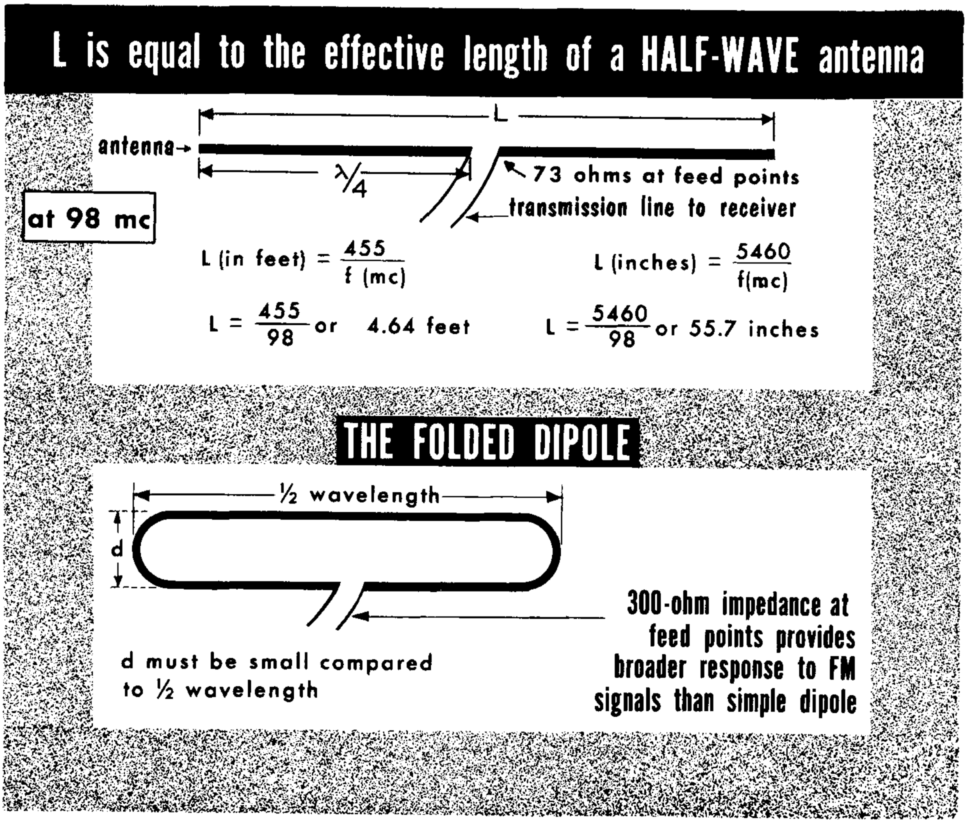


THE HALF-WAVE DIPOLE—BASIC FM ANTENNA

at lower power than AM stations, and reception is frequently limited to distances within sight of the transmitting antennas. (Antennas are discussed in detail in Volume 6.) The half-wave dipole antenna for the FM receiver, is tuned: it provides optimum reception at a frequency whose half-wavelength dimension equals the length of the antenna. The dipole antenna, is composed of two quarter-wavelength sections in an in-line arrangement, each quarter-wavelength section insulated from the others and from ground. The half-wave dipole antenna is highly directional, and its positioning is sometimes critical. Because the transmitting FM antenna is positioned horizontally, the receiving antenna should also be positioned horizontally for optimum reception.

Receiving Antenna Dimensions

The FM antenna provides maximum signal voltage to the receiver input. Unlike most broadcast AM receiver antennas, FM receiver antennas act as resonant lines having standing waves on them. FM antennas are cut to the required length to receive a signal of sufficient strength to drive the r-f amplifier. If a single frequency is to be received, the antenna may be designed for maximum response at that frequency. If, however, a band of frequencies is to be received by one antenna, the antenna length must represent



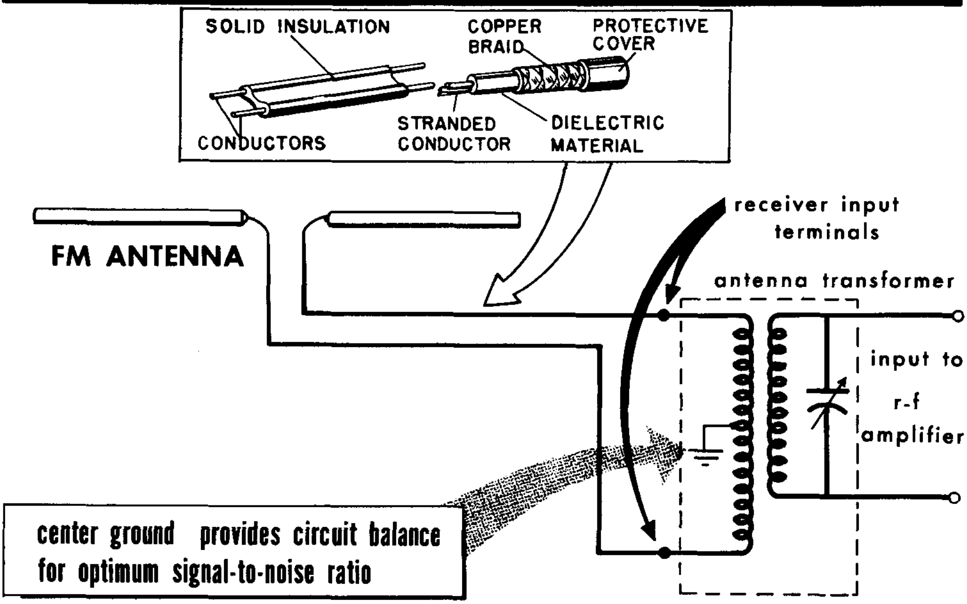
a compromise. Usually, the length is in resonance at the geometric center of the band. The geometric center, or mean, is equal to $\sqrt{\lambda_1 \lambda_2}$, where λ_1 and λ_2 are wavelengths at the two ends of the band. Most FM antennas are simply cut for the midfrequency of 98 mc.

The resistance of a half-wave dipole at its center or feed points, is about 72 ohms. The transmission line connecting the antenna with the receiver should also have a characteristic impedance of 72 ohms in order to operate as a non-resonant transmission line with no standing waves. The transmission feeds the signal to the receiver via a matching or antenna transformer at the input to the r-f amplifier.

Transmission Lines

As with antennas, the transmission line used for AM broadcast reception is relatively simple often consisting of nothing more than a length of wire. At frequencies used for broadcasting FM, special transmission lines are used to conduct a signal from the antenna into the receiver input terminals of the FM receiver. Many types of transmission lines are commercially available; two are used predominantly – twin lead parallel lines and flexible coaxial

TRANSMISSION LINES USED WITH FM ANTENNAS

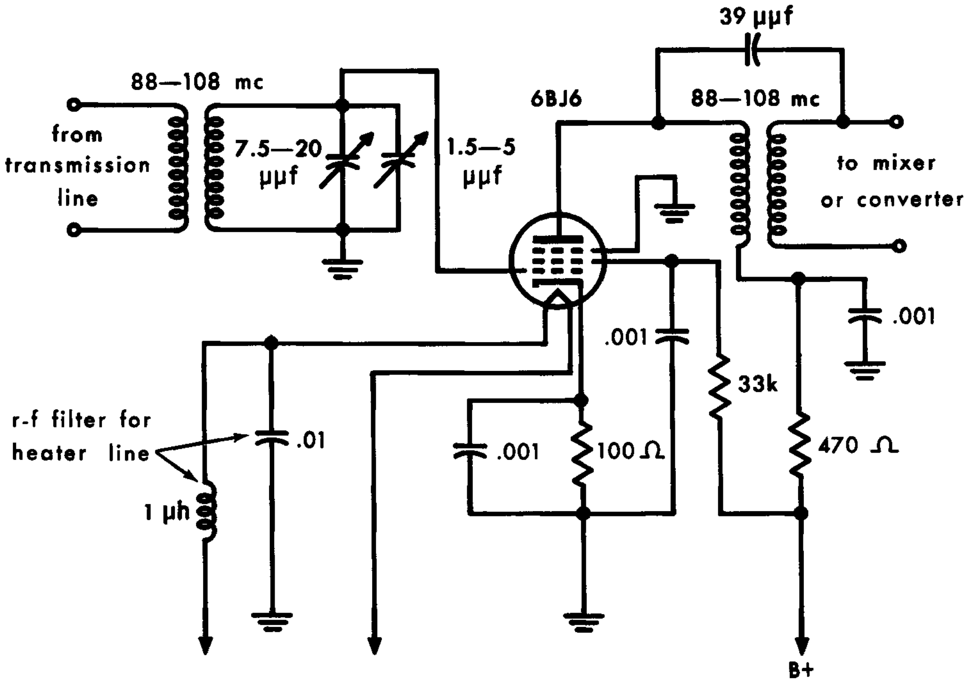


cable. As with antennas, transmission lines have their own characteristic impedance (see Volume 6). What is important at the moment is that we realize the characteristic impedance of the transmission line must match the characteristic impedance of the antenna for optimum FM reception; hence, the simple dipole antenna should use a 72-ohm cable. The folded dipole has a characteristic impedance of about 300 ohms and should be used with a 300-ohm transmission line. In areas where signal strength is very high, these considerations are not particularly critical. However, in fringe areas, even tiny signal losses in the transmission line may reduce the signal-to-noise ratio to a point where the reception would be poor.

Transmission lines connect to the antenna coil in the receiver input which is usually designed for 300-ohm or 72-ohm impedance. The receiver input impedance is usually stated on the cabinet or chassis of an FM receiver. For optimum reception, the antenna, transmission line, and receiver input impedances should be the same. Special impedance-matching devices are available where it is necessary to connect a 300-ohm device to a 72-ohm device.

R-F Amplifier Circuits

The typical tuned r-f circuit used at FM frequencies resembles an AM circuit in the schematic, but there are major differences. There are fewer turns in the coil; the primary may have only one turn, 3/4-inch in diameter, and the secondary only two and one-half turns. The tuning capacitor's value may be only 7.5–20 μμf, the trimmer's 1.5–5 μμf. The frequencies involved re-



FM R-F CIRCUIT

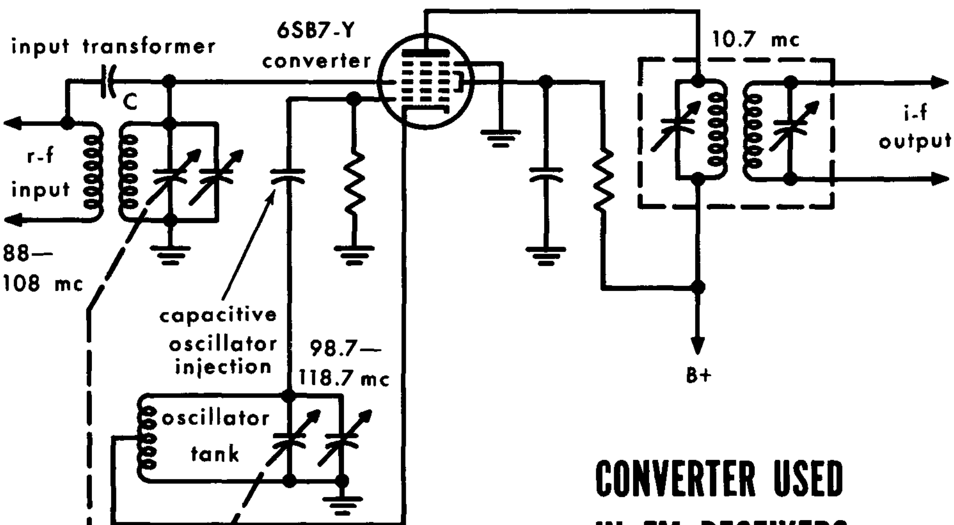
quire small values of inductance and capacitance, bringing about serious problems of stray inductance and capacitance. Layout and wiring of FM receiver circuits must be done with care. As in an AM receiver, an antenna transformer is used to couple the antenna signal to the grid circuit.

Relatively wide bandwidth response is essential due to the wide range of frequencies, (a 200 kc bandwidth for each channel). The typical high-gain, high-Q circuit used in an AM receiver cannot be used for FM. A low-gain, low-Q, wide-bandpass circuit must be used.

Often a tuned r-f stage ahead of the converter in an AM receiver is not needed due to the strong signals used to reduce interference in AM broadcasting. Most broadcast AM receivers use only the tuned r-f circuit at the r-f grid of the converter. Tuned r-f amplifier stages in FM receivers are desirable because of the weaker signals in FM broadcasting. These stages improve the signal-to-noise ratio, sensitivity, selectivity and image rejection.

Mixer and Converter Circuits

The mixer-oscillator and converter circuits found in FM sets are similar to those used in AM receivers. The purpose of heterodyning to produce an intermediate frequency is the same in FM as in AM. However, again due to the frequencies used in FM, certain circuit modifications must be made to produce the proper i-f signal. A typical converter system is shown. The type 6SB7-Y tube used in this circuit is especially designed for high-frequency work. The circuit is more or less the same as that of an AM converter system. For instance, the first grid is used as the oscillator grid in a Hartley oscillator circuit. The third grid is the r-f signal input grid and the r-f signal input is either fed directly from the antenna circuit or from a preselector stage. There are, however, certain circuit changes or additions made in these circuits which stem from the high frequencies involved. The coils contain fewer windings because of the need for smaller values of inductance, but other changes are also noticeable. For example, the i-f transformer has a broad frequency response characteristic; at the high end of the band the re-



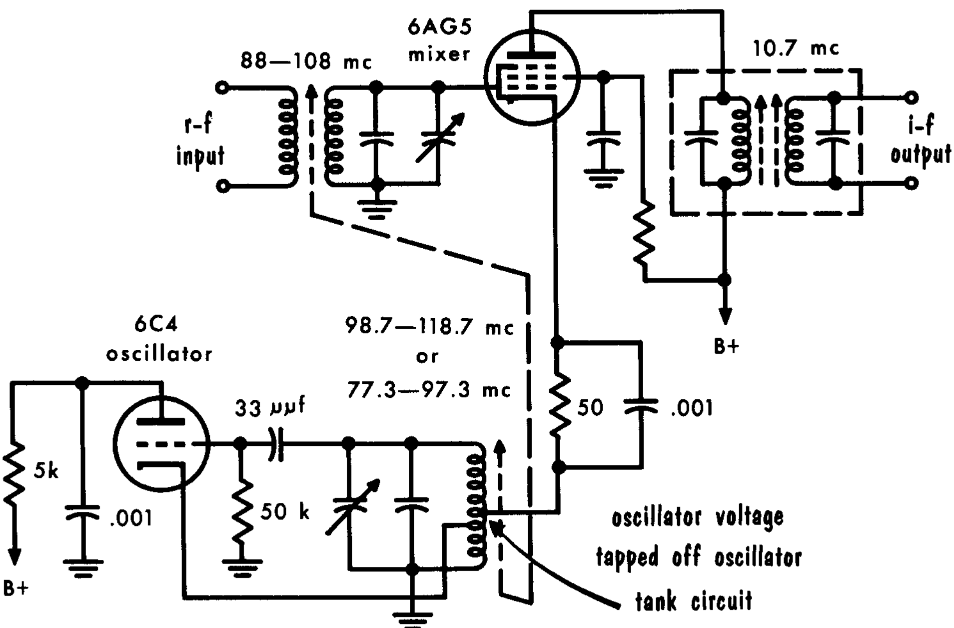
CONVERTER USED IN FM RECEIVERS

sponse (gain) falls off. There is a loss of amplification at these frequencies. Capacitor C is inserted between the high side of the primary and secondary of the input transformer to increase the coupling at the higher frequencies and thus improve the amplification at these frequencies. As the frequencies are increased the capacitor offers less reactance and consequently the high-frequency currents take this reactive path because it offers a lower impedance than the transformer. Thus the overall frequency response is equalized. Such capacitors sometimes are used in the shortwave band of AM receivers.

Mixer and Converter Circuits (Cont'd)

At the frequencies employed in the commercial FM band, the stability of the local oscillator becomes a major problem. The local oscillator tends to become synchronized with the incoming signals, causing a loss of the i-f output entirely. For maximum frequency stability, a separate oscillator tube is used. Even in a normal, well designed FM receiver changes in internal capacitances of the oscillator tube (or oscillator section of a tube), or the expansion of coil windings and capacitor plates during warmup may cause the local oscillator frequency and consequently the i-f to drift considerably. A small shift in oscillator frequency may shift the i-f signal beyond the range of the i-f stages with a loss in output signal. Various methods are used to combat oscillator drift. For example, the second harmonic of the local oscillator

FM MIXER-OSCILLATOR CIRCUIT



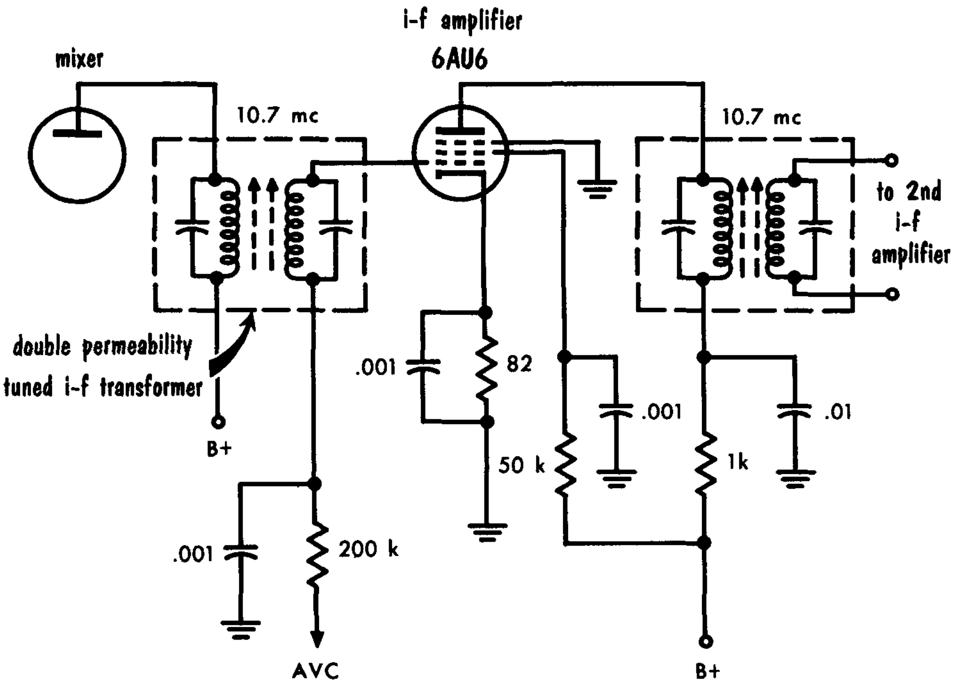
frequency is sometimes used for mixing. In this case, the local oscillator may be operated at a lower fundamental frequency where frequency stability is improved. Another method is to use capacitors having a negative temperature coefficient. These are connected in shunt with capacitors having a positive temperature coefficient, to counteract the change in capacitance when the temperature of the oscillator stage varies. Frequency stability of the local oscillator in the standard FM band makes it advantageous to operate the local oscillator at a frequency below that of the incoming signal. However, if the local oscillator is operated above the frequency of the incoming signal, it is not as likely to interfere with television receivers in the same vicinity that are operating on the lower TV channels. Therefore, FM receivers will be found operating both below and above the incoming signal frequency.

FM I-F Amplifiers

I-f amplifiers in FM receivers (as in AM receivers) provide a great deal of the receiver sensitivity and selectivity. The i-f amplifier gain in an FM receiver is relatively low. One reason is the use of a 10.7-mc i-f. At this high frequency it is difficult to get the gain per stage possible at 455 kc. A bandpass of about 200 kc is required as compared to some 10 kc in AM i-f amplifiers. This is achieved by using broadly tuned low-Q circuits, making for low gain per stage. As a result of this, FM receivers generally have more stages of i-f amplification than AM receivers.

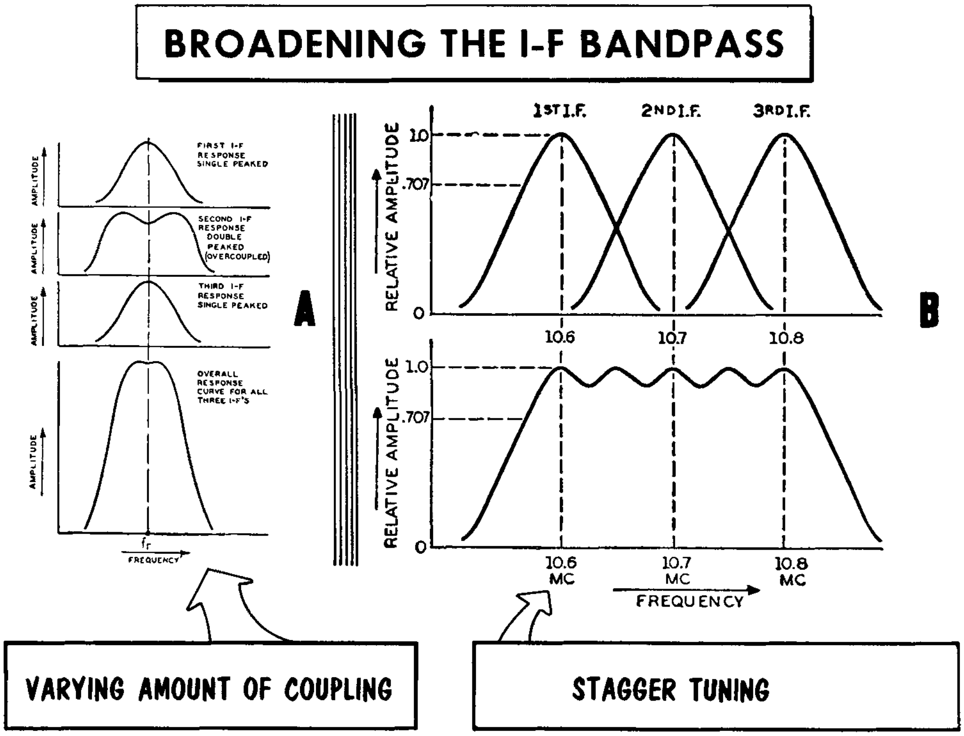
The i-f amplifier serves the function of increasing the signal strength of the FM signal to a level where it can be applied to the demodulator for removal of the audio component. It generally employs double-tuned transformers having

TYPICAL FM 10.7-MC I-F AMPLIFIER



equal primary and secondary inductances. As in the case of AM i-f amplifiers, the transformers may be either capacitively or permeability tuned. A low value of i-f is undesirable because local oscillator drift might force the receiver to operate outside the i-f range. The frequency of 10.7 mc is highly desirable as an i-f because it is beyond the range of image frequency interference. (Twice its value is equal to 21.4 mc, which is outside the 20-mc width of the FM band.)

FM I-F Amplifiers (Cont'd)



Relatively high i-f gain and broad selectivity can be obtained in a number of ways. In most of these, three i-f transformer networks are used. The number of i-f stages is a determining factor in the amount of i-f gain, but the type of coupling and Q of the i-f transformer circuits is the determining factor in the bandwidth. Three principal types of i-f transformer coupling arrangements are used. In the first, all three low-Q i-f transformers are single peaked, somewhat under critical coupling, to the same resonant frequency. In the second, the first and third i-f transformers are single peaked just under critical coupling, and the second i-f transformer is overcoupled to produce a double-peaked response curve. All three have the same resonant frequency. In the third, all three transformers are single peaked but the resonant frequency of each is slightly different. Usually the first transformer has the lowest resonant frequency, the third has the highest, and the second is between the other two. This system is known as "stagger tuning".

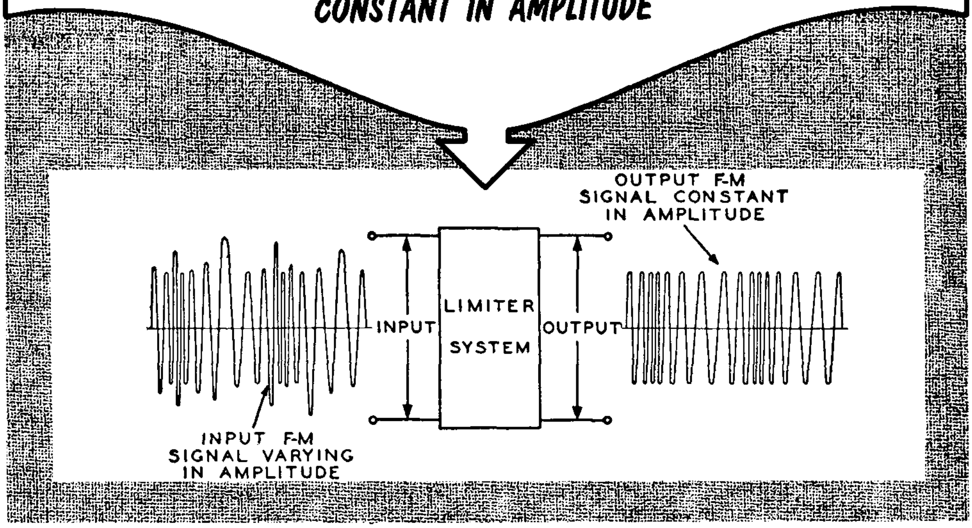
In illustration (A) we see that by using three i-f stages in which two are single peaked below critical coupling and the other double peaked (overcoupled), a broad, flat-top overall response curve can be obtained. The response for the complete i-f system is obtained by combining the individual curves of each stage. Illustration (B) shows three single-peaked i-f stages stagger-tuned. Their individual curves overlap to produce a broad bandwidth response.

Limiting

The limiter in an FM receiver removes amplitude modulation and passes to the discriminator an FM signal of constant amplitude. As the FM signal leaves the transmitting antenna it varies in frequency with an audio modulating signal, but it has essentially a constant amplitude. As the signal travels between the transmitting and receiving antennas, however, natural and man-made noises (static) disturbances, are combined with it to produce variations in the amplitude of the modulated signal. Other variations are caused by fading of the signal. Still other amplitude variations are introduced within the receiver itself due to a lack of uniform response of the tuned circuits.

All these undesirable variations in the amplitude of the FM signal are amplified as the signal passes through the successive stages of the receiver up to

THE LIMITER IN AN FM RECEIVER CLIPS OFF THE AMPLITUDE VARIATIONS OF THE INPUT SIGNAL, MAKING ITS OUTPUT CONSTANT IN AMPLITUDE



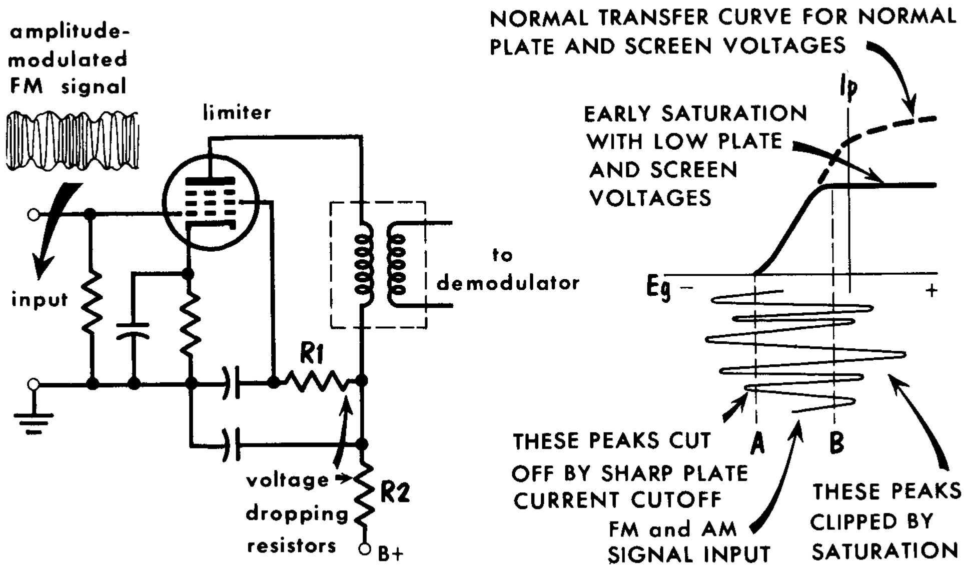
the input of the limiter. This condition in which both FM (desired) and AM (undesired) are present is shown in the illustration. The character of the signal after leaving the limiter should be such that all amplitude variations have been removed, leaving a signal that varies only in frequency.

The output of the limiter is fed into the FM demodulator. One of the demodulator circuits in particular, the ratio detector, is quite insensitive to amplitude variations, and when this circuit is used, it is unnecessary to use a limiter stage; instead, the output of the i-f amplifier is fed directly into the ratio detector circuit. Limiting action is obtained essentially by overdriving a tube, and "cutting off" the positive and negative peaks of the FM signal.

Plate Circuit Limiting

Limiting action may be produced by reducing the plate and screen voltages of a pentode. A sharp-cutoff pentode such as a 6SJ7 or a 6AC7 makes an excellent limiter for both positive and negative amplitude swings of the signal amplitude. The input voltage to the limiter comes from the i-f amplifier which precedes it. The circuit appears as a straightforward amplifier. Note re-

THE LIMITER REMOVES ALL AMPLITUDE MODULATION FROM THE FM SIGNAL

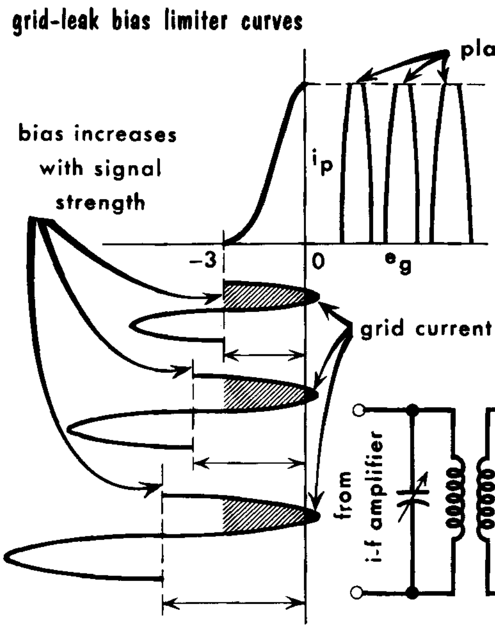


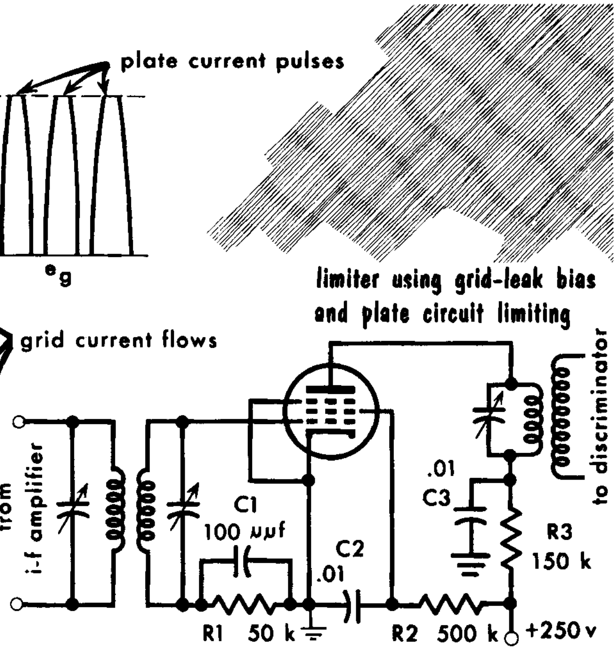
sistors R1 and R2. They are sufficiently high in value to drop both the screen and plate voltages to about 30 volts or less. When this is done, the grid transfer characteristic curve changes from its normal shape to the flattened one shown. We see that plate saturation occurs at much lower values of plate current than previously. This makes it possible for much smaller input voltages to bring out the saturated condition since the linear portion of the transfer curve is much shorter than it is in a normally operating amplifier.

The net effect of saturation is that if the signal swings go further negative than A and more positive than B, all the peaks outside these lines are clipped and have absolutely no effect on the output plate current of the tube. In the negative case these peaks lie outside the cutoff point; in the positive case the input peaks exceed the saturation voltage, and since maximum plate current has already been attained, it can cause no further change. Hence the positive and negative clipping results in an output voltage in which all amplitude changes have been removed, leaving only FM. A plate circuit limiter, although removing amplitude variations, contributes very little to the gain of the receiver because of the low voltages.

Grid-Leak Bias Limiting

The manner in which a grid-bias limiter functions is shown by the i_p-e_g curve. Grid-leak bias is used so that with varying signal amplitudes, the bias can adjust itself automatically to a value that allows just the positive peaks of the signal to drive the grid positive and cause grid current to flow. If a signal having a peak amplitude greater than the cutoff bias is impressed on the grid of the tube, a bias voltage having a magnitude approximately equal to the peak value will be developed. Accordingly, grid current will flow for a very small part of the positive half cycle at the peak of the signal swing. Plate current flows for almost the entire positive half cycle. When the signal amplitude increases, a greater bias is developed, but the grid cutoff voltage remains the same and the average plate current changes very





little. Thus, the amount of plate current flow in the limiter stage is approximately constant for all signals having an amplitude great enough to develop a grid-leak bias voltage that is greater than the cutoff voltage.

The frequency variations in the FM signal are maintained in the output because the plate current pulses are produced at the signal frequency and excite the plate-tuned tank circuit which has a relatively low Q and a wide bandpass. Because of the flywheel effect, a complete a-c waveform is passed to the secondary of the discriminator transformer for each input cycle. When the peak amplitude of the signal is less than cutoff voltage, the limiting action fails. Thus, the stages preceding the limiter must have sufficient gain to provide satisfactory limiting action on the weakest signal to be received.

In a frequency-modulated wave, the instantaneous frequency varies about the carrier frequency in proportion to the amplitude of the modulating signal.

The variations in instantaneous frequency are determined by the frequency of the modulating wave; the higher the modulating frequency, the greater the number of deviations in a given time period.

The amplitudes of the sidebands, as well as their frequencies, depend on the amplitude and frequency of the modulating signal and the frequency deviation of the transmitter.

For a sinusoidally-modulated signal, the sidebands are distributed in symmetrical pairs on either side of the carrier frequency at integral multiples of the modulation frequency.

The ratio of the frequency deviation to the frequency of the modulating signal is called the modulation index.

Frequency modulation is less susceptible to interference than AM because most interference is in the form of amplitude modulation, which can be eliminated without affecting the intelligence contained in the FM.

FM receivers are generally of the superheterodyne type.

The i-f amplifier provides all the selectivity and most of the sensitivity of the FM receiver. The selectivity depends on the tuned circuits used in the i-f voltage amplifiers.

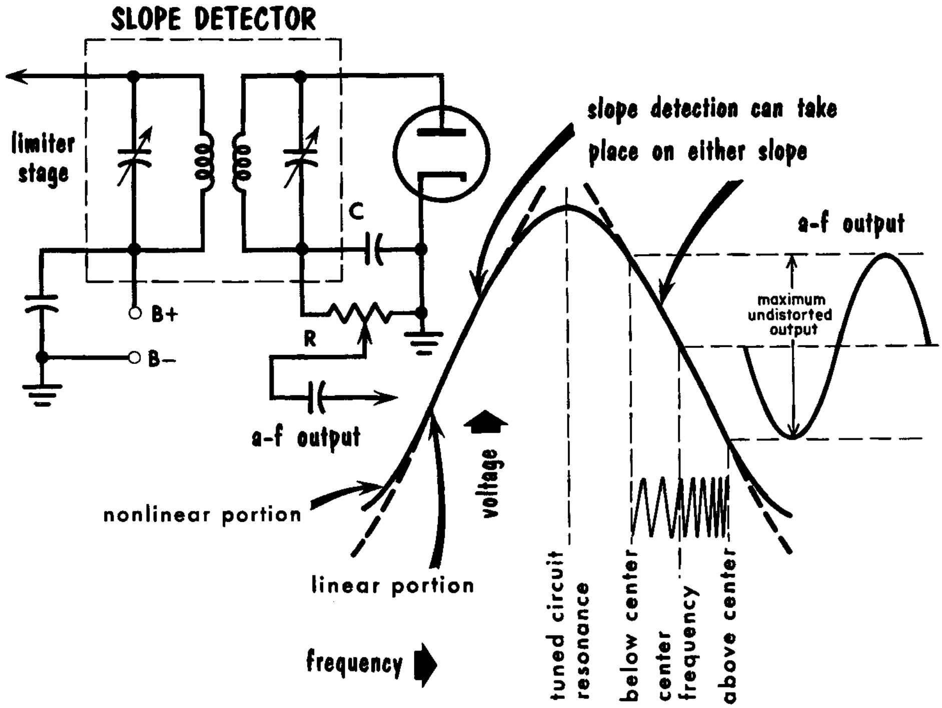
Transformer-coupled circuits are used in i-f amplifiers for high-gain, broad frequency response, and sharp adjacent-channel selectivity. The stability of the i-f amplifier depends on the shielding, parts placement, and amount of feedback through the grid-plate capacitance.

Limiter circuits are generally i-f amplifiers operated with low screen voltage so that they overload easily on strong input signals. They eliminate AM, and also some of the noise that appears along with the desired signal.

REVIEW QUESTIONS

1. What happens to the frequency deviation of an FM wave when the amplitude of the modulating signal is increased?
2. If the frequency of the modulating signal is doubled, what happens to the FM wave?
3. What effect does the amplitude of the modulating signal have on the FM signal?
4. What is meant by the modulation index?
5. What is the purpose of pre-emphasis?
6. What is the function of the limiter stage?
7. Explain briefly the operation of a limiter.
8. Describe the three principal types of i-f transformer-coupling arrangements?
9. What is a common intermediate frequency used in FM receivers?
10. What is the frequency range of the standard FM broadcast band?
11. What characteristics make frequency modulation "noise-free"?
12. What is the maximum FM carrier shift permitted by the FCC in commercial broadcasting?

FM Demodulators – The Slope Detector



Just as AM receivers translate amplitude variations into audio signals, FM receivers translate frequency variations into audio signals. One of the oldest and simplest FM demodulators is the slope detector. This circuit is based on the carrier frequency of the FM signal falling on the sloping side of the r-f response in an AM receiver. The frequency variations of the FM carrier signals are converted into amplitude variations. This conversion results from the unequal response above and below the carrier center frequency. Thus, when the incoming FM signal is less than the center frequency, the output voltage is in a positive direction; when the incoming signal swings above the center frequency, the output voltage is in a negative direction. The resultant AM signal may be coupled to a regular AM detector where the original audio voltage can be reproduced.

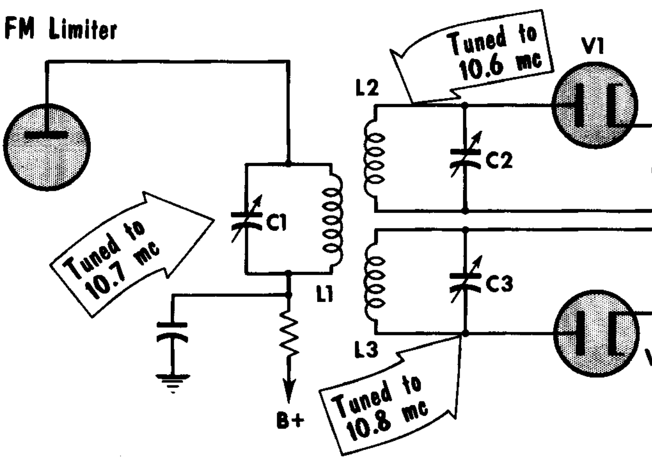
The slope detector is a simple discriminator in that it responds differently to various radio frequencies. This circuit is used sometimes in inexpensive FM receivers where the resonant circuits of the detector input transformer are tuned slightly off the center frequency, but still close enough that the center frequency falls on the linear portion of the response curve. Either linear position can be used. The slope detector, though simple in design, cannot handle large signals because the linear portion of the response curve is too short for large-signal operation. Signals beyond the linear portion of the curve are distorted. In addition, operating on the slope of the response curve of a tuned circuit means that less than maximum gain is being obtained.

Double-Tuned Discriminator

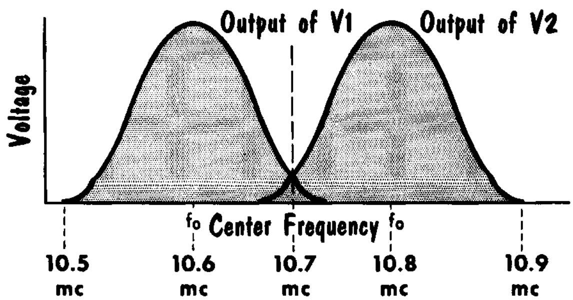
The FM demodulating or detecting circuit must be designed to have its output vary directly (linearly) with the frequency applied. One of the basic circuits is called the double-tuned or Travis discriminator. The tuned circuit of L1 and C1 is the plate load of the limiter circuit. It is tuned to the intermediate frequency of 10.7 megacycles. In the secondary, L2 and C2 are tuned to 10.6 mc, 100 kc below the i-f. The output of L2, C2 is fed to a diode detector circuit consisting of V1, R1, and C4. The circuit of L3 and C3 is tuned to 10.8 mc, 100 kc above the i-f. The output of L3, C3 is fed to a diode detector circuit consisting of V2, R2, and C5.

The normal response curve of each of the tuned circuits in the secondary is shown. As the signal at the output of the limiter shifts in frequency from 10.7 mc towards 10.6 mc, the signal inductively coupled to L2, C2 produces a large current flow. The output detected by V1 develops a voltage across R1. The flow of current, determined by the polarity of V1, is such as to cause an IR drop across R1, with the top of the resistor positive with respect to the bottom. When the signal shifts from 10.7 mc towards 10.8 mc, the signal inductively coupled to L3, C3 produces a large current flow. The output detected by V2 develops a voltage across R2. The flow of current, determined by the polarity of V2, is such as to cause an IR drop across R2, with the top of the resistor negative with respect to the bottom. The r-f voltages developed across R1 and R2 are filtered by C4 and C5.

DOUBLE-TUNED DISCRIMINATOR



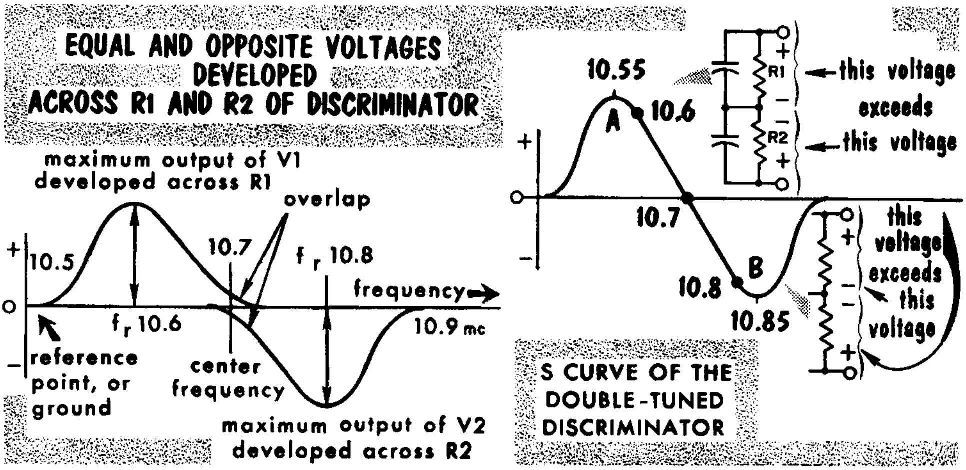
DOUBLE-TUNED DISCRIMINATOR RESPONSE CURVE



Double-Tuned Discriminator (Cont'd)

The lower frequencies at 10.6 mc develop an output voltage across R1 positive with respect to ground. The higher frequencies of 10.8 mc develop an output voltage across R2 negative with respect to ground. The conduction of diode V2 causes an output voltage indicated by a response curve in a negative direction, because the output voltage developed across R2 is negative with respect to ground. The output is zero at the i-f of 10.7 mc. Each tuned circuit conducts slightly, because there is a small degree of overlap of each response curve at 10.7 mc. The resultant low signal voltage output is developed equally across both R1 and R2. However, being of opposite polarity, the two equal-value voltages oppose or buck each other, resulting in zero output.

As the input to the discriminator shifts above and below the i-f center frequency at an audio rate, it produces voltages across R1 and R2. The resultant output voltage from the top of R1 to ground follows the variations in frequency, producing an audio output signal. As the output frequency of the



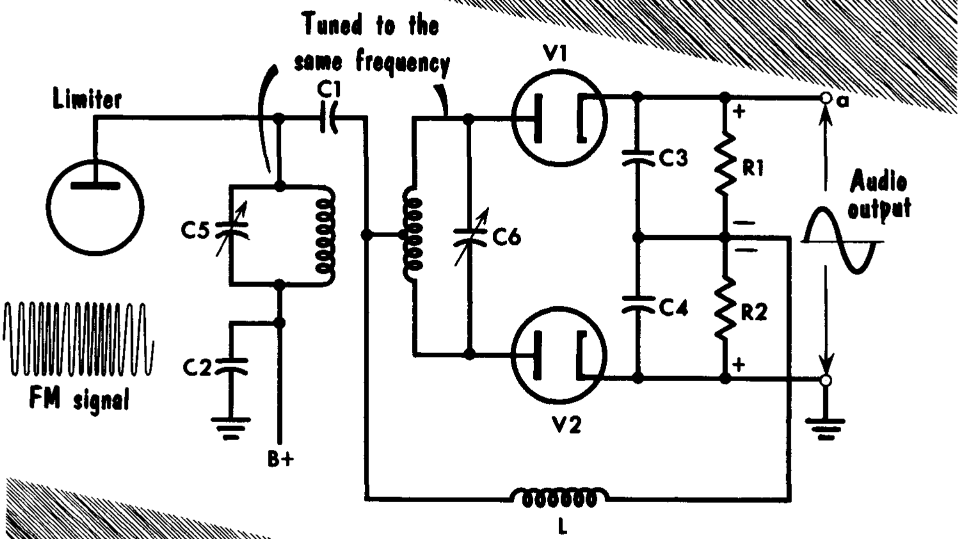
limiter swings further above and below the i-f, the output voltage across R1 and R2 increases. Maximum output is produced when the signal is at either 10.6 mc or 10.8 mc; this corresponds to maximum amplitude in the modulating signal whereby the broadcast station's carrier frequency shifts farthest above and below its center value.

Actually, the response curve of each tuned circuit in the secondary is represented as a continuous S curve. The resonant frequency is usually extended to provide a wider range of frequencies than from point A to B. This is the linear portion of the response curve, in which the conversion of FM to audio takes place. Bandwidth limits may extend as much as ± 120 kc, and the resonant frequency of each tuned circuit may be as low as 10.55 mc and as high as 10.85 mc. This provides a linear portion of the response curve capable of covering a total bandwidth of 240 kc. The extended range of the linear portion of the curve also covers drift in the tuned circuits or the local oscillator which may cause the i-f to be above or below the normal value of 10.7 mc.

The Foster-Seeley Discriminator

The double-tuned discriminator just discussed is effective in performance but comparatively costly to manufacture and is cumbersome to adjust. The presence of two separately tuned secondary circuits makes alignment difficult because of the interaction between the two resonant circuits. A modern dis-

SCHEMATIC OF THE FOSTER-SEELEY DISCRIMINATOR

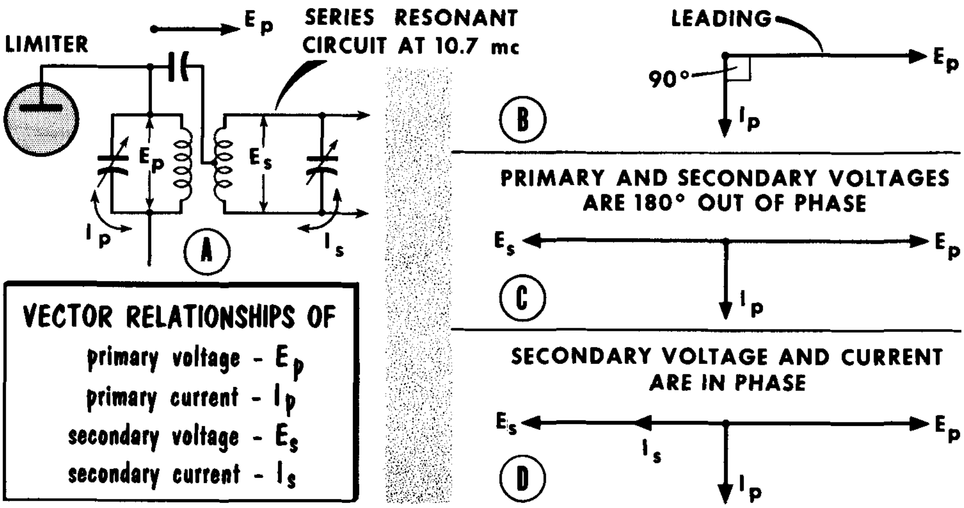


criminator circuit which resembles the double-tuned arrangement closely is the Foster-Seeley discriminator, named after its inventors. This discriminator gained immediate popularity because of its easy adjustability.

Minor variations of this basic circuit are in almost every FM receiver and TV sound system, but the operation of the discriminator may always be analyzed by the same set of vectors used in the following discussion. As can be seen from the schematic of the Foster-Seeley discriminator, there is only one secondary tuned circuit. Where, in the double-tuned circuit, the total response was dependent upon the individual differences between two circuits tuned to different frequencies, in this arrangement, the phase differences between the secondary and primary voltages are used to affect the diodes in a manner which produces usable output. A complete understanding of the process can be had by a step-by-step vector analysis.

Foster-Seeley (I-F At Center Frequency)

We will analyze the vector development of the Foster-Seeley discriminator step-by-step, starting with the i-f signal at the center frequency. With the signal unmodulated (or modulated but passing through dead center frequency at the instant of consideration), the voltage E_p developed across the primary winding of the discriminator transformer (A) is taken as a reference voltage



and drawn as a horizontal vector pointing toward the right, away from the point of origin. Since current through any coil lags the voltage across the coil by 90°, the current, I_p, in the primary winding is shown as a vector (B) lagging the E_p vector by 90°.

The primary current I_p induces a voltage in the secondary winding (C). Since the greatest voltage is generated at the time when the rate of change of current in the primary is greatest – and this occurs when the current wave is passing through the zero axis – the secondary induced voltage vector, E_s, must be shown displaced from the I_p vector by 90°. This places the E_s vector 180° out of phase with vector E_p.

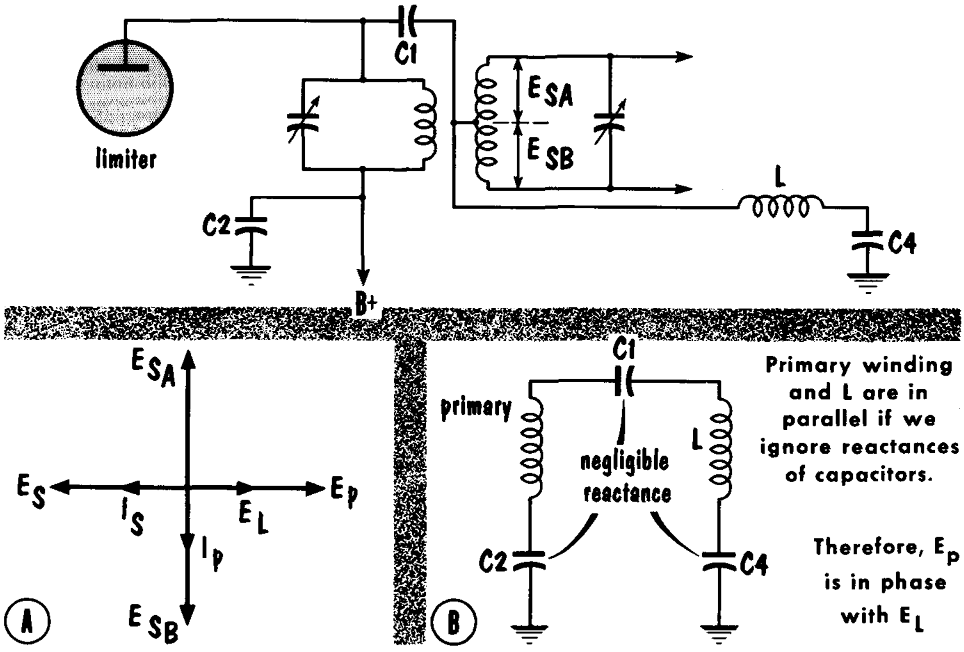
Since this is a perfectly resonant circuit (i-f signal at center frequency) at this instant, the inductive and capacitive reactances cancel each other, leaving the series circuit around the tank purely resistive. In a resistive circuit, current is in phase with the driving voltage. Hence, the series circulating secondary current I_s must be in phase with the secondary induced voltage E_s and is shown as a vector superimposed on E_s (D).

When, as a result of the induced voltage, secondary current I_s flows in the split secondary winding, it produces a secondary voltage drop which leads I_s by 90°. This secondary voltage is a reactive voltage drop resulting from the flow of secondary current. It should not be confused with E_s, a voltage generated by induction.

Foster-Seeley (Cont'd)

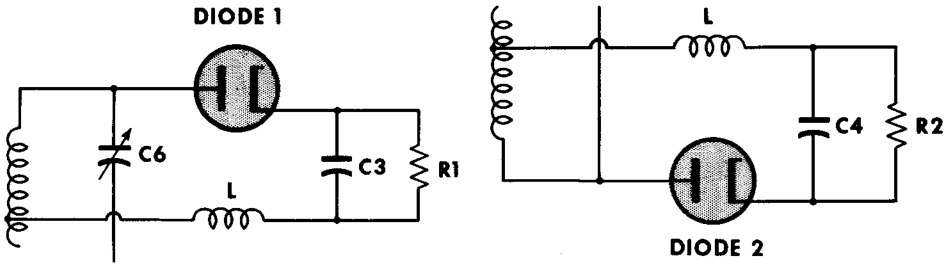
Since the secondary winding of the transformer is centertapped, the secondary voltage may be considered as being made up of two separate voltages (E_{SA} and E_{SB}) of equal magnitude but opposite in phase (A) with respect to the center. This is a valid consideration because, effectively, the centertap simply moves the zero reference point from one end of the winding to its middle, thus providing two distinct voltages with reference to this point; E_{SA} which is shown to be leading E_p by 90° and E_{SB} which is shown lagging E_p by 90°. (The symbols SB and SA are used to represent, respectively, secondary-bottom half and secondary-top half so that later reference to the transformer secondary halves will be clearer.

Consider now the voltage across the coil L. This coil is effectively in parallel with the primary of the discriminator transformer (B), since the capacitors C1, C2, and C4 which complete the connections for parallelism are large enough to have a negligible reactance at the frequencies used in FM. Since it is in parallel with the primary, the voltage E_L developed across the

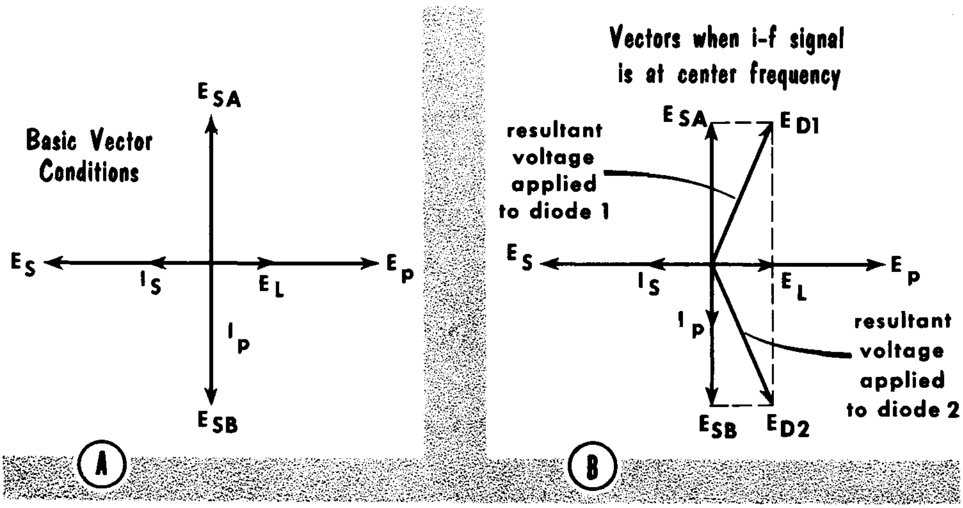


coil L is in phase with the primary voltage. (Note: Actually to all intents and purposes, the primary voltage itself may be considered to be applied to the centertap of the secondary directly. Coil L is used to provide a d-c return path, through its low resistance, for each diode circuit and for all direct currents through the diodes. The high reactance of this coil prevents the primary voltage from being shorted out, offering a load across which the primary signal is developed.

Foster-Seeley (Cont'd)



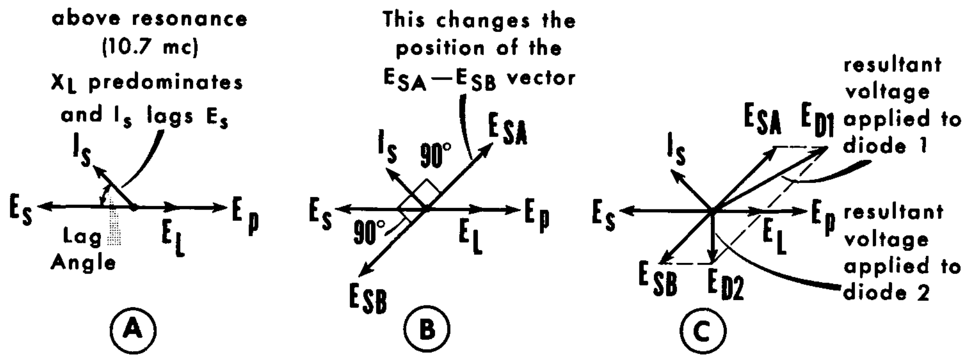
L is in series with both diode circuits



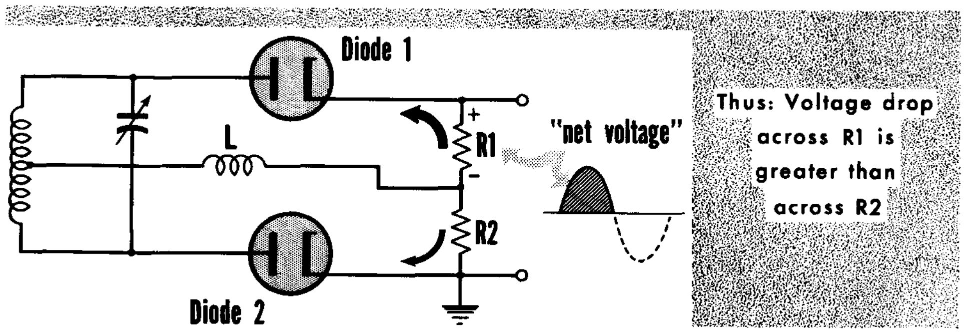
Vector illustration (A) shows the new voltage E_L superimposed on the primary voltage vector E_p. As mentioned, coil L is in series with each half of the secondary winding in separate circuits through the two diodes. This may be shown by tracing either circuit; from the diode plate, through the top half, then through L, up through R1, back to the cathode. The identical circuit exists through the lower diode and the lower half of the secondary. In a series circuit like this, the voltages across the two coils must add vectorially (B). Due to the 90° phase relationship between E_L on the one hand, and E_{SA} and E_{SB} on the other, the resultant r-f voltages which drive direct currents through the diodes must form the diagonal of the parallelograms of the vectors. These resultant vectors are labelled E_{D1} (top diode) and E_{D2} (bottom diode). The vectors E_{SB} and E_{SA} are equal in magnitude because of the symmetry of the coil, so that the driving voltages E_{D1} and E_{D2} are also equal. Thus the direct currents in the diodes and the d-c voltage drops across diode resistors R1 and R2 are equal. They thus cancel each other, and the output d-c voltage between ground and the top of R1 is as it should be for center-frequency operation; exactly zero.

Foster-Seeley (I-F Signal Above Center Frequency)

When a signal is higher than the resonant frequency of the series circuit to which it is fed, the series circuit acts inductively (A). This is true because, for the higher frequency, X_L and X_C no longer cancel each other. X_L is larger, and X_C smaller, than at resonance. Thus, X_L predominates and the circuit is said to be inductive. When the circuit was in resonance, the in-



I-F SIGNAL ABOVE CENTER FREQUENCY



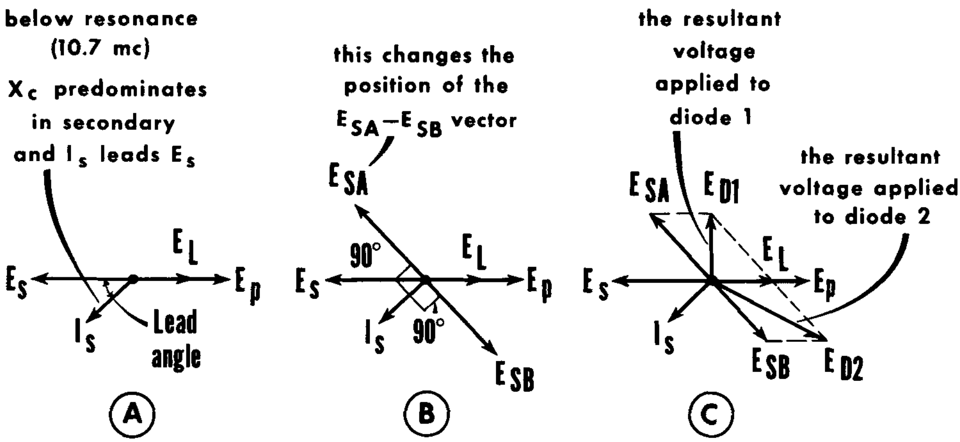
duced secondary series circulating current I_s was in phase with the secondary induced voltage E_s; now, however, with the circuit predominantly inductive, I_s lags E_s.

The lag of I_s reflects itself in the changing vector picture (B) by shifting the directions of the E_{SA} and E_{SB} vectors to new positions. These voltages (E_{SA} and E_{SB}) are produced by an IX_L drop ($I_s \times X_L$) of either half of the secondary and must therefore always be out of phase with the current I_s by 90°.

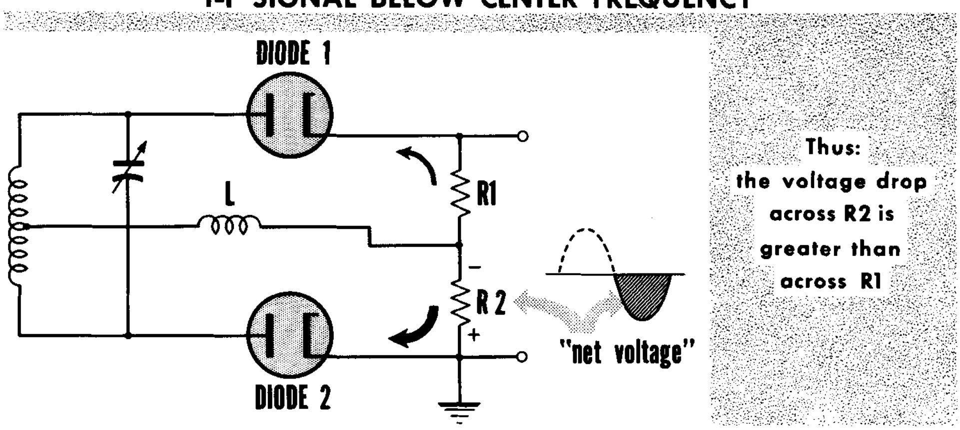
Adding E_L to each of the two secondary voltages (E_{SA} and E_{SB}) vectorially (C), the summed resultants E_{D1} and E_{D2} are again obtained. But this time E_{D1} is much greater than E_{D2} so that a larger direct current flows through R1 than through R2. Thus, a net output voltage (positive in this case) appears across the series load resistors R1 and R2. Again, this is the expected result since the i-f signal is above the center frequency. It is frequency modulated so that an output voltage must appear if detection is to take place.

Foster-Seeley (I-F Signal Below Center Frequency)

The vector diagrams for the condition in which the i-f signal is below the center frequency are shown in parts (A), (B), and (C). It must be remembered that the circuit is capacitive rather than inductive, and that the current I_s leads the secondary induced voltage E_s. We see that E_{D2} is greater than



I-F SIGNAL BELOW CENTER FREQUENCY



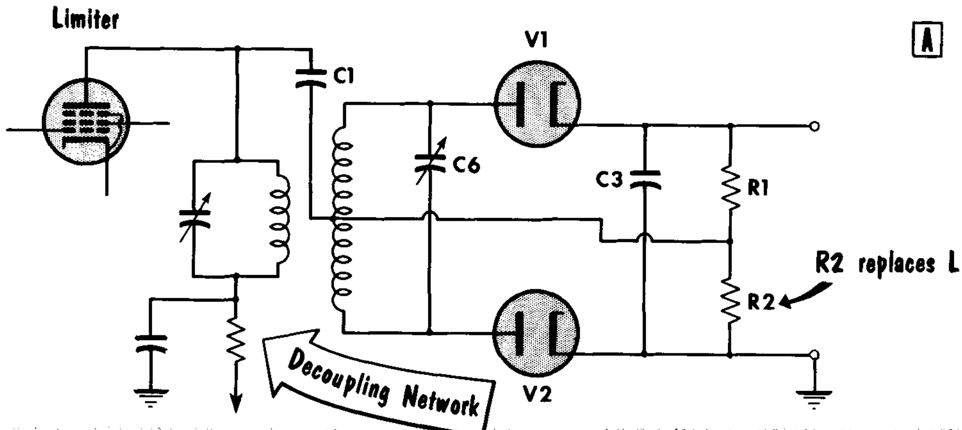
E_{D1} for this condition. Now, a larger current flows through R2 than through R1. Again, a net output voltage (negative in this case) appears across the load resistors.

In summary, it may be seen that a voltage appears across the diode load which varies in exact accordance with the frequency deviation of the i-f signal. The greater the frequency excursion on either side of the center frequency, the greater the magnitude of the voltage developed across the diode load and the louder will be the sound from the loudspeaker of the receiver. The frequency of this voltage will vary as the rate of the frequency deviation, thus giving the detected signal.

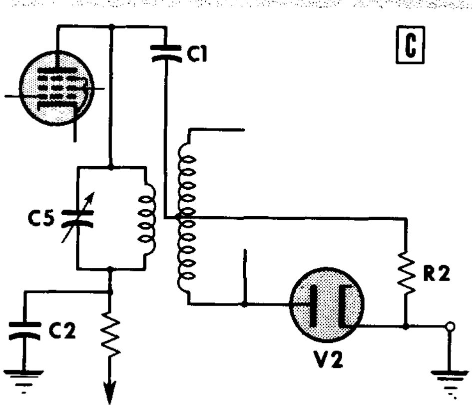
Simplified Foster-Seeley Discriminator

The original Foster-Seeley discriminator may be simplified by omitting the return choke L and one of the two diode load bypass capacitors without affecting its operation. At first glance the reference voltage of the primary seems to have been removed from the secondary diode-detector circuits.

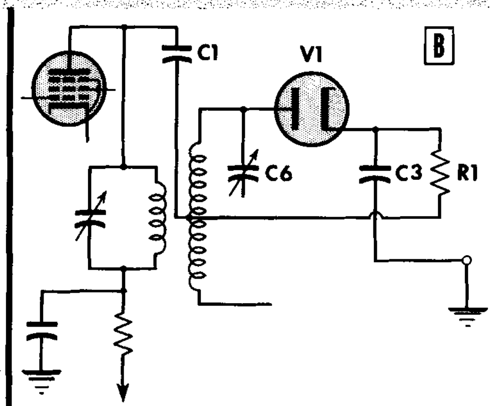
Redrawing one part of the circuit we find the primary voltage coupled through C1 to R2, placing the primary voltage in series with E_{SB}. The resultant voltage is applied to V2 as before. In the other diode circuit, the primary voltage is coupled through C1 to R1, with the return path through C3. This places the primary voltage developed across R1 in series with E_{SA}. The resultant voltage is applied to V1 as before. The value of C3 is chosen to act as a filter for the produced output voltage.



MODIFICATION OF THE FOSTER-SEELEY DISCRIMINATOR, REMOVING L AND REPLACING C3 AND C4 WITH A SINGLE CAPACITOR, C3



DIODE V2 CIRCUIT WITH MODIFICATION

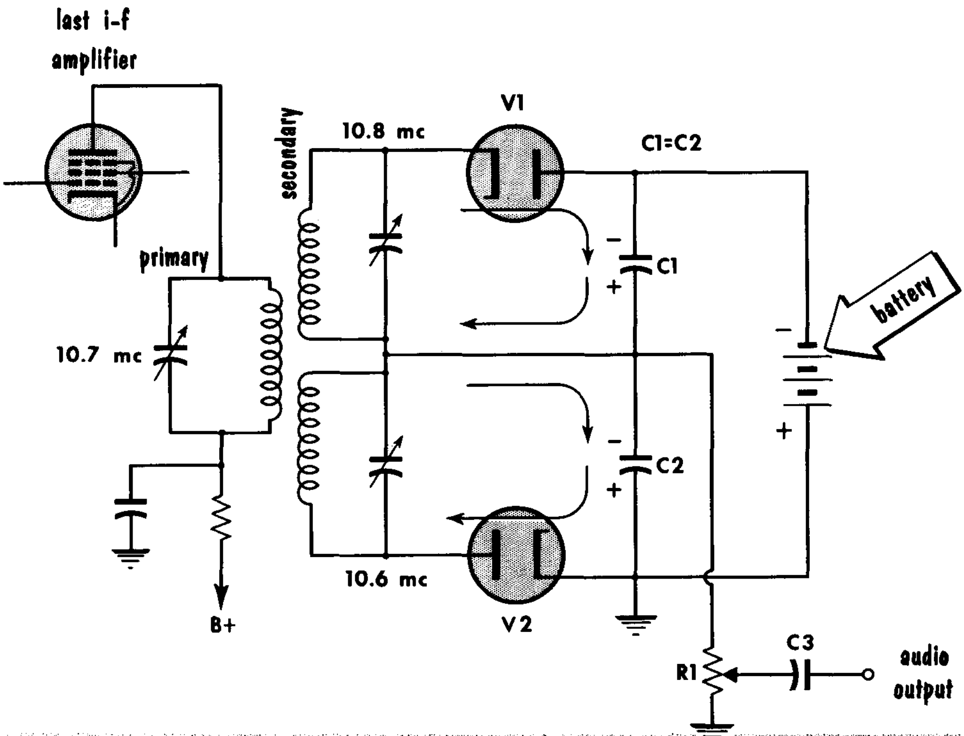


DIODE V1 CIRCUIT WITH MODIFICATION

The Ratio Detector

Limiter circuits are costly; they must have sufficient signal amplitude to operate properly, requiring additional i-f and r-f amplifier stages. A detector circuit insensitive to amplitude variations is needed. The ratio detector circuit meets these requirements.

The diagram shows a basic ratio detector using a double-tuned transformer and a battery. Note the arrangement of diodes V1 and V2. The negative

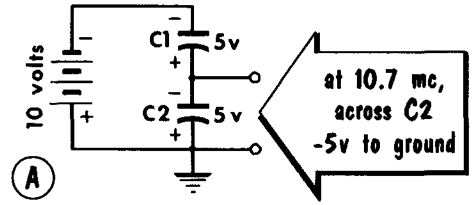


BASIC RATIO DETECTOR SCHEMATIC

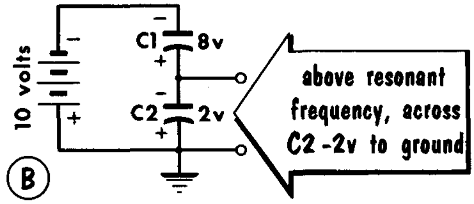
terminal of the battery is connected to the plate of V1, the positive terminal to the cathode of V2. This isolates the battery voltage from the tuned circuits.

Since the values of C1 and C2 are the same, the battery charges each to one-half the value of the battery voltage. The frequency applied to the secondary tuned circuits determines which diode will conduct heavier. This, in turn, charges the capacitor in series with the diode to a higher value. The total voltage across C1 and C2 cannot exceed the fixed voltage of the battery. Each capacitor may, at any instant, differ in amount of charge from the other capacitor, but when totaled, they equal the full value of the battery voltage. The total remains the same – only the ratio of voltage of C1 to C2 may change.

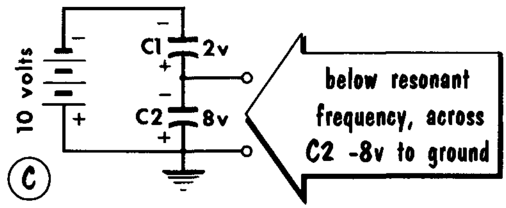
The Ratio Detector (Cont'd)



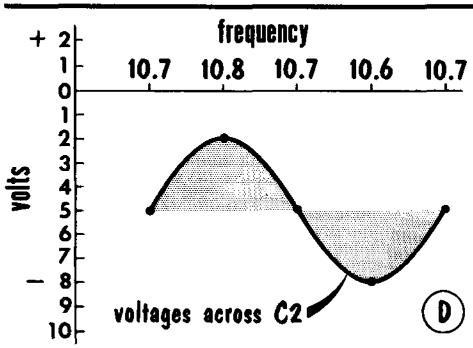
RATIO DETECTOR AT I-F (10.7 mc)

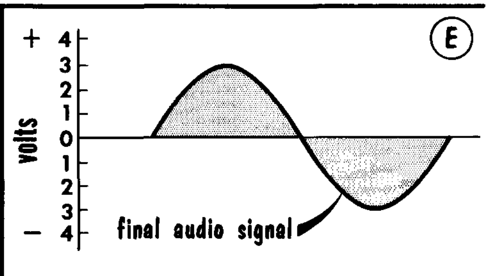


RATIO DETECTOR AT 10.8 mc



RATIO DETECTOR AT 10.6 mc





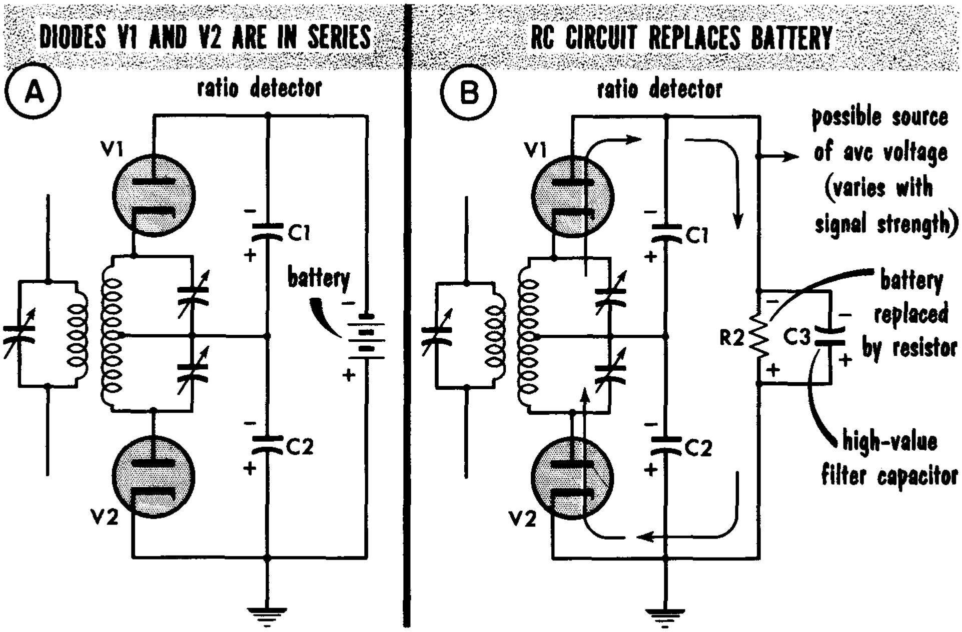
AUDIO OUTPUT OF RATIO DETECTOR

As the action of the double-tuned detector at the i-f of 10.7 mc results in zero output, the voltage values across C1 and C2 will be equal (A). When the signal shifts from 10.7 mc toward 10.8 mc, diode V1 conducts heavier than V2, charging C1 to a higher value. The output of diode V2 is reduced, resulting in a smaller charge across C2 (B). The total voltage of C1 and C2 still equals the battery voltage, but the division of voltage across each capacitor has changed. The ratio of voltage of C1 to C2 is varying. When the signal shifts from 10.7 mc towards 10.6 mc, diode V2 conducts heavier, charging C2 to a higher value. The output of diode V1 is reduced, resulting in a smaller charge across C1 (C). The total voltage of C1 and C2 still equals the battery voltage. It is only the ratio of voltage of C1 to C2 that has changed.

Assume a signal has shifted from 10.7 mc to 10.8 mc and back to 10.7 mc, then to 10.6 mc and again back to 10.7 mc. Plotting the various voltages developed across C2 (D) results in a voltage that varies from -5 to -2, to -5 to -8, and back to -5, at an audio frequency determined by the modulating signal. This varying voltage, when passed through a coupling capacitor and applied to the audio amplifier, appears as in (E).

The Ratio Detector (Cont'd)

The value of the battery voltage is important. With the ratio detector redrawn as in (A), we see that the diodes are connected in series to prevent the battery current from flowing through the tuned circuits. With the battery polarity as shown, there is a negative voltage at the plate of V1 and a positive voltage at the cathode of V2. Before either diode can conduct, the signal voltage must be strong enough to overcome the battery voltage. A high battery voltage would require a strong signal to overcome the battery voltage and permit the diodes to conduct. A low battery voltage would result in low values of voltage across C1 and C2. This would not permit the voltage of C2 to vary by a large

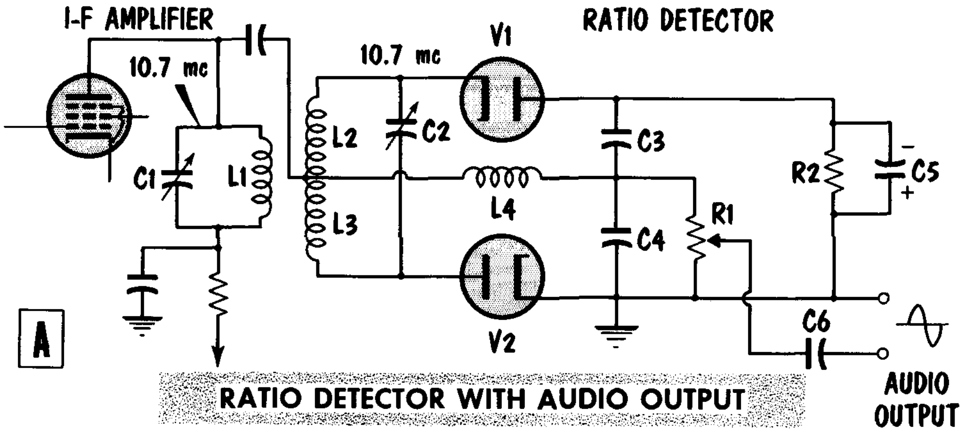


amount, resulting in a low-voltage audio-output signal. To overcome this difficulty, the average value of the signal the receiver is tuned to is used to determine this voltage.

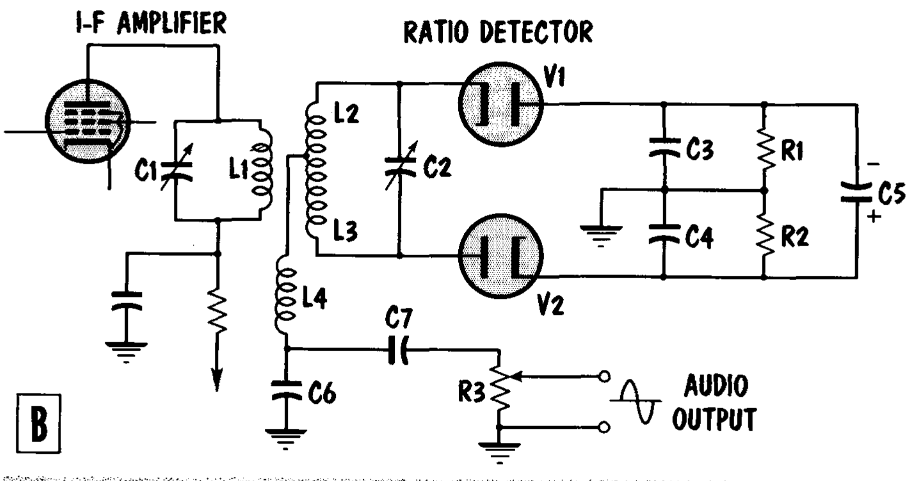
By using R2 in place of the battery (B), signal current flowing in the secondary tuned circuits is rectified by V1 and V2. The rectified signal develops a voltage drop across R2 which is filtered by C3. The values of R2 and C3 are such that it takes a relatively long time to charge C3 – approximately 1/10 to 1/4 of a second. This prevents a sudden change of voltage from quickly charging C3. The voltages across C1 and C2 must equal the total voltage across C3. The voltage across C3 cannot change; hence, the voltages across C1 and C2 will not change. Since interference consists mainly of short-duration, sharp noise pulses with insufficient time to charge C3, the voltage of C3 remains substantially constant, and the output does not vary, eliminating the interference.

The Ratio Detector (Cont'd)

A practical form of ratio-detector would be based on the phase discriminator, (A). Diodes V1 and V2 conduct in accordance with the vector sum of the voltage across L4 in series with the induced voltage of L2 or L3. The resultant causes C3 or C4 to charge to a higher or lesser value, resulting in an audio output as previously explained for ratio-detector operation.



RATIO DETECTOR WITH AUDIO OUTPUT



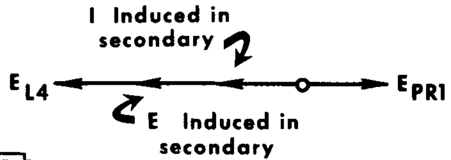
MODIFYING THE RATIO DETECTOR, RFC CHOKE L4 IS REPLACED BY A SMALL WINDING INDUCTIVELY COUPLED TO L1

A modified version of this circuit is shown in (B). The resistor used to develop a voltage for C5 is divided, with the centertap grounded. Now L4 is not capacitively coupled to the primary circuit. Instead of being an r-f choke, L4 now consists of several turns of wire, closely coupled to primary winding L1 for maximum induced signal. The complete path for this signal is through C6 to ground.

The Ratio Detector (Cont'd)

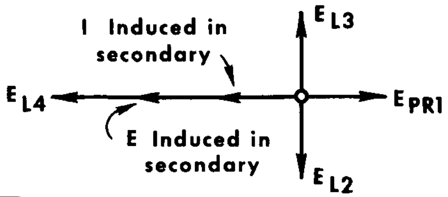
A

VECTOR DIAGRAM of MODIFIED RATIO DETECTOR



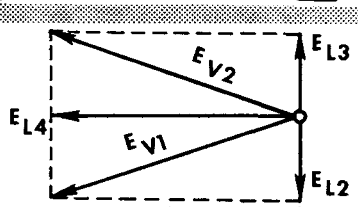
B

Voltages E_{L2}, E_{L3} added to ABOVE



C

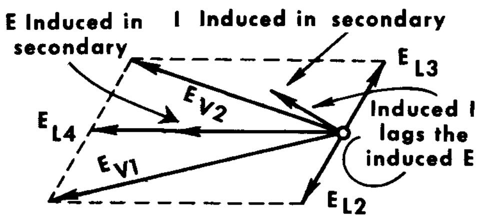
Ratio Detector Vectors AT 10.7 mc



Resultant vector diagram

D

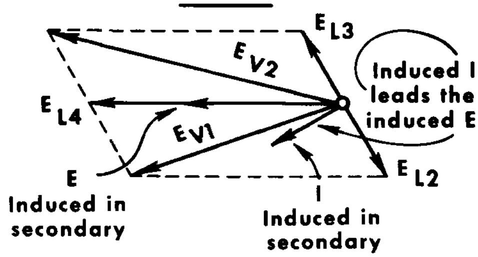
Ratio Detector Vectors ABOVE 10.7 mc



Frequencies ABOVE RESONANCE result in INDUCTIVE CIRCUIT vector diagram

E

Ratio Detector Vectors BELOW 10.7 mc



Frequencies BELOW RESONANCE result in CAPACITIVE CIRCUIT vector diagram

The voltage induced in L4 is 180° out of phase with that of L1. When indicated on a vector diagram, as shown in (A), the induced voltage of L4 is now in the same direction as the induced voltage and current of the secondary. Vector diagram (A) indicates the resonant frequency of 10.7 mc. Adding to this diagram the voltages developed across L2 and L3 results in the vector diagram in (B). Vector diagram (C) shows the relationship of voltages of L4, L3, and L2, and the resulting vector sum applied to diodes V1 and V2, in (C). At resonance, diodes V1 and V2 conduct equally, resulting in zero output.

With a modulating signal causing the frequency to shift above the resonant value of 10.7 mc, the tuned circuit of L2, L3, and C2 becomes predominantly inductive. The current lags the voltage, (D). The resultant vector sum causes V1 to conduct heavier. With a modulating signal causing the frequency to shift below the resonant value of 10.7 mc, the tuned circuit of L2, L3, and C2 becomes predominantly capacitive, and the current leads the voltage, (E). The resultant vector sum causes V2 to conduct heavier.

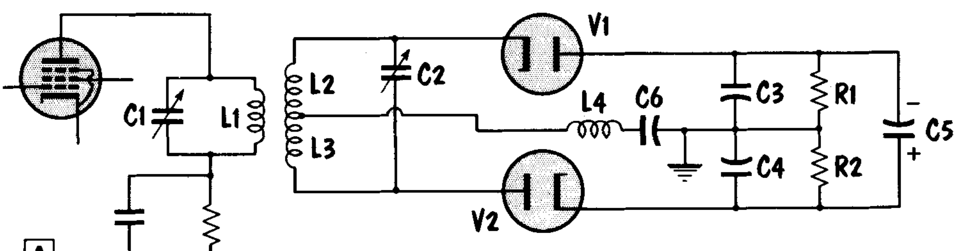
The Ratio Detector (Cont'd)

Another major difference in the ratio detector is the development of the audio signal. The circuit has been redrawn in (A). C7 and potentiometer R3 (page 4-100) are omitted since they only couple and vary the value of audio signal to the audio amplifier. With the circuit at 10.7 mc, V1 and V2 conduct equally. Tracing the current path of V1, we find that C6 is in series with the current flow (B).

Tracing the current path of V2, we also find that C6 is in series with the current flow (C). As both diodes conduct equally at resonance, the current flow through C6 is of equal but opposite value. The object is to leave C6 with no charge, resulting in zero voltage across it. This is as desired since, at resonance, there is zero output.

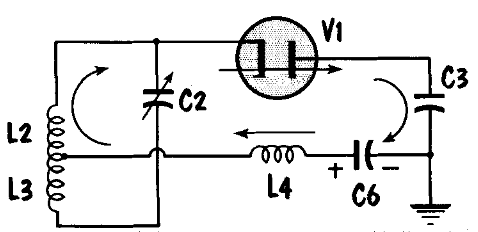
When the signal frequency shifts above or below resonance, the current flow of one diode differs from the other. The diode conducting more heavily produces a charge across C6; its polarity is determined by the heavier-conducting diode. A stronger modulating signal causes larger frequency shifts. When detected by this circuit, it results in more or less current flow through either diode. This results in proportionately larger voltages developed across C6. The frequency of the varying voltage developed across C6 varies directly with the frequency at which the signal shifts above and below resonance. The result is a varying audio signal voltage across C6, coupled through C7 and R3 to the a-f amplifier.

The value of C6 is important; to develop the IX_C drop, a typical value of .002 µf is used. This value has an X_C at 10.7 mc low enough to be effectively a short circuit. At audio frequencies, however, the value of X_C permits a sufficient IX_C voltage to be developed.

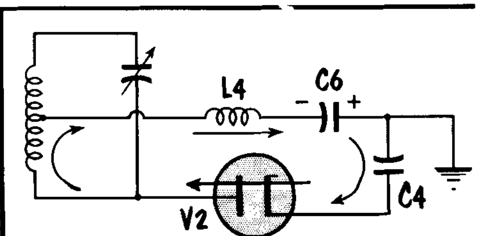


A

RATIO DETECTOR REDRAWN

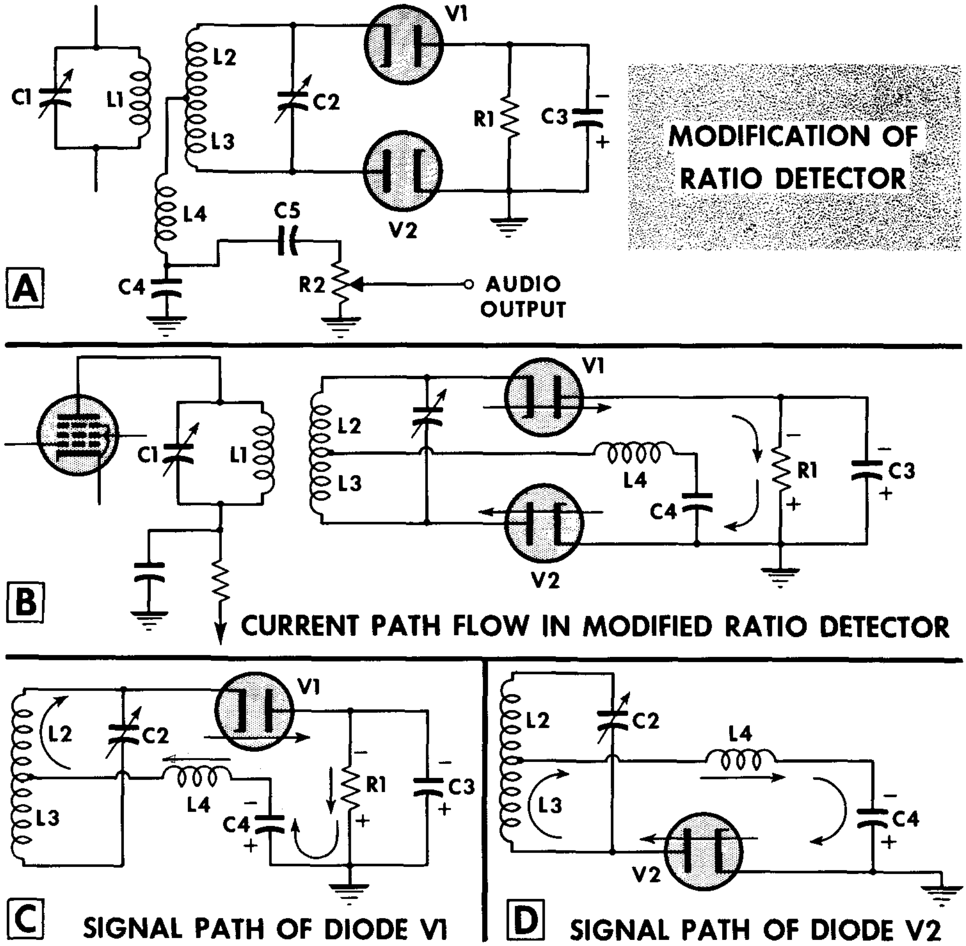


B CURRENT PATH OF DIODE V1



C CURRENT PATH OF DIODE V2

The Ratio Detector (Cont'd)



MODIFICATION OF RATIO DETECTOR

CURRENT PATH FLOW IN MODIFIED RATIO DETECTOR

SIGNAL PATH OF DIODE V1

SIGNAL PATH OF DIODE V2

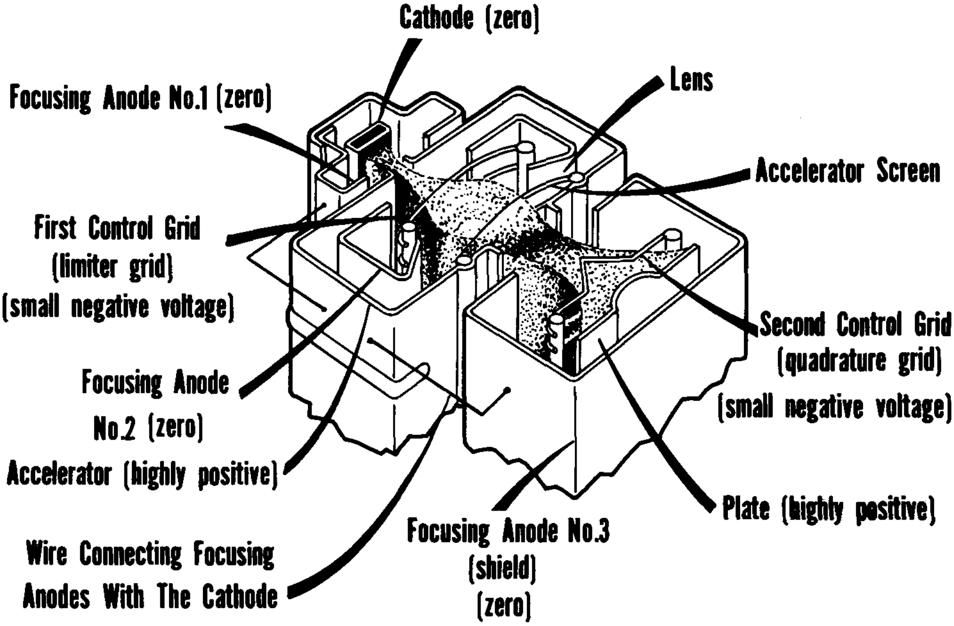
The modification of the ratio detector in (A), is probably the most popular version. For ease of explanation, the coupling circuit of C5 and R2 is omitted. The circuit may then be redrawn as (B). The signal developed across L2, L3, and C2 causes a rectified current to flow through V1, R1, and V2, resulting in a voltage across C3. The voltage value is determined by the signal strength of the station received.

In addition, there are the signal paths (C) and (D). At the frequency of 10.7 mc, the reactance of C3 is low enough to constitute a short circuit. As in the previous ratio detector, the two signal current paths flow through C4, developing an audio output voltage across it. The modified ratio detector retains the control or battery voltage, making it insensitive to signal amplitude changes. The ratio of voltage of two capacitors for an output signal, is replaced by a single capacitor whose voltage and polarity depend on the conduction of diodes V1 and V2.

The Gated-Beam Detector

An interesting and popular approach to limiting and demodulation of FM signals is through the gated-beam detector. To understand this circuit we must study the gated-beam tube. Note the unusual shape and placement of the various electrodes. As the electrons are attracted from the cathode, focus anode 1, at cathode potential, keeps the electrons from scattering. The electrons pass through anode 1 in the form of a beam. The high positive attrac-

INTERNAL CONSTRUCTION AND VOLTAGES OF GATED-BEAM TUBE



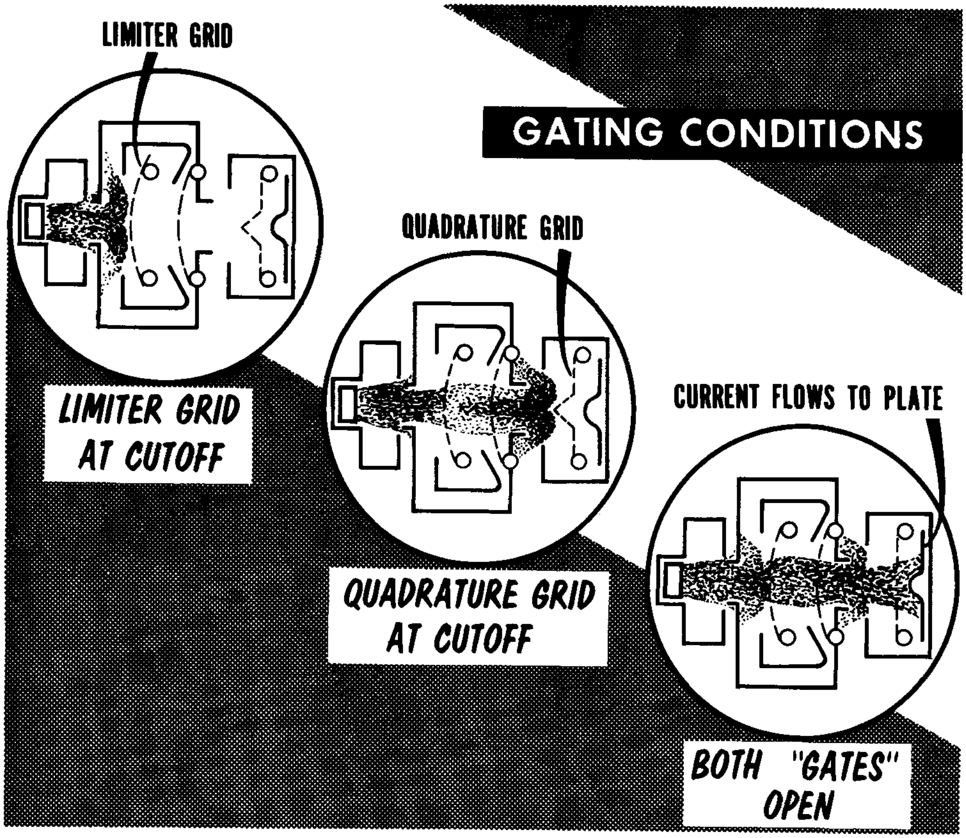
tion of the accelerator screen grid and accelerator assembly speed up the movement of the electrons leaving focus anode 1. After the accelerated beam passes through focus anode 1, it tends to spread out to the highly positive accelerator assembly.

Focus anode 2 is placed inside the accelerator assembly at cathode potential to repel the electrons and keep them in a beam. The accelerated beam of electrons passes through grid 1, which has a small, negative d-c voltage on it. The beam continues past the accelerator screen grid, where the plate becomes the attracting force. Once again the electron stream is accelerated. To prevent the beam from scattering on its way to the plate, focus anode 3 at zero voltage repels the electrons, keeping them in the shape of a beam. Before the electron stream reaches the plate, it passes through control grid 2, which has a small negative d-c potential on it. The beam of electrons finally arrives at, and is collected by, the highly positive plate.

Double-Gate Action

In a triode or pentode, when plate current cutoff occurs because of a highly negative control grid, cathode current also ceases to flow because the electrons leaving the cathode fall back on it, making the net current flow equal to zero.

Should the first control (limiter) grid in the gated-beam tube receive a negative potential greater than cutoff, electrons still flow back to the power supply via the accelerator assembly so that, although there may be no plate current,



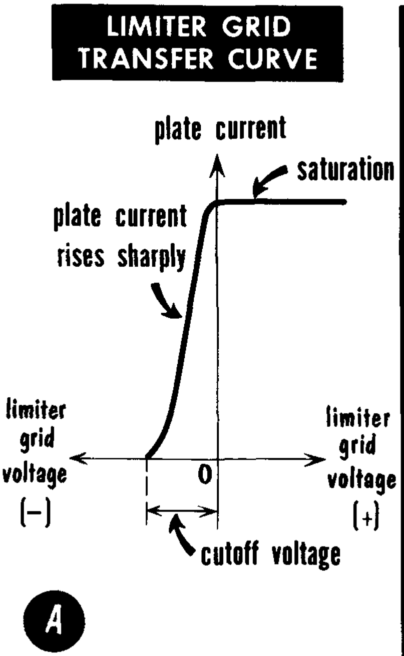
cathode current is not greatly affected. Similarly, stoppage of plate current due to a highly negative second control grid (quadrature grid) causes the return current to flow through the accelerator assembly again, maintaining a relatively constant cathode current.

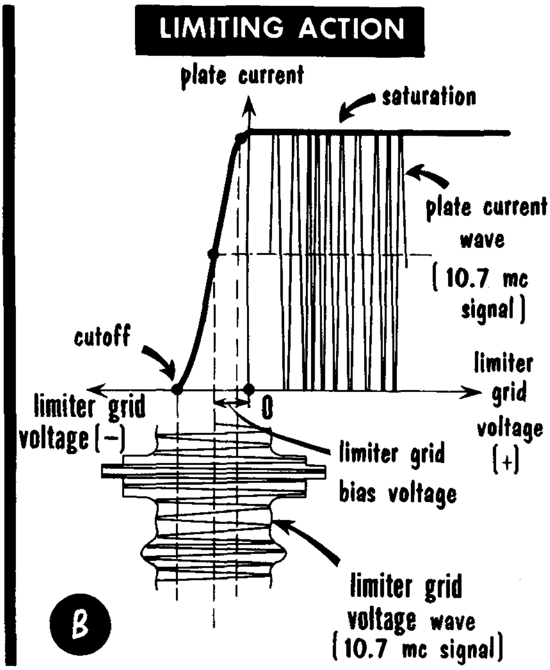
The operation of the system may be viewed as a switching action. The electron beam current is switched to the accelerator when either of the two gates—the limiter grid or the quadrature grid—is closed, and to the plate when both gates are open.

Limiter Grid Transfer Characteristic

With normal voltage relationships established on the static electrodes, the limiter grid voltage may be varied while plate current is recorded to obtain the limiter grid transfer characteristic curve (A). Starting with a cutoff potential on this grid, the plate current is, of course, zero, but the cathode current may be quite high since the first gate is closed and the beam current has been switched to the accelerator. As the negative bias on the limiter grid is lowered past cutoff, beam current suddenly switches to the plate. This accounts for the extremely sharp rise in plate current. Saturation is reached quickly, the plate current levels off near zero limiter grid voltage.

A limiter requires that the tube cut off sharply with relatively small negative grid voltages and that saturation be reached quickly. In the previous circuits, these conditions were achieved by using grid-leak bias and low plate voltages which provided limiter action at relatively low gain. Here, however, we have



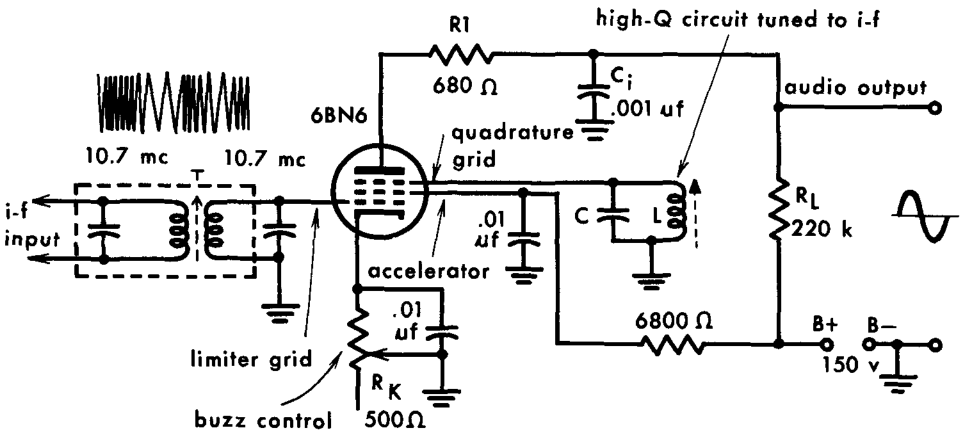


the essential kind of curve needed without the accompanying loss of gain common in standard limiter circuits. Diagram (B) shows the limiting action of the 6BN6 with an applied signal of 10.7 mc, having both amplitude and frequency modulation. Note that the limiter grid is biased near the center of the linear portion of the curve and that limiting action is good without loss of gain. In most standard circuits using the 6BN6 gated-beam tube, the limiter grid bias is obtained by a variable cathode resistor, called a "buzz" control. Careful adjustment of this control makes almost perfect limiting possible, removing all forms of AM.

Gated-Beam Tube as an FM Detector

By connecting the circuit of the 6BN6 as shown, the tube serves as both a limiter and as an FM detector. The i-f signal is fed to the primary of T, and limiter action occurs as described. A resonant L-C circuit (also tuned to the i-f) is connected between the quadrature grid and ground, and an output circuit from the plate of the tube has been added. When an electron stream modulated or varying at 10.7 mc passes the quadrature grid, excitation is supplied to the resonant circuit. To visualize what happens at the quadrature grid and in the L-C resonant circuit, study the changing electric field around the quadrature grid and its effect on the resonant circuit. Electrons flowing

BASIC GATED-BEAM CIRCUIT

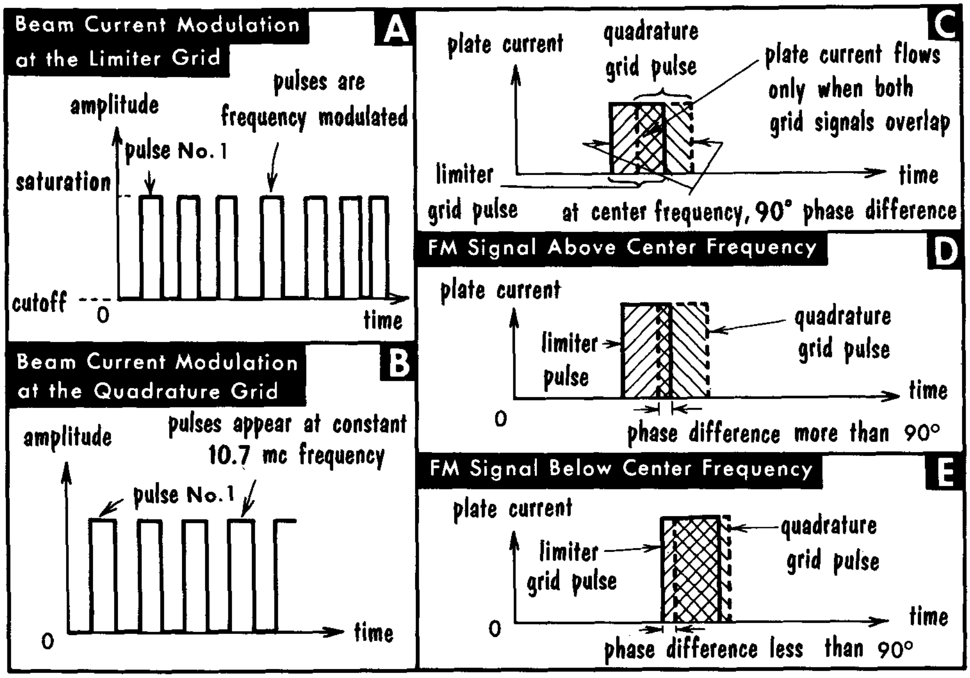


from one element in a tube to another produce electric fields of varying intensities around each of the intermediate elements in the path of the stream. The strength of the field produced is a function of the concentration of electrons in that area at that time.

Assume that a relatively large plate current is flowing in the 6BN6 as a result of a sufficiently low bias voltage applied to the first (or limiter) grid. As the concentrated electron stream approaches and reaches the quadrature grid, electrons are driven out of this grid into the L-C circuit. A quadrature grid current, and a voltage across the tank, result from the charge induced by the moving electrons. This current leads the limiter grid voltage by 90°. When C has charged to peak voltage, the exciting current has dropped to zero and C can discharge through L. If nothing else happens, L and C produce a damped oscillatory wave which will gradually decay to zero. But if the resonant circuit is tuned to the same frequency as the varying electron stream in the tube, it receives a timed pulse of energy causing it to build up once again to peak voltage. In this manner the electron stream in the tube keeps the L-C circuit oscillating.

Gated-Beam Tube as an FM Detector (Cont'd)

We can now examine the combined effect of limiter and quadrature grids with respect to the incoming signal. Due to the peak-shearing effect of the limiter grid, the electron beam current between the limiter and quadrature grids varies in a square-wave manner (A). Like the incoming signal, the wave is frequency modulated. The voltage at the quadrature grid due to the L-C oscillation would by itself have an effect on the beam current (B). Suppose that an FM signal at 10.7 mc is received. Pulse 1 at the limiter grid and at the quadrature grid can be represented on a single axis (C). Remembering that these grids behave as gates, and that plate output can be obtained only when both gates are open simultaneously, we can see that plate current flows only during the overlapping portion of the pulses. This is the only time when both grids are sufficiently positive to permit passage of the beam to the plate.

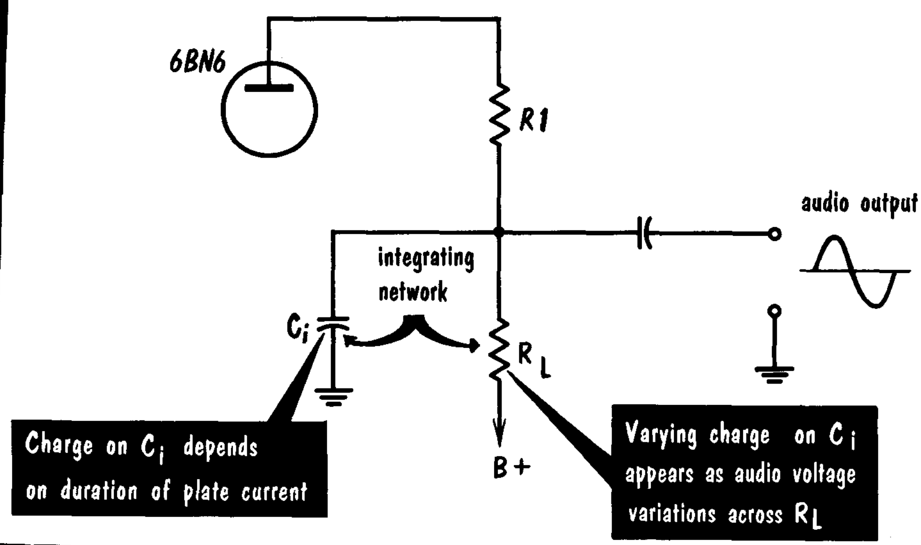


At the next instant, there may be a positive frequency deviation. This increase in frequency means a shorter interval between pulses. With pulses arriving more quickly than they were at the center frequency (but with the quadrature grid pulses still occurring at the same rate), the limiter pulse arrives at the quadrature grid earlier than it did before (D). As the diagram shows, both gates are open together during an overlap interval which is thinner than it was before, indicating that the plate current of the tube is flowing for a shorter time. Similarly, if the incoming signal becomes lower in frequency, the limiter grid pulse arrives at the quadrature grid later than before, thus providing a greater overlap of pulses (E). This means that the plate current flows for a longer time per pulse.

Audio Output From the Gated-Beam Detector

No mention has been made of the function of capacitor C_i. Regardless of the relative occurrence of the gate openings, the plate current magnitude is always the same. The length of time per cycle during which it flows depends entirely upon coincidence of the limiter and quadrature pulses or lack of it; this, in turn, is a function of the frequency of the received signal. A fre-

OUTPUT CIRCUIT OF GATED-BEAM DETECTOR

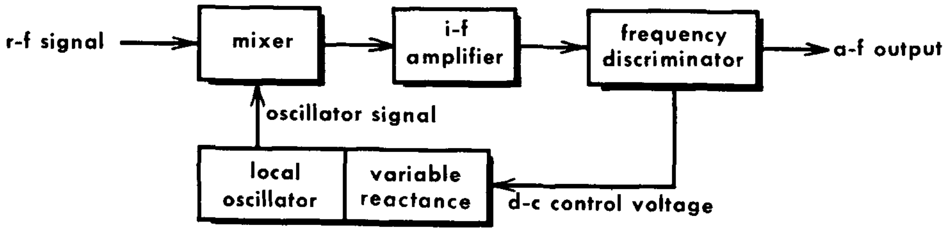


quency higher than that of the resting frequency results in a short-time flow and a frequency lower than resting yields a plate current flow of longer duration.

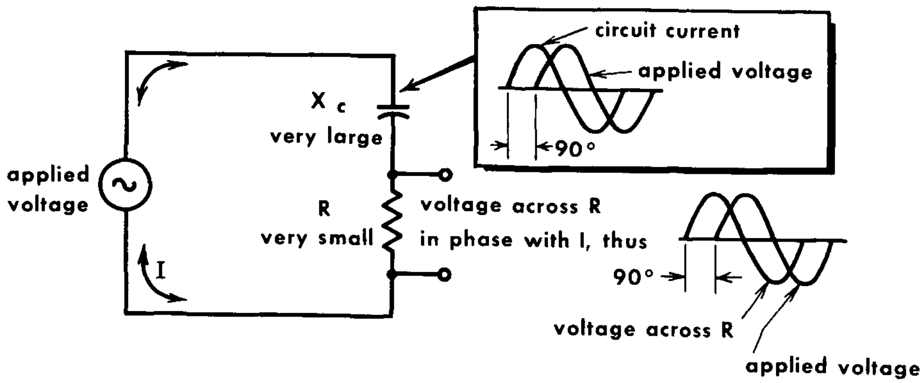
Variations in current time may be translated into proportional variations in output voltage by an integrating network (C_i and R_L). Since C_i charges through R_L, it builds up a greater voltage when the current flows for a longer time. This is reflected as a changing voltage across the load resistor, R_L. The changing voltage represents the audio signal. The time constant of R_L-C_i must be large enough to smooth out i-f pulsations but small enough so that the audio modulation frequency signal can be built up across R_L. (Sometimes C_i is omitted, in which case the circuit and tube shunt capacitance provides the filtering or integration required.) The function of R1 is to improve the linearity of the audio output signal. If this resistor is not used, output increases somewhat, but linearity suffers.

Automatic Frequency Control

BASIC AUTOMATIC FREQUENCY CONTROL SYSTEM



R-C SERIES CIRCUIT



The automatic frequency control (afc) circuit in a receiver controls the local oscillator frequency so that the correct i-f will always be produced when a signal is being received. Two things are necessary for any afc system; a frequency discriminator capable of changing a frequency variation into the form of a d-c voltage that can be used for control purposes, and a variable reactance whose value can be controlled by the d-c voltage changes due to the frequency discriminator. The variable reactance connects to the oscillator circuit to control its frequency.

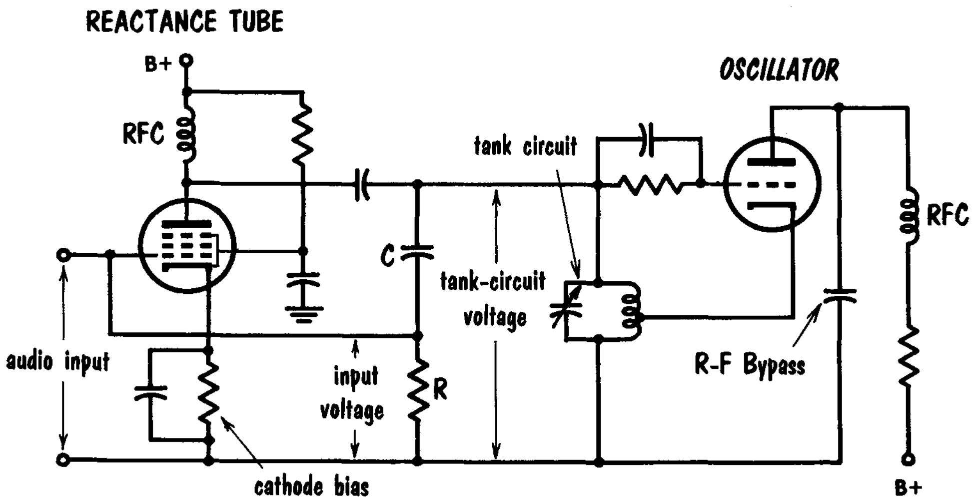
A reactance tube can be almost any tube connected in a circuit so that its a-c impedance from plate to ground is similar to the impedance of an inductive or capacitive reactance. The impedance of a reactance causes the current flowing through it to be 90° out of phase with the voltage impressed across it. By reproducing this relationship, a tube can simulate a reactance.

In an R-C series circuit, C is a low-value capacitor whose reactance is much larger than the resistance of R. The current in this R-C circuit is determined predominantly by C, and leads the applied voltage by 90°. Since the voltage across a resistor is in phase with the current through it, the voltage across R will also lead the applied voltage by 90°.

Automatic Frequency Control (Cont'd)

In our reactance tube – oscillator circuit, any r-f current flowing from cathode to plate in the reactance tube will also flow through the tank circuit of the oscillator. The r-f current must flow 90° out of phase with whatever voltage is applied between plate and ground (across the oscillator tank circuit). The oscillator voltage is also applied across the R-C circuit. We just saw that the voltage across R led the applied voltage by 90°. But the voltage across

REACTANCE TUBE CIRCUIT APPEARS AS CAPACITOR ACROSS TANK CIRCUIT

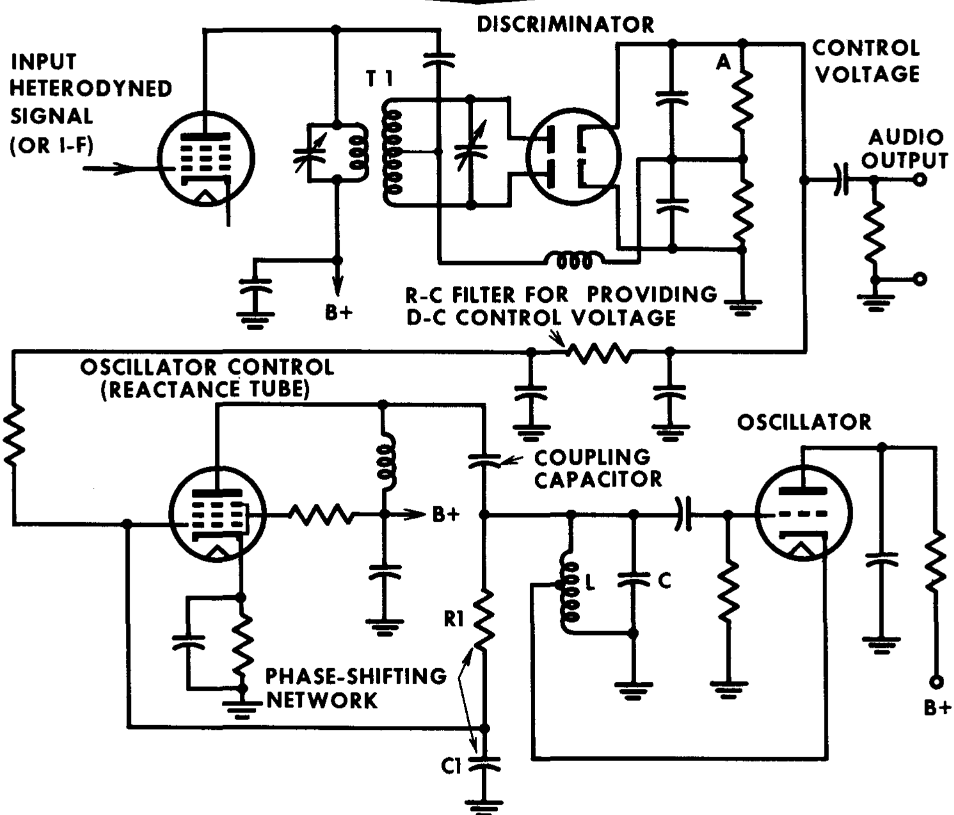


R is also the input voltage between grid and ground. Thus, the grid voltage, and therefore the plate current, will be 90° out of phase with the applied voltage from the oscillator. Since the plate current leads the applied or plate voltage by 90°, the reactance tube appears to the oscillator tank circuit as a capacitor.

The amount of a-c plate current in the reactance tube depends on the gain of the tube. The gain of the tube can be varied by varying the grid bias. This controls the amount of plate current drawn by the tube, and thus controls the value of the "reactance". In determining the resonant frequency of the oscillator, the reactance of the reactance tube plays just as important a role as the conventional L-C circuit components. The reactance tube – oscillator circuit is adjusted so that for a given bias voltage on the reactance tube, the oscillator will operate at a desired frequency. As we will see, by varying this bias (either increasing or decreasing it), we will change the amount of reactance "injected" into the oscillator circuit, and thus affect the oscillator frequency.

Automatic Frequency Control (Cont'd)

AUTOMATIC FREQUENCY CONTROL CIRCUIT



We can now examine a complete afc system. Note the reversal of R and C; the injected reactance will be inductive, with plate current lagging the applied voltage. If the oscillator suddenly became unstable and drifted to a lower frequency, the operating i-f would also decrease in frequency. The discriminator circuit then functions so that at point A, a positive voltage appears, instead of zero voltage when the oscillator is operating properly. The magnitude of the voltage depends upon the amount of detuning that took place. The positive voltage decreases the reactance tube bias. This increases the mutual conductance (g_m) of the tube, and the injected inductance is decreased. If the injected inductance is decreased, the total effective inductance is also decreased, and the oscillator frequency increases to its normal value. This continues until the correction voltage falls to zero. Should the oscillator drift to a higher frequency, the correction or error voltage wouldbe negative, and inject more inductance to lower the resonant frequency and correct the i-f.

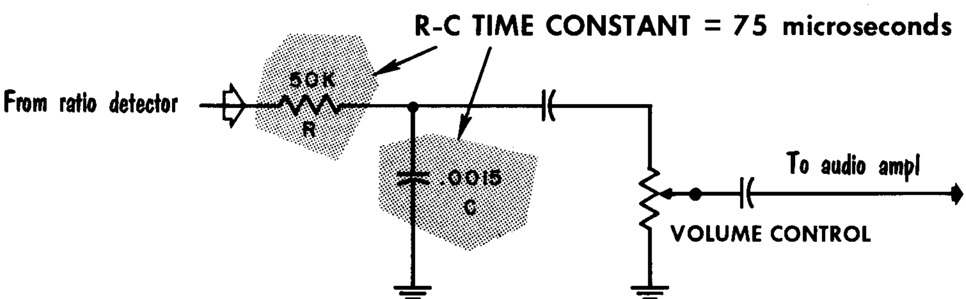
De-emphasis

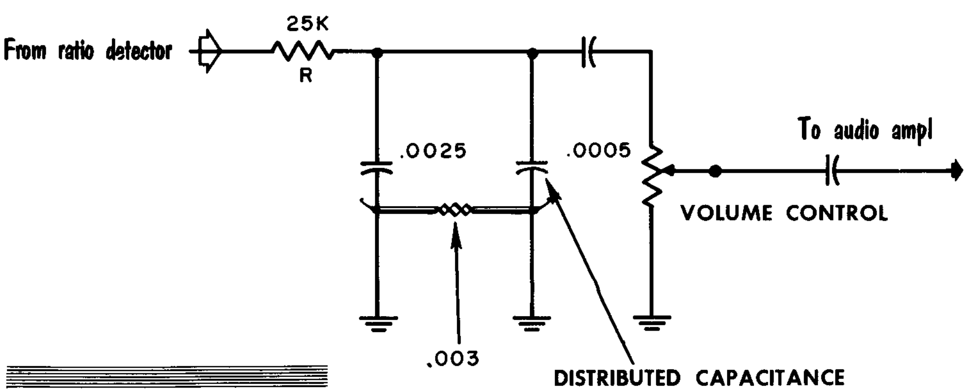
The transmitted audio signal is pre-emphasized at the transmitter. If this is not corrected at the receiver, the sound signal has a heavy treble effect. To compensate, the FM receiver contains a de-emphasis network – a low-pass filter. As in the case of the high-pass filter, the de-emphasis network also has a 75-microsecond time constant.

The de-emphasis network can be placed in several places in the FM receiver, but practice finds it located at the output of the FM detector circuit. An R-C filter shunts the higher audio frequencies to ground while affecting the lower frequencies to a lesser extent. The overall effect of the pre-emphasis – de-emphasis maneuver is to provide a noise-free signal at the output of the FM receiver that is a replica of the sound signal at the transmitting studio.

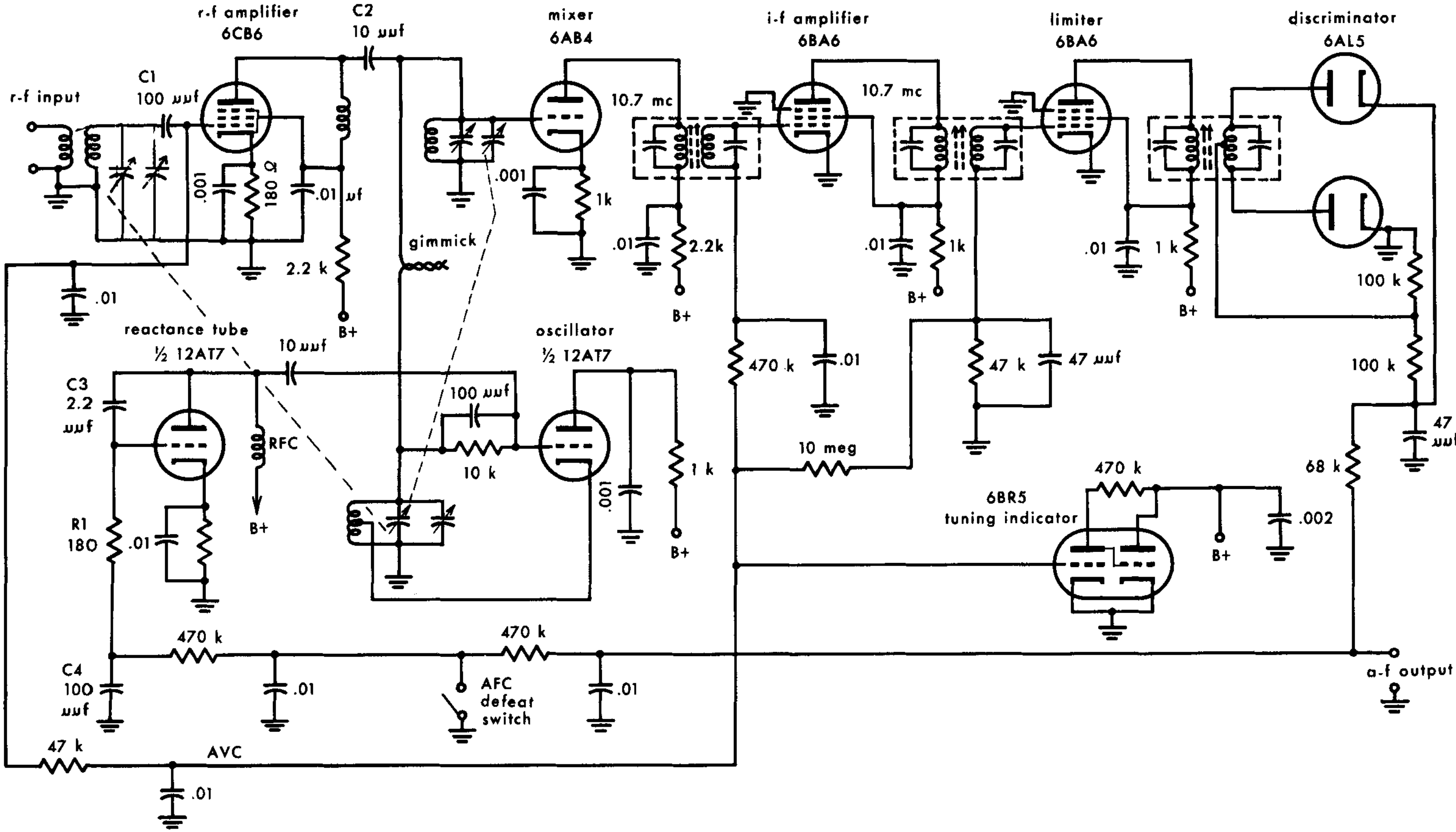
In designing a de-emphasis network, engineers take into consideration the distributed capacitances. Thus, the "lumped" R-C time constant may not be precisely equal to 75 microseconds.

DE-EMPHASIS NETWORKS





FM TUNER



r-f amplifier
6CB6
C2
10 μμf
mixer
6AB4
i-f amplifier
6BA6
limiter
6BA6
discriminator
6AL5
r-f input
C1
100 μμf
10.7 mc
10.7 mc
.001
180 Ω
.01 μf
.001
1k
.01
2.2k
.01
1k
.01
1 k
gimmick
2.2 k
B+
B+
B+
B+
.01
100 k
reactance tube
½ 12AT7
10 μμf
oscillator
½ 12AT7
470 k
.01
47 k
47 μμf
100 k
C3
2.2
μμf
100 μμf
RFC
10 k
10 meg
47
μμf
1 k
.001
68 k
470 k
6BR5
tuning indicator
R1
180
.01
B+
B+
.002
B+
C4
100
μμf
470 k
470 k
.01
AFC
defeat
switch
.01
a-f output
47 k
AVC
.01

Basic FM Tuner

Having discussed the various circuits involved in the reception of FM signals, let us look at a typical complete FM tuner. The power supply has been omitted. Any conventional half-wave or full-wave power supply can be used. In addition, the audio amplifier has been omitted, in that any simple or elaborate amplifier can be used.

The input of the tuner is a broadband r-f amplifier. This stage is broadly tuned to the 88-108 mc FM band. C1, a coupling capacitor, prevents the avc line from being grounded out. The output of the r-f amplifier is coupled through C2 to a tuned circuit of the mixer stage. As discussed previously, C2 provides broader bandpass. The triode mixer is used with the output of the 12AT7 oscillator, being coupled to it through a "gimmick." This is nothing more than two insulated wires twisted together. Effectively, this provides capacitive coupling. Afc is used to stabilize the local oscillator. The control voltage is obtained from the output of the discriminator. There are a-f variations in this output, but these are filtered by the RC filter. In addition, this tuner contains an afc "defeat" switch. When this switch is closed, the control voltage line is short-circuited to ground and the reactance tube does not exert control over the oscillator. The RC phase-shifting network used in this reactance tube is C3 and R1. R1 is effectively grounded through C4, a 100 $\mu\mu$f capacitor. C4 has little reactance at the extremely high frequency of the oscillator and the feedback signal easily follows the path through C4 to ground. C4 has little effect on the phase shift.

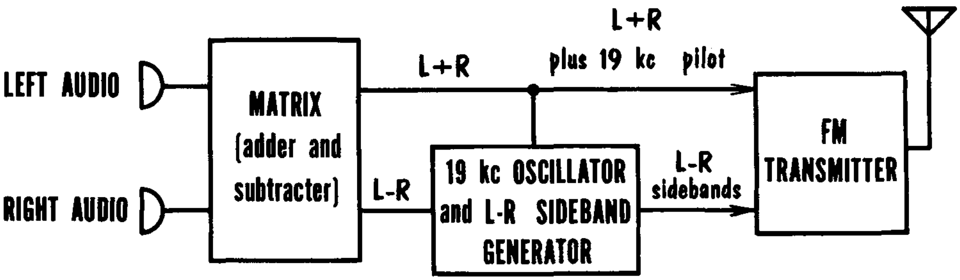
The output of the 6AB4 mixer is coupled to the i-f amplifier stage. Here, the 10.7 mc signal is amplified and fed to the input of the 6BA6 limiter stage.

The limiter stage serves two functions. It removes amplitude variations while providing a driving signal for the discriminator and, in addition, an RC network connected to the grid line develops the avc voltage fed to the input of the i-f amplifier and the input of the r-f amplifier. The discriminator output also serves two functions. It provides the demodulated FM signal or the a-f output and, in addition, delivers a control voltage to the reactance tube for stabilization of the oscillator. Another interesting circuit in this tuner is the 6BR5 tuning indicator. This circuit receives the avc voltage which varies with signal strength. Hence, it provides an indication of when a station is properly tuned.

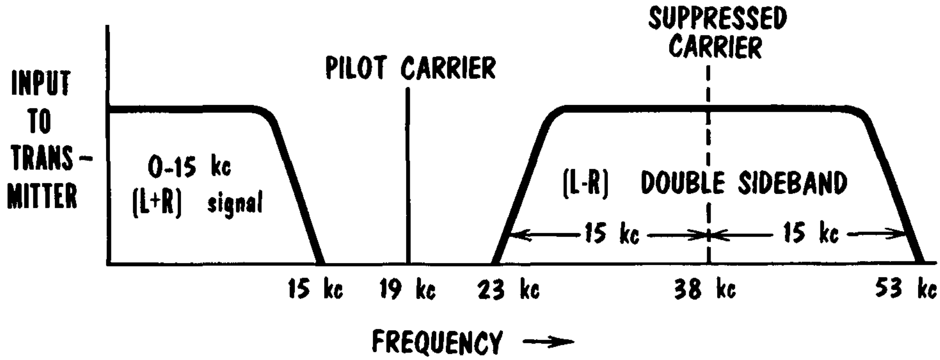
This tuner is especially interesting in that it shows the closed interrelationship of the various circuits. For instance, should the oscillator drift in frequency, an i-f other than 10.7 mc would be produced. This would reflect as lower gain and, hence, a lower avc voltage. In turn, this would be indicated by the tuning indicator. At the same time, the oscillator drift would produce an "error" voltage at the output of the discriminator which would be converted into a d-c control voltage fed to the reactance tube. Finally, this tube would inject the proper reactance to produce a reversal of the oscillator drift. Variations in this circuit are widely used, with one major change – the ratio detector is used in place of the limiter and discriminator circuits. When this is done, the ratio detector provides the demodulated audio output signal, the avc voltage, and the control voltage for the reactance tube.

Multiplex Stereo FM

FM STEREO TRANSMITTING ARRANGEMENT

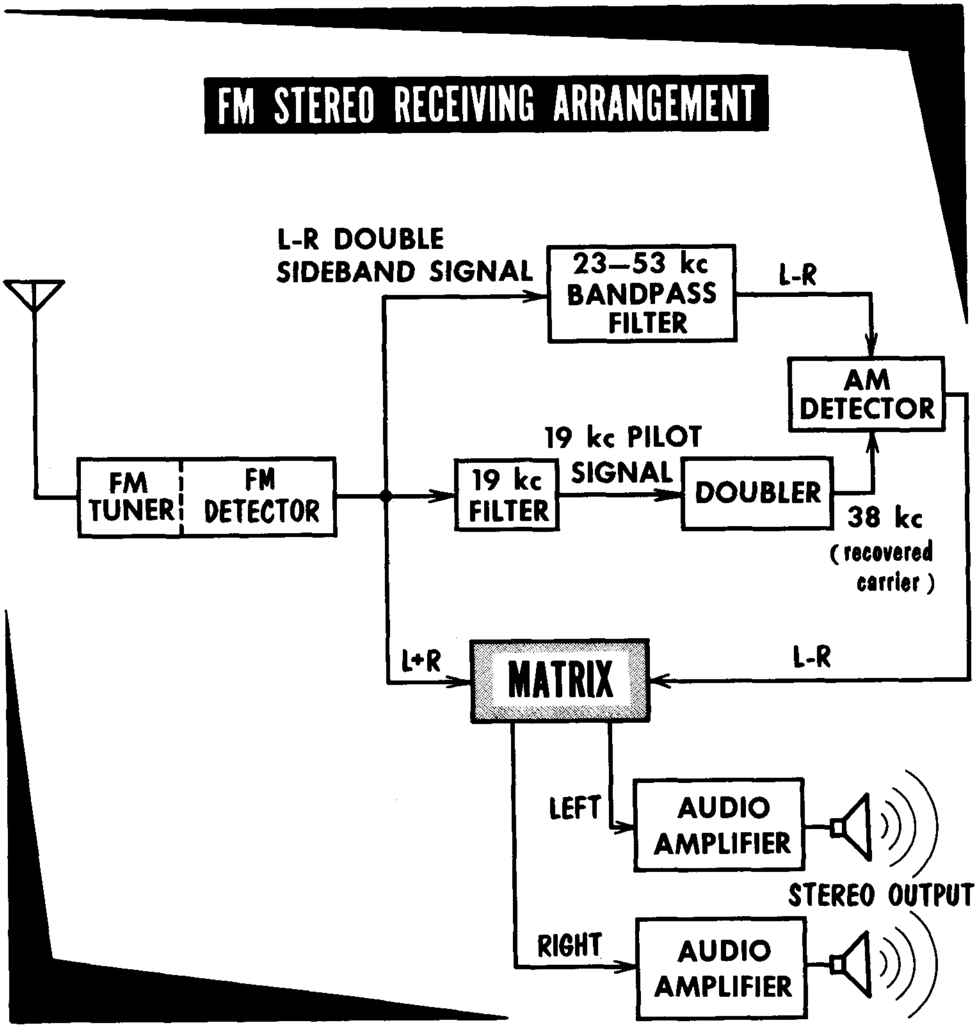


FM STEREO SPECTRUM ORGANIZATION



A new FCC approved system of multiple transmission (multiplex) of stereo audio signals by FM stations is shown in block diagram. The left (L) and right (R) stereo inputs are applied to a matrix section where a sum (L+R) and difference (L-R) signal are developed. The sum (L+R) is fed to the FM modulator directly while the (L-R) signal is routed to a balanced modulator. A 38-kc subcarrier combines with the L-R signals to produce L-R sidebands centered around 38 kc. The subcarrier is eliminated or suppressed during transmission. For detection at the receiver, an accurate 38-kc subcarrier is required. A 19-kc pilot carrier is sent out to either synchronize a 38-kc oscillator in the receiver or develop a 38-kc signal by frequency-doubling techniques. Thus, the spectrum being fed to an FM stereo transmitter might appear as shown. The pilot carrier is set at 19 kc rather than 38 kc since it is evident from the diagram that selective filtering is simple, since 4 kc exists between the pilot carrier and either L+R or L-R signals.

Multiplex Stereo FM (Cont'd)



At the receiver or "adapter", the discriminator output is routed as follows: the L+R signal is fed to a matrix; the 19-kc pilot signal is removed by a sharply selective trap and doubled to 38 kc for use as a subcarrier during L-R detection; the L-R signal plus the reinserted 38-kc are detected and applied to the same matrix as the L+R signal. The matrix accepts L+R and L-R signals and delivers the original L and R signals. Standard pre-emphasis and de-emphasis networks are used at both the transmitter and receiver for L and R channels to improve signal-to-noise ratio.

The left-right sideband signals would not be heard in a conventional monophonic FM receiver, nor would the pilot signal be heard. The system described is completely compatible with present FM broadcasting standards.

FM demodulation must distinguish between frequency variations as compared to AM demodulations which must distinguish between amplitude variations.

In the double-tuned discriminator, one tuned circuit operates above the center frequency, the other tuned circuit operates below the center frequency. The combined output of each tuned circuit provides a continuous "S" curve.

Discriminators are characterized by fair linearity, ease of alignment, and easily derivable afc and avc voltages.

In the Foster-Seeley discriminator there is only one secondary tuned circuit.

In the Foster-Seeley discriminator, the phase difference between the secondary and primary voltages is used to affect the diodes in a manner which provides usable output. The Foster-Seeley discriminator requires a preceding stage of limiting.

Ratio detectors require no limiters, since they are relatively insensitive to AM. The sensitivity is higher than the limiter-discriminator combination, but the circuit is more difficult to align and has less rejection of impulse noise.

A ratio detector splits the rectified voltages in such a way that their ratio is directly proportional to the ratio of the applied i-f voltages which vary with frequency.

The gated-beam tube operates as an excellent limiter, and as a fair limiter-discriminator combination. The response is linear over a wide range, with good rejection of AM. It is one of the easiest of all detectors to align, and has sensitivity comparable to the standard limiter-discriminator.

The afc system used in some receivers takes a voltage from the discriminator detector and applies it to a reactance modulator. This modulator then changes the frequency of the local oscillator to maintain proper tuning.

The automatic-frequency control circuit maintains the incoming signal in proper tune, regardless of drift in the receiver oscillator.

REVIEW QUESTIONS

1. What is the principal disadvantage of the double-tuned discriminator?
2. How is this overcome in the Foster-Seeley discriminator?
3. How does the ratio detector provide immunity from AM superimposed on the FM signal?
4. How does the sensitivity of the ratio detector compare with the limiter-discriminator?
5. Describe the passage of the electron stream in the gated-beam tube.
6. What are the principal advantages of the gated-beam tube as a limiter?
7. Describe the gating action of the two grids of the gated-beam tube, and show how audio voltage is derived.
8. How is the audio output derived from a Foster-Seeley discriminator?
9. How is the audio output derived from the ratio detector?
10. How can an avc voltage be obtained from the ratio detector?
11. Explain the operation of the automatic-frequency-control circuit in FM receivers.
12. Explain how a tube can be made to appear as a reactance in afc circuits.

GLOSSARY

Align: To adjust the tuned circuits of a receiver for maximum signal response.

Amplitude Modulation: A system of superimposing intelligence on a carrier wave by causing the amplitude to vary in accordance with the audio.

Antenna: A device used to radiate or absorb r-f energy.

Antenna Circuit: The inductances, r-f transformers, traps, filters, etc., that couple the antenna to the input of the first r-f stage.

Automatic Frequency Control (AFC): A circuit that holds a radio receiver on the frequency of the station to which it is tuned.

Automatic Volume Control (AVC): A method of automatically regulating the gain of a receiver so that the output tends to remain constant though the incoming signal may vary in strength.

Bandpass: A group ot frequencies passed by a circuit with relatively little attenuation. The bandpass includes those frequencies whose voltage is not less than about 70% of the maximum.

Beat Frequency: A frequency resulting from the combination of two different frequencies. It is numerically equal to the difference between or the sum of these two frequencies.

Broadcast Band: A name given to a band of frequencies extending from 540 kc to 1600 kc. Since modulation up to about 5 kc is used in this band, the entire bandwidth, including sidebands, extends from 535 kc to 1605 kc.

Carrier: The r-f component of a transmitted wave upon which an audio signal or other form of intelligence can be impressed.

Characteristic Impedence: The ratio of the voltage to the current at every point along a transmission line on which there are no standing waves.

Continuous Waves (CW): Radio waves which maintain a constant amplitude and frequency.

Converter Tube: A multi-element vacuum tube used both as a mixer and as an oscillator in a superheterodyne receiver. It generates a local frequency and combines it with an incoming signal to produce an intermediate frequency.

Critical Coupling: The degree of coupling that provides the maximum transfer of energy between two resonant circuits at the resonant frequency.

Crystal Diode: Mineral or crystalline material which allows electrical current to flow more easily in one direction than in the opposite direction, thus converting ac into pulsating dc.

Crystal Filter: Circuit using a crystal as a selective component. Used to discriminate against all signals except those at the center frequency of the crystal.

Decouple: The use of L, C, and R components to control the passage of signals in one circuit and away from another circuit.

De-Emphasis Network: A low-pass R-C filter usually connected at the output of the demodulator that shunts the higher audio frequencies to ground while affecting the lower frequencies to a lesser extent.

Demodulation: See Detection.

Detection: The process of separating the modulation component from the received signal.

Discriminator: A receiver circuit that removes the desired signal from an incoming FM radio signal by changing modulations in terms of frequency variation into amplitude variation.

Fidelity: The faithfulness with which a signal is reproduced by an amplifier. High fidelity is synonomous with low distortion.

Frequency Modulation: A system of superimposing intelligence on a carrier wave by causing the frequency to vary in accordance with the audio.

Ground: A metallic connection with the earth to establish ground potential. Also, a common return to a point of zero r-f potential, such as the chassis of a receiver.

Harmonic: An integral multiple of a fundamental frequency. (The second harmonic is twice the frequency of the fundamental or first harmonic.)

Heterodyne: To beat or mix two signals of different frequencies.

Image Frequency: An undesired signal capable of beating with the local oscillator signal of a superheterodyne receiver which produces a difference frequency within the bandwidth of the i-f channel. It is equal to the r-f signal plus twice the i-f signal.

Intermediate Frequency: The fixed frequency to which r-f carrier waves are converted in a superheterodyne receiver.

Limiting: Removal by electronic means of one or both extremities of a waveform at a predetermined level.

Local Oscillator: The oscillator used in a superheterodyne receiver, the output of which is mixed with the desired r-f carrier to form the i-f.

Loose Coupling: Less than critical coupling; coupling providing little transfer of energy.

Mixer: A vacuum tube or crystal and suitable circuit used to combine the incoming and local oscillator frequencies to produce an i-f.

Modulation: The process of varying the amplitude (AM), the frequency (FM), or the phase (PM) of a carrier wave in accordance with other signals to convey intelligence.

Multiplexing: The simultaneous transmission of two or more signals over the same radio channel.

Padder: A variable capacitor connected in series with the main tuning capacitor to provide frequency adjustments at the low end of a band.

Preselector: A name given to the r-f amplifier section of a receiver. It indicates the stages that precede the mixer or converter stage.

Ratio Detector: An FM demodulator that splits the rectified voltages in such a way that their ratio is directly proportional to the ratio of the applied i-f voltages which vary with frequency.

Reactance Tube: A tube connected across the tank circuit of an oscillator that can act as an inductance or capacitance, and thus cause variations in oscillator frequency. This is used in afc circuits.

Receiver: Any circuit or group of circuits designed to select a particular frequency or group of frequencies, and obtain from them intelligence.

S Meter: A meter connected in a receiver circuit to indicate signal strength.

Saturation Limiting: Limiting the maximum output voltage of a vacuum tube circuit by operating the tube in the region of plate-current saturation (not to be confused with emission saturation).

Selectivity: The degree to which a receiver is capable of discriminating between signals of different carrier frequencies.

Sensitivity: The degree of response of a circuit to signals of the frequency to which it is tuned.

Shielding: A metallic covering used to prevent magnetic or electrostatic coupling between adjacent circuits.

Sidebands: Frequencies, in addition to the carrier frequency, produced by modulating the carrier with an audio signal.

Single Sideband: A system of radio transmission in which one set of sidebands (either upper or lower) is completely suppressed, and the carrier frequency is partly or completely suppressed.

Squelch Circuit: Circuit for preventing a radio receiver from producing a-f output in the absence of a signal having predetermined characteristics.

Stagger Tuning: Method of aligning the i-f stages of a superheterodyne receiver to produce wide bandwidth. This is accomplished by peaking alternate i-f transformers at slightly different frequencies.

Superheterodyne: A receiver in which the incoming signal is mixed with a locally generated signal to produce a predetermined intermediate frequency.

Tight Coupling: Degree of coupling in which practically all the magnetic lines of force produced by one coil link a second coil.

Trimmer Capacitor: A small capacitor connected in parallel with the main tuning capacitor to provide small variations in total capacitance.

Tuning: The varying of an inductance or capacitance to control the resonant frequency of a circuit.

INDEX TO VOL. IV

(Note: A cumulative index covering all six volumes in this series is included at the end of Volume VI.)

basic radio

by MARVIN TEPPER

Electronic Services Division
Raytheon Company

Author of
FUNDAMENTALS OF RADIO TELEMETRY

VOL. 5

JOHN F. RIDER PUBLISHER, INC., NEW YORK

Copyright September 1961 by John F. Rider Publisher, Inc.

All rights reserved. This book or any parts thereof may not be reproduced in any form or in any language without permission of the publisher.

Library of Congress Catalog Card Number 61-11229

Printed in the United States of America

Fifth Printing, 1968

PREFACE

The purpose of this book is to fill a need for a text stating in plain, everyday language, what many people consider a complex technical subject. A technical subject need not be complex. Careful filling in of the background with essential information, and then leading step by step to the final explanation, provides a logical method of explaining the most difficult subject.

It would be impossible to cover in a single book or series of books, the immense scope implied in the word *electronics*. However, an understanding of radio circuits serves as a foundation for advanced study in all fields of electronics, such as television, radar, computers, etc. For teaching radio, the all-important basic tool of electronics, most available textbooks are woefully inadequate. One type contains information so brief as to acquaint rather than instruct. Another type is based on the premise that teaching a student to design a circuit is the best method of having him understand that circuit's operation.

Basic Radio represents the neglected middle ground. It is a course in radio communications, as distinct from a general course in electronics. The text deals with the circuitry and techniques used for the transmission and reception of intelligence via radio energy. Assuming no prior knowledge of electricity or electronics, the six volumes of this course "begin at the beginning" and carry the reader in logical steps through a study of electricity and electronics as required for a clear understanding of radio receivers and transmitters. Illustrations are used on every page to reinforce the highlights of that page. All examples given are based on actual or typical circuitry to make the course as practical and realistic as possible. Most important, the text provides a solid foundation upon which the reader can build his further, more advanced knowledge of electronics.

The sequence of *Basic Radio* first establishes a knowledge of d-c electricity. Upon this is built an understanding of the slightly more involved a-c electricity. Equipped with this information the reader is ready to study the operation of electron tubes and electron tube circuits, including power supplies, amplifiers, oscillators, etc. Having covered the components of electronic circuitry in Volumes 1 through 3, we assemble these components

in Volume 4, and develop the complete radio receiver, AM and FM. In Volume 5 we recognize the development of the transistor, and devote the entire volume to the theory and circuitry of transistor receivers and semiconductors. The last volume of the course, Volume 6, covers the long-neglected subject of transmitters, antennas, and transmission lines.

No prior knowledge of algebra, electricity, or any associated subject is required for the understanding of this series; it is self-contained. Embracing a vast amount of information, it cannot be read like a novel, skimming through for the high points. Each page contains a carefully selected thought or group of thoughts. Readers should take advantage of this, and study each individual page as a separate subject.

Whenever someone is presented with an award he gives thanks and acknowledgement to those "without whose help . . ." etc. It is no different here. The most patient, and long-suffering was my wife Celia, who typed, and typed, and typed. To her, the editorial staff of John F. Rider, and others in the "background", my heartfelt thanks and gratitude for their assistance and understanding patience.

MARVIN TEPPER

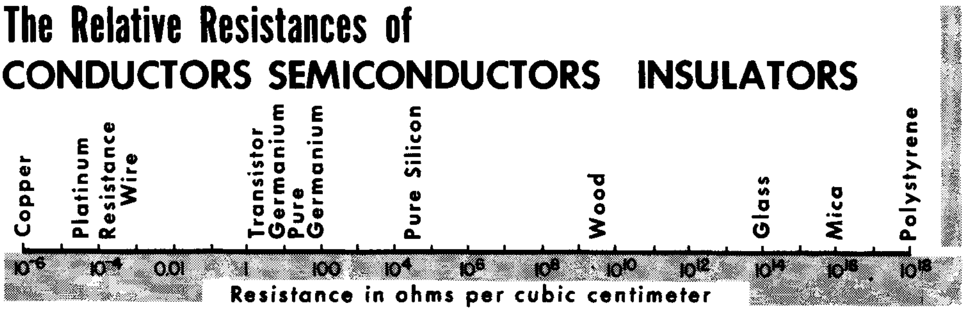
Malden, Mass.
September 1961

CONTENTS

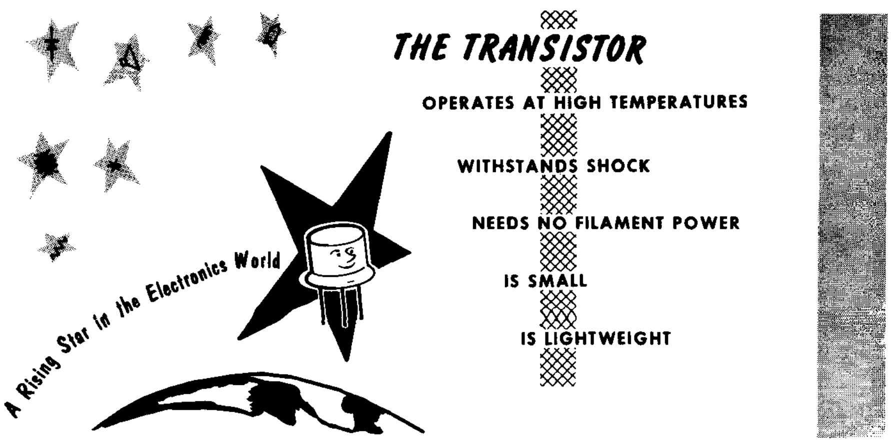
VOL. V — BASIC RADIO

The Transistor--Introduction

Developed after many years of laboratory research, the transistor has grown in popularity at an astounding rate. Transistors are made of semiconductor materials; that is, materials whose ability to conduct an electric current falls somewhere between that of an insulator such as mica or glass, and that of a conductor such as silver or copper. Silicon and germanium are the most widely-used materials in the manufacture of transistors. These materials, which in their pure state are very poor conductors, gain important electrical conductivity characteristics with the addition of certain impurities.



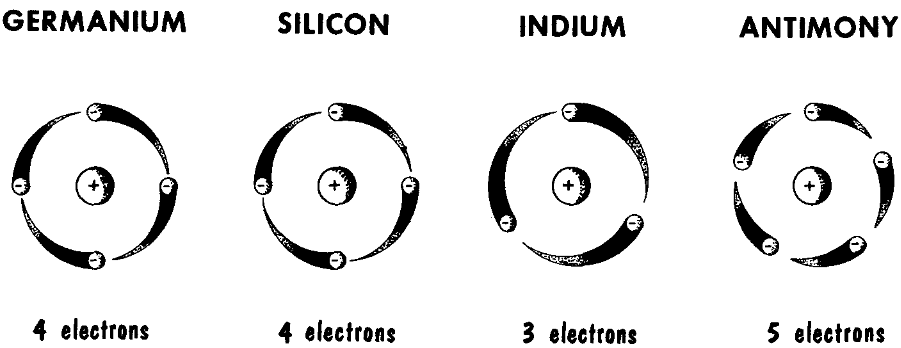
The transistor was at first a laboratory curiosity, of interest mainly because it was a "solid-state" device that could be made to amplify. Improvements, particularly in the manufacturing process, rapidly made it a practical device. The transistor has already achieved a prominent place in the electronics field and is replacing the vacuum tube in numerous applications. In comparison to vacuum tubes, the transistor is extremely small, lightweight, does not require filament power, is virtually immune to mechanical shock and, in general, can operate well at relatively high temperatures. In the following pages, we will discuss semiconductors, transistors, and practical radio circuits containing transistors. We start with a study of semiconductors because this is the "stuff" of which transistors are made.



Atomic Structure of Semiconductors

An atom of any substance consists of a nucleus made up of neutral particles (neutrons) and positively-charged particles (protons) surrounded by a group of negatively-charged particles (electrons). The electrons travel in orbits around the nucleus. In a neutral atom, there are as many electrons as protons.

SIMPLIFIED DIAGRAMS OF ATOMS



(Only outermost orbits are shown)

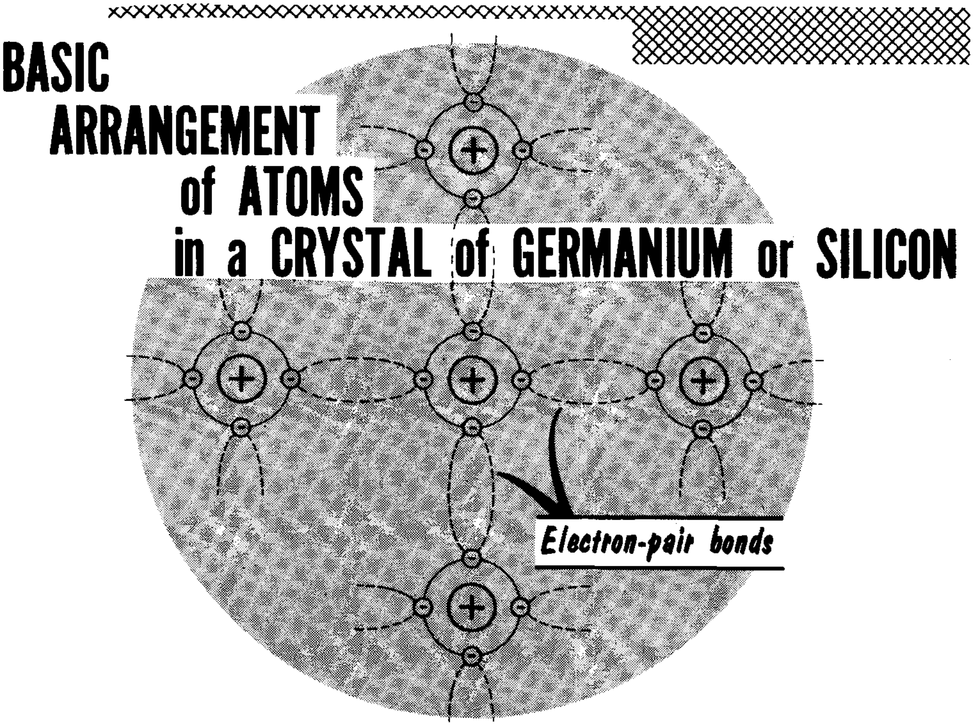
In our study of semiconductors, we will be primarily concerned with the silicon, germanium, indium, and antimony atoms. The silicon atom has a total of 14 electrons in 3 orbits. Four of these electrons are in the outermost orbit. The germanium atom has 32 electrons in 4 orbits. Again, there are four electrons in the outermost orbit. The indium atom contains 49 electrons in 5 orbits, with three electrons in the outermost orbit. Of the 51 electrons in the antimony atom (also arranged in 5 orbits), five are in the outermost orbit. The outermost orbit of these atoms can hold 8 electrons. Thus, none of the above atoms has a complete outer orbit.

In dealing with semiconductors, we are concerned only with the nucleus and the outermost electrons. When shown in the simplified diagram, the atoms of germanium and silicon appear identical because they have the same number of electrons in their outermost orbits. We are concerned primarily with the electrons in the outer ring or orbit (called valence electrons) because they determine the chemical characteristics of the atom or element. Since germanium and silicon have the same number of valence electrons, they have similar chemical characteristics.

Semiconductor Crystals

In a single crystal of pure germanium or silicon, the individual atoms are aligned in a structure which we descriptively call a lattice. The lattice effect is created because all adjacent atoms are symmetrically equidistant. Each atom is held in place by bonds formed between each of its outermost electrons to each of the four adjacent atoms. Each outermost electron thus becomes linked in an electron-pair bond. Thus, a perfect germanium crystal has neither an excess nor a deficiency of electrons. All of the electrons in the crystal are either bound to the nuclei of the atoms or to each other in electron-pair bonds.

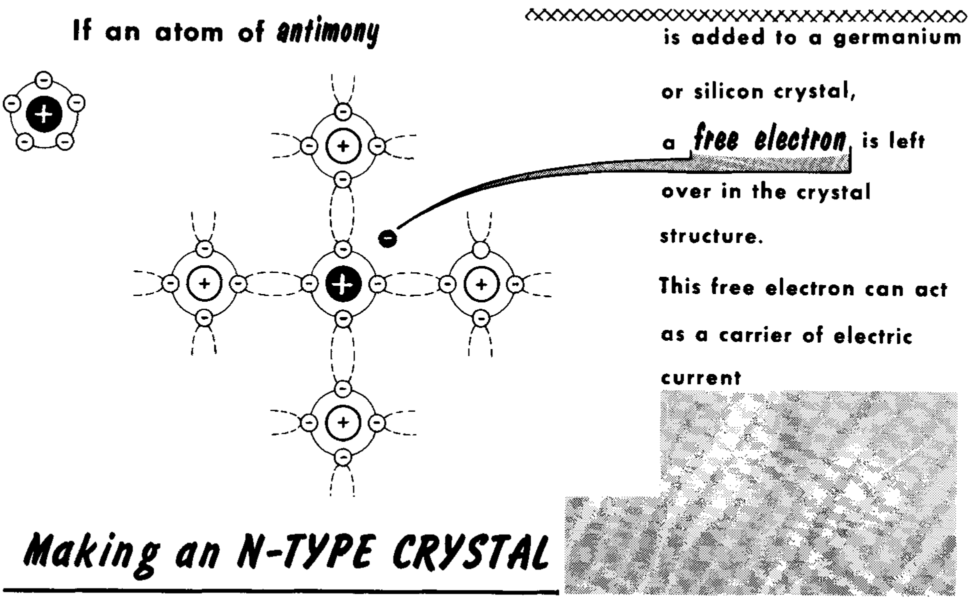
From the illustration, it might appear that the electrons in the outer ring can easily be displaced by the application of a voltage. This is not the case. By forming covalent bonds, or electron pairs, between neighboring atoms, the atoms behave in many respects as though their outer ring were complete (8 electrons). This makes them extremely inactive since there are no free electrons; their outer orbits each contain four electrons, and they "share" four additional electrons.



A perfect germanium crystal is an electrical insulator and of very little use in transistor work. However, with the addition of a small amount of chemical impurities, heat, or light energy, the conductivity of the germanium is improved and it becomes a semiconductor.

N-Type Semiconductors

If we introduce an impurity such as an atom of antimony (which has five electrons in its outer orbit) into a crystal of germanium or silicon, the crystal lattice as a whole is not changed. The antimony atom merely replaces a germanium or silicon atom, with four of its outermost electrons joining in covalent bonds with those of adjacent atoms. The remaining electron is free to move through the crystal. This free electron aids the conduction of electricity; the antimony-enriched germanium or silicon crystal conducts current more easily than does a pure crystal. Because only small quantities (about 1 part in 100,000,000) of impurities are added, the crystal does not become a good conductor; rather it falls somewhere between a conductor and an insulator. For this reason, these crystals are called semiconductors.

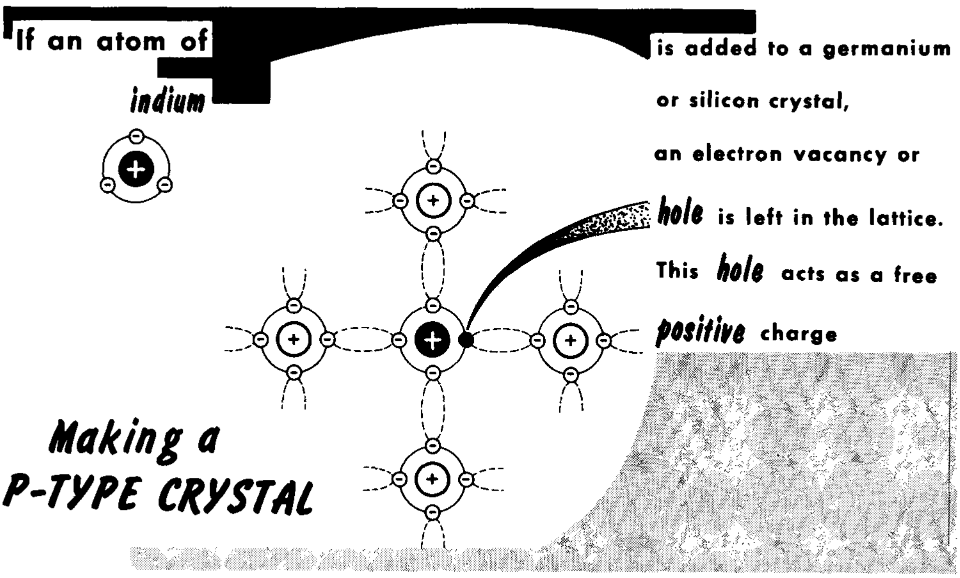


Since the introduction of antimony adds a free electron to the crystal, antimony is called a donor. Although the crystal itself remains neutral, it has available an unbound negative charge and, therefore, is called an n-type semiconductor. The adding of controlled amounts of impurities to a pure crystal is called doping. Because antimony has 5 electrons in its outer ring, it is called a "pentavalent" material. Another pentavalent substance frequently used is arsenic.

When a voltage is applied across a piece of n-type semiconductor material, there will be an electron flow through the crystal, constituting an electric current. The free electrons will flow toward the positive terminal and be repelled away from the negative terminal. If the voltage were reversed in polarity, all conditions would remain the same except that the direction of current would be reversed. N-type material is a "linear" conductor.

P-Type Semiconductors

If we add a single atom of indium to a crystal of germanium or silicon, a different type of crystal results because indium has only three electrons (trivalent) in its outer orbit. When the indium atom replaces one of the germanium or silicon atoms in the lattice, the crystal becomes deficient in one electron. To fill the vacancy thus created, an electron must be borrowed from some other part of the crystal. The borrowed electron moves into the hole created by adding the indium. In moving to fill the hole, the borrowed electron leaves another hole behind it. Thus, while the electrons move in one direction, the holes left by them move in the opposite direction. Because the loss of an electron (which is a negative charge) results in effect in a positively-charged particle, holes may be thought of as free positive charges that can act as current carriers.



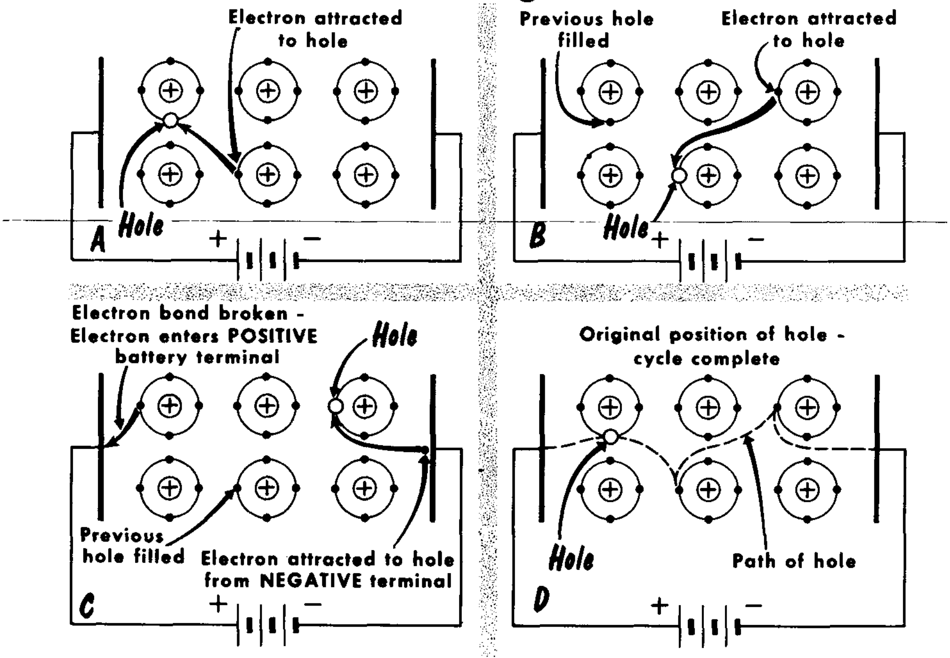
Because the indium atom takes an electron from some other atom in the crystal, it is called an acceptor. Since the crystal has a free positive charge available for electrical conduction, it is called a p-type semiconductor. The crystal as a whole remains electrically neutral.

When a voltage is applied across p-type material, the holes, having an effective positive charge, will be attracted to the negative terminal and repelled from the positive terminal. This drift of holes constitutes an electric current, and is equivalent to a flow of electrons in the opposite direction. Each time a hole reaches the negative terminal, an electron is emitted from the negative terminal into the hole in the crystal to neutralize it. At the same time, an electron from a covalent bond enters the positive terminal to leave another hole in the crystal. This hole then begins its drift toward the negative terminal.

Characteristics and Properties of Holes

For an understanding of transistors, we must think of a hole as a specific particle. Holes in motion are the same as electrons in motion--they both make up an electric current. However, we must consider the differences that do exist. A hole can exist only in a semiconductor material such as silicon or germanium because it depends for its existence on a specific arrangement of electrons (electron-pair bonds) and atoms as found in crystal materials. Holes do not exist in conductors such as copper. Another important consideration is that holes are deflected by electric and magnetic fields in exactly the same way as electrons. However, since holes possess a charge equal and opposite to that of the electron, the direction of deflection of the hole is opposite to that of the electron.

MOVEMENT of a HOLE through a SEMICONDUCTOR

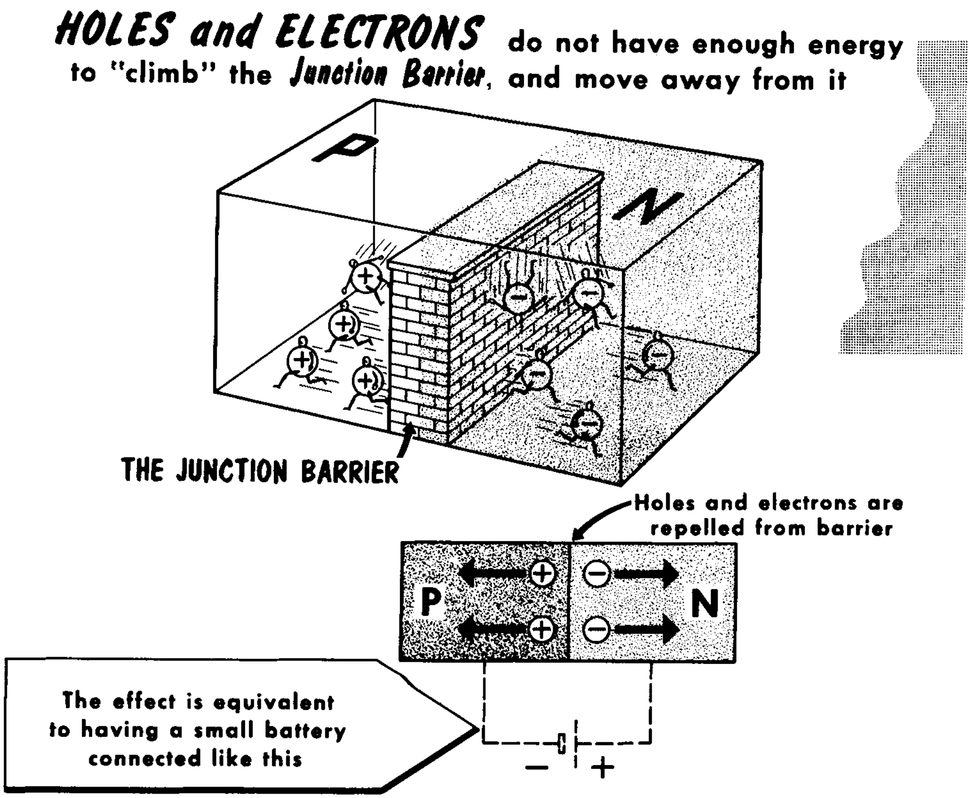


When a hole is filled by a free or excess electron from an adjacent electron-pair bond, the hole no longer exists. When a voltage is applied across a piece of p-type germanium, holes are repelled by the positive terminal and attracted to the negative terminal. When a hole reaches the negative terminal, an electron from the battery (see illustration) enters the germanium and fills the hole. When this occurs, an electron from an electron-pair bond in the crystal near the positive terminal breaks its bond and enters the positive terminal. The breaking of the bond creates another hole which begins to drift toward the negative terminal. This action provides a continuous flow of electrons in the external circuit and holes in the germanium.

The Junction Barrier

If an n-type crystal is joined to a p-type crystal, electrons do not flow across the junction to fill or neutralize the holes. The antimony atoms (actually, ions) in the n-type crystal are positively charged because they have given up electrons. Thus, they repel the positive holes in the p-type crystal. Similarly, the indium atoms in the p-type crystal, in accepting an electron, become negatively charged and repel electrons in the n-type crystal. Free electrons and holes in the n- and p-type material therefore move away from the p-n junction rather than toward it; the junction may be thought of as a barrier to the passage of current carriers. The combined crystal thus acts as if a small equivalent battery were placed across the junction, thereby establishing a small voltage across it. If the current carriers are to pass, the barrier must be eliminated.

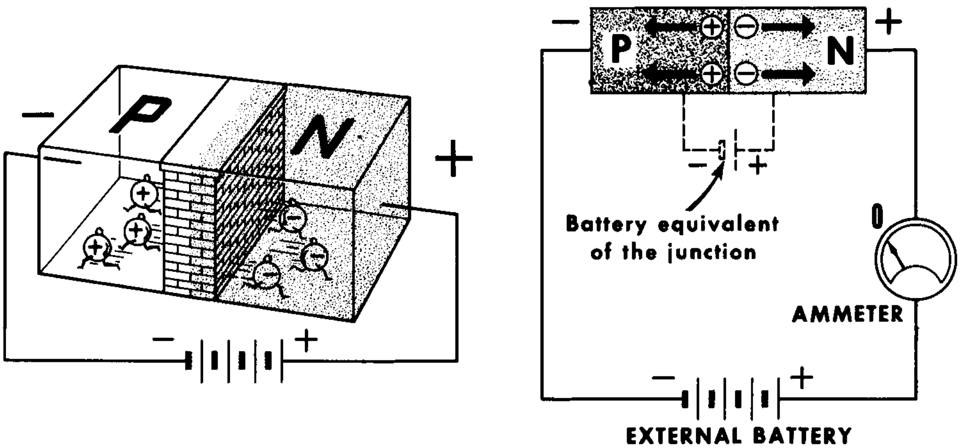
A p-n junction cannot be formed merely by placing a piece of n-type material against a piece of p-type material. The junction is formed by taking a single crystal of pure germanium and treating one section with a trivalent impurity and one section with a pentavalent impurity. The semiconductor device we get from this process is called a junction diode. As we shall see, this p-n junction is no longer a linear device. Current will flow much easier in one direction than in the other.



Operation of P-N Junction--Reverse Bias

Let us review an important point. It might be thought that in time, the electrons in the n-type germanium would, by diffusion, occupy the holes in the p-type germanium, thereby neutralizing the entire crystal. This does not occur because the electrons and holes tend to drift apart. In n-type material, the atoms of the pentavalent impurity have a positive charge; in p-type material, the atoms of the trivalent impurity carry a negative charge. These relatively fixed atoms repel the charges in the opposite piece of material--the positive atoms in the n-type material repel the holes or positive charges in the p-type material, and vice-versa. This action provides a battery equivalency across the p-n junction.

REVERSE BIAS RAISES the JUNCTION BARRIER
and NO current flows

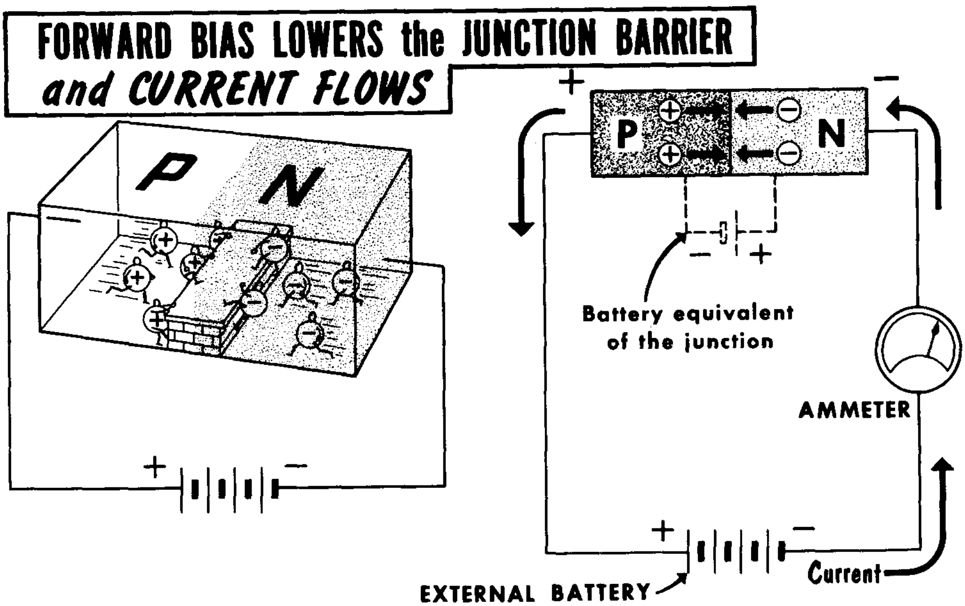


Let us now connect an external battery across our p-n crystal with the same polarity as the junction-barrier voltage. The external battery voltage will add to the equivalent junction voltage. The negative terminal is connected to the p-type germanium and the positive terminal to the n-type germanium. The positive holes are attracted toward the negative terminal and the negative electrons toward the positive terminal. Note that in both attractions, the holes and the electrons are attracted away from the p-n junction. This action effectively increases the junction barrier height. With electrons and holes repelled away from the junction, there will be no current flow of electrons or holes through the germanium. As we will see, current flow takes place only when the holes and electrons pass through the junction barrier.

With the battery connected as shown, the p-n junction is biased in the non-conducting or reverse-bias direction. Should the reverse bias be excessively high, the crystal structure may break down and be damaged permanently.

Operation of a P-N Junction--Forward Bias

We will now connect the positive terminal of the external battery to the p-type germanium and the negative terminal to the n-type germanium. The holes are now repelled from the positive terminal of the battery and drift toward the p-n junction. The electrons are repelled from the negative terminal of the battery and also drift toward the junction. Under the influence of the battery voltage, the holes and electrons penetrate the junction and combine with each other. For each combination of an electron and a hole, an electron from the negative terminal of the external battery enters the n-type germanium and drifts toward the junction. Similarly, an electron from an electron-pair bond in the crystal, near the positive terminal of the external battery, breaks its bond and enters the positive terminal of the battery. For each electron that breaks its bond, a hole is created which drifts toward the junction. Recombination around the junction region continues as long as the external battery is connected.



Note that there is a continuous flow of electron current in the external circuit. The current in the p-type germanium consists of holes; the current in the n-type germanium consists of electrons. In this condition, the p-n junction is said to be biased in the forward direction. If the forward bias is increased, the current is increased.

In forward bias, the external battery voltage opposes and overcomes the junction-barrier voltage, which may be only a few tenths of a volt. This lowering of the junction barrier permits a free flow of current. Of course, excessive forward bias would produce excessive current, with a possibility of a crystal-structure breakdown.

The Junction Diode as a Rectifier

We have seen that the p-n junction is a unilateral device; that is, when forward biased, it will permit current to flow, and when reverse biased, it will not permit current to flow. These, then, are the basic ingredients of a diode. Below, we see a curve showing current flow through a junction diode as the bias voltage is varied and reversed in polarity. Note that current flow in the forward-bias direction is quite high--measured in milliamperes. However, current flow in the reverse-bias direction, although very low and measured in microamperes, is not zero. The reverse-bias current flow occurs because some acceptor ions and their associated holes occur in the n-type germanium, and some donor ions and their associated excess electrons occur in the p-type germanium. The holes found in n-type germanium and the excess electrons in the p-type germanium are called minority carriers because they are so few in number, compared with the holes found in p-type material and the electrons in n-type germanium, which are called majority carriers.

RECTIFIER ACTION in VACUUM TUBE and JUNCTION DIODE

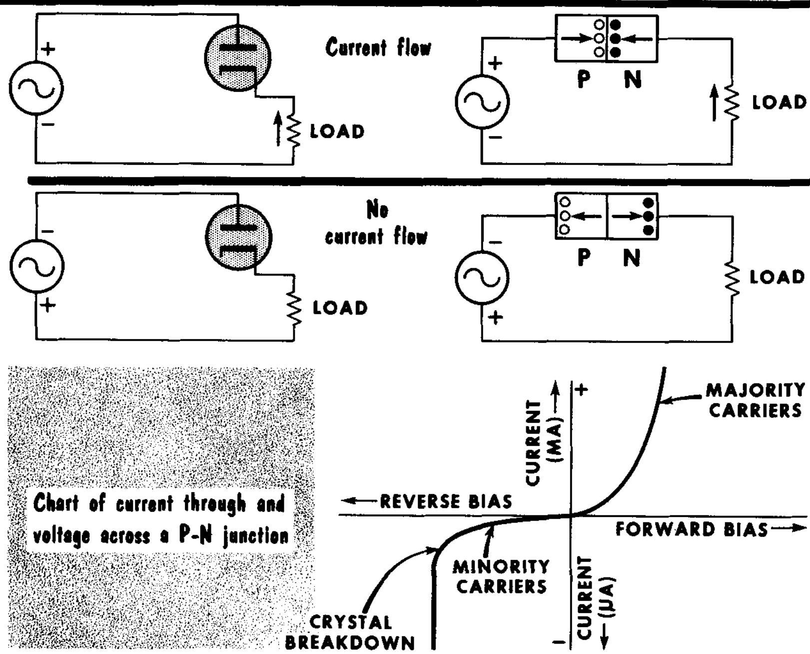
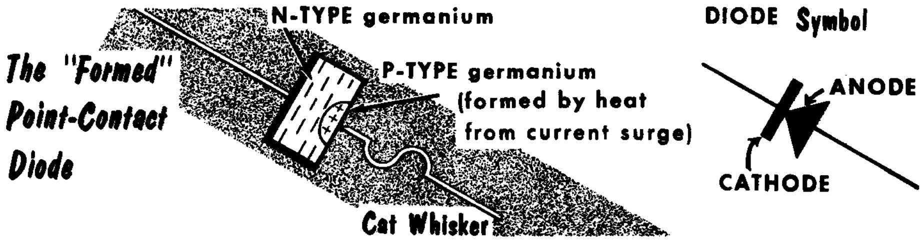


Chart of current through and voltage across a P-N junction

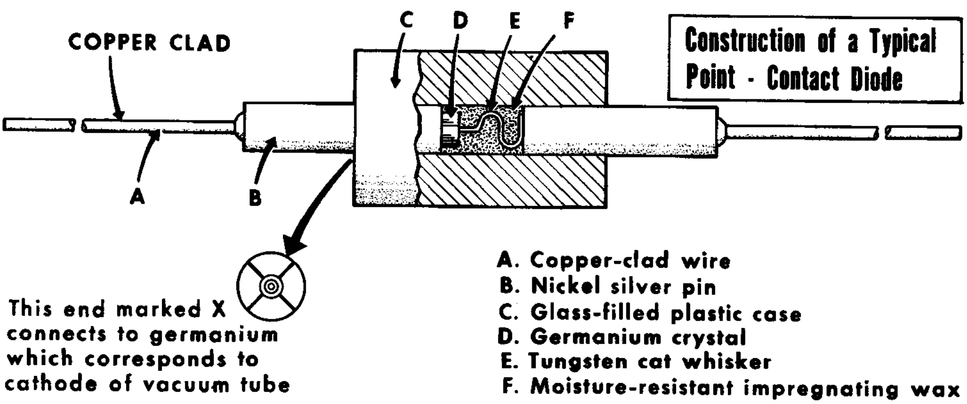
Note that when a very high reverse bias is applied, a high reverse current flows. This current is not due to the minority carriers, but to a breakdown of the crystal structure. The point at which the reverse voltage is high enough to break down covalent bonds and cause current flow is called the Zener breakdown voltage. This voltage has the same importance as the inverse voltage rating of a vacuum tube, since it defines the maximum reverse voltage that can be applied to a junction without excessive current flow.

The Point-Contact Diode

Up until this point, we have given all our attention to the junction diode. There is a second type commonly used--the point-contact diode. This unit is an outgrowth of the old galena crystal detector, which consisted of a piece of galena (a lead ore) mounted so that an irregular surface was exposed to the point of a wire (cat whisker). The wire was moved around to find a sensitive spot for optimum radio detection.



The modern counterpart of this is the germanium crystal point-contact diode. Here, the germanium replaces the galena, and the cat whisker consists of a length of wire such as tungsten about 0.005" diameter. In a modern unit such as the 1N34 or 1N81, the cat whisker is fitted to the germanium crystal at the factory, and the entire unit is sealed. In practice, the germanium consists of n-type material. The unit is "formed" by passing a large momentary surge of current across the junction of the wafer and the whisker. The heat produced by this current forces some electrons away from the area of the point, leaving holes. This produces a small p-type region around the point of the cat whisker.

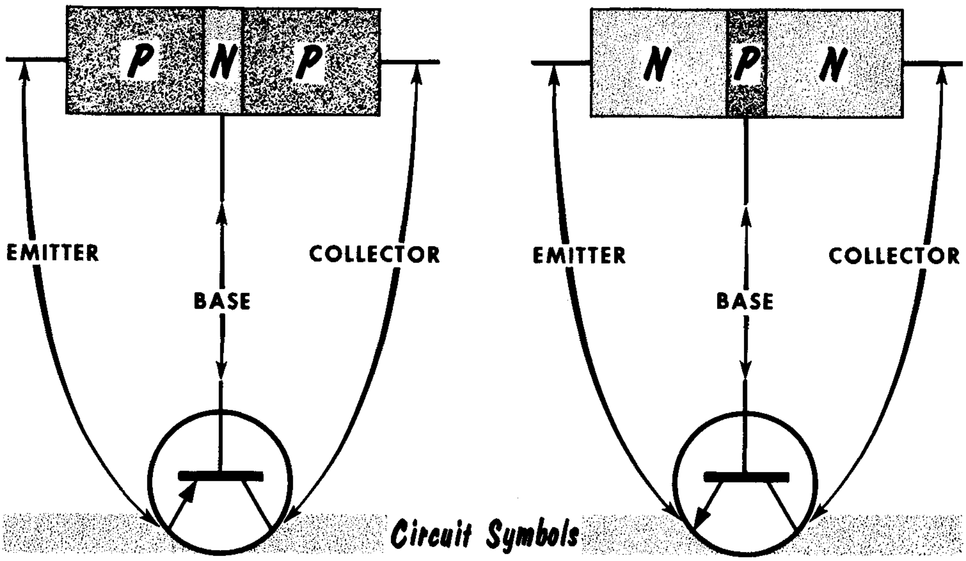


Thus, we have all the ingredients of a junction diode--a p- and an n-region. However, since the p region is so tiny, there is very little capacitance across the junction. A typical shunt capacitance might be as low as 0.8$\mu\mu f$. This makes the point-contact diode highly desirable for high-frequency work such as in video detectors and microwave mixers.

What is a Transistor?

So far, we have discussed p- and n-type materials and their actions. Going one step further, we have observed the action of p-n junctions under conditions of forward and reverse bias. We are now prepared to study the transistor. We can consider a junction transistor as being composed of two separate p-n junctions "tied" together. One basic type of junction transistor is known as the p-n-p. In this transistor, there is a very thin layer of n-type semiconductor sandwiched between two much thicker layers of p-type semiconductor material. The second basic type of junction transistor is the n-p-n. Here, the opposite condition exists. A thin layer of p-type semiconductor material is sandwiched between two much thicker layers of n-type material. As we will see, it is very important that the center layer of semiconductor material be made extremely thin.

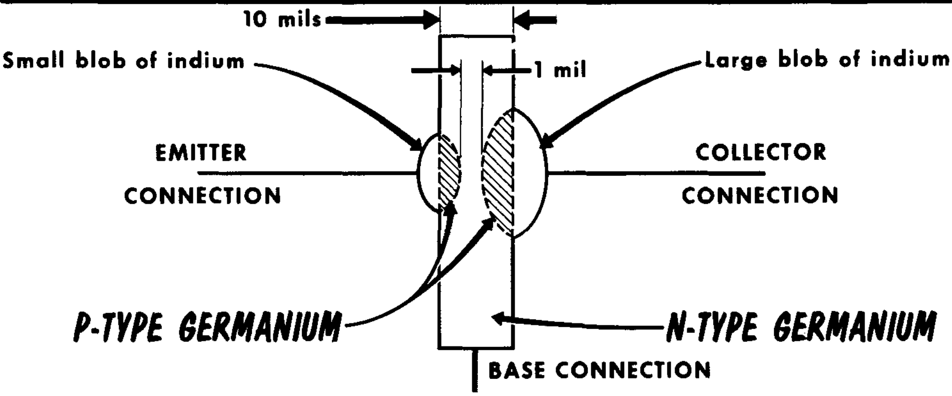
P-N-P and N-P-N TRANSISTORS

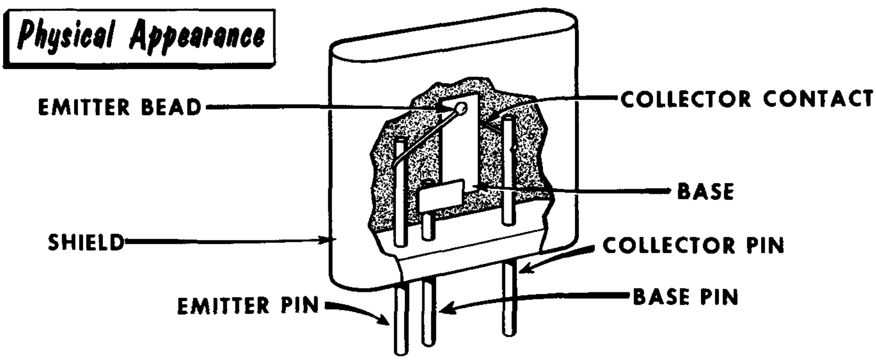


Once a p-n-p or n-p-n junction transistor is formed, electrical connections are made to the layers. The connection to the thin center layer is called the base, and can be compared roughly to the control grid of a vacuum tube. The connection to one of the outer layers is called the emitter, often compared to the cathode of a vacuum tube. The connection to the other outer layer is called the collector, often compared to the plate of a vacuum tube. In some transistors, the outer layers are symmetrical and can be used either as the emitter or collector, depending upon circuit biasing voltages. However, it is much more common practice to make the junctions asymmetrical, and the manufacturer will identify each lead as being either the emitter, the base, or the collector. The illustration shows the physical representation of p-n-p and n-p-n transistors together with their circuit symbols.

The Diffused Alloy-Junction Transistor

The ALLOY-JUNCTION TRANSISTOR (P-N-P)



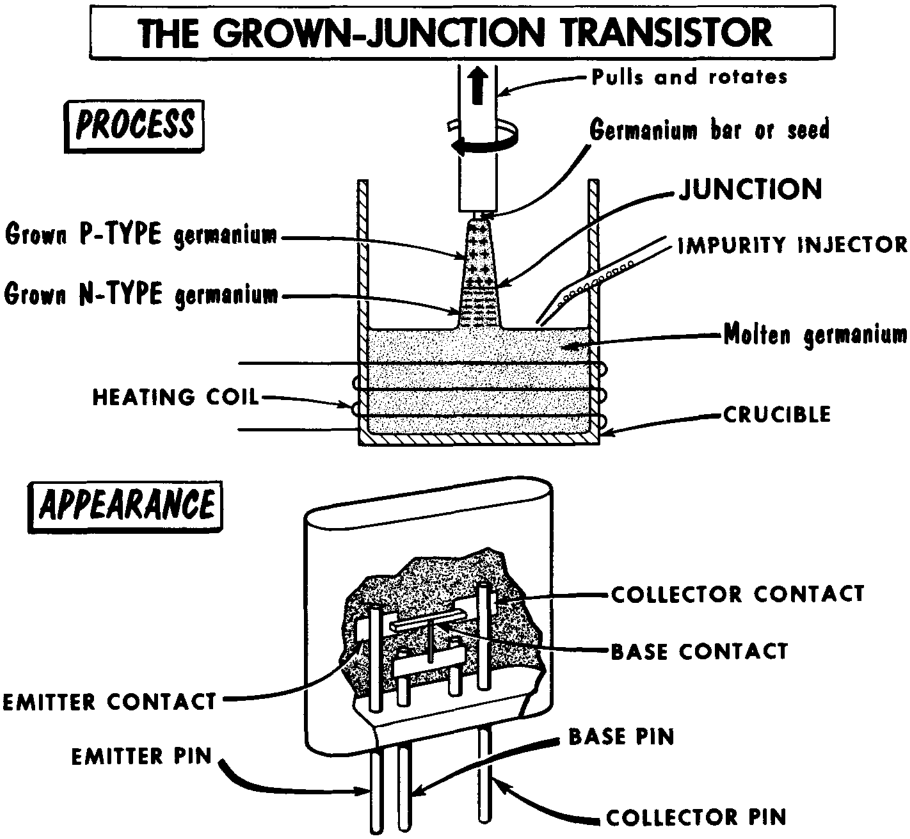


A p-n junction cannot be made merely by joining a pellet of p-type crystal to a pellet of n-type crystal. Special techniques must be used and have been developed. For example, to make a p-n-p alloy-junction transistor, the process starts with a piece of germanium about 10 mils or .01" thick. A small pellet (impurity) of trivalent material, such as indium, is then placed near the center of the two opposing faces of the germanium wafer. This combination is then heated in an oven to a temperature somewhere between the melting temperature of germanium and indium. This causes the indium pellets to diffuse into, or alloy with, the germanium wafer, to form areas or regions of p-type conductivity separated from each other by a thin layer having n-type conductivity of the original wafer. Connecting leads are then soldered to each indium pellet and to the germanium crystal. The assembly is then hermetically sealed in a lightproof container. Since the collector is usually required to have a greater heat dissipation than the emitter, the indium pellet used as the collector is larger than the emitter.

To make n-p-n alloy-junction transistors, the same technique is used except that the impurity added is usually arsenic, a pentavalent element.

The Grown-Junction (Rate-Grown) Transistor

The grown-junction or rate-grown transistor is made, basically, by pulling a bar of germanium from a bath of molten germanium. In this process, both acceptor and donor impurities are present in nearly equal amounts throughout the entire crystal. One is made preponderant over the other by changing the conditions of growth, particularly the temperature and the rate of withdrawal of the crystal.

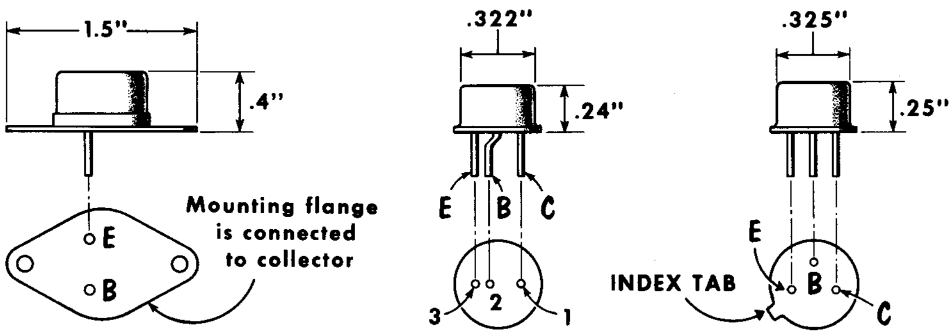


A small crystal or "seed" is dipped into the molten germanium and slowly raised. The molten germanium "sticks" to the seed and forms a long, single crystal. During the process, the crystal is rotated and impurities added so that the germanium changes from p-type to n-type, depending upon the temperature of the molt and the rate of pulling. As acceptor and donor impurities are alternately added, the p- and n-type layers are formed. The bar is then sawed into pieces with -- for n-p-n transistors -- a thin p-type layer in the center and long n-type layers on each side. An advantage of rate-growing is the excellent control over the base region, which can be made extremely thin--a few tenths of a mil thick. These junctions are very good for high-frequency work.

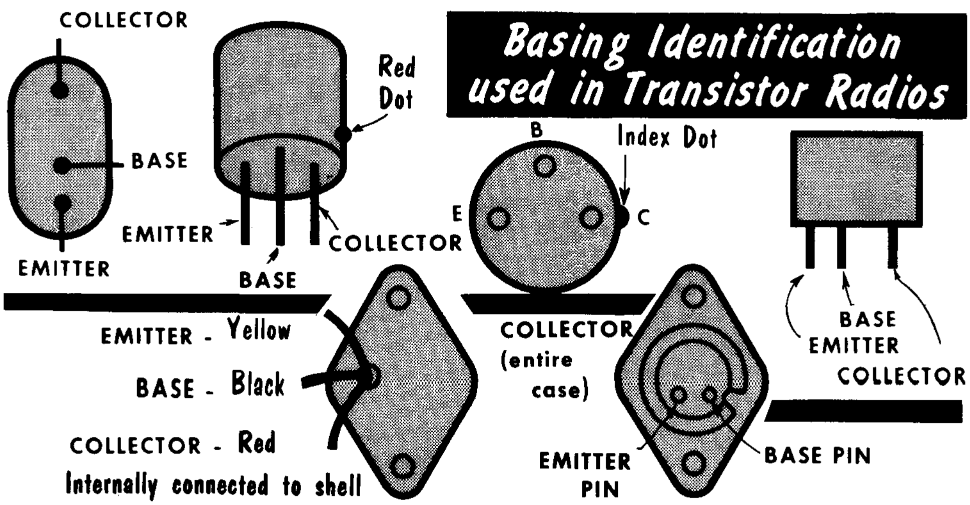
Transistor Basing and Construction

Several transistor outline drawings are shown on this page to illustrate the various physical shapes to be found in the transistor field, together with the location of the emitter, base, and collector leads. Among the typical transistors shown, note that the diameter may be as little as 0.322", or less than one-third inch. The leads are generally thin, tinned wires that must be handled carefully. On some large power transistors (to be discussed later), the collector may be connected to the case and the entire unit bolted to the chassis.

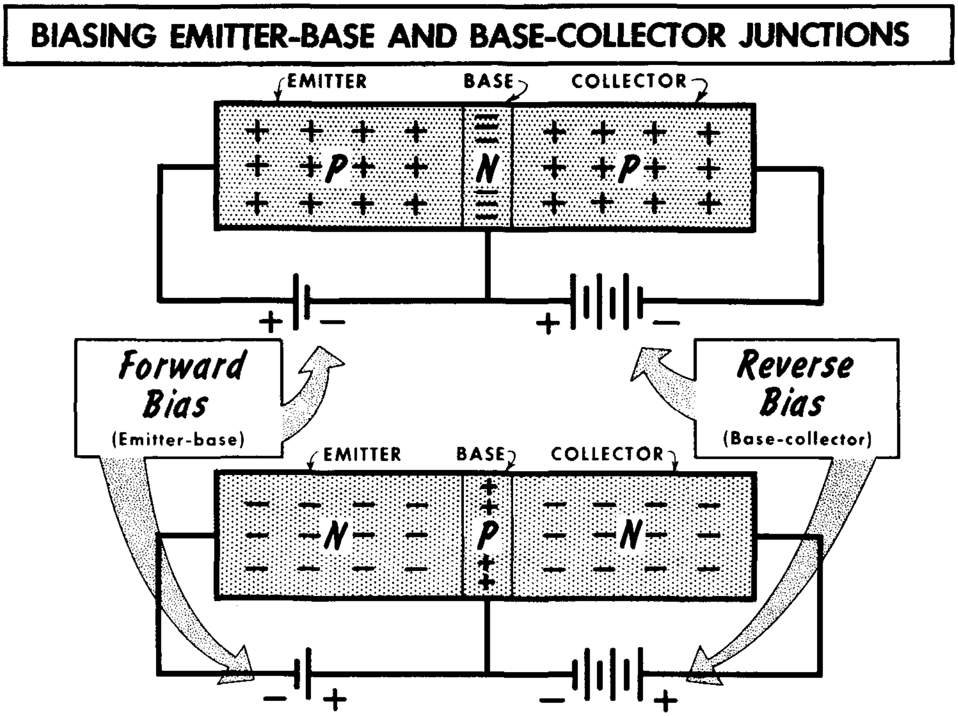
Viewed from the Side or Bottom, the *TYPICAL TRANSISTOR* Looks Like This



To identify the leads, an index or color dot is often placed next to the collector terminal. Sometimes, a small index tab is used. It is always good practice to check the manufacturer's schematic for transistor lead identification before any testing is done.



Operation of the P-N-P Transistor

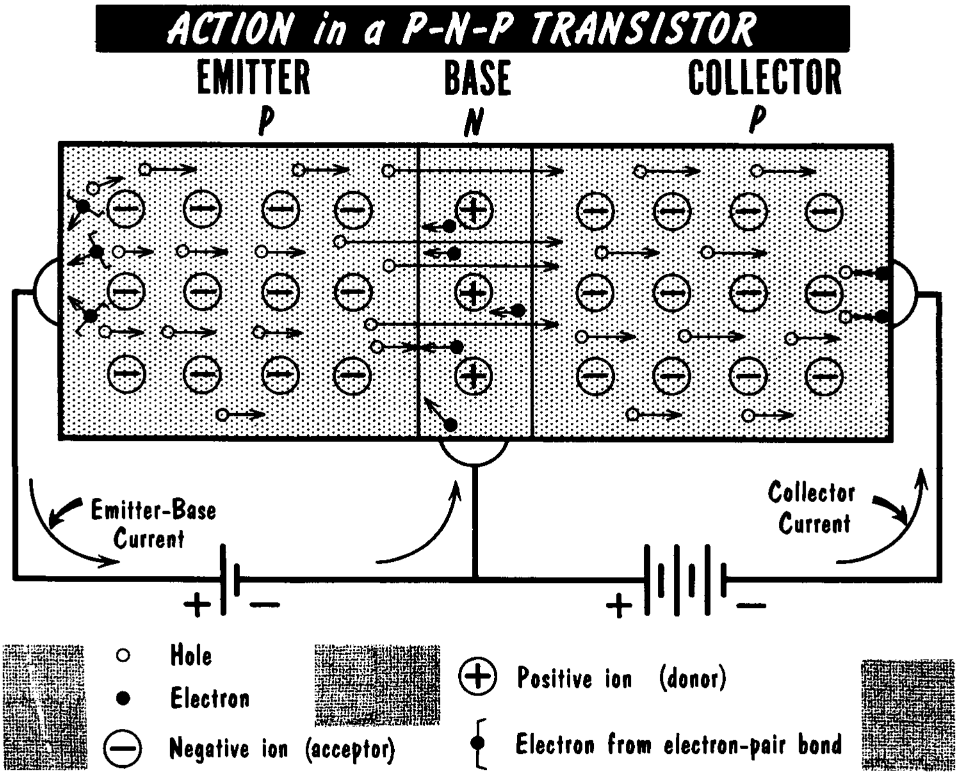


Let us observe the action of a p-n-p transistor. In this unit, there are two p-type sections, the emitter and collector, separated by a thin layer of n-type material, the base. Note that the emitter-base p-n junction is biased in a forward direction. In this connection, the external battery opposes the internal (barrier junction) voltage developed at this junction. In short, with the emitter biased positively with respect to the base, we say this p-n junction is forward biased. The p-n collector-base junction is biased in the opposite direction. In this way, the external battery aids the internal junction voltage, and we say that the collector is biased negatively (reverse bias) with respect to the base.

If no voltage were applied to this transistor, the holes in the emitter would move to the left and the electrons in the base would move to the right, both because of the internal junction barrier voltage. However, when an emitter-base forward-bias voltage is applied, the holes in the emitter and the electrons in the base move toward the junction, both moving in the direction of the battery terminal which attracts them. At the emitter-base junction, some holes and electrons combine with each other and are neutralized. However, because of the extreme thinness of the base layer, and because of the attraction of the relatively high negative collector voltage, almost all of the holes pass or "diffuse" through the base and produce a "hole current" between the emitter and the collector.

Operation of the P-N-P Transistor (Cont'd)

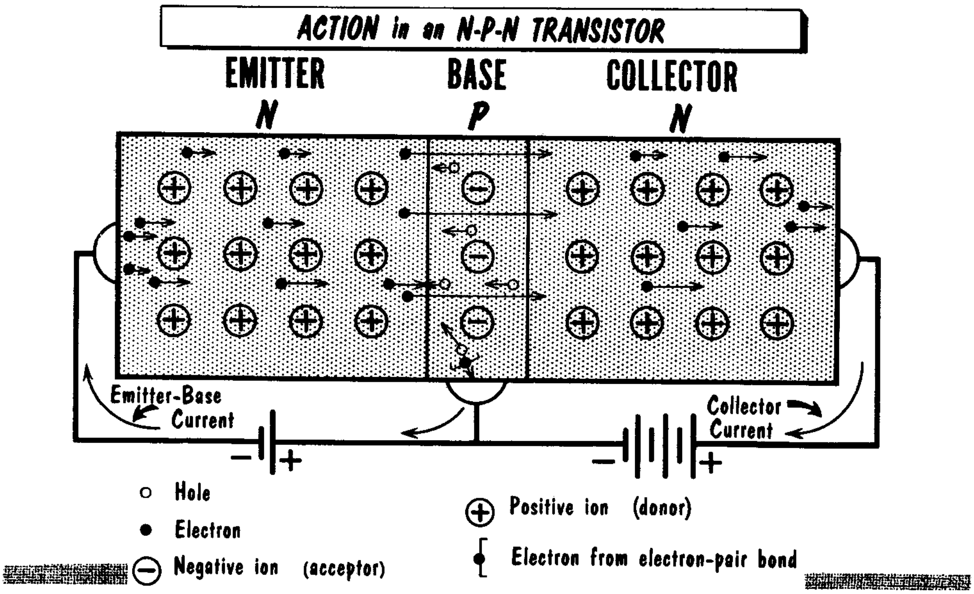
The amount of holes combining with electrons in the emitter-base junction region is quite small--usually less than 5%. However, because of this combination, there is a small emitter-base current, and the collector current is less than the emitter current by that amount. For example, if the emitter current is 1 ma, the collector current would be about 0.95 ma. Holes reaching the negative battery terminal at the collector combine with electrons from the battery. At the same time, new holes are formed at the emitter by electrons breaking their electron bonds and entering the positive battery terminal. Thus, while the current carriers within the transistor consist of holes, the external current consists of electrons.



As we shall see, maximum effectiveness of the transistor takes place when all the holes or current carriers from the emitter pass through the base region and on to the collector. Loss of some of these current carriers is unavoidable because of the electrons in the base region. However, by making the base sufficiently thin, the percentage of the holes that diffuse through without combining with electrons may reach as high as 99%. The holes that combine with base electrons set up an "undesirable" emitter-base current, with electrons leaving the emitter to flow into the emitter-base battery, and electrons leaving the battery to flow into the base.

Operation of the N-P-N Transistor

The functioning of an n-p-n transistor is very similar to that of the p-n-p type. The important difference is that the current carriers are now electrons instead of holes. Here again, the internal junction barrier voltage in the emitter-base region is offset by the externally applied bias voltage which forces the electrons to the right and the holes to the left. The electrons from the emitter



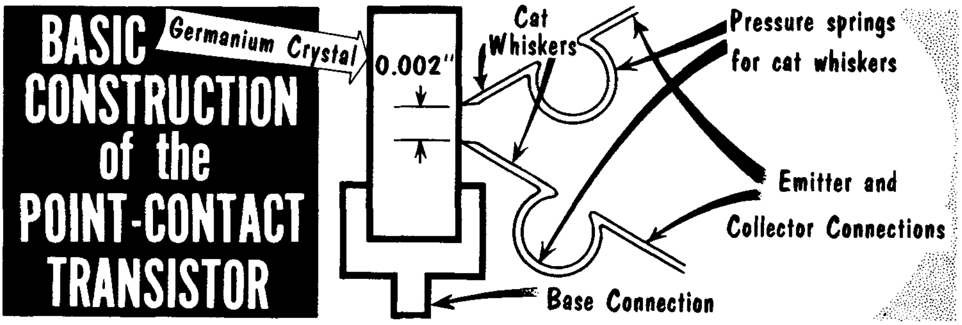
enter the center or base region, which has a p-type conductivity, and here, a small percentage of them combine with holes. Those electrons that do not combine with holes pass, or diffuse through the base region into the n-type collector under the influence of the high positive collector voltage. These electrons constitute the collector current.

Since a certain number of electrons combine with holes at the emitter-base junction, the collector current will be somewhat less than the emitter current. Actually, the amount of emitter current that enters into combination is usually less than 5% of the total current. Generally, the collector current is proportional to the collector voltage for small collector voltages. However, for a given emitter current, the number of current carriers is constant; above a certain collector voltage, the collector current is nearly constant.

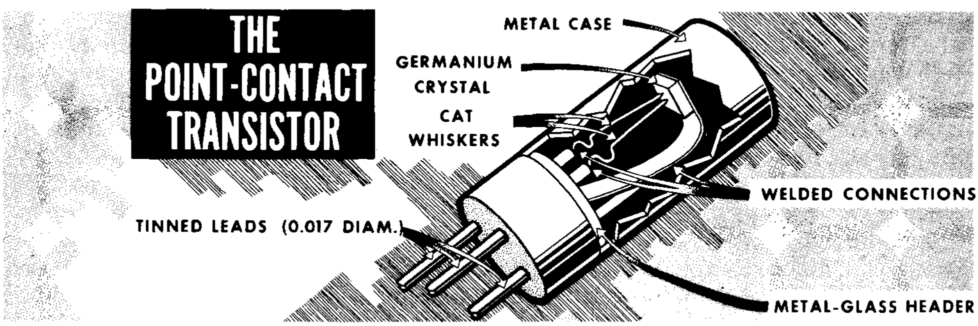
The fundamental differences between n-p-n and p-n-p operation are as follows: The emitter-to-collector current carrier in the p-n-p transistor is the hole; in the n-p-n transistor, the emitter-to-collector current carrier is the electron. Also, the bias voltage polarities are reversed, resulting from a reversal of n-type and p-type semiconductor materials.

The Point-Contact Transistor

Fundamental operation of the point-contact transistor is as follows: Assuming an n-type unit, the emitter and collector points make contact with the n-type germanium base. The free electrons in the base tend to congregate near the surface of the crystal and are called "surface bound." Immediately below the surface are the layers of positively-charged donor (pentavalent) atoms. These two layers constitute the equivalent of a voltage similar to that of the junction barrier in the junction diode.

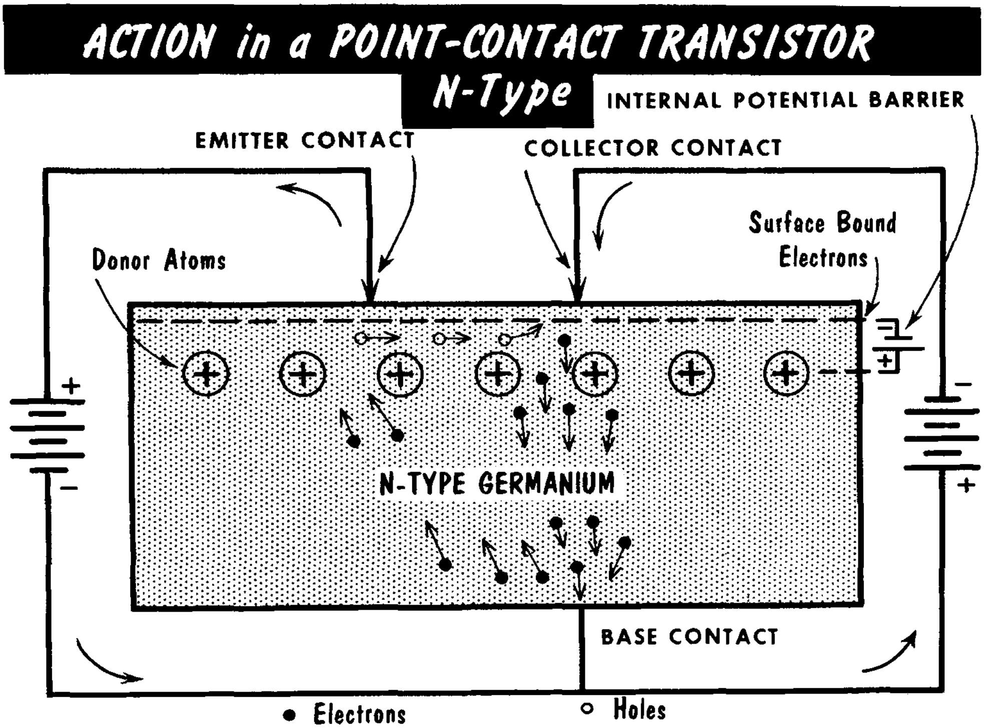


With the emitter positive (forward biased), some surface-bound electrons are removed by the emitter. For each electron thus removed, there is another one which enters the crystal from the negative terminal of the bias source. This provides a steady, small emitter current. To make the current larger, covalent bonds must be broken. The intense electric field produced by the tiny emitter point makes it an excellent device for producing this breakdown.



The liberated electrons are immediately attracted to and enter the positive emitter terminal. These electrons are the emitter current carriers, and each leaves a hole behind. This creation of holes is called hole injection, since the effect is the same as if the holes were injected into the transistor through the emitter. The holes immediately diffuse toward the collector because of the negative potential at that terminal. Some of these holes combine with free electrons and cease to exist--an undesirable situation.

The Point-Contact Transistor (Cont'd)

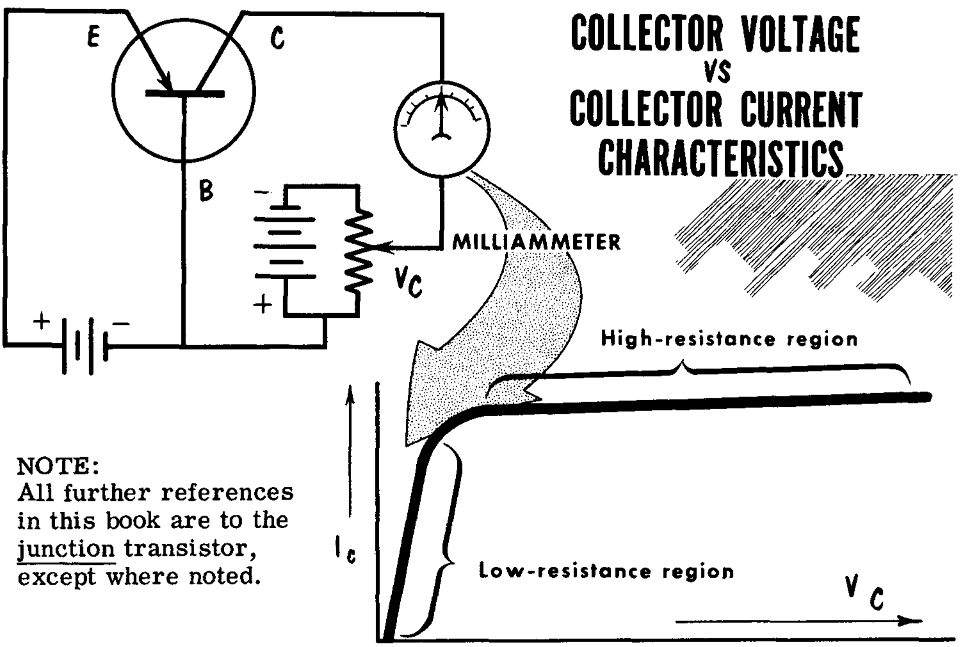


At the collector, the surface-barrier potential produced by the surface-bound electrons limits the current flow between collector and base. However, holes that reach the collector area combine with the surface-bound electrons in that area and reduce the surface-barrier potential. This permits the collector to inject more electrons into the germanium, increasing the collector current. Thus, holes injected by the emitter affect collector current without acting as current carriers themselves. The electrons that flow between the collector and the base are the collector current carriers. The point-contact transistor is made by placing two sharp, metallic wire points called "cat whiskers" against the face of a pellet of n- or p-type germanium. The wires are very fine and usually made of some springy, conducting material such as phosphor-bronze. They each contain a kink, to force them against the crystal surface under pressure. The points of the cat whiskers make contact with the semiconductor crystal about .002" apart. The connection to the crystal is called the base lead; the two wire leads become the emitter and the collector.

The current flow in the collector circuit is greater than the flow in the emitter circuit, the difference representing the gain of the transistor. While the maximum current gain in the junction transistor is 1, a typical point-contact transistor may have a current gain of 2.5. An advantage of these transistors is their ability to operate well at high frequencies due to the low capacitances between the point contacts and the base.

Collector Voltage-Collector Current Characteristics

If we connect a simple transistor circuit as shown, we can vary the collector voltage while observing its effect on the collector current. As we start from zero collector voltage and increase it, there is a rapid and constant increase in the movement of free current carriers, with a resultant increase in collector current. A further increase in collector voltage will produce still more current carrier activity with a greater increase in collector current. However, a point soon is reached where a still further increase in collector voltage will result in very little change in collector current. This condition can be compared with that of plate current saturation in a vacuum tube. Actually, in transistors, a point is reached where the number of free current carriers available (electrons in n-type material and holes in p-type material) is no longer great enough to permit a significant increase in collector current. From this point on, the curve is almost horizontal.



COLLECTOR VOLTAGE VS COLLECTOR CURRENT CHARACTERISTICS

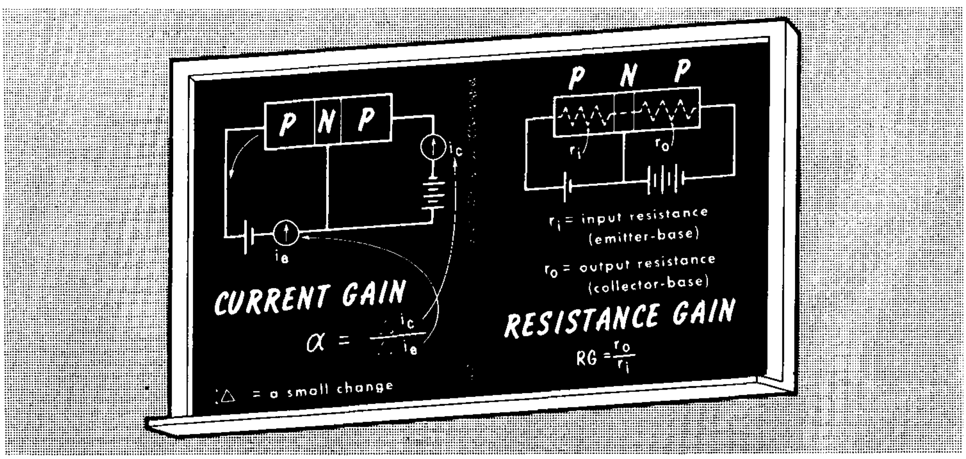
NOTE:
All further references in this book are to the junction transistor, except where noted.

We know from Ohm's law that $R = \frac{E}{I}$. We learned also that the plate resistance of a tube can be measured by E_p/I_p. We can therefore find the collector resistance by Vc/Ic. (In transistor terminology, voltage is expressed as "V" rather than "E".) In the very early portion of the curve, the rise is very great, indicating a low collector resistance. After the "bend", there is very little change in collector current for a given change in collector voltage, indicating a very high collector resistance. This is similar to the high plate resistance of the pentode. In practice, the collector is almost always biased in this region of high resistance. The resistance in this area may be as high as 1 or 2 megohms.

Current Amplification Factor--Alpha

The current amplification of a transistor can be compared with the voltage amplification factor (μ) of a vacuum tube. This, as you recall, is a measure of the relative ability of the grid voltage and the plate voltage to produce an equal change in the plate current ($\Delta e_p / \Delta e_g$). Note that the voltage in the output circuit is directly related to the input voltage. For this reason, vacuum tubes are considered as voltage-operated devices. Similarly, it is characteristic of transistors that the current flowing in the output circuit is directly related to the input current. Therefore, transistors are considered as current-operated devices. Thus, in discussing transistors, we think in terms of input and output current. This does not mean that transistors cannot be used as voltage amplifiers. Actually, they are, and we will see how voltage gain is obtained in transistor amplifiers.

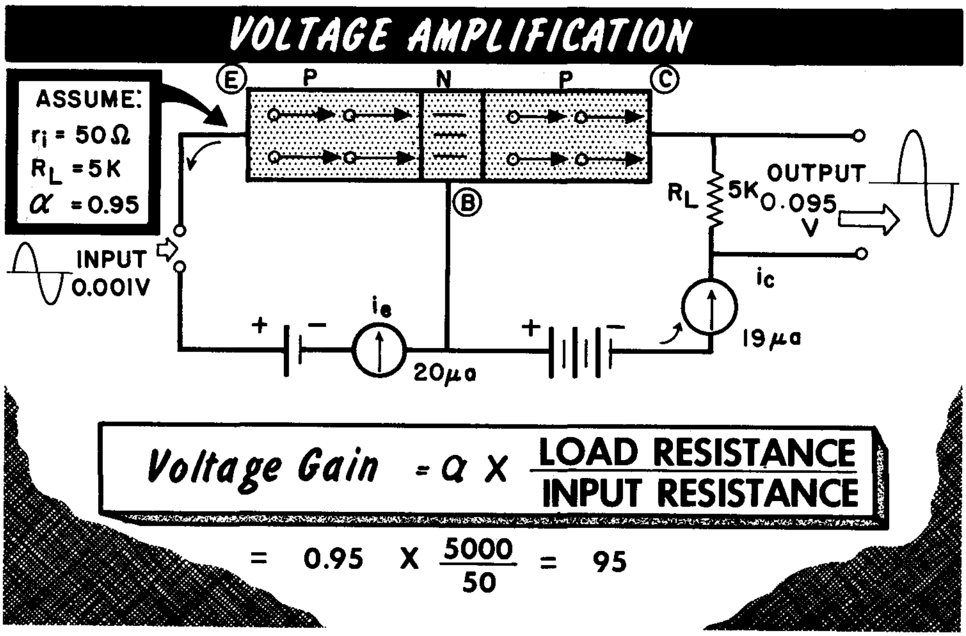
The change in collector current caused by the change in emitter current, assuming a constant collector voltage, is called the current gain of a transistor. This current gain, or alpha (α), is expressed: $\alpha = i_c/i_e$. Since every transistor will have some emitter-base current, alpha cannot reach unity or 1 in a conventional junction transistor. However, alphas of between 0.95 and 0.99 are obtainable. For a point-contact transistor, alpha can be as high as 2.5. As we will see, current gain (α) applies primarily to common-base amplifiers.



Another item of interest related to transistors is resistance gain. With the emitter-base circuit biased in the forward direction, the internal resistance of the input circuit is low. When the base-collector circuit is biased in the reverse direction, it has, as we have seen, a very high internal resistance. As we shall see, this resistance gain is very important. The resistance gain of a transistor is expressed as the ratio of the emitter-base internal resistance divided by the collector-base internal resistance. Said in another way, it is the output resistance divided by the input resistance ($RG = \frac{R_o}{R_i}$).

Transistor Voltage Gain

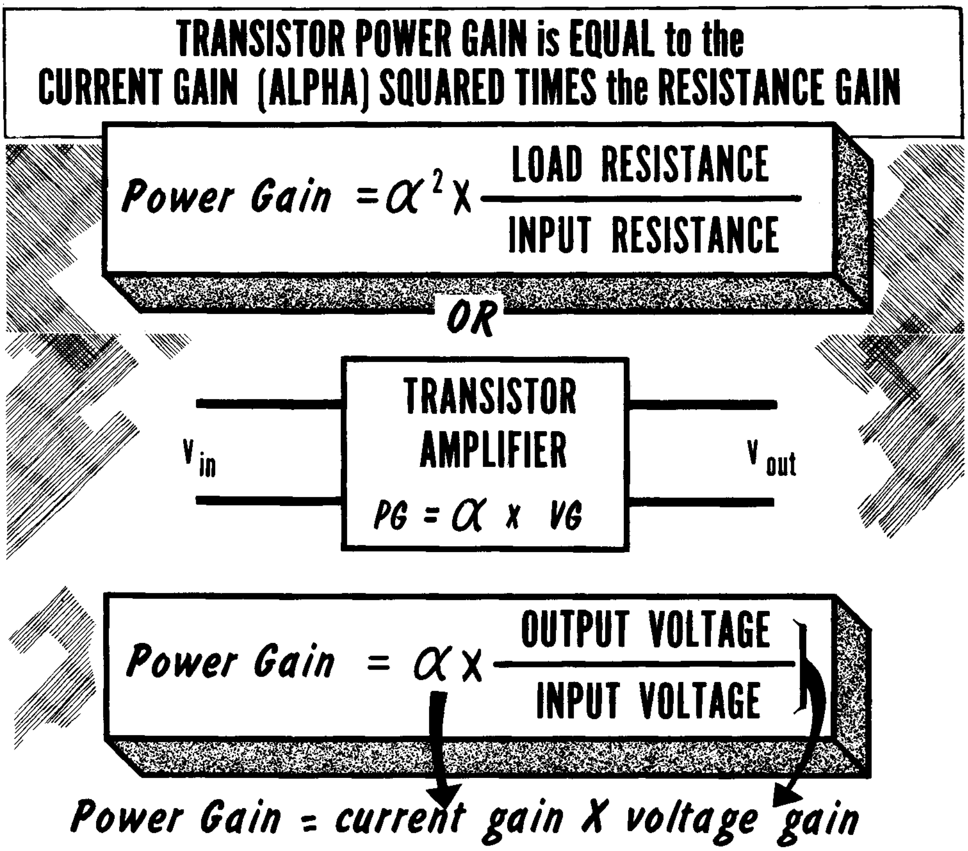
In our study of current gain (alpha), we learned that any change in emitter current produces a change in collector current alpha times as great. Thus, in a junction transistor, current "amplification" is generally limited to about 0.95 or so. Therefore, current "gain" in the usual sense is not obtainable. However, voltage gain is. For purposes of explanation, let us consider a junction transistor having a current gain of unity, or 1. Thus, any a-c component in the emitter circuit will produce the same a-c current fluctuation in the collector circuit.



Let us now assume that an input signal voltage of 1 millivolt is applied to the input (emitter-base) circuit. Because of the forward bias, the input circuit resistance is low--say 50 ohms. Now we can compute the emitter current (i_e) flow, using Ohm's law: i_e = .001/50, or 20 microamperes. Since the current gain of this transistor is 1, the collector current i_c also would be equal to 20 μa. However, because of the reverse bias, the collector circuit has a very high internal resistance. This high resistance is important because it permits a high-value load resistor to be used in the collector circuit without affecting the output current. This, then, is the key to voltage amplification in transistor circuits. Since the output voltage across the load resistor is equal to the output current times the load resistance, say 5000 ohms, the output voltage is equal to 20 μa times 5000 ohms, or 0.1 volt. Since the input voltage was 0.001 volt, this represents a voltage gain of 100. Thus, we can sum up voltage gain by saying it is equal to alpha times the ratio of load resistance to input resistance. In the above example, if alpha was 0.95, voltage gain would have been 0.95 × 100, or 95.

Transistor Power Gain

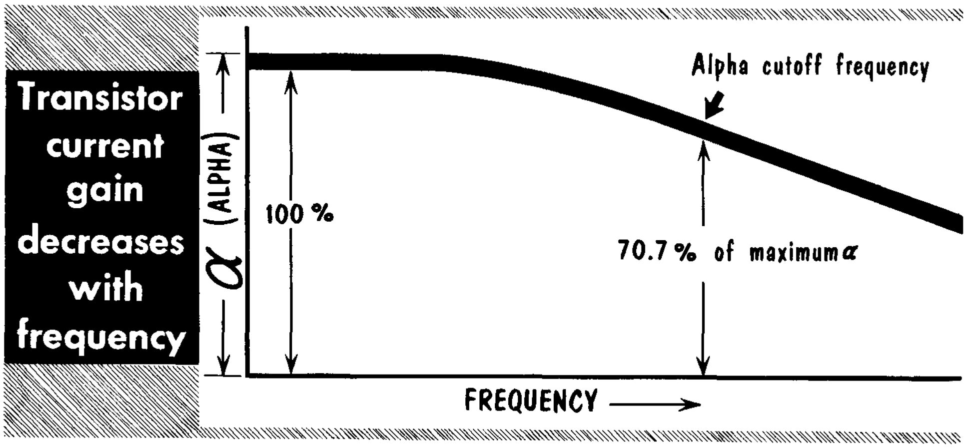
From our discussion of voltage gain in a transistor, we can move easily to an understanding of power gain. The input power delivered to the input circuit of a transistor can be obtained from Ohm's law formula, $P = E^2/R$. In terms of our transistor circuit, this would be equal to the input voltage squared divided by the input resistance (the resistance of the emitter-base circuit), or V_{in}^2/r_i. The power delivered to the external collector load circuit would be equal to the output voltage squared divided by the load resistance, or V_{out}^2/R_L. The power gain of the amplifier is, then, the ratio of output power divided by input power.



Since voltage gain is equal to α times R_L/R_i, it follows that power gain would be equal to α^2 times R_L/R_i, since power is also equal to I^2R. Simplifying it still further, we can state that power gain is equal to α times V_{out}/V_{in}. This is reasonable, for if the voltage gain of an amplifier is the ratio of V_{out}/V_{in}, and if the current gain is equal to α, the power gain must be the product of these two; that is, α V_{out}/V_{in}. Since the power gain in a junction transistor is equal to alpha times the voltage gain, we can see that the power gain will always be slightly less than the voltage gain.

Transistor Frequency Response--Alpha Cutoff

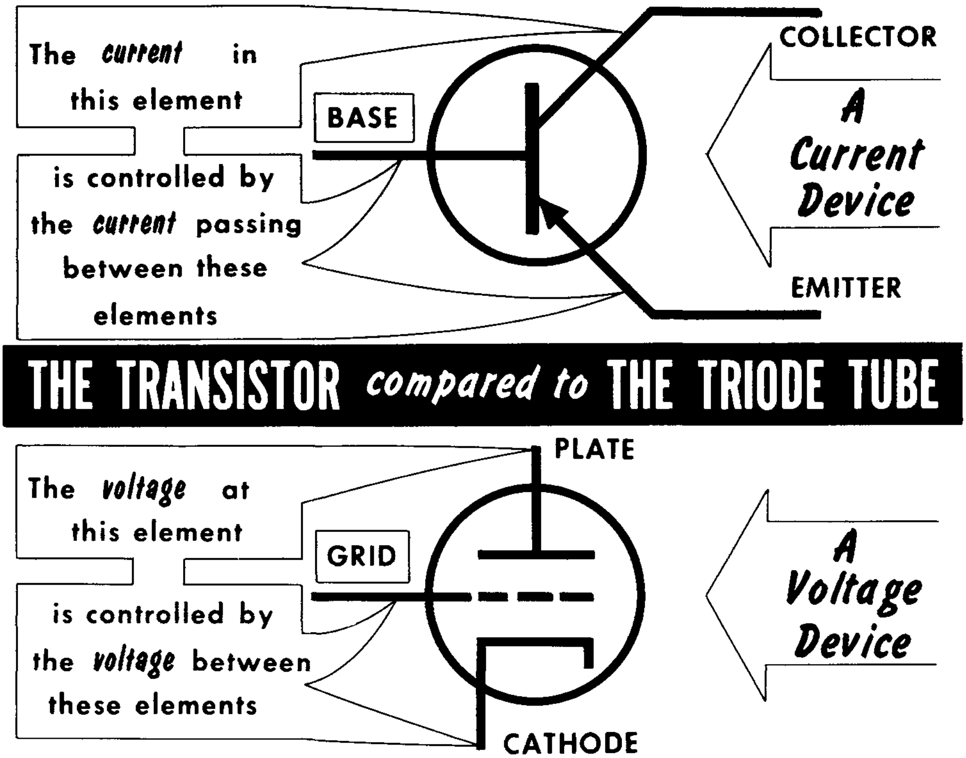
The time it takes electrons and holes to pass from the emitter to the collector (input to output circuit) is called the transit time. The transit time of these current carriers is one of the major factors limiting the high-frequency response of transistors. The movement of holes or electrons from the emitter through the base layer to the collector requires a short but finite time. In the transistor, the electron does not have a clear or unimpeded path from emitter to collector. As a result, the transit time is not the same for all electrons injected into the emitter at any one instant. If the transit time for all electrons were the same, there would be a simple delay in time between the input and output signals. Since the injected carriers do not all take the same path through the transistor, those produced by a single pulse at the emitter do not all arrive at the collector at the same time. The resulting difference is very small and unimportant in the audio-frequency stage. However, at the higher frequencies, the difference becomes a significant part of an operating cycle and causes partial cancellation between carriers. This causes a reduction in amplitude of the higher frequencies. The decrease in the output signal means a decrease in alpha, the current gain. In addition, the degradation in frequency response becomes steadily worse as the operating frequency is increased, until eventually there is no relationship, and no gain, between input and output waveforms.



Another limitation of high-frequency response is the capacitance between transistor regions, as well as between transistor leads. As the frequency increases, capacitive reactance decreases, and there is a tendency for signals to be shunted across the emitter-collector circuits.

The alpha cutoff frequency is the frequency at which current amplification has fallen to 0.707, or a loss of 3 db, from its low-frequency value, usually measured at 1 kc or lower. The alpha cutoff frequency is generally considered as the highest "useful" frequency amplified by a transistor, and is determined largely by the size of the emitter and collector electrodes, and the thickness of the base region.

Transistor-Vacuum Tube Analogy

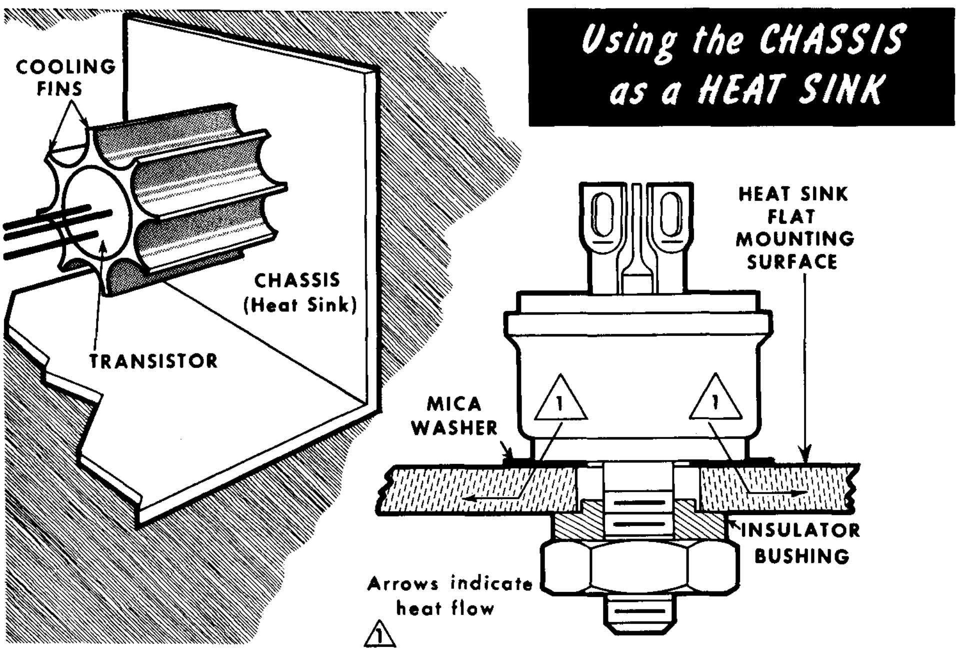


There are certain comparisons that could be made between the transistor and the vacuum tube. If we consider the triode vacuum tube, we have first the cathode, an element that emits electrons. At the other end of the tube is the plate or anode, at a positive potential, which attracts and collects the emitted electrons from the cathode. Between them, close to the cathode, is the control grid. This is a voltage-operated element that controls the flow of electrons between cathode and plate by means of an electrostatic charge between these elements. By applying comparatively small voltage changes between grid and cathode, we are able to obtain large voltage changes in the output circuit between plate and cathode.

In the transistor, we again have three elements. Here, the emitter is the supplier of current carriers, either electrons or holes. The collector collects the electrons or holes, and the base controls the flow of these charges by controlling the charge concentration in the base region. Thus, in a very general sense, we can compare the operation of a transistor with that of the vacuum tube. However, we must keep this in mind: that while the vacuum tube is primarily a voltage amplifier, the transistor is basically a current amplifier. As such, certain important differences exist. These will become more apparent in our further study of transistors.

Temperature Effects

Transistors are generally quite sensitive to increases in temperature. As the temperature of a semiconductor rises, an increasing number of electrons in the crystal are freed, producing an increase in collector current unrelated to the emitter or base currents. (This leakage current is normally quite small.) Leakage current flow can cause a further rise in the collector temperature, which releases still more electrons, again increasing the leakage current. Once started, this process (called thermal runaway) continues to repeat itself, regenerating until the transistor overheats and destroys itself. Because of this danger, the power-dissipation ratings of transistors are always given for a particular range of temperatures. Silicon transistors are usable at far higher temperatures than are their germanium equivalents.



Power transistors are often mounted on a heat sink (usually a separate piece of metal or the chassis of an amplifier), to carry off the heat and to prevent the temperature of the transistor from rising above that of its surroundings. One power transistor, for example, requires a 6" square of 1/8" aluminum as a heat sink when dissipating 8.5 watts. If no heat sink is used, its maximum permissible dissipation in air is 1.5 watts. In many power transistors, the collector is connected internally to the case or shell. This acts as a considerable aid toward heat dissipation. Because the current in a transistor increases with increasing temperature, a thermistor (a component whose resistance changes inversely with temperature) may be used in the emitter bias circuit to limit the current to a safe level for the temperature at which the transistor is operating.

The smallest part of an element which can take part in chemical changes is called an atom.

Atoms consist of positively-charged particles called protons, negatively-charged particles called electrons, and uncharged particles called neutrons.

A semiconductor has a resistivity between that of a conductor and an insulator. Examples of semiconductors are germanium and silicon.

In a crystal, the atoms are arranged in a specific pattern called a lattice.

Electrons shared by adjacent atoms in a crystal form electron-pair bonds (covalent bonds).

N-type germanium contains donor impurities--materials having five valence electrons. One of these electrons cannot form an electron-pair bond, and is called an excess electron.

Common donor materials include arsenic, antimony, and boron.

P-type germanium contains acceptor impurities--materials having three valence electrons. Since four valence electrons are needed to complete all adjacent electron-pair bonds, a hole is created.

A hole can be considered a positive charge which diffuses or drifts through a crystal. The drift of holes constitutes a current.

Forward bias of a p-n junction causes heavy current (flow of majority carriers). Reverse bias causes very low current (flow of minority carriers).

Holes constitute the principal current through the p-n-p transistor. Electrons constitute the principal current through the n-p-n transistor.

The emitter, base, and collector of the transistor are comparable to the cathode, grid, and plate of the vacuum tube.

The emitter-base junction is normally biased in the forward (low resistance) direction.

The base-collector junction is normally biased in the reverse (high resistance) direction.

Collector current depends upon the emission of carriers from the emitter-base barrier.

Alpha (α) is an expression of current amplification, and is a measure of a change in collector current to a change in emitter current, with the collector voltage kept constant. It is used primarily in the common-base amplifier.

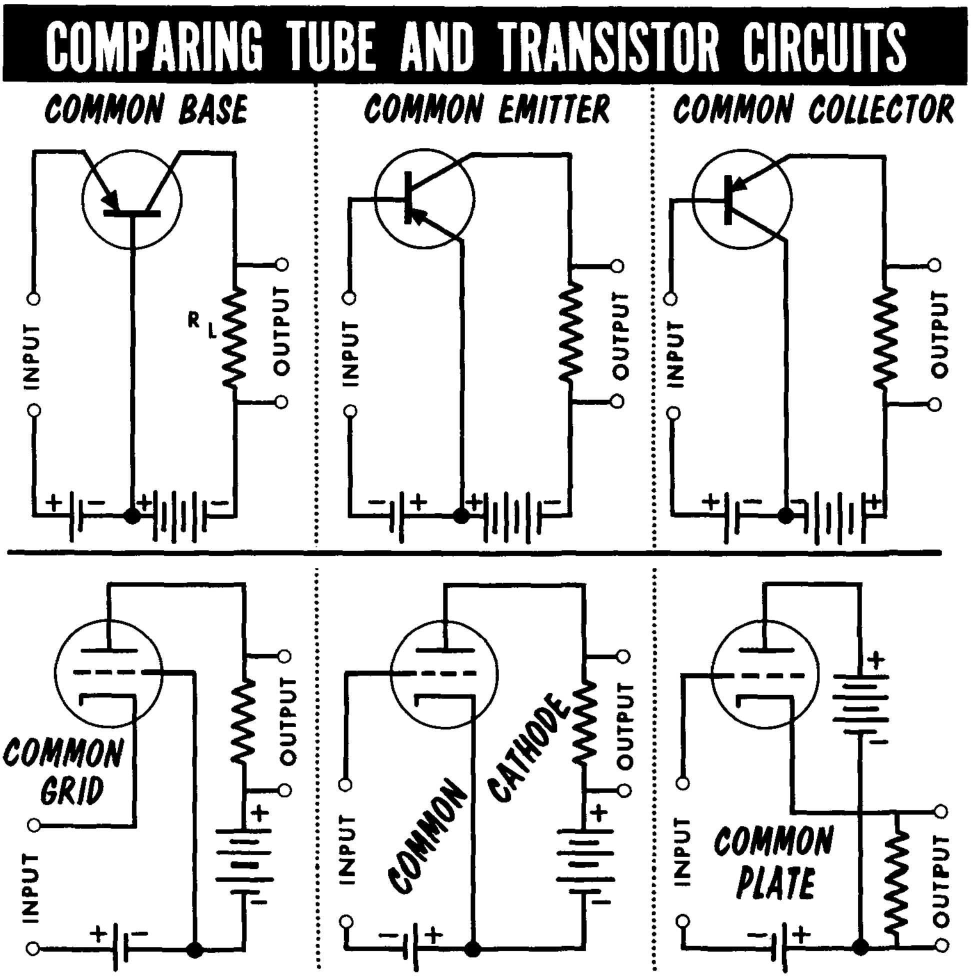
The alpha cutoff frequency indicates the point where the gain of a transistor has fallen to 0.707 of its maximum gain.

REVIEW QUESTIONS

1. What is meant by covalent bonds?
2. Explain the differences between n-type and p-type semiconductors.
3. What is a junction barrier?
4. Explain the operation of a p-n junction during conditions of forward bias and reverse bias.
5. How does a junction diode act as a rectifier?
6. Explain the basic construction of a transistor.
7. Explain the operation of p-n-p and n-p-n transistors.
8. What is the meaning of alpha and what is its significance?
9. How is a voltage gain acheived in a junction transistor circuit?
10. What is meant by the alpha cutoff frequency?

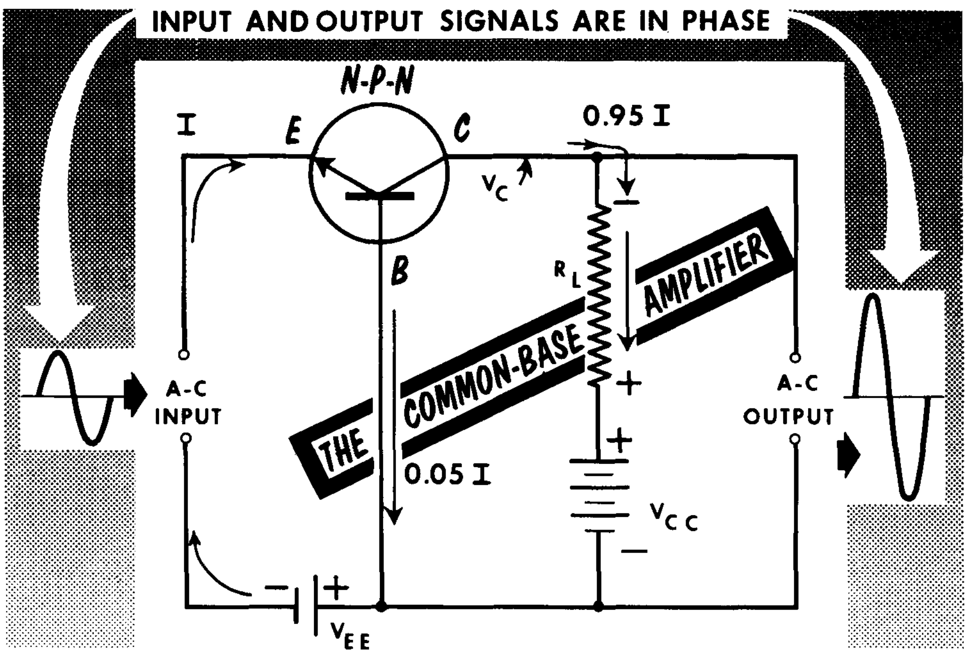
Three Basic Transistor Circuits

The three-element transistor we have been studying (p-n-p or n-p-n) can be connected into three basic type circuits--common base, common emitter, and common collector. Since there are only three connections to a transistor, and since each transistor circuit must have an input circuit requiring two input leads and an output circuit requiring two output leads, it follows that one of the three transistor leads must be common for both the input and output circuits. Very often, the common lead is used as a reference point for the entire circuit, and is thus connected to chassis or ground. This gives rise to the expressions grounded base, grounded emitter, and grounded collector. Common and grounded mean the same thing. In some instances, the element--emitter, base, or collector--will be connected directly to ground. Where this occurs, the element is at both a-c and d-c ground. Where the element goes to ground through a battery or resistor that is bypassed by a capacitor, the element is at a-c ground only.



The Common-Base Amplifier

We will begin our study of transistor amplifiers with the common-base circuit, since it is this configuration we are most familiar with, having used it from the beginning of this book. With the emitter-base circuit forward biased and the collector-base circuit reverse biased, current will flow as shown in the diagram. Using an n-p-n transistor, note that the collector current is 95% of the emitter current, with the remaining 5% flowing in the base circuit as a result of electrons from the emitter combining with holes in the p-type base. The details of transistor action have already been discussed.



With no signal input, a certain collector current will flow, producing a voltage drop across R_L. This voltage drop is in opposition to the collector battery voltage V_{CC}, and places the collector voltage V_C at some value lower than the battery voltage. Now, we inject a signal into the emitter-base circuit. We will assume that the first half cycle is positive-going and the second half cycle negative-going. Since the emitter is negative with reference to the base, the positive-going half cycle will oppose the negative bias voltage. With a reduction in forward bias, the emitter current will be reduced with consequent resulting reduction of collector current. Since the collector current is reduced, the voltage drop across R_L is reduced. We said that the collector voltage V_C is equal to the battery voltage V_{CC} minus the voltage drop across R_L. Thus, since the voltage drop across R_L is decreased, the collector will become more positive. Hence, we see an important point: as the input cycle varies through its positive half cycle, the output signal developed at the collector also varies through a positive half cycle.

The Common-Base Amplifier (Cont'd)

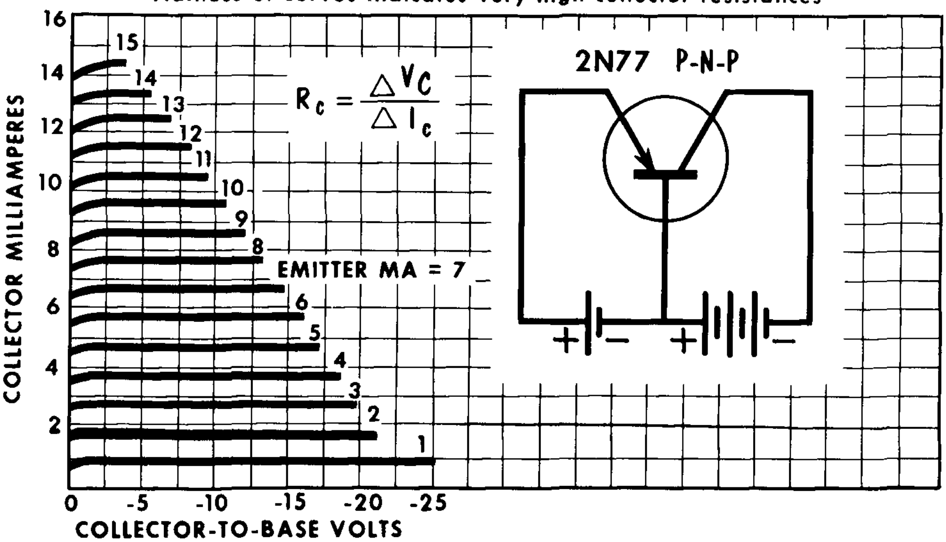
Now we observe the input signal as it goes through its negative half cycle. As the emitter goes negative, forward bias is increased, collector current is increased, and the voltage drop across R_L is increased. Since this voltage opposes the positive potential of V_{CC}, the collector goes more "negative", or less positive. We can thus conclude: In a common-base circuit, the input and output voltages are in phase--there is no phase reversal. This is exactly the same relationship that exists in the grounded or common-grid vacuum tube circuit.

As we have seen, the current gain (α) of the common-base amplifier is low, always being less than unity. However, by proper design of the center layer of semiconductor material, values as high as 0.98 are commonly reached. A factor which compensates considerably for the low current gain is the extremely high resistance gain of this circuit. The common-base amplifier has a very low input impedance of from 30 to 150 ohms, and a very high output impedance of 300 K to 1 megohm. Since approximately the same current flows in the emitter and collector circuits, a very high load resistance (R_L) can be placed in the collector output circuit, resulting in considerable voltage gain. Actually, voltage gains up to 1000 are not unusual in this circuit arrangement. Relatively good power gains are also available, with gains of 20 to 30 db (100 to 1000) common. A disadvantage of the common-base circuit is the difficulty involved in matching impedances because of the extremes in input and output resistances. However, it is ideal for applications such as amplifying the output of a low-impedance magnetic phono pickup, where no matching transformer would be required.

COLLECTOR CHARACTERISTIC CURVES for a COMMON-BASE AMPLIFIER

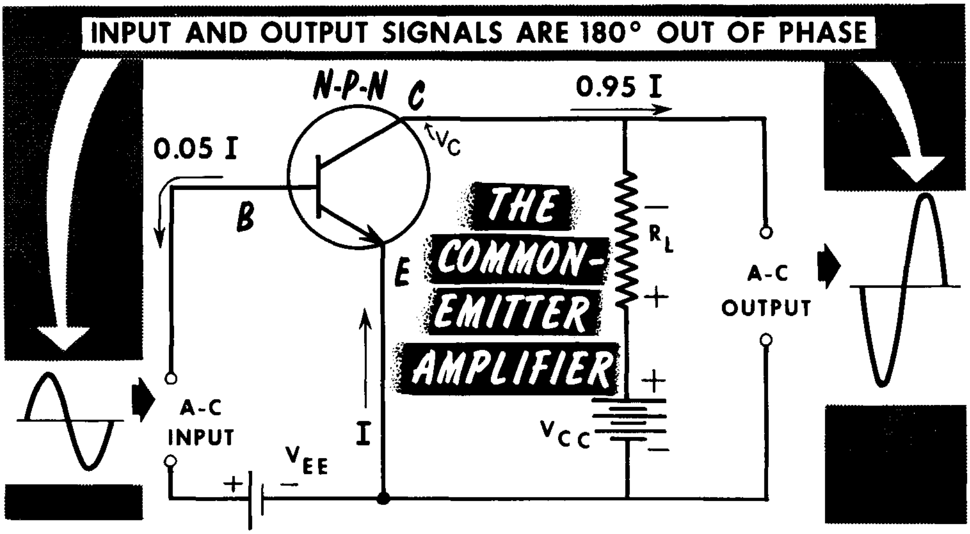
Note that the collector current never exceeds the emitter current

Flatness of curves indicates very high collector resistances



The Common-Emitter Amplifier

In the common-emitter circuit, once again we have a forward-biased base-emitter circuit and a reverse-biased collector-emitter circuit. Note that the emitter lead is now common to both the input and output circuits. Using the same n-p-n transistor circuit, the current in the collector circuit is again 95% of the emitter current, with the balance flowing in the base circuit.



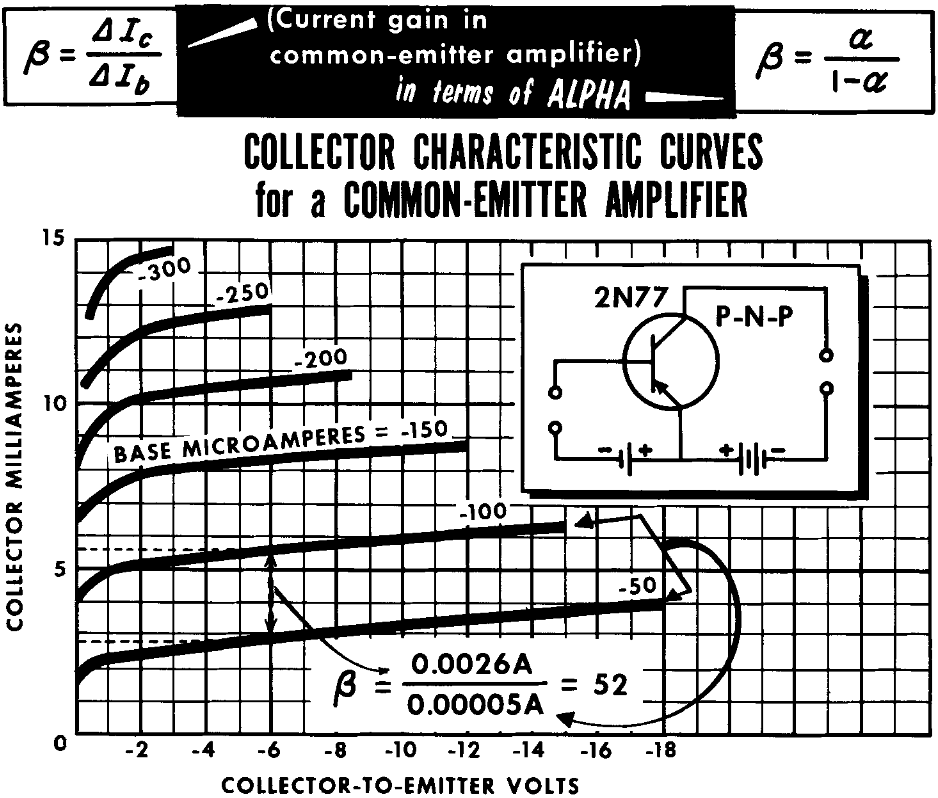
Applying the same input signal, note how the positive half cycle aids the input circuit forward bias. As this positive-going half cycle is applied between base and emitter, it increases forward bias with a resulting increase in collector current. This produces an increased voltage drop across R_L, which subtracts from V_{CC}, making the collector voltage V_C less positive (more negative).

During the negative half cycle, the input signal opposes the forward bias of the input circuit, thereby reducing the emitter and collector current. With a drop in collector current, the voltage drop across output load resistor R_L decreases, making the collector more positive. We now see the phase relationship in the common-emitter circuit--the input and output voltages are 180° out of phase. This is exactly the same relationship that exists in the common-cathode vacuum tube circuit.

The input and output impedances of the common-emitter circuit are considerably less severe than that of the common-base circuit, and impedance matching is much simpler. For instance, the input resistance to the common-emitter circuit may range from 500 to 1500 ohms, while the output resistance is usually in the range of 50 K ohms. Since the resistance gain of the common-emitter is so much less than that of the common-base circuit, it would seem that the available voltage gain would be much less. This is not the case.

The Common-Emitter Amplifier (Cont'd)

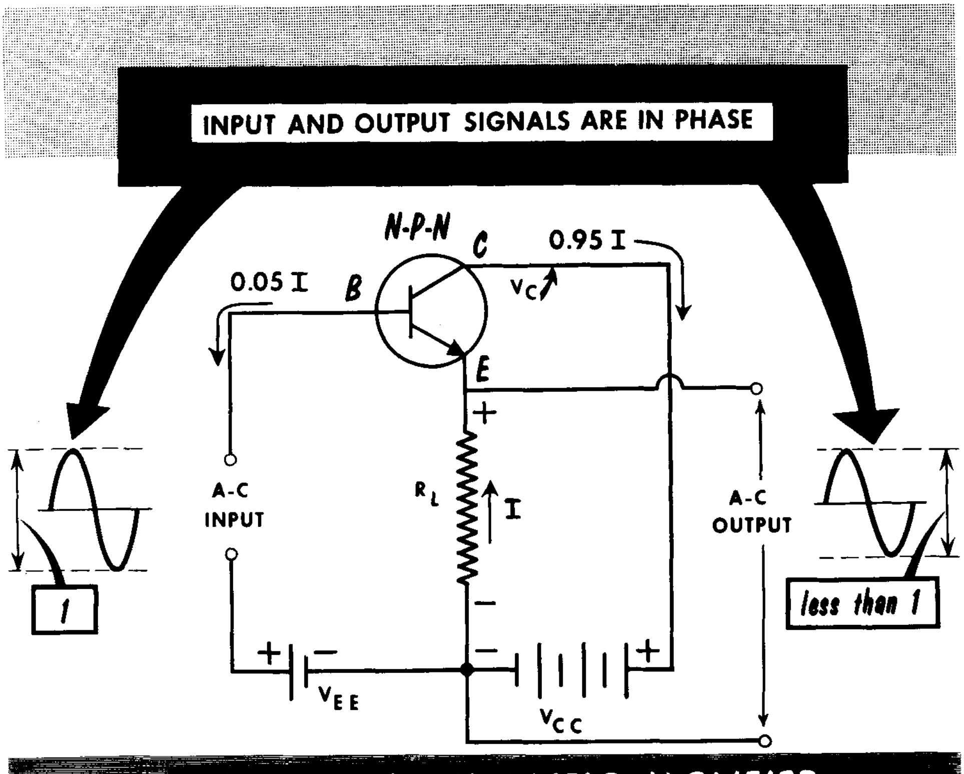
In the common-emitter circuit, current gain is measured by control of the collector current by the base current; the term for current gain in this circuit is beta (β). The beta is equal to a change in collector current divided by a change in base current, or β = delta I_c/delta I_b. In its more common form, the current gain of the common-emitter circuit is stated in terms of alpha; that is, beta = alpha/(1 - alpha).



From this, we can see that extremely large beta gains are possible when the alpha characteristic of a transistor is high. For instance, the beta current gain of a common-emitter transistor when alpha is 0.95 is 0.95/(1 - 0.95), or 19. If alpha is as high as 0.98, then the beta current gain is 0.98/(1 - 0.98), or 49. Since an alpha of 0.98 is somewhat high, the current gain of most common-emitter amplifiers is in the range of 35, with values up to 60 attainable. Because of this very high current gain, and even though the resistance gain is very low compared to the common-base amplifier, the voltage gain is still quite respectable, being usually slightly lower than that of the common base. Common-emitter voltage gains of 250 to 500 are attainable. Since power gain is a function of the current squared, very high power gains are possible in this high-current gain circuit. Power gains of up to 40 db (10,000) are common.

The Common-Collector Amplifier

In the common-collector circuit, the collector lead is common to both the input circuit (base-collector) and the output circuit (emitter-collector). Note that the load resistance R_L is in the emitter lead and the output is taken from the emitter. You will recall this being the same situation that exists in the vacuum tube cathode-follower circuit. In our circuit, using an n-p-n transistor, when we apply a positive-going signal, the input circuit forward bias



THE COMMON-COLLECTOR AMPLIFIER

is increased, resulting in an increase in collector current. This current flows through R_L in the emitter circuit, making the emitter more positive. When the incoming signal is negative going, the input circuit forward bias is reduced, lowering the emitter and collector currents. This reduces the voltage drop across the emitter or load resistor, making the emitter negative going, or less positive. Since the output signal is taken across the load resistor between emitter and ground (or common), we see that the output voltage will vary in step with the input signal. From this, we make the observation: <u>In the common-collector circuit, the input and output voltages are in phase</u>. This, of course, is exactly the same situation as exists in the common-plate or cathode-follower circuit.

The Common-Collector Amplifier (Cont'd)

The common-collector circuit is interesting in that it is so different from the two previous circuits. The input resistance of the common-collector circuit is extremely high, ranging from 100 K to 500 K ohms, while the output resistance is low, in the range of 100 to 1000 ohms. The input resistance is high due to the high resistance of the base-collector circuit; the output resistance is low due to the low resistance of the emitter-collector circuit. This circuit makes an excellent isolation network between a high-impedance and a low-impedance circuit, such as a transmission line. In so doing, of course, it is acting as an impedance-matching device.

As in the case of the vacuum tube cathode follower, the voltage gain of the common-collector amplifier will always be less than unity (1), and no real gain is obtained. This is due to the degenerative effect of the load in the emitter circuit. Since the input and output signals are in phase, the output signal acts to oppose changes in input circuit bias produced by the incoming signal. Practically, we can say this circuit is capable of unity voltage gain.

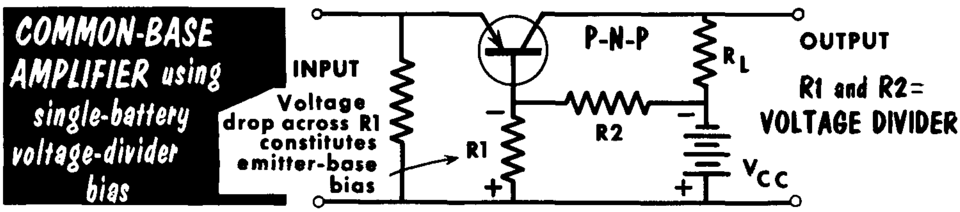
SUMMARY OF TRANSISTOR CIRCUIT CHARACTERISTICS

	INPUT IMPEDANCE	OUTPUT IMPEDANCE	CURRENT GAIN	VOLTAGE GAIN	POWER GAIN
COMMON BASE	30-150 Ω	300,000Ω - 1 Megohm	Less than 1	300 - 1000	20-30 db
COMMON EMITTER	500-1500 Ω	50,000 Ω	35	250-300	40 db
COMMON COLLECTOR	100,000 - 500,000 Ω	100 - 1000 Ω	35	Less than 1	15-30 db

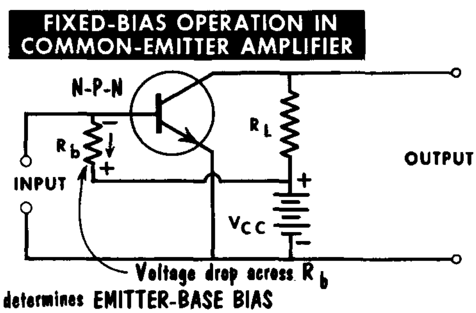
The value of current gain for a common-collector amplifier is not very different from that for a common-emitter amplifier. When alpha is nearly unity, $\alpha/(1 - \alpha)$ is not very different from $1/(1 - \alpha)$. In spite of this high current gain, however, the common-collector gives less than unity voltage gain because the input resistance is so high compared with the load resistance. Current gains of 35 or so are commonplace. The power gain of this circuit is very low, falling in the range of 15 to 30 db.

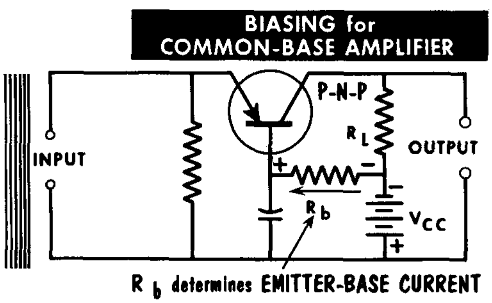
Transistor Biasing

All transistor circuits discussed so far have used two voltage sources (batteries)--one for the emitter-base bias and one for the base-collector bias. Obviously, these two bias voltages are necessary. However, there are practical means by which we can eliminate the need for two batteries. In doing so, we must be careful that our biasing methods provide the same voltage stabilization provided by batteries. In the comparatively seldom-used common-base amplifier, a simple voltage divider can be used to establish emitter-base bias. Two resistors connected across the collector supply will provide the necessary voltage distribution. Note that with the p-n-p transistor, the voltage drop is of such a polarity as to make the base negative with respect to the emitter; this is as it should be for forward bias. A disadvantage of this type of bias is its inefficiency. The constant voltage-divider bleeder current represents a waste of power.



The extremely popular common-emitter circuit requires more consideration in the design of its bias arrangement. One method, often called fixed base-current bias, has the base resistor connected directly between the base and the "high" side of the collector battery. Base-bias current will flow in this circuit through Rb and produce a voltage drop opposite in polarity to that of the collector supply; the difference is applied between the emitter and base to provide the necessary input forward bias.

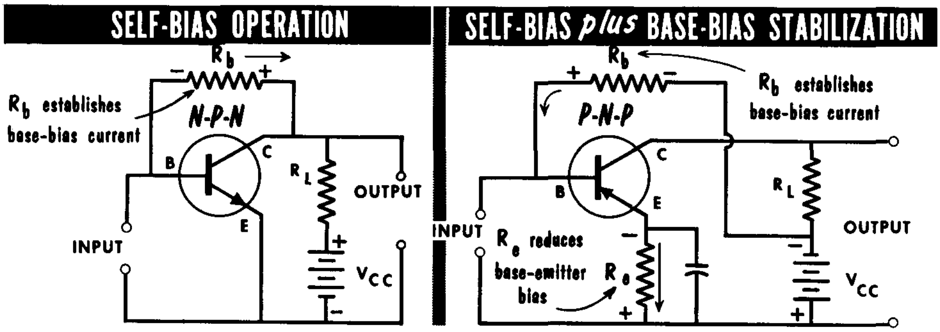




Another biasing arrangement for common-base amplifiers involves the placing of a resistor in the base circuit between the base and the high side of the collector supply. This biases the emitter positively with respect to the base, and establishes the proper emitter current. Normally, there would be signal degeneration as a result of this resistor, but this is avoided by placing the base at a-c ground potential through a bypass capacitor.

Transistor Biasing (Cont'd)

The connection of the fixed-bias circuit permits a degree of instability to enter, in that the bias voltage changes with changes in ambient temperature which in turn cause transistor current variations. To improve this condition, we use a circuit called self bias. Here, the base resistor is connected directly between base and collector. In this circuit, we improve stability by introducing degeneration (negative feedback) similar to that produced by an unbypassed cathode resistor in vacuum tube circuits. This degeneration, in transistor circuits, is a form of automatic control of the base bias. In this



circuit, base-bias resistor R_b establishes the value of collector current and prevents excessive shifts in the collector d-c operating point due to temperature change or transistor replacement. In the self-bias circuit, a change in temperature may cause an increase in collector current. This would cause an increase in voltage drop across the load resistor, reducing the collector voltages. This, in turn, would produce a decrease in the base-bias current, thus compensating for the change. Should there be a decrease in collector current, causing a rise in collector voltage, base-bias current would increase.

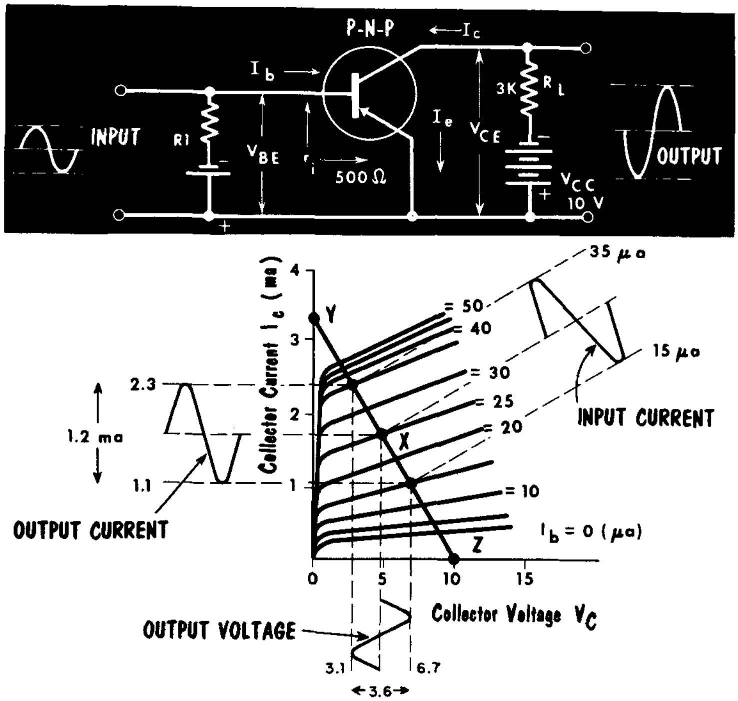
The self-bias circuit offers much higher stability than the fixed-bias arrangement. However, the self-bias circuit produces a-c feedback through the bias network which reduces the gain slightly. This feedback is often reduced by using two series resistors in place of R_b, and bypassing the tap to ground through a capacitor.

Still another popular method of providing bias stabilization in transistor circuits is through a bias resistor placed in the emitter circuit. It is another form of self bias, and is generally used in conjunction with base bias. If the collector current flowing through the emitter resistor tends to increase, say as the result of an increase in temperature or a change in transistors, the voltage drop across R_e would increase. This would reduce the base-emitter bias voltage, resulting in a collector current decrease. The opposite would occur should the collector current decrease. The voltage drop across R_e reproduces a signal loss through degeneration. The emitter stabilizing resistor is often bypassed however, to avoid a-c signal degeneration. In this respect, the network operates similar to that of a bypassed cathode resistor in vacuum tube circuits.

Transistor Analysis Using Characteristic Curves

We are familiar with the construction of load lines in vacuum-tube circuitry and the analysis we can make from them. In transistors, a similar situation exists: we can construct a load line and use it for circuit analysis. Using a common-emitter circuit, we will calculate the current, voltage, and power gain from the output characteristic curves. These curves plot the collector current against the collector voltage for different values of base current. We will assume the following conditions: collector supply voltage (V_{CC}) is 10 volts; load resistor (R_L) is 3000 ohms; emitter-base input resistance (r_i) is 500 ohms; peak-to-peak input current is 20 microamperes. The operating point (X) is 25 μa base current.

COMMON-EMITTER AMPLIFIER and its OUTPUT CHARACTERISTIC CURVES with LOAD LINE



Our first step in constructing a loadline for R_L on the output characteristic curve is as follows: when I_C is zero, there is no voltage drop across R_L, and the collector voltage (V_C) is equal to the full 10 volts. We make this point Z on the collector voltage axis. When V_C is zero, the total V_{CC} is being dropped across R_L. Under this condition, I_C is equal to 10/3000 or 3.3 ma. We call this point Y on the I_C axis. Connecting points Y and Z establishes the load line. We now establish our desired operating point on the load line where I_b equals 25 μa. Vc is equal to 4.8 volts.

Transistor Analysis Using Characteristic Curves (Cont'd)

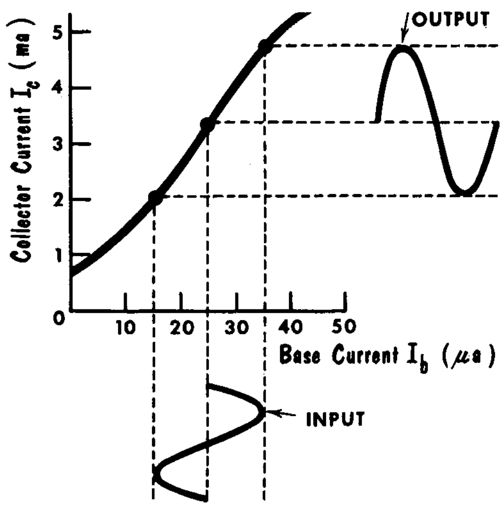
Since the input current is 20 μa peak-to-peak, it will deviate between 10 μa and below the operating point. Thus, the input current waveform will vary between 15 μa and 35 μa. Following this, we can construct the output current waveform by projecting across to the collector-current axis. The output current will swing between 2.3 ma and 1.1 ma. Projecting downward, we get the output voltage waveform and note that it swings between 3.1 and 6.7 volts.

Current Gain: In the common-emitter amplifier, this is the ratio of the change in collector current to a change in base current. From our construction curve, we find this is 1.2 ma/0.02 ma, or a current gain (amplification) of 60.

Voltage Gain: In the common-emitter amplifier, this is the ratio of the change in collector voltage to a change in base voltage. We can determine the change in base voltage since it is equal to a change in input current (20 μa) multiplied by the input impedance (500 ohms), or 0.01 volt. The voltage gain is thus equal to the ratio of 3.6 volts (ΔV_c)/0.01 volt (ΔV_b), or an amplification of 360.

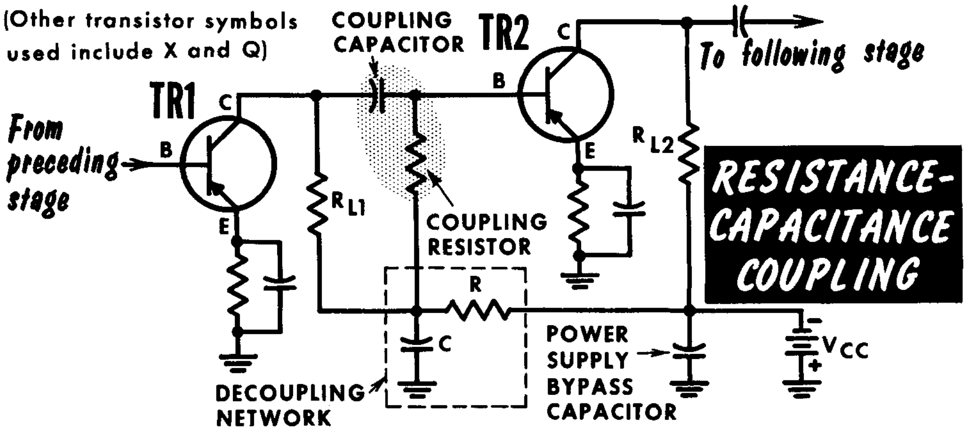
Power Gain: This is equal to the voltage gain multiplied by the current gain (360 times 60), or 21,600. The power input is thus increased 21,600 times in going through the transistor. In terms of decibels, this is approximately 33 db.

DYNAMIC TRANSFER CHARACTERISTIC CURVE showing linear (class A) operation



Another important curve for studying transistor characteristics is the dynamic transfer characteristic curve, in which collector current is plotted against base current for a given collector voltage. This curve is similar to the I_p-E_g vacuum tube curve. When the proper operating point is determined, and if the change of base current is within the linear portion of the dynamic transfer characteristic curve, the transistor operates linearly. This is called class-A operation, and the output signal will be an exact replica of the input signal.

Interstage Coupling



We have examined the basic types of transistor circuit arrangements. However, to obtain sufficient amplification or gain, more than one amplifier is required. As in tube circuits, two or more circuits can be cascaded for greater gain. The most popular methods of coupling transistor circuits are through R-C and transformer arrangements.

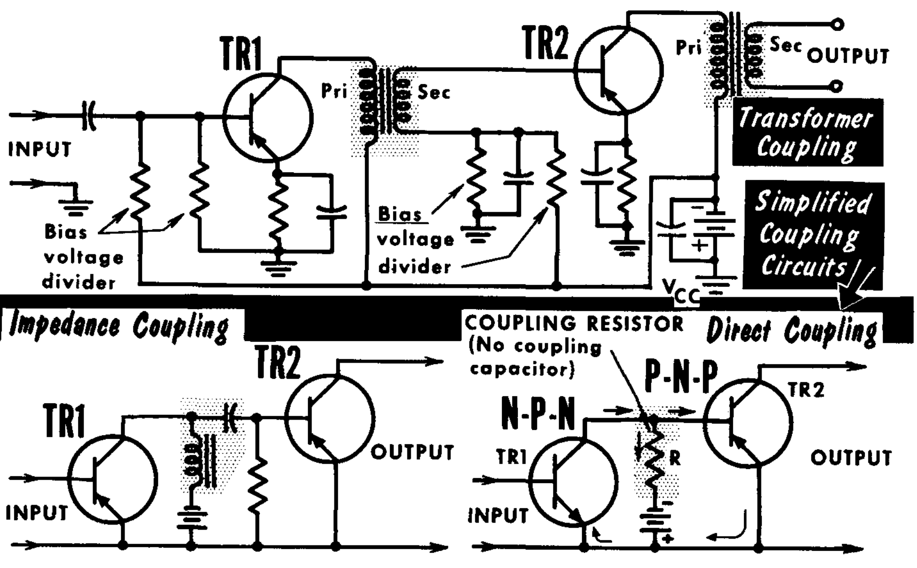
A typical two-stage R-C coupled common-emitter amplifier is shown. This type of coupling is desirable for the amplification of low-level or weak signals, where transformers would be more susceptible to hum pickup. Another feature is the fact that resistors and capacitors take up much less room than transformers. In our R-C coupled circuit, fixed bias is used together with emitter stabilization. The collectors of TR1 and TR2 are connected to the battery through load resistors RL1 and RL2. Since the coupling capacitor must be able to pass the lowest frequencies, and the input and output resistances are relatively low, the coupling capacitor must be fairly large in order to present a low reactance. Values of 1 to 10 μf are common. These capacitors are generally electrolytic types, something rarely ever used for coupling in vacuum tube circuits. Being electrolytic, polarity always must be observed. Fortunately, the larger leakage currents of these capacitors are not as critical as in tube circuits.

To prevent feedback, a decoupling network is often used. This can be done merely by placing a resistor in series with the base resistor and then bypassing the resistor. The time constant of this network must be long enough to fully bypass the lowest frequency. Since R must be kept small so as not to lower the battery voltage to the preceding stages, a very large decoupling capacitor of about 50 to 100 μf is frequently used. To prevent feedback because of the voltage drop across the internal resistance of the battery, a bypass capacitor is placed across it. Another popular means of coupling transistor circuits is with transformers. Since the collector impedance of a transistor is high compared to the base input impedance, a transformer offers an excellent means of matching impedances and thus providing maximum power gain.

Interstage Coupling (Cont'd)

Although a stepdown transformer is used, this does not mean that there will be a signal loss. Since the transistor is a current device, the voltage stepdown transformer will actually provide a current gain for the signal. This action is similar to the output transformer in an amplifier that feeds a loudspeaker. A typical circuit might include a voltage divider for base bias and an emitter-stabilizing resistor, bypassed to prevent signal degeneration. The primary winding (including the a-c reflected load from the secondary) is the collector load impedance of TR1. The secondary winding introduces the a-c signal to the base and also acts as the base d-c return path.

Because there is no collector load resistor to dissipate power, the power efficiency of the transformer-coupled amplifier is high. However, the frequency response of this type of coupling is not as good as that of the R-C coupled stage. The primary winding shunt reactance at low frequencies reduces the low-frequency response. High-frequency response is reduced by collector capacitance and leakage reactance between primary and secondary.



The impedance-coupled amplifier offers a compromise: an inductor replaces the collector load resistor. Thus, the d-c power loss is virtually eliminated; however, the low-frequency response is reduced by the shunt reactance of the inductor. The high-frequency response is reduced by the collector capacitance. Frequency response is better than that of the transformer arrangement, but not as good as that of the R-C coupled amplifier.

The direct-coupled amplifier is used for amplification of d-c and low-frequency signals. Its principal feature is that it retains the d-c component of a signal. Note that coupling capacitors are eliminated. Coupling resistor R acts as both the collector load resistor for TR1 and the bias resistor for TR2.

Since there are both n-p-n and p-n-p transistors, a separate graphical symbol is used to represent both.

The transistor can be connected as a common-base, common-emitter, or common-collector, amplifier.

The circuit of the common-emitter amplifier is similar to that of the vacuum tube common-cathode amplifier.

The circuit of the common-base amplifier is similar to that of the common or grounded-grid vacuum tube amplifier.

The circuit of the common-collector amplifier is similar to that of the common-plate or cathode follower vacuum tube amplifier.

Each transistor arrangement may be biased with either two batteries or a single battery.

The common-base and common-collector amplifiers do not provide a phase reversal of the input voltage signal.

The common-emitter amplifier provides a 180° phase reversal of the input voltage signal.

The d-c electron flow in the base lead may be toward or away from the base region, depending on the relative magnitudes of the base-emitter current and the saturation current.

Reverse-bias collector current (I_{cbo}), also called saturation current, increases rapidly at high temperatures and causes increased emitter current.

D-c negative feedback can be used to minimize variations in emitter current caused by temperature changes.

The current, voltage, and power gain of a transistor amplifier can be calculated from the output static characteristic curves on which a loadline has been drawn.

The loadline indicates the way in which the collector supply voltage is divided between the load and the collector under various conditions of collector current.

The dynamic transfer characteristic curve may be used to determine the linearity and nonlinearity of the output signal to the input signal for a specific operating point and a specific load resistance.

REVIEW QUESTIONS

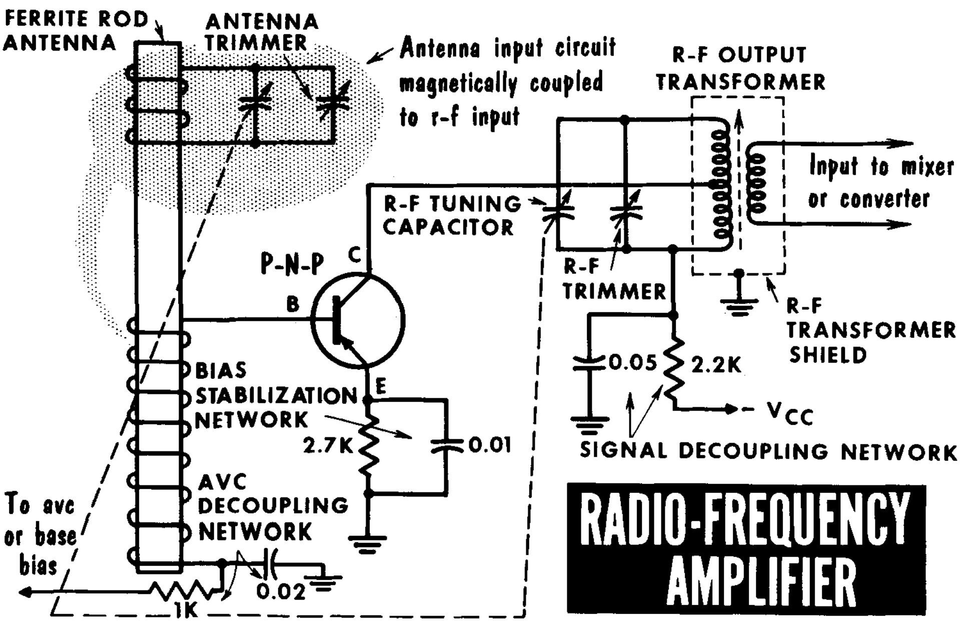
1. Name the three basic-type circuits in which transistors can be connected.
2. A very low input impedance and a very high output impedance are characteristic of what type of amplifier configuration?
3. Give one important disadvantage of the common-base amplifier.
4. What is the phase relationship between the input and output voltages in the common-emitter amplifier?
5. In what type of circuit configuration are very high power gains possible?
6. Which circuit arrangement has a very high input resistance?
7. Which circuit arrangement always gives less than unity voltage gain?
8. What is meant by fixed base-current bias?
9. What is meant by self bias?
10. What is the function of the emitter-stabilizing resistor?
11. Give an advantage and a disadvantage of transformer coupling in transistor circuitry.

The Transistor Receiver--R-F Amplifier

Basically, the transistorized receiver is the same as the vacuum-tube type; it performs stage by stage all the functions of the superheterodyne receiver. We will now study the transistorized receiver, starting with the r-f amplifier stage, and examine the circuitry.

Transistor r-f amplifiers, like their vacuum tube counterparts, are most often used for improving the gain, overall signal-to-noise ratio, and selectivity characteristics of a multistage circuit. The design is basically the same as that of an i-f amplifier. The chief problem is in the selection of a transistor having a sufficiently high alpha cutoff. Early r-f amplifiers were used only for broadcast band work up to 1600 kc. However, recent improvements have permitted the design of shortwave transistor receivers in which the r-f amplifier operated at frequencies up to 18 mc and higher. Generally, the r-f amplifier in a transistor radio is considered a luxury, and is found most often on multiband receivers requiring high r-f gain.

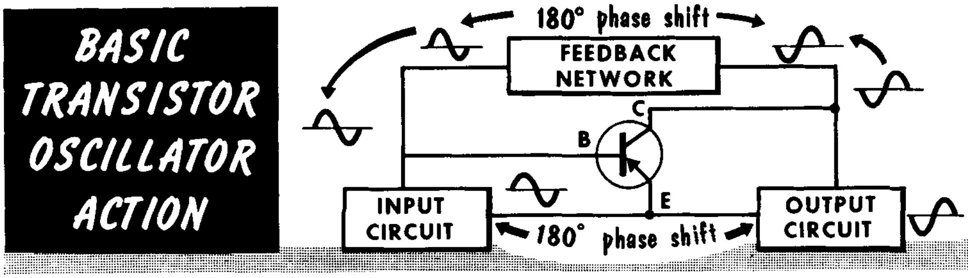
A typical r-f amplifier circuit is shown. Using a high-gain ferrite-rod antenna in the antenna tuning circuit, the input signal is fed to the transistor base circuit through magnetic coupling. Through transistor action, the signal is amplified and then transformer-coupled to the input of the following mixer or converter stage. The stepdown r-f output transformer provides an impedance match between the high impedance of the r-f amplifier collector circuit and the low impedance of the following base or emitter circuit. Note that the collector is connected to some point on the output transformer primary. This is a design feature which permits finding the optimum point to present the best output load impedance for the collector.



RADIO-FREQUENCY AMPLIFIER

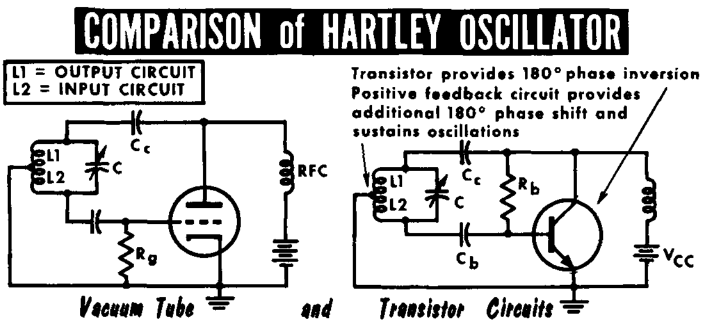
Transistor Oscillators

Oscillators have two principal sections; a frequency-determining section and an amplifying section. The frequency-determining (feedback) section usually consists of an L-C or R-C network. The amplifying section is usually a tube or transistor amplifier having sufficient gain to compensate for losses in the frequency-determining section. This arrangement provides us with positive feedback, which in turn produces oscillation. Thus, we can say that a tran-



sistor oscillator is nothing more than an amplifier that has a portion of its output signal feeding back, in proper phase, to the input.

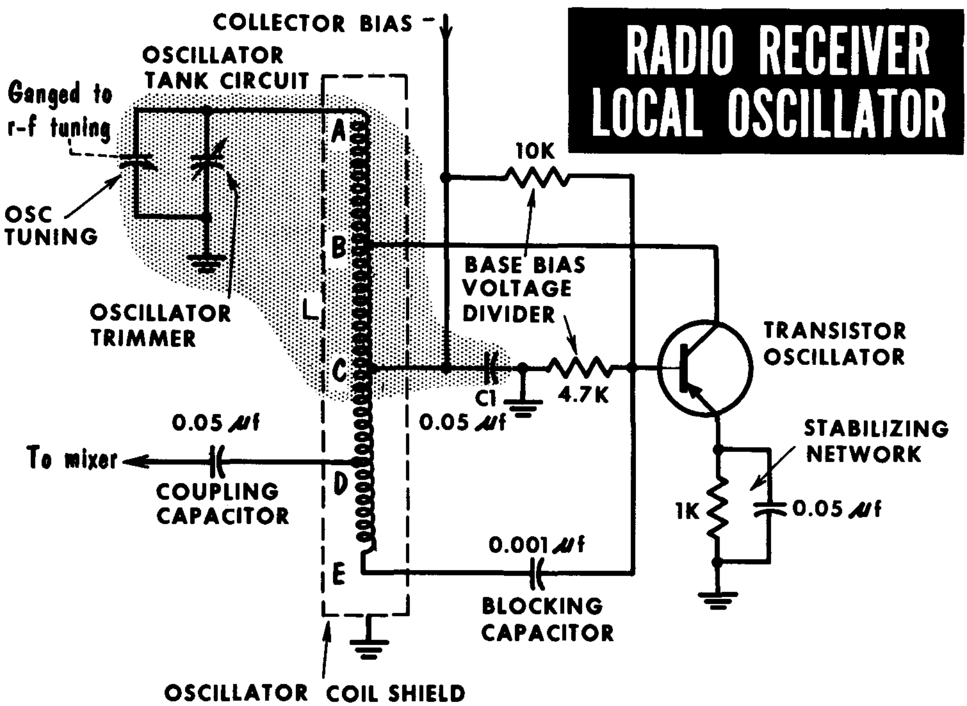
With the transistor acting as an amplifier, the input signal is amplified, with a portion of the output energy fed back to the input to supply the necessary input power to overcome circuit losses. When this is done, the transistor supplies its own input signal and oscillates at a frequency determined by the value of the feedback components. The transistor oscillates because any small current change in either the input or output circuit is transferred from one to the other through the transistor and feedback network.



Let us compare the popular vacuum-tube Hartley oscillator with its transistorized counterpart. In the tube circuit, positive feedback is accomplished by arranging the resonant tank to be common to both the input grid and output plate circuits. The equivalent transistor circuit, using a common-emitter connection, provides positive feedback by placing the resonant tank so that it is common to both the input-base and output-collector circuits.

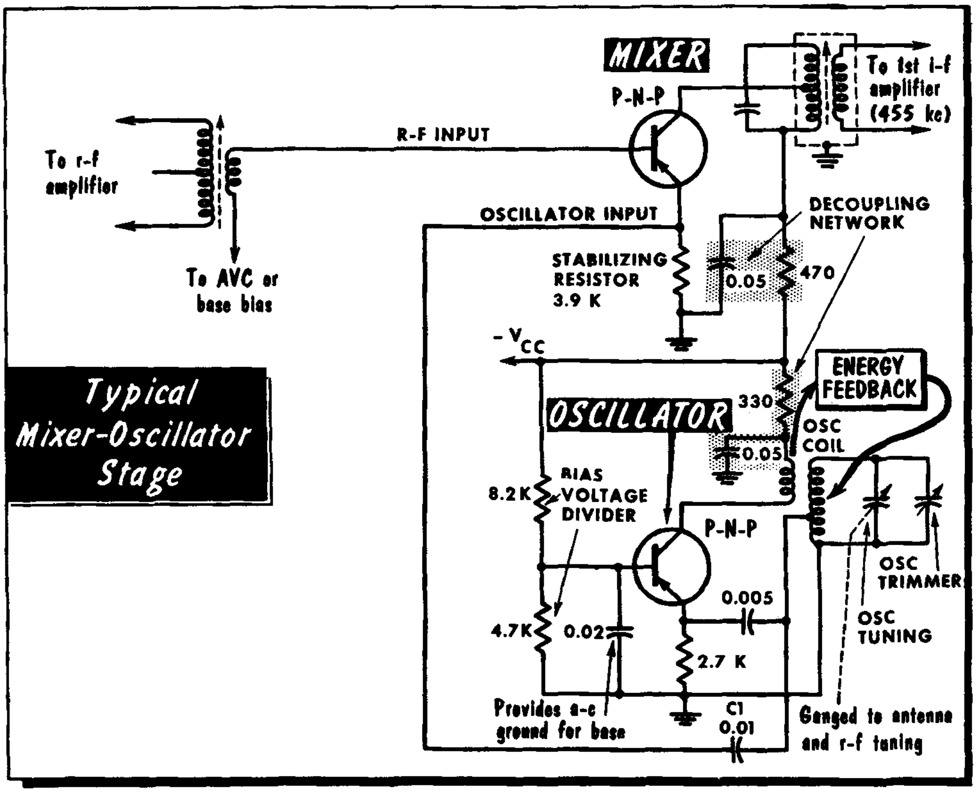
A Practical Local Oscillator

The common-emitter oscillator shown is a popular Hartley-oscillator type used in many transistor radios. It generates the unmodulated signal against which the modulated incoming signal is heterodyned in the mixer stage. The frequency-determining network consists of coil L and the oscillator tuning and trimmer capacitors. The oscillator coil, from point A to point C, resonates with the capacitors. Effectively, point C is at ground potential for a-c due to the low reactance of C1 at radio frequencies. This capacitor also prevents the collector bias from shorting to ground. The portion of the coil between points B and C forms the output circuit since it falls between the collector and the emitter. The portion of the coil between C and E forms the input circuit since it falls between the base and the emitter. The emitter, of course, is grounded through the stabilizing bypass capacitor and thus, is effectively at point C. Point D is used as a takeoff point for feeding the output of the oscillator to the mixer stage through capacitive coupling. Its exact position is determined by the circuit designer to provide the best combination of maximum signal output with optimum impedance match.



Fixed base bias is obtained by a voltage-divider network comprised of a 10K- and a 4.7 ohm resistor in parallel, with emitter stabilization provided by the 1000-ohm emitter resistor. Signal degeneration across this resistor is prevented by bypassing it with a .05-µf capacitor. The .001-µf blocking capacitor permits the signal to be fed to the base from point E, while blocking the full bias voltage from the base through points D and E in coil L.

The Mixer

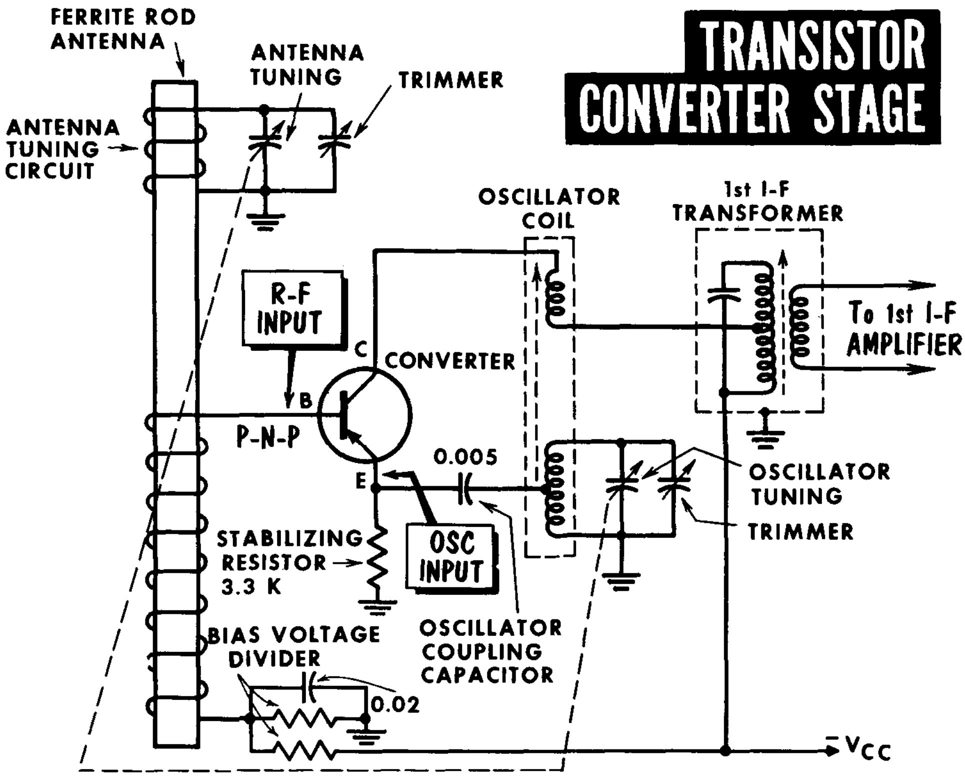


The function of the mixer stage is to heterodyne or mix the unmodulated local oscillator output signal with the incoming signal from the r-f amplifier. Many frequencies will be present in the output circuit as the result of the heterodyning process, and it remains for the i-f amplifier to select the desired frequency. The mixer stage, basically an amplifier having a tuned output circuit, is biased on a nonlinear portion of its characteristic curve (linear amplification does not produce heterodyning). By having a separate oscillator stage, the oscillator is usually apt to be more stable and unaffected by changes in other circuits.

In the typical circuit, the output from the r-f amplifier (or directly from the antenna) is fed to the base of the mixer circuit. The output from the oscillator is fed to the emitter of the mixer circuit. Thus, effectively, the two signals are mixed in series. Coupling capacitor C1 permits the passing of oscillator energy while blocking the low resistance of the oscillator coil from shunting the stabilizing resistor. The oscillator is a common tickler-feedback type, with energy from the collector circuit inductively coupled back to the emitter. The base of the oscillator is effectively at a-c ground. The usual bypass capacitor across the stabilizing resistor is eliminated to prevent the oscillator signal from being shunted to ground.

The Converter

In this circuit, the functions of mixer and oscillator stages are combined in a single unit. Its principal advantage is in the saving of a separate transistor which would have to be used as an oscillator. As in the mixer, the r-f input is fed into the base and the oscillator input is fed into the emitter. However, in the converter, the output from the collector serves two functions: it provides an output signal to the i-f transformer; and it feeds some of the output signal back into the emitter, or oscillator-input circuit. The energy fed back sustains oscillations. Once again, the stabilizing resistor is unbypassed to avoid shunting the oscillator coil.

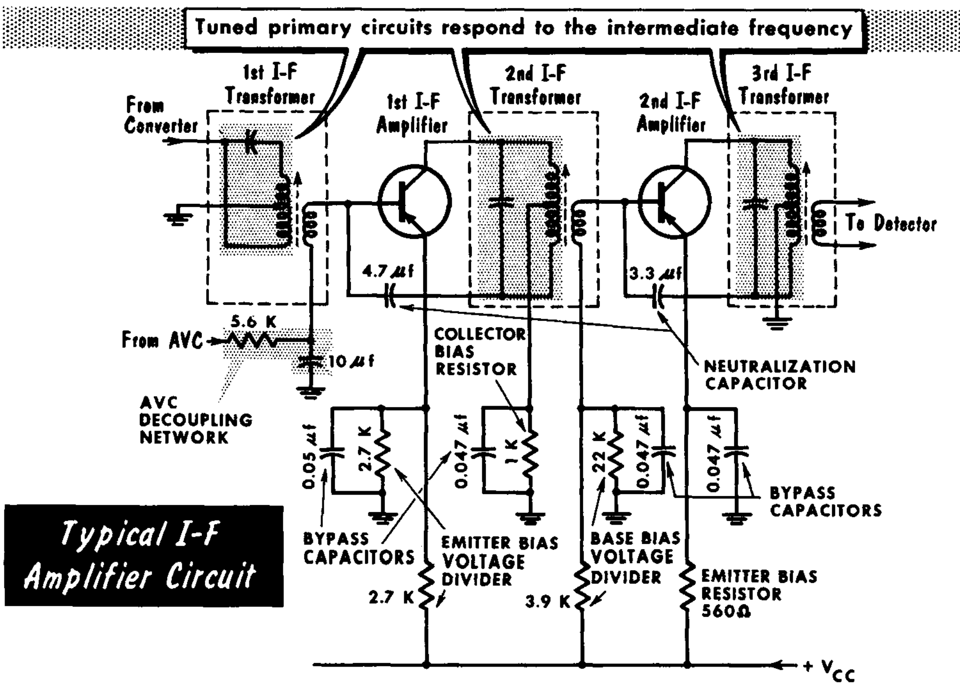


This circuit is popular since most radios do not require a stage of r-f amplification. Thus, this circuit and variations of it are found in virtually all "economy-type" receivers. There can be many variations of the oscillator circuit, but almost all operate on the principle that there is positive feedback from the collector-output circuit to the emitter-input circuit. Biasing is arranged so that the base-emitter characteristic is nonlinear, resulting in heterodyning action. The 1st i-f transformer is tuned to the intermediate frequency, usually about 455 kc. This transformer is a stepdown unit providing both a current gain and an impedance match between the converter collector circuit and the first i-f amplifier input circuit.

The I-F Amplifier

As we have learned from our study of vacuum-tube circuits, the i-f amplifier is nothing more than a fixed-tuned r-f amplifier. Its function is to select one of the many signal frequencies present in the output circuit of the converter stage, and to produce high amplification of this signal. The selected signal represents the difference frequency between the incoming r-f signal and the oscillator output. Being fixed-tuned, the i-f amplifier need not amplify a wide range of frequencies, but merely be tuned to one--the intermediate frequency (usually 455 kc). As such, all tuned circuits in the i-f amplifier (usually permeability tuned) are adjusted to peak at one single frequency, and to provide maximum gain at that frequency. The Q of the coils used in these stages can be quite high, since the band of frequencies passed need be only 10 kc ($\pm$ 5 kc from the i-f).

Because of the high gain of the i-f amplifier, the stages often tend to become unstable and oscillate, causing howling in the audio output. To avoid this, many i-f amplifiers use negative feedback; that is, they return a portion of the output (out of phase) to the input. In the circuit shown, negative feedback, or collector neutralization, is accomplished through a capacitor connected between collector and base in each stage. This cancels any positive feedback voltages developed. Two stages of i-f amplification are typical for most transistor receivers. In some circuits, the avc voltage is applied to both stages; in others it is applied to only one. Note that in this circuit, the emitters are returned to a positive voltage and the collectors to ground. This indicates that the negative side of the battery is grounded.

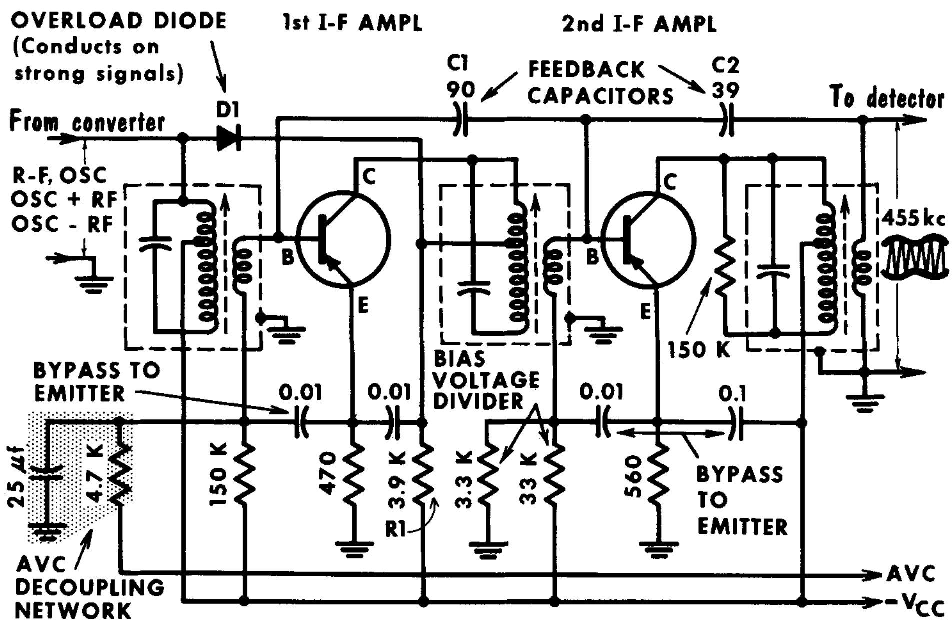


Typical I-F Amplifier Circuit

I-F Amplifier With Overload Diode

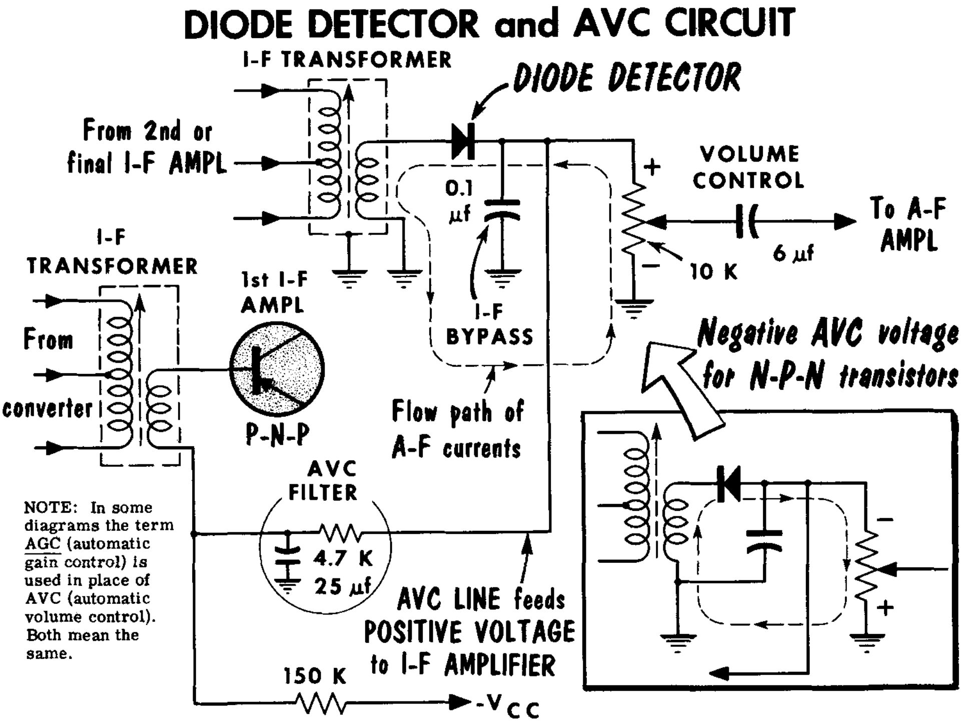
The two-stage i-f amplifier shown here incorporates an interesting feature found in many transistor radios--the use of an overload diode. Indicated in the diagram as D1, it is connected from the collector of the converter to a tap on the primary winding of the first i-f amplifier output transformer. In this position, the diode is effectively connected across a portion of the primary of the converter output transformer, and acts as a variable r-f load on the transformer. D1 is biased so that it will not conduct on weak signals. As we will see, this diode prevents overloading on strong signals.

DIODE PREVENTS OVERLOAD on STRONG SIGNALS



Using p-n-p transistors, the base is always biased negatively with respect to the emitter. When a strong signal is received, the positive avc voltage applied to the base of the first i-f amplifier is increased, thus decreasing the forward bias on that transistor. The decrease in forward bias results in a decrease in collector current, which in turn lowers the voltage drop across R1 and increases the collector voltage on the first i-f amplifier. This produces a reduction in bias across D1. On strong signals, the i-f amplifier collector voltage will approach the collector voltage of the converter stage. When this occurs, the overload diode bias is cancelled and D1 begins to conduct and load down the primary of the first i-f transformer. This damping, or loading down, lowers the gain of the circuit and compensates for the very strong signal. Also note in this circuit the feedback capacitors. This neutralization or feedback is not required for all circuits and depends upon transistor and circuit characteristics.

The Detector -- AVC



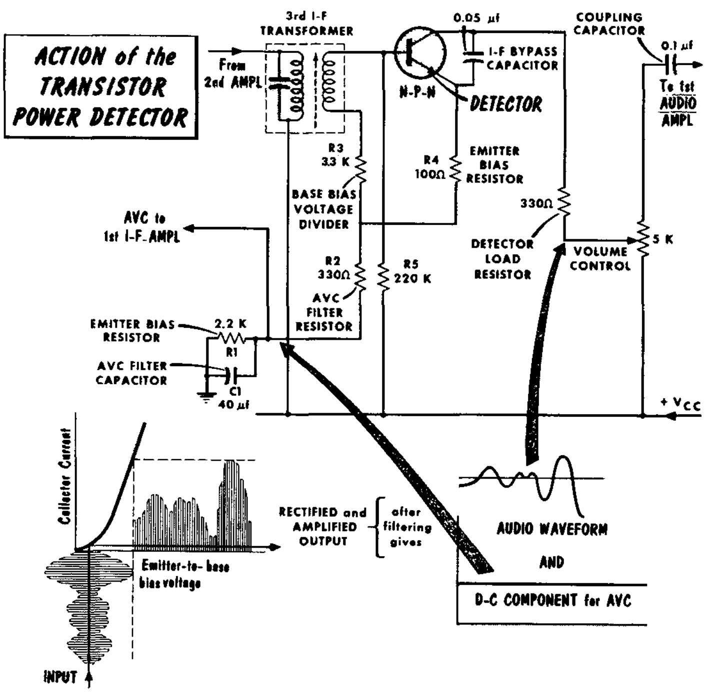
NOTE: In some diagrams the term AGC (automatic gain control) is used in place of AVC (automatic volume control). Both mean the same.

The purpose of the detector is to obtain from the modulated i-f an undistorted copy of the modulation waveform, which represents the signal intelligence. Frequently, the detector also has to produce an output signal proportional to the carrier amplitude which can be used for automatic volume control. Both diodes and transistors can be used for detection and agc. Agc is used to minimize the effects of fading, and to obtain approximately the same output volume from signals of different strengths. This control is obtained by feeding back to one or more i-f stages a voltage proportional to the carrier strength at the detector load. The control voltage has the effect of reducing the gain of the controlled stages. The most popular method of doing this is by reducing the emitter current in these stages.

In our typical circuit, the i-f signal is rectified by the diode detector, the i-f component being bypassed by the 0.1-µf capacitor. The audio component is then developed across the 10,000-ohm volume-control potentiometer. When p-n-p transistors are used as i-f amplifiers, the base is biased negatively with respect to the emitter. Thus, since the function of the agc voltage is to reduce the gain of the amplifier, the agc voltage fed to the base must be positive so as to reduce the forward bias on the transistor. In n-p-n transistors, the agc voltage must be negative. This can be done merely by reversing the polarity of the diode connection. The agc filter removes signal variations and tends to keep the agc voltage steady.

The Power Detector

The transistor detector offers the advantage of detection plus amplification over the diode detector. When used as a detector, the transistor is biased at or near cutoff, thus operating in a nonlinear portion of its characteristic curve. We can think of detection as taking place in the input or base-emitter portion of the transistor, with amplification occurring in the output or emitter-collector circuit. Operating as a class-B amplifier, this circuit provides a significant audio power gain, as well as the avc voltage. When the modulated i-f signal enters the detector, the signal is rectified and amplified. The i-f component is removed by a bypass capacitor, and the resulting audio signal is applied to the audio amplifier stage. In addition, the d-c component of the signal remains in the output and is used as the avc voltage.



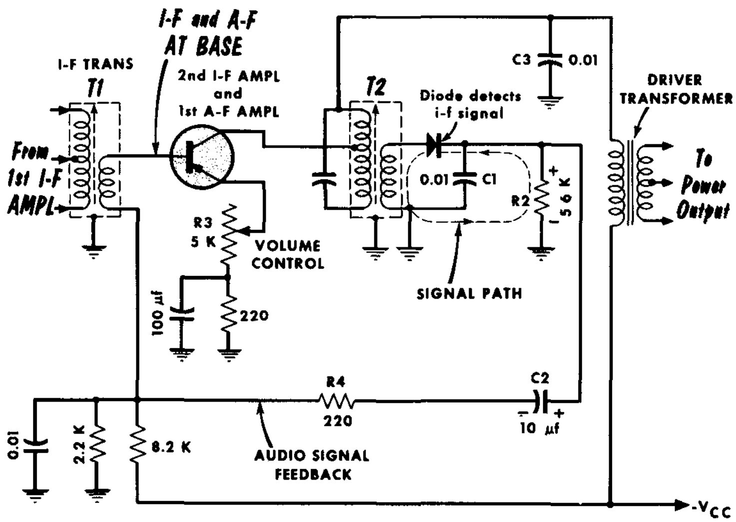
In the circuit shown, cutoff bias is applied to the base of the detector through R1, R2, and voltage divider R3-R5. When the incoming signal strength increases, the signal fed to the detector base-emitter circuit increases. This produces an increase in detector current during the conducting half cycle, and current will flow up from ground through R1, R2, R3, and then from emitter to collector in the detector. This increase in current through R1 produces an increase in positive bias which can be applied to the emitter of an n-p-n transistor or the base of a p-n-p transistor, reducing the forward bias and lowering the gain of the i-f stage. C1 and R2 act to present the proper time constant for avc filter action. The opposite will occur when a weaker signal arrives at the detector.

The Reflexed I-F Audio Amplifier

A single transistor stage, used to amplify both the intermediate and the audio frequencies, is the feature of an unusual but popular circuit called a reflexed amplifier.

In the circuit shown, the third stage operates as a standard transformer-coupled i-f amplifier. The i-f signal in the secondary of T2 is detected by a crystal diode in the following manner: When the i-f signal is positive with respect to ground, the diode conducts and charges capacitor C1. When the i-f signal reverses direction, the diode does not conduct and C1 discharges through resistor R2. Thus, the voltage across R2 is the rectified i-f signal (audio component). The amplitude of this voltage depends upon the setting of volume control R3 and upon the strength of the received signal. This voltage is fed to the base of the second i-f stage through C2, R4, and the secondary of T1. The audio signal and the i-f signal appear simultaneously at the base of the transistor, and are amplified together. Volume control R3 (in the emitter circuit) controls the gain of the i-f and the a-f signals.

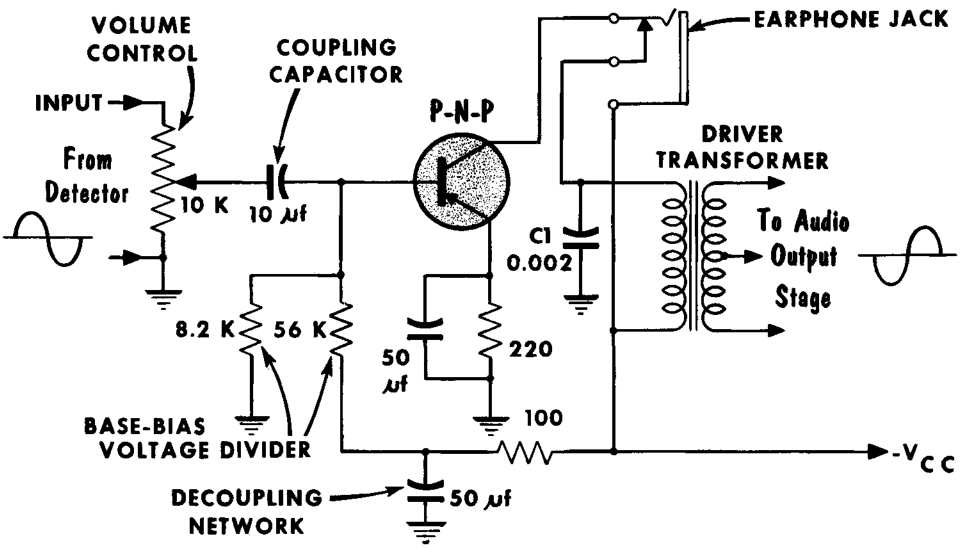
The REFLEXED I-F - A-F AMPLIFIER



No interaction takes place between the two signals because they use separate input and output loads. The driver transformer has many more turns and a higher inductance than the i-f transformer. Thus, the small i-f transformer presents little impedance to the audio signal, while the audio acts as though the i-f transformer primary was shorted. As far as i-f is concerned, C3 presents a low impedance to the i-f, decoupling the i-f from the primary of the audio transformer. Thus, only audio appears in the primary of the driver transformer. The transformers therefore do not offset each other and may be connected together.

The Audio Amplifier

The Single-Stage Audio Amplifier



The output of the detector stage consists of an extremely weak audio signal that represents the modulation of the incoming r-f signal. What is required at this point is a stage of audio amplification to build up this small audio signal to a point where it can drive the power output stage. The power output stage, usually push-pull, requires far more driving power than can be delivered by the detector stage directly, except in some cases where a power detector is used. In most cases, one stage of audio amplification is required; however, a two-stage resistance-coupled amplifier is not unusual.

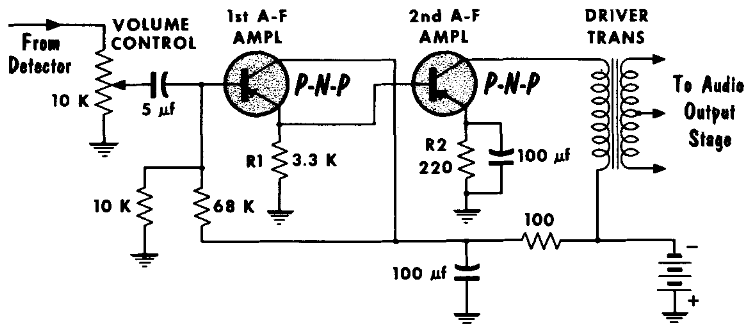
In the typical audio driver stage shown, the input signal is taken from the sliding contact of the volume control and fed to the base of the audio amplifier through a 10-µf coupling capacitor. This capacitor isolates or blocks the d-c base-bias voltage of the audio stage from the detector and volume-control circuitry. The base is biased through conventional voltage-divider bias, and the emitter contains a 50-µf capacitor bypassing a 220-ohm stabilizing resistor. Transformer coupling is used to match the high collector output-circuit impedance of the audio amplifier to the low input impedance of the output stage.

An interesting feature found on many sets is the earphone jack in the audio stage. While the output of this stage is insufficient to drive a loudspeaker, it can actuate earphones. When the phones are not inserted, the jack is short-circuited and there is normal circuit action. When the phones are inserted, the primary of the driver transformer is open and the phone impedance forms the collector load of the audio stage. C1 is used to stabilize the collector circuit and prevent oscillation; it also acts to improve tone response.

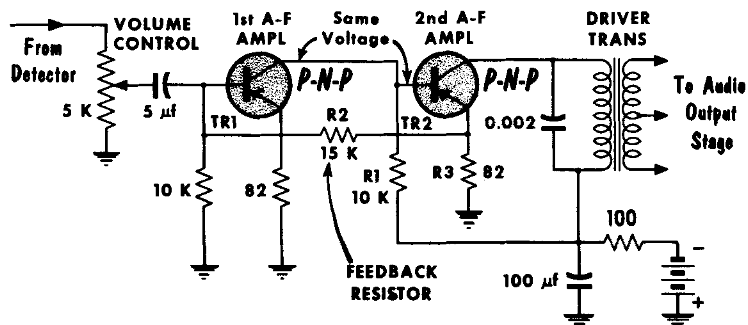
Direct-Coupled Amplifiers

To obtain high audio amplification together with good low-frequency response and simple circuitry, direct-coupled amplifiers sometimes are used. In the first circuit, using two p-n-p transistors, the first amplifier stage is an emitter-follower. The signal developed across load resistor R1 is directly coupled to the base of the second stage. Since the input impedance of the emitter-follower is high, a much better impedance match is obtained between the detector and the audio driver, reducing signal distortion. Since R1 is the load resistor, it cannot be bypassed. However, R2, which is not the load resistor for the second stage, is bypassed to avoid degeneration. The voltage fluctuations across R1 represent the signal input to the base of the second audio stage.

In the second circuit, also using two p-n-p transistors, both stages are common-emitter class-A amplifiers, with the input to TR1 180° out of phase with the input to TR2. R1 lowers the collector voltage of TR1 so that the base of TR2 is less negative than its collector. R2 provides negative feedback from TR2 to TR1. This reduces the gain of TR1 and cuts down on high signal peaks. A negative-going signal applied to the base of TR1 produces an increase in collector current and a voltage drop across R1, making the base of TR2 less negative. This reduces conduction in TR2 and the voltage drop across R3. Thus, a small positive-going voltage is fed back through R2 to the base of TR1. This positive change opposes the negative voltage that was fed to TR1, thus reducing distortion and stabilizing d-c operating conditions.

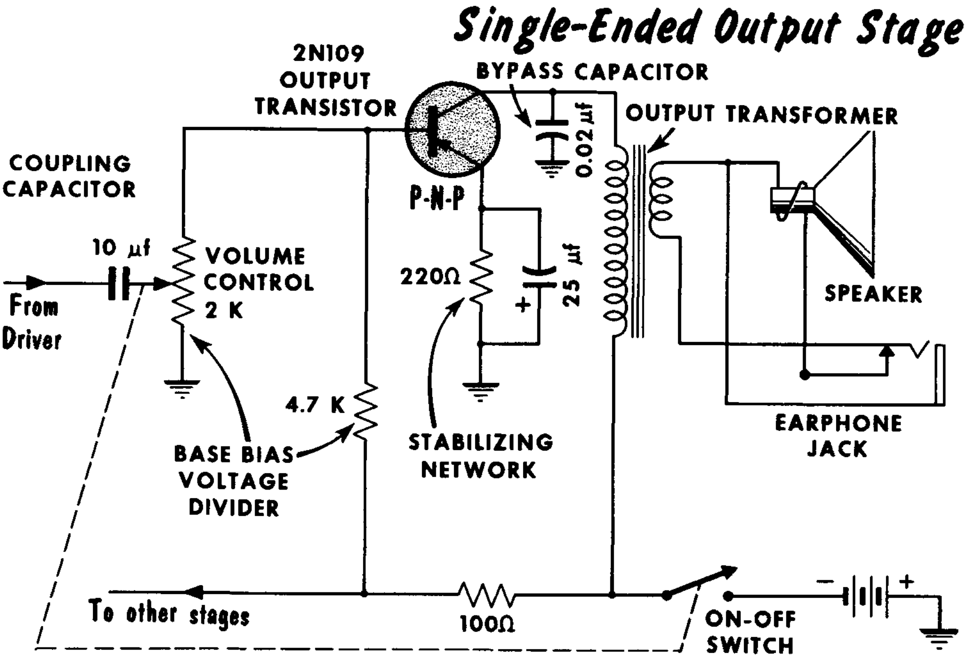


TWO-STAGE DIRECT-COUPLED AMPLIFIERS



Single-Ended Audio Output Stage

Most transistor radios use a push-pull output stage to supply the necessary power needed to drive the loudspeaker. However, it is not unusual to see a single-ended stage. To prevent distortion, it is necessary that this stage

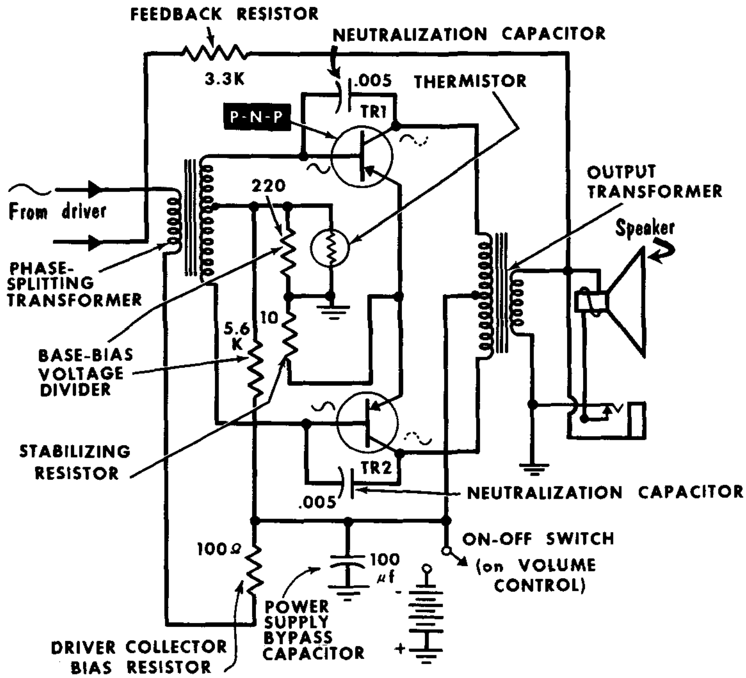


be driven as a class-A amplifier, where collector current flows throughout the entire input cycle. While termed a "power" amplifier, the typical portable receiver using a single transistor in its output stage is only capable of about 100 milliwatts output--small by vacuum tube standards but adequate for small 3-inch speakers. Where an appreciable amount of heat is to be dissipated by a power transistor, a <u>heat sink</u> is required. This is a metal conductor that will remove the excess heat. Often, the metal chassis is used as a heat sink. Without an adequate heat sink, <u>thermal runaway</u> may develop. This condition occurs when excessive heat causes increased collector current, which in turn causes still more heat and still more current.

In the typical single-transistor audio output stage, the input can come directly from the detector, or from a previous audio amplifier. The signal is coupled to the base of the output stage, with the volume control forming part of the base-bias voltage divider. Here again, the emitter stabilizing resistor is bypassed to avoid signal degeneration. A conventional output transformer matches the collector impedance to the speaker coil impedance. The transformer primary is bypassed to ground to avoid noise voltages and prevent high-frequency distortion. The jack permits listening with earphones and muting the speaker.

The Push-Pull (Double-Ended) Amplifier

DOUBLED-ENDED (PUSH-PULL) OUTPUT STAGE



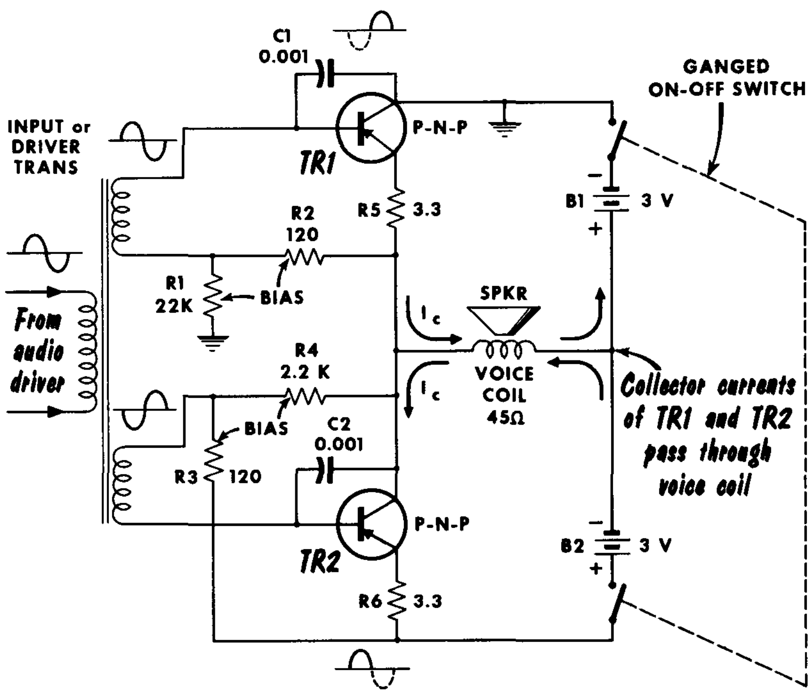
This amplifier consists of two transistor circuits operating 180° out of phase, but with their outputs combined. It can be operated either class A or class B, with the latter offering the advantage of efficiency in that collector current flows in each transistor only during alternate half cycles. A feature of push-pull circuitry is the elimination of even-order harmonics and the d-c component in the load. In class-B operation, the output power can be as much as four times the collector dissipation of each transistor.

In our circuit, each transistor is biased near cutoff (class B), and a split-secondary input transformer feeds out-of-phase signals to TR1 and TR2. Amplification takes place during the half cycle that each transistor conducts, and the outputs are combined in the secondary of the output transformer. Note the thermistor connected in parallel with one of the base-bias resistors. The thermistor resistance decreases as the temperature of this resistor increases, thus lowering the resistance of the parallel combination (which in turn reduces the base-bias voltage). The reduction in base bias limits current flow in the collector circuits, keeping the power dissipation of the transistors within the proper limits for the temperature at which they are operating. The feedback loop from one end of the output transformer secondary to the driver-stage emitter provides some inverse feedback which stabilizes overall operation of these circuits.

Asymmetric Push-Pull Amplifier

This interesting and popular push-pull circuit has the output transistors connected in series across the collector supply battery, and the load resistance (loudspeaker voice coil) connected between their junction and a centertap on the battery. Separate secondary windings are needed on the input transformer for the two bases because of the special base-bias connections. No output transformer is required.

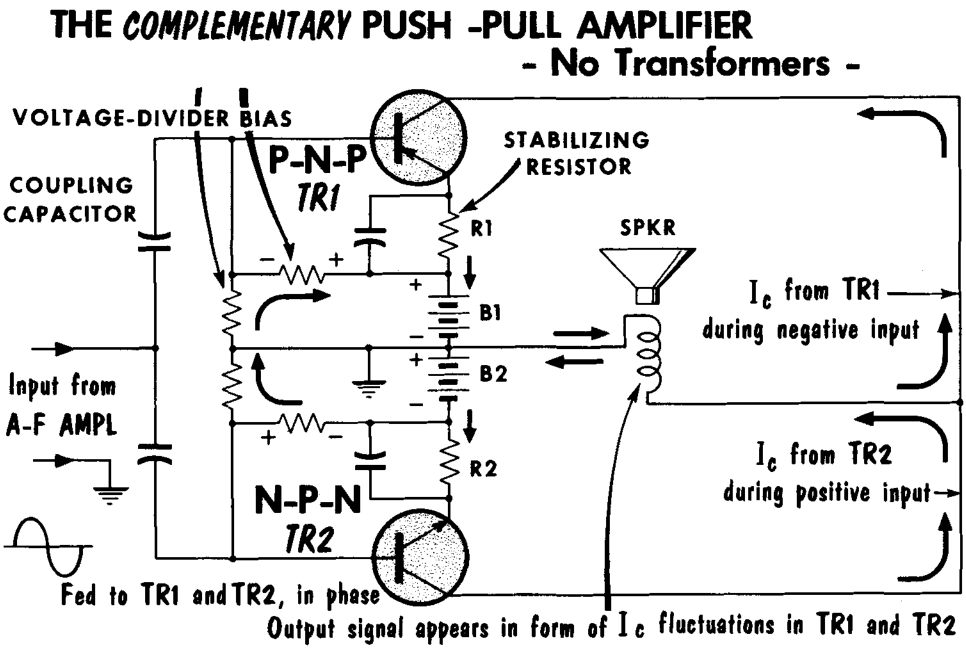
PUSH-PULL AMPLIFIER REQUIRING *NO* OUTPUT TRANSFORMER



Both transistors are biased near cutoff so that with no signal input, neither is conducting. Out-of-phase signals are fed to the base circuit of each transistor from the driver transformer. Each transistor thus conducts on alternate half cycles of the incoming signal. As a result, the collector currents of TR1 and TR2 flow alternately through the speaker voice coil. Base bias for TR1 is produced by current flow from ground (-) through R1 and R2, the speaker voice coil, and to the positive (+) terminal of the battery. Bias for TR2 is produced by current flow from the positive (+) terminal of B2, through the voice coil, and through R4 and R3 to the negative (-) terminal of B2. Resistors R5 and R6 provide the necessary d-c stabilization. The proper value of these resistors permits the d-c operating point to be fairly independent of changes in transistors and temperature fluctuations. Capacitors C1 and C2 provide degenerative feedback to improve the overall frequency response. The principal advantage of this circuit is that no output transformer is required.

Complementary Symmetry Push-Pull Circuit

This unusual circuit has no parallel in vacuum-tube circuitry. It depends upon the complementary nature of transistors. That is, an n-p-n and a p-n-p transistor can be made with identical characteristics except for the difference in polarities required. This permits the use of a push-pull circuit having neither an input nor an output transformer. This amplifier contains an n-p-n

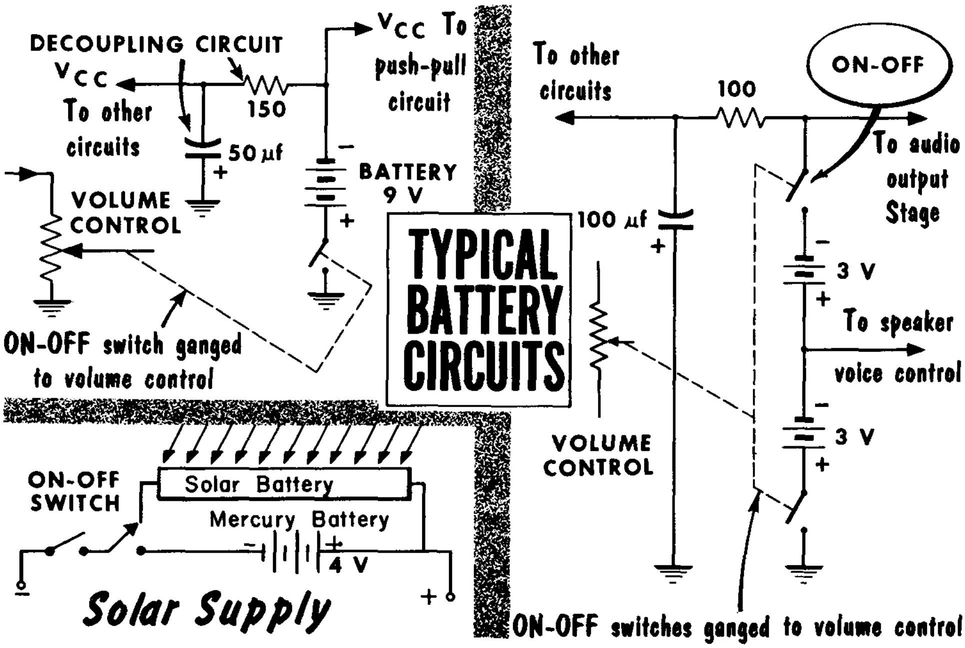


and a p-n-p transistor, each connected in a common-emitter stage. Conventional voltage-divider base bias is used for each transistor, with bypassed stabilizing resistors in each emitter circuit. A single input feeds the same signal to the base of each transistor, and a high-impedance loudspeaker voice coil acts as the output load.

We can analyze this circuit by considering it as a class-A amplifier. Thus, with zero-signal input, both transistors are forward biased and collector current flows. Under this condition, there is perfect balance, and current flows from B2 (-) through R2, TR2, TR1, R1, and back to B1 (+). Since this now acts as a balanced Wheatstone bridge, no current will flow through the load or speaker coil. With a positive-going signal applied to both transistors, forward bias on the p-n-p transistor decreases, while forward bias on the n-p-n unit increases. Since the forward bias on TR2 is now greater than that on TR1, collector current from TR2 will flow through the speaker coil. When a negative-going signal is applied, the opposite occurs, and TR1 collector current flows through the coil. Thus, the current fluctuations through the coil are determined by the amplitude and frequency of the incoming signal, and audio signals are reproduced by the loudspeaker.

Battery-Operated Power Supplies

Virtually all transistor radios are battery operated. These batteries are made up of several cells connected in series or a single battery. The most common voltages used are 6 and 9 volts, while voltages of 3, 4, 5, and 9 are not unusual. The most popular cell is the 1.5-volt zinc-carbon dry cell. Another popular source of power is the mercury cell, having a voltage rating of about 1.34 volts. This also can be connected in series for higher voltage ratings. Because of the low current drain from a typical portable transistor radio--often less than 10 ma--battery life often may be as long as 400 hours using large cells such as the D-size zinc-carbon unit. Another popular cell is the 1.25 volt nickel-cadmium type. This is a secondary cell and can be recharged often by plugging it into a 117-volt a-c supply. Another popular innovation is the solar cell, which produces a voltage when exposed to the sun or artificial light. Conventional solar batteries produce a voltage of 4 to 5 volts that can power a transistor radio directly, or can charge a battery which in turn powers the radio.



It is most important that battery polarity always be observed. Whether n-p-n or p-n-p transistors are used, a battery can have either its positive or negative terminal grounded or common. Failure to observe this may cause instant damage to the transistors. For instance, a p-n-p transistor requires that the collector be negative with respect to the emitter. If the positive battery terminal is grounded, the emitter will be connected directly or through a network to ground. If the negative terminal is grounded, the collector circuit will connect to ground. Most transistor receivers have an on-off switch ganged to the volume control, and connected so as to open and close the battery circuit.

A Typical Transistorized Receiver

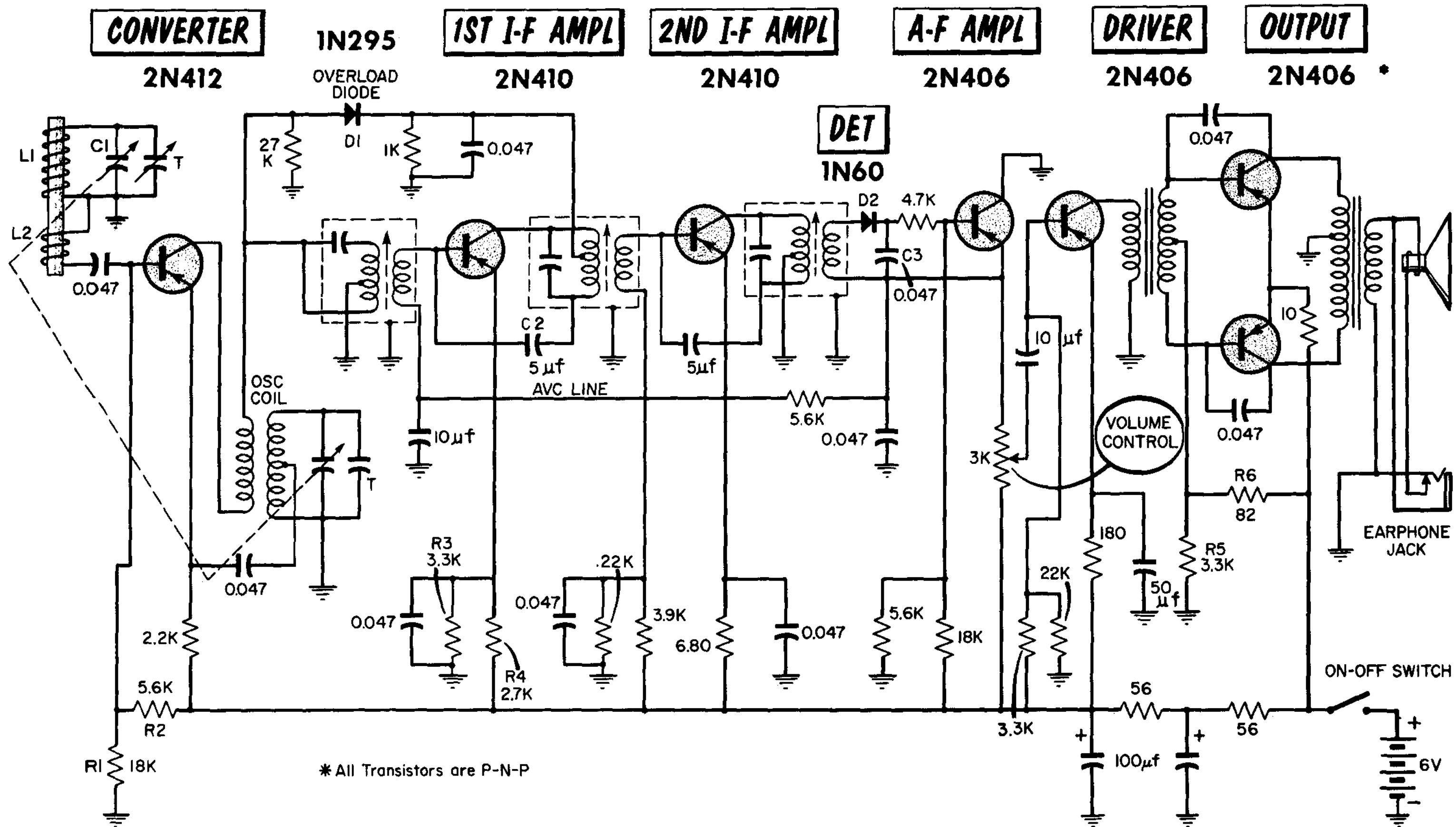
This receiver uses all p-n-p type transistors and has the negative side of the battery grounded. Antenna coil L1 is of the ferrite-rod type and is tuned to the incoming signal by C1. Coil L2 is inductively-coupled to L1 and applies the incoming signal to the base-emitter circuit of the converter through a .047 μf coupling capacitor. Voltage divider R1-R2 provides base bias to this stage. Collector current flows through the oscillator coil and the first i-f input transformer. The tuned portion of the oscillator circuit consists of the oscillator coil in parallel with the oscillator tuning capacitor. This tuned circuit is ganged with the antenna tuning capacitor and is fed in-phase energy from the collector portion of the oscillator coil. This energy is amplified after being capacitively coupled to the emitter of the converter. Thus, the collector circuit of the converter contains both the incoming r-f signal and the local oscillator signal. The "beat", or i-f signal, is selected by the primary of the first i-f input transformer.

Two stages of transformer-coupled i-f amplification are used. The first i-f amplifier is base-biased by the avc line from the diode detector. In addition, fixed-emitter bias is provided by voltage divider R3-R4. R3 is bypassed to prevent signal degeneration. An overload diode (D1), which effectively shunts a portion of the first i-f transformer primary, varies the gain of the incoming signal should it exceed a certain amplitude. Negative feedback is provided in this stage by C2. The second i-f amplifier uses conventional voltage-divider base bias, and transformer-couples its output signal to crystal diode detector D2.

The detector circuit is connected to the base-emitter input of the audio amplifier, with the avc line taken from the emitter lead. Capacitor C3 bypasses the i-f component of the incoming signal. The audio amplifier is a grounded-collector type, with the volume control acting as the load for this stage. The base is voltage-divider biased. A 10-μf capacitor couples the audio signal from the volume control to the input circuit of the driver. The driver stage is transformer coupled to the push-pull output stage, and supplies the necessary power for the operation of the output stage.

The secondary of the driver transformer is centertapped, with equal and opposite signals being fed to the base circuits of each push-pull transistor. Voltage divider R5-R6 provides base bias for the output stage, with an unbypassed 10-ohm emitter resistor providing stabilization through negative feedback. Both output transistor circuits are stabilized by capacitors connected between base and collector. The output transformer couples and matches the high impedance of the collector circuits to the low impedance of the loudspeaker. Provision is made for connection of an earphone. When the earphone is connected, the loudspeaker circuit is open, and only the earphone is in the circuit. This push-pull output circuit is operated class B, and battery current will increase significantly with an increase in input signal strength. The entire receiver is powered by a 6-volt battery.

A TYPICAL TRANSISTORIZED RECEIVER



The higher the collector voltage, the higher the noise factor of a transistor.

Interstage coupling may make use of R-C, impedance, or direct coupling.

Conventional push-pull amplifiers require a centertapped secondary winding input transformer or a phase inverter as a driver.

Complementary symmetry eliminates the need for a centertapped secondary-input transformer and a phase-inverter driver.

Because the transistor is essentially a power amplifier (even at high frequencies), impedance matching between the output of one stage and the input of the following stage is important for maximum gain.

Impedance matching is achieved by selecting the proper turns ratio of the primary and secondary winding in the transformer coupling.

Positive feedback occurs in a transistor amplifier through the collector-base junction capacitance. At high frequencies, the amount of feedback may be sufficient to sustain oscillations.

Neutralizing circuits are used to eliminate the possibility of oscillations in high-frequency tuned amplifiers.

Automatic volume control of a transistor amplifier is usually achieved by controlling the emitter d-c current or the collector d-c voltage.

The mixer stage is fundamentally an amplifier having a tuned output circuit, and biased on a nonlinear portion of its characteristic curve.

To avoid oscillation, many transistor i-f stages use negative feedback.

To prevent overloading on strong signals, an overload diode is often used.

Avc works in transistor radios to reduce the gain of the controlled stages by reducing the emitter current in these stages.

The transistor detector offers the advantage of detection plus amplification.

The reflex amplifier is a single transistor stage used to amplify both the intermediate and the audio frequencies.

Direct-coupled amplifiers provide good audio amplification together with good low-frequency response and circuit simplicity.

Where an appreciable amount of heat is to be dissipated by a power transistor, a heat sink is required.

In class-B operation, the output power can be as much as four times the collector dissipation of each transistor.

In the asymmetric push-pull amplifier, no output transformer is required.

Popular cells used in transistor radios include the zinc-carbon, mercury, nickel-cadmium, and the solar type.

REVIEW QUESTIONS

1. What is the function of the transistor r-f amplifier stage?
2. Explain the basic operation of the transistor oscillator.
3. What is the function of the mixer stage?
4. Basically, how does the converter differ from the mixer?
5. Explain the operation of the overload diode.
6. What is the advantage of the power detector over the diode detector?
7. Explain the operation of the reflexed amplifier.
8. What are the advantages of direct-coupled amplifiers?
9. What is meant by thermal runaway and how is it controlled?
10. Explain the operation of the complementary symmetry push-pull circuit.
11. Why is it important to observe battery polarity in a transistorized circuit?

GLOSSARY

Acceptor: A substance with three electrons in the outer orbit of its atom which when added to a semiconductor crystal, provides a hole in the lattice structure of the crystal. An acceptor (indium or gallium) is a p-type impurity.

Alpha (α): The current gain factor of a transistor when connected in a common-base circuit. Alpha is equal to the ratio of collector current change to emitter current change for a constant collector voltage.

Barrier: The electric field between the acceptor ions and the donor ions at a junction.

Barrier Height: The difference in potential from one side of a barrier to the other.

Base: The center semiconductor material of a double junction (n-p-n or p-n-p) transistor. The base is comparable to the grid of an electron tube.

Beta (β): The current gain factor of a transistor connected in a common-emitter circuit. Beta is equal to the ratio of a change in collector current to a change in base current for a constant collector voltage.

Bias: The d-c operating voltage or current applied to an element of a transistor. Bias current establishes the operating point of a transistor.

Collector: The end semiconductor material of a double junction (n-p-n or p-n-p) transistor that is normally reverse-biased with respect to the base. The collector is comparable to the plate of an electron tube.

Common-Base Amplifier: A transistor amplifier in which the base is common to the input and output circuits. This circuit is comparable to the grounded-grid triode circuit.

Common-Collector Amplifier: A transistor amplifier in which the collector is common to the input and output circuits. This circuit is comparable to the cathode follower electron-tube circuit.

Common-Emitter Amplifier: A transistor amplifier in which the emitter is common to the input and output circuits. This circuit is comparable to the conventional common-cathode electron-tube circuit.

Complementary Symmetry: An arrangement of p-n-p and n-p-n transistors that provides push-pull operation from a single input signal. Such a circuit makes use of the similar, but opposite characteristics of p-n-p and n-p-n transistors.

Cutoff Frequency: The frequency at which the alpha of a transistor falls below 0.707 times the maximum gain.

Donor: A substance with electrons in the outer orbit of its atom which, when added to a semiconductor crystal, provides a free electron in the lattice structure of a crystal. A donor is an n-type impurity. Typical donors are antimony and arsenic.

Electron-Pair Bond: A valence (covalent) bond formed by two electrons, one from each of two adjacent atoms.

Emitter: The end semiconductor material of a double junction (p-n-p or n-p-n) transistor that is forward-biased with respect to the base. The emitter is comparable to the cathode of an electron tube.

Forward Bias: An external potential applied to a p-n junction so that the barrier is lowered and relatively high current flows through the junction.

Heat Sink: A mass of metal or other good heat conductor used to rapidly dissipate the heat energy produced by a transistor.

Hole: A mobile vacancy in the electonic valence structure of a semiconductor. The hole acts similarly to a positive electronic charge having a positive mass.

Impurity: A substance added to a semiconductor to give it a p-type or n-type characteristic.

Junction: A point or area of contact between n- and p-type semiconductors.

Junction Transistor: A device having three alternate sections of p-type or n-type semiconductor material.

Lattice Structure: In a crystal, a stable arrangement of atoms and their electron-pair bonds,

Majority Carriers: The holes in p-type semiconductors or free electrons in n-type semiconductors.

Minority Carriers: The holes in n-type semiconductors or excess electrons in p-type semiconductors.

N-P-N Transistor: A device consisting of a p-type section and two n-type sections of semiconductor material, with the p-type in the center.

N-Type Semiconductor: A semiconductor crystal into which a donor impurity has been introduced. It contains free electrons.

P-N Junction: The area of contact between n-type and p-type semiconductor materials.

P-N-P Transistor: A device consisting of an n-type section and two p-type sections of semiconductor material, with the n-type in the center.

Point Contact: A physical connection made by a metallic wire on the surface of a semiconductor.

P-Type Semiconductor: A semiconductor crystal into which an acceptor inpurity has been introduced. It provides holes in the crystal lattice structure.

Reverse Bias: An external potential applied to a p-n junction to raise the barrier and prevent the movement of majority current carriers.

Saturation (Cutoff) Current: The current flow between the base and collector or between the emitter and collector, measured with the emitter lead or the base lead open.

Semiconductor: A conductor whose resistivity is between that of metals and insulators. It exists in crystalline form.

Stabilization: The reduction of variations in voltage or current due to undesirable circuit changes.

Transistor: A semiconductor device capable of transferring a signal from one circuit to another and producing amplification.

Zener Diode: A p-n junction diode reverse-biased into the breakdown region, used for voltage stabilization.

INDEX TO VOL. V

(Note: A cumulative index covering all six volumes in this series is included at the end of Volume VI.)

basic radio

by MARVIN TEPPER

Electronic Services Division
Raytheon Company

Author of
FUNDAMENTALS OF RADIO TELEMETRY

VOL. 6

JOHN F. RIDER PUBLISHER, INC., NEW YORK

Copyright September 1961 by John F. Rider Publisher, Inc.

All rights reserved. This book or any parts thereof may not be reproduced in any form or in any language without permission of the publisher.

Library of Congress Catalog Card Number 61-11229

Printed in the United States of America

Fifth Printing, 1968

PREFACE

The purpose of this book is to fill a need for a text stating in plain, everyday language, what many people consider a complex technical subject. A technical subject need not be complex. Careful filling in of the background with essential information, and then leading step by step to the final explanation, provides a logical method of explaining the most difficult subject.

It would be impossible to cover in a single book or series of books, the immense scope implied in the word *electronics*. However, an understanding of radio circuits serves as a foundation for advanced study in all fields of electronics, such as television, radar, computers, etc. For teaching radio, the all-important basic tool of electronics, most available textbooks are woefully inadequate. One type contains information so brief as to acquaint rather than instruct. Another type is based on the premise that teaching a student to design a circuit is the best method of having him understand that circuit's operation.

Basic Radio represents the neglected middle ground. It is a course in radio communications, as distinct from a general course in electronics. The text deals with the circuitry and techniques used for the transmission and reception of intelligence via radio energy. Assuming no prior knowledge of electricity or electronics, the six volumes of this course "begin at the beginning" and carry the reader in logical steps through a study of electricity and electronics as required for a clear understanding of radio receivers and transmitters. Illustrations are used on every page to reinforce the highlights of that page. All examples given are based on actual or typical circuitry to make the course as practical and realistic as possible. Most important, the text provides a solid foundation upon which the reader can build his further, more advanced knowledge of electronics.

The sequence of *Basic Radio* first establishes a knowledge of d-c electricity. Upon this is built an understanding of the slightly more involved a-c electricity. Equipped with this information the reader is ready to study the operation of electron tubes and electron tube circuits, including power supplies, amplifiers, oscillators, etc. Having covered the components of electronic circuitry in Volumes 1 through 3, we assemble these components

in Volume 4, and develop the complete radio receiver, AM and FM. In Volume 5 we recognize the development of the transistor, and devote the entire volume to the theory and circuitry of transistor receivers and semiconductors. The last volume of the course, Volume 6, covers the long-neglected subject of transmitters, antennas, and transmission lines.

No prior knowledge of algebra, electricity, or any associated subject is required for the understanding of this series; it is self-contained. Embracing a vast amount of information, it cannot be read like a novel, skimming through for the high points. Each page contains a carefully selected thought or group of thoughts. Readers should take advantage of this, and study each individual page as a separate subject.

Whenever someone is presented with an award he gives thanks and acknowledgement to those "without whose help . . ." etc. It is no different here. The most patient, and long-suffering was my wife Celia, who typed, and typed, and typed. To her, the editorial staff of John F. Rider, and others in the "background", my heartfelt thanks and gratitude for their assistance and understanding patience.

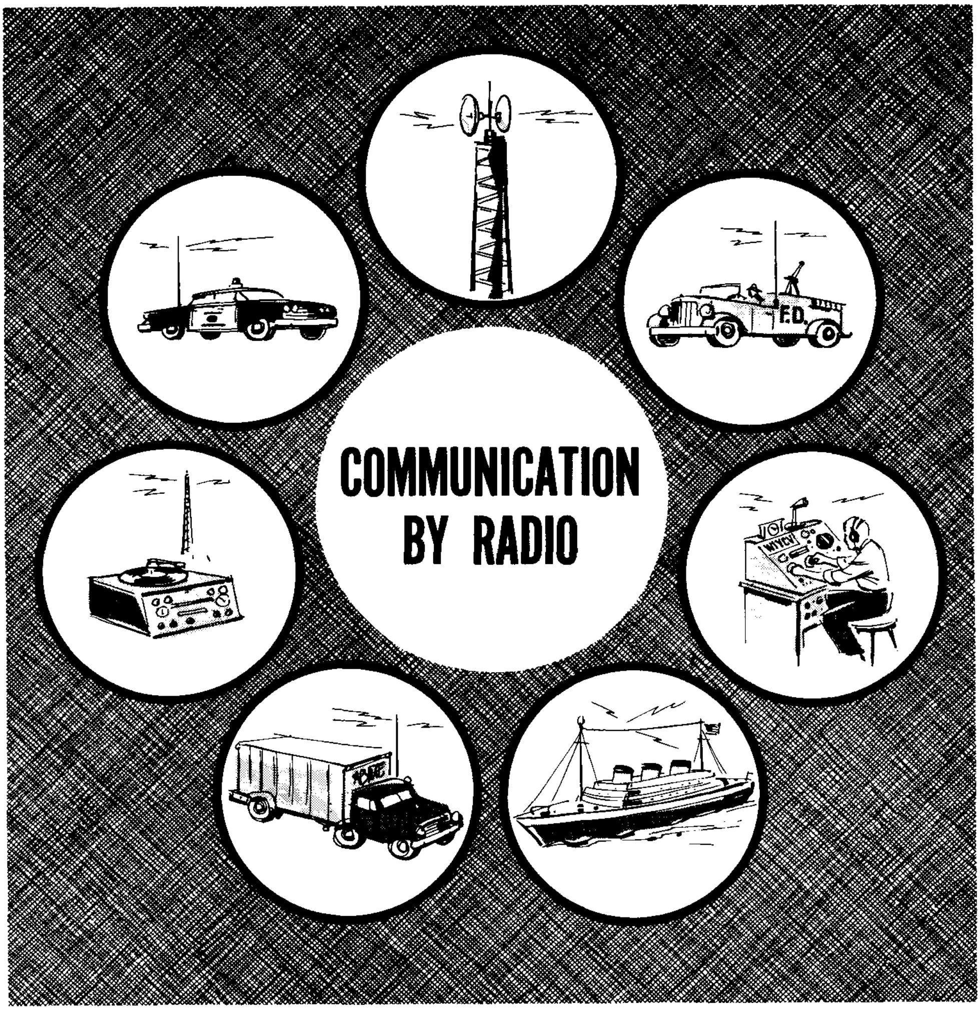
MARVIN TEPPER

Malden, Mass.
September 1961

CONTENTS

VOL. VI — BASIC RADIO

Communication by Radio

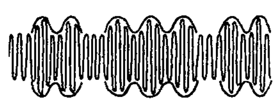
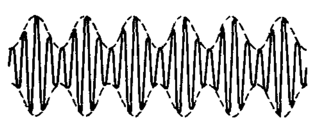


In the broadest sense, "communication by radio" means the transfer of intelligence from one point to another through space, using radiated electromagnetic energy (radio waves) in the frequency spectrum of from about 10 kc to about 30,000 mc. The transmitting antenna radiates these waves and the receiving antenna intercepts them. This extremely broad span of frequencies embraces many different techniques which come within the meaning of radio communications. In some of them the nature of the intelligence transmitted differs greatly from that dealt with in this book. But even so, there are many similarities in the organization and operation of the different transmitters used in the various radio services. Understanding the theory of operation of the conventional radio transmitter will help in comprehending the functioning of all other kinds of transmitters.

Kinds of Radio Signals

We have already established that communication by radio is accomplished by electromagnetic energy (electromagnetic waves) which travel from a transmitting antenna to the receiving antenna. This broad statement does not, however, identify the kinds of radio signals that come within the boundaries of electromagnetic waves. The differences among radio signals arise prin-

KINDS of RADIO SIGNALS

Emission Designation	Pictorial Representation	Description
A0		Continuous oscillations of constant amplitude and frequency. Nonmodulated waves.
A1	 Dot Dash Dot Dash	Continuous oscillations of constant amplitude and frequency, interrupted (keyed) in a sequence which corresponds to an International code.
A2	 Dot Dash Dot	Modulated telegraphy in which an audio tone amplitude modulates a carrier. The audio tone is keyed in accordance with an International code.
A3		Telephony. Constant-frequency oscillations modulated by the intelligence being transmitted.

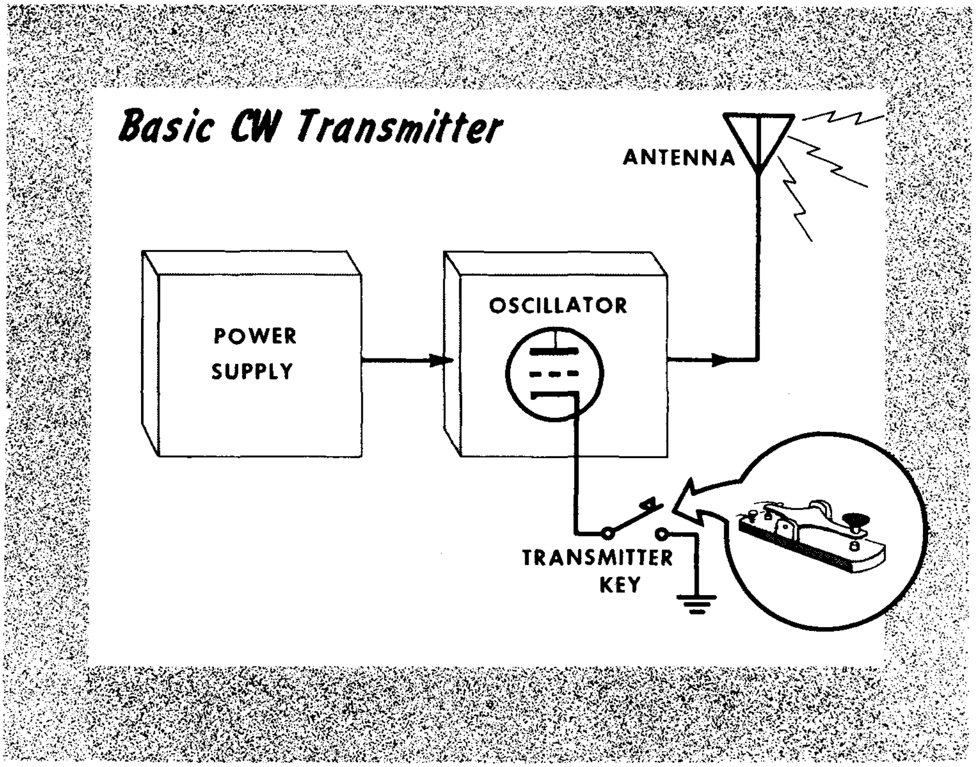
cipally from the techniques employed in making the intelligence part of the radiated signal prior to radiation.

These differences give rise to different identifications. The identity of each type of radio signal is called a type of emission. Modern communication technology has given rise to a great many kinds of modulation; hence, kinds of emission. This being a course in basic radio, only the broadly applicable techniques are discussed. Where variations warrant mention, they are dealt with separately.

The Basic Radio Transmitter (Radiotelegraphy)

The most fundamental radio transmitter is the radiotelegraph transmitter. It consists of a vacuum-tube generator of radio-frequency energy (the oscillator); a source of operating voltages (the power supply); a signaling key whereby the generation of oscillations can be started and stopped at will, and an antenna. In the strict sense the antenna is not part of the transmitter, but it is a necessary adjunct, without which a transmitter is useless.

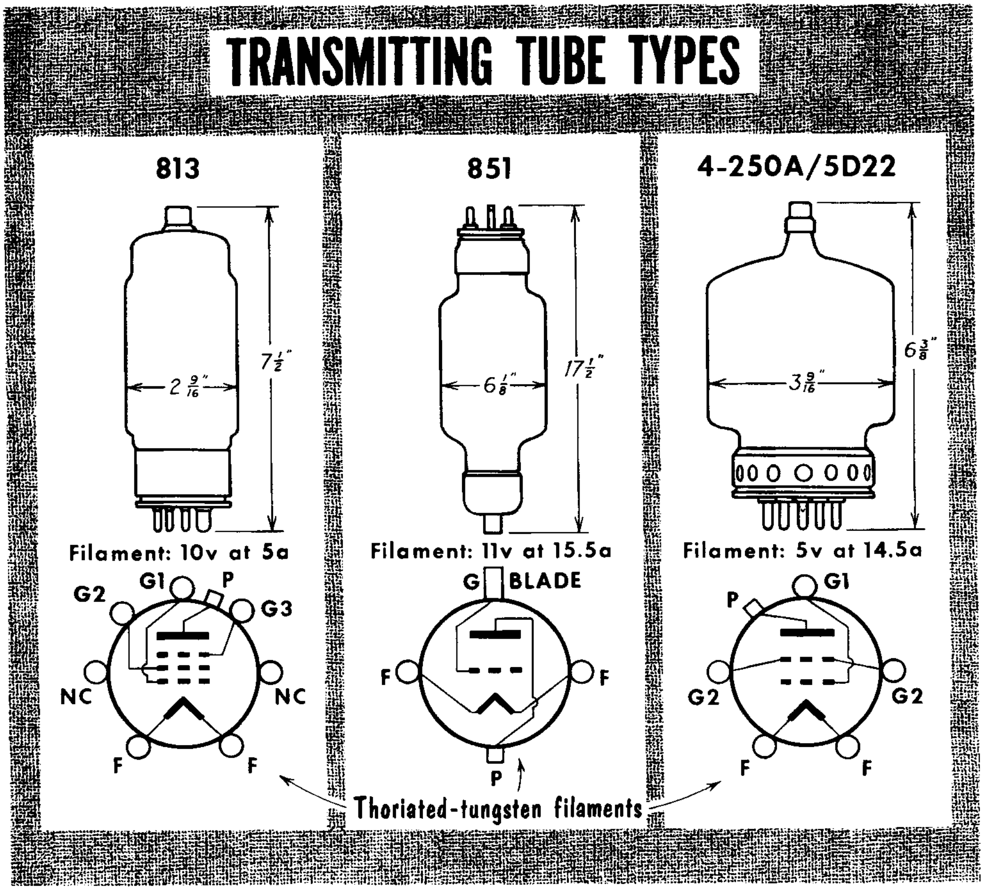
The basic transmitter made up of the abovementioned parts has many shortcomings. Its greatest deficiency is its inability to keep the oscillator constant on a precise frequency setting; in other words, the radiated signal is not constant in frequency, and consequently the received signal is difficult to



"read". Also, the signal power derived from the oscillator is low, which limits the distance over which communications can be carried on.

Manipulating the signaling key to close and open the circuit in accordance with the radiotelegraphy code causes the oscillator to generate corresponding bursts of energy. This energy radiates from the antenna to "transmit" the message. Transmitters used for radiotelegraph purposes are called continuous wave or CW transmitters.

Transmitting Tubes: The Emitter



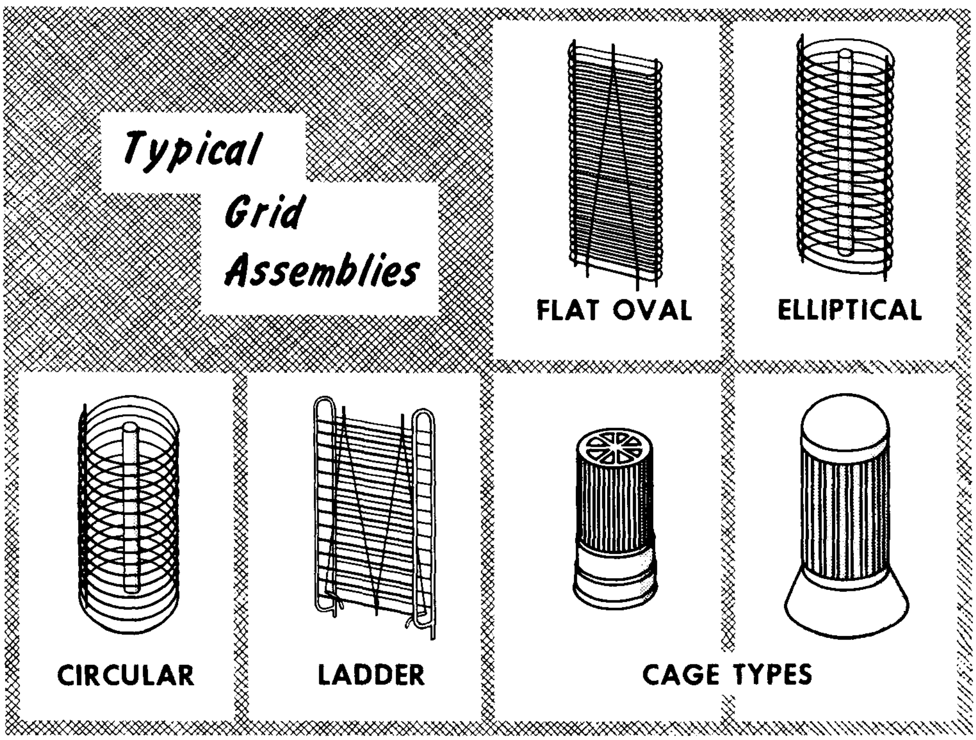
Many of the tubes used in radio receivers are also used in radio transmitters. This is especially true in low-power units. However, due to the many high-power circuits used in transmitters, a separate "breed" of tubes known as transmitting types are used. Generally, they are of special construction to handle the larger amounts of power. They are larger, specially shaped, have rugged plate, grid, and filament elements, and special basing arrangements.

The high plate current generally found in transmitting tubes necessitates the use of emitters rich in sources of electrons. Both indirectly-heated cathodes, and directly-heated filaments are used. The cathode emitters are barium-oxide or strontium-oxide coated surfaces, brought to electron-emitting temperature by an associated heater. Directly-heated filaments are made of tungsten, with thorium oxide impregnated in the metal during manufacture, or nickel, coated with an emitting material. The thoriated-tungsten filament is extremely popular, but must be operated within 5% of its specified voltage rating. For this reason, many transmitters have a "filament voltage" meter and adjustment to permit monitoring of this voltage.

Transmitting Tubes: Grids

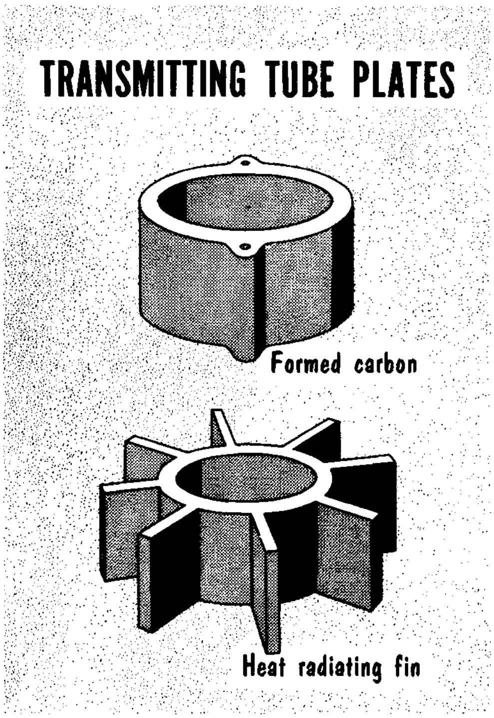
The control grid in a transmitting-tube type resembles its receiving tube counterpart – a spiral form, or ladder structure. In some tubes the control grid is a cage-like structure that surrounds the emitter. Materials used for grids include pure metals such as tungsten, molybdenum, or tantalum, or various alloys of these metals. With the grids located between heat radiators such as the filament and plate, they may emit primary electrons when bombarded by ions from gas in the tube. To prevent this the grids are often coated with gold or platinum. Platinum is preferred because it can withstand higher temperatures without vaporizing. Carbon is also commonly used.

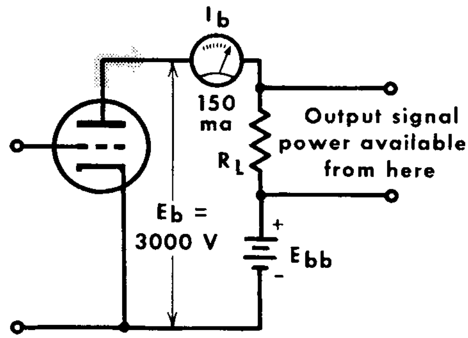
Control-grid electrodes are subject to several ratings. One is the grid-bias voltage, which may be negative or zero, depending on the kind of tube and how it is used. Another, is the safe peak amplitude of the input-signal voltage. Excessive input-signal voltage can not only impair tube performance, but can damage the tube. Another rating encountered in certain amplifying applications in transmitters is called driving power or r-f excitation, and is expressed in watts. If the application demands that the input-signal voltage exceed the applied negative grid bias, a unidirectional or d-c grid current flows between the emitter and the control grid inside the tube, and around the control grid – emitter circuit outside the tube. The product of the grid current and the peak positive grid – emitter voltage is the driving power (in watts) that is consumed in the grid circuit.



Transmitting Tubes: The Plate

The plate or anode of a transmitting tube is the point of greatest difference between receiving and transmitting tubes. In transmitting tubes the plate is subjected to tremendous impact by tube current, and accounts for much of the heat produced in the tube. To withstand high temperatures, the plate must radiate the heat, and a variety of materials and constructional features are





Total Plate Power Input

$$= \text{D-c plate voltage} \times \text{d-c plate current}$$

$$= E_b \times I_b$$

$$= 3000 \times 0.150$$

$$= 450 \text{ watts}$$

used. Materials used for plates are such metals as nickel, tantalum, and molybdenum. Nickel may be coated with carbon, and sometimes the entire plate is made of graphite. The carbonized coating aids the heat-radiating ability of the metal. In some tubes, pure copper plates are used. The usual run of plates are large, some with radial fins attached, or with radiating surfaces formed of edgewise wound metal ribbon. Often metal plates operate while glowing cherry red. For tubes up to about 250 watts (power input to plate), ordinary air circulation is sufficient for cooling; for ratings up to several thousand watts, forced air-draft cooling is used. For very high power tubes, circulating distilled water is often used for drawing off heat.

A phrase often used in conjunction with transmitter operation is plate-power input. This is the total amount of d-c power (in watts) supplied to the plate circuit. It is from this that the output-signal power delivered to the load of a tube is derived, as well as the power lost or dissipated as heat in the plate circuit inside the tube.

Transmitting-Tube Ratings (CCS, ICAS, and Typical Operation)

Exclusive of the filament or heater voltage and current, both of which are fixed for any given tube regardless of use, the safe maximum operating voltages applied to the control grid, screen grid (where applicable), and the plate electrodes, as well as the safe maximum amounts of current permitted to flow in these circuits, are determined by two standard conditions of use of the tube. They are called CCS ratings and ICAS ratings. Both are usable for guidance. The letters CCS stand for Continuous Commercial Service, which means continuous, around-the-clock operation consistent with maximum tube life and dependability. In contrast, is the ICAS rating. These letters stand for Intermittent Commercial and Amateur Service, by which is meant alternate periods of "on" and "off" operation, each theoretically lasting for five minutes. In practice, time is disregarded during the "on"-"off" periods. The philosophy behind the ICAS rating is highest signal output from the tube with only limited concern for long tube life.

In addition to the CCS and ICAS ratings of safe maximum values of voltage and current, there is always still another set of operating conditions known as typical operation. The typical operating values can conform with CCS operation or with ICAS operation, depending upon the circumstances. But whichever it is, the typical operating voltage and currents are in almost all instances below the CCS or the ICAS ratings. We show, for comparative purposes, CCS and ICAS ratings, and typical operating conditions for a standard transmitting tube. Those tubes not intended for ICAS service bear only CCS and typical operating specifications.

EXAMPLES of MAXIMUM CCS and ICAS RATINGS and TYPICAL OPERATION of 814 TUBE used as R-F POWER AMPLIFIER

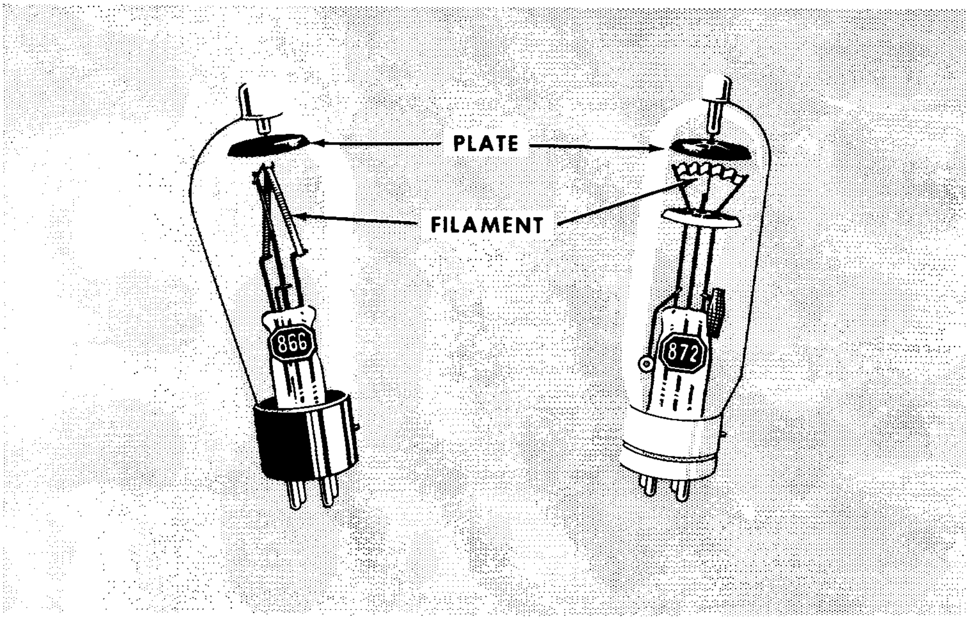
	Maximum		Typical
	CCS	ICAS	Operation
D-c Plate Voltage	1250	1500	1250
D-c Screen-Grid Voltage	400	400	300
D-c Control-Grid Voltage	-300	-300	-80
D-c Plate Current	150 ma	150 ma	144 ma
D-c Control-Grid Current	15 ma	15 ma	10 ma
Plate Input	180 watts	250 watts	180 watts
Screen-Grid Input	10 watts	10 watts	10 watts
Plate Dissipation	50 watts	65 watts	45 watts
Driving-Power Control Grid	1.5 watts	1.5 watts	1.5 watts

Transmitting-Type Rectifiers (Organization of Gaseous Tubes)

The gas-filled or gaseous rectifier is a two-element tube (diode) consisting of an indirectly-heated cathode, or a filament and a plate (anode) in a glass envelope. Also within the envelope is a small amount of gas such as argon, neon or xenon among others, or a small amount of liquid mercury. The rectifier that uses the liquid mercury is known as a mercury-vapor tube because the liquid mercury vaporizes when the hot emitter raises the temperature of the envelope sufficiently. The xenon-filled, and the mercury-vapor diodes are standard half-wave rectifiers.

Gaseous rectifiers are used in transmitter power supplies when the d-c voltage required is at least 1000 volts, and the current requirement is at least 150 milliamperes dc. They are used in rectifying circuits similar to those

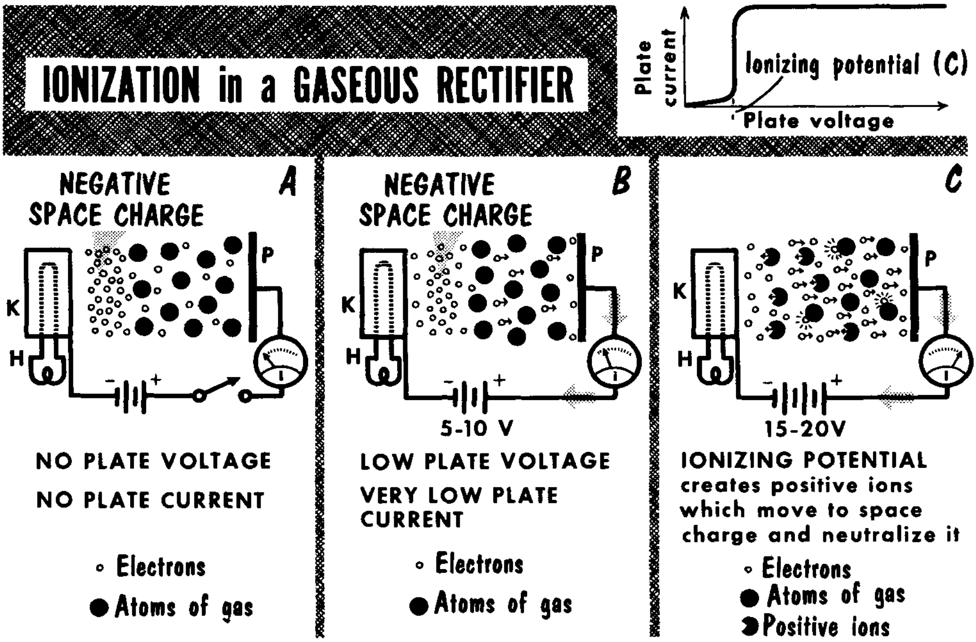
Internal Structure of Mercury-Vapor Rectifiers



used with high-vacuum tubes, namely, half-wave, full-wave, and bridge arrangements. Because of the manner of functioning, the gas-filled rectifier does not permit the simultaneous "on" and "off" application of the heater and plate voltages. When the tube is placed into service for an operating period, the heater voltage must be applied first, without application of the plate voltage, for at least 30 seconds. Then the plate voltage is applied. The time interval stated is necessary to allow the tube to heat sufficiently, and especially to allow the mercury to vaporize. Once the emitter has reached the proper temperature, the plate voltage can be applied or removed at will. The heater voltage must remain on all the time the power supply is in service, even though the plate voltage is not applied.

Transmitting-Type Gaseous Rectifiers (Theory)

Assume a half-wave gaseous rectifier with the hot cathode emitting electrons, and no voltage on the plate. The gas or the mercury vapor content of the tube corresponds to a cloud of atoms of the substance dispersed throughout the tube. The electrons from the cathode form a space charge around the



emitter. Now assume 5 to 10 volts positive applied to the plate. Electrons are drawn out of the space charge and advance to the plate as plate current, just as in the high-vacuum tube. At the same time, an equal number of electrons leave the emitter and enter the space charge.

If now the positive-polarity plate voltage is increased to perhaps 15 to 20 volts, electrons drawn out of the space charge have a higher velocity, collide with the atoms of gas, and knock electrons out of the atoms, thus making them positive ions. The voltage at which the electrically neutral atoms are changed to positive ions is called the ionizing potential. The positive ions are repelled by the positively-charged plate and are attracted to the space charge which, consisting of electrons, is the equivalent of a negatively-charged cloud. When the positive ions enter the space charge, the space-charge electrons are attracted to the positive ions, thus making the latter electrically neutral, and neutralizing the space charge. In effect, there is no space charge. Now all the emitted electrons move to the plate as plate current. If the positive-plate voltage falls below the ionizing potential, de-ionization of the gas occurs and space charge reappears. Gaseous rectifiers have exceptionally low internal plate resistance. This makes them useful for high current rectifiers.

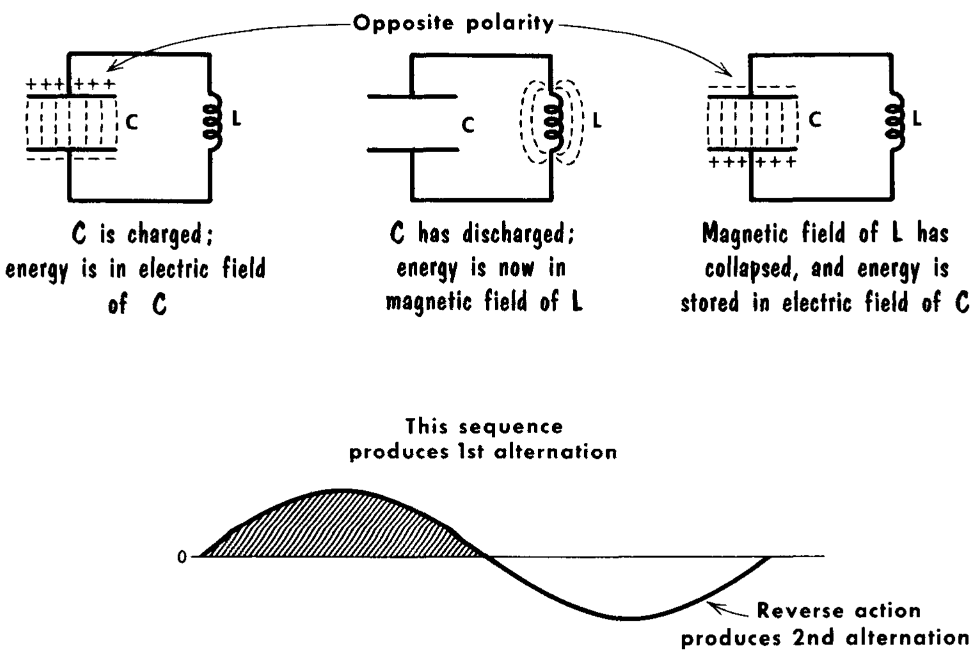
The Oscillator

Oscillators are the heart of a transmitter, often the transmitter itself. It would be best to begin by a review of the basics of oscillator circuits. An inductance and a capacitance is the basis of an oscillatory circuit. Assuming C to be fully charged, it will discharge, building a magnetic field about L. When C is fully discharged the energy of the circuit is in the magnetic field about L. With C discharged there is no longer any difference of potential to keep the current flowing through the circuit. The field about L collapses, inducing a back EMF. The induced back EMF causes the current to again flow in the original direction. The collapsing field then recharges C, but in the opposite polarity to that with which it was originally charged.

When the field about L has completely collapsed, all the energy of the circuit is once again stored in C. The capacitor again discharges. The action of building a field about L is repeated, as is the collapsing field that again charges C. The polarity of the charge on C now is the same as that with which it started, completing the cycle.

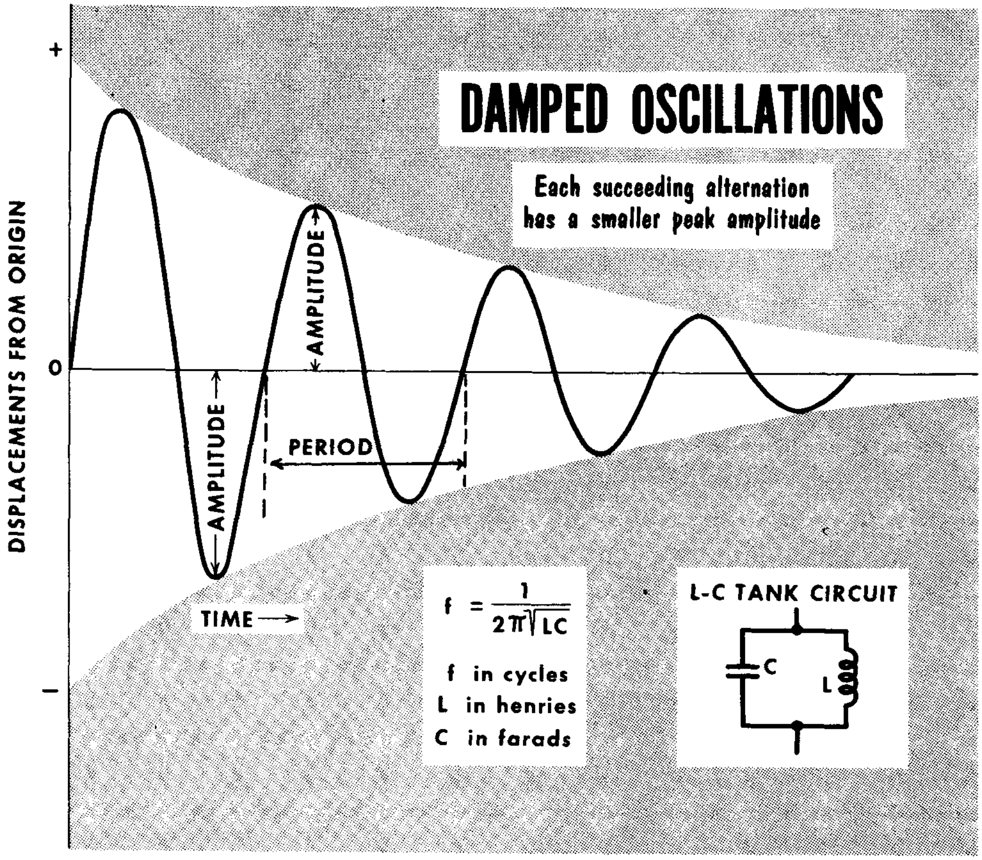
Plotting the capacitor current flow from the initial point where no current is flowing, through the point where maximum current is flowing as C discharges, back to the point where no current is flowing as C is recharged in the opposite polarity, provides one alternation of a sine wave. When the cycle is repeated, it provides an alternation in the opposite polarity. Combining the two alternations results in a sine wave of current flowing in the circuit.

OSCILLATORY DISCHARGE PRODUCES SINE WAVE



Oscillator Losses

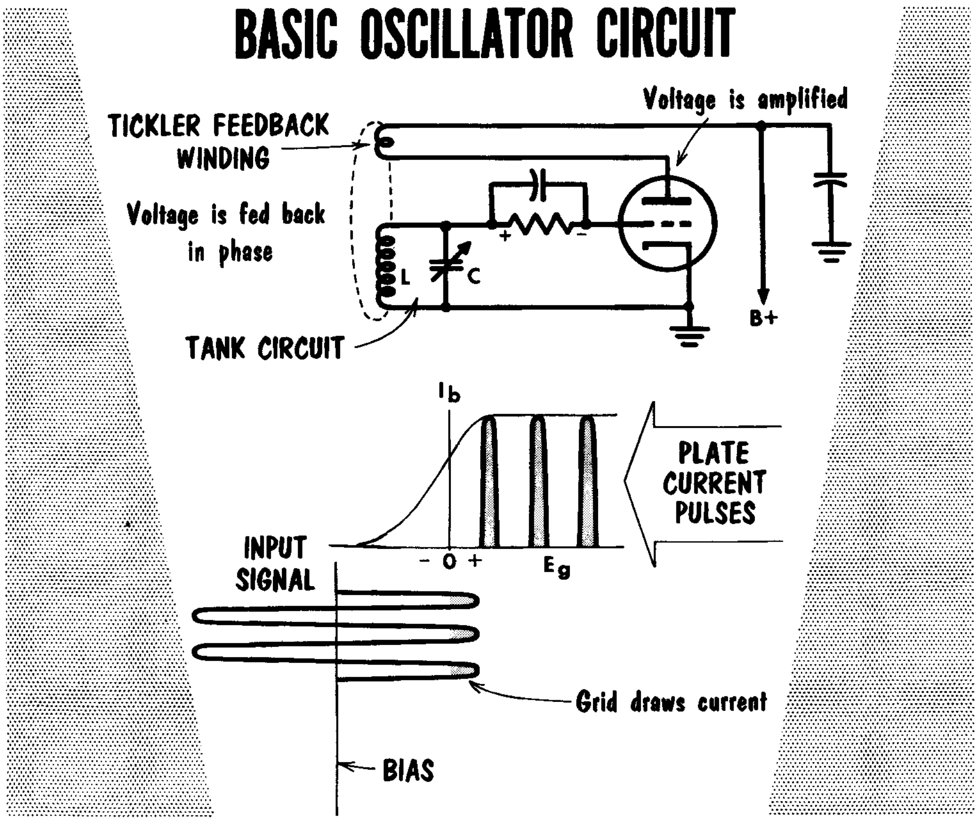
Assuming perfect capacitor and inductors, there would be no loss in an L-C circuit, or "tank" circuit; the oscillations would continue indefinitely. The frequency of the oscillations is dependent upon the values of L and C. A large value C takes more time to charge and discharge, producing a low frequency. A large value of inductance offers greater impedance to the flow of



current, thus also lowering the time of each alternation. The frequency of oscillation then, is determined by the value of L and C.

No coil is a perfect inductor; it must have the d-c resistance of the turns of wire making up the coil. This resistance is the major source of loss in an oscillator tank circuit. If not overcome, this loss causes the oscillations to die out, or become "damped". We have learned that the Q of a coil is equal to X_L/R. Hence, a low-Q coil has a high value of resistance compared to its reactance; it will quickly "dampen" the train of oscillations. A high-Q coil having a low value of R compared to its X_L, permits oscillations to continue for a longer period of time.

Oscillator Losses (Contd.)



To maintain constant amplitude of the oscillations in the tank circuit the damping effect of the resistance of the coil must be overcome. Applying the oscillations of the tank circuit to the grid of a vacuum-tube amplifier makes available at the plate circuit an increased amount of power at the same frequency as that of the tank circuit. A controlled amount of energy from the amplified output is fed back to the tank circuit. This energy is fed back in phase, called regenerative feedback. The output of the plate circuit being 180° out of phase with the grid circuit, it must be shifted another 180° to place it in phase with the grid circuit.

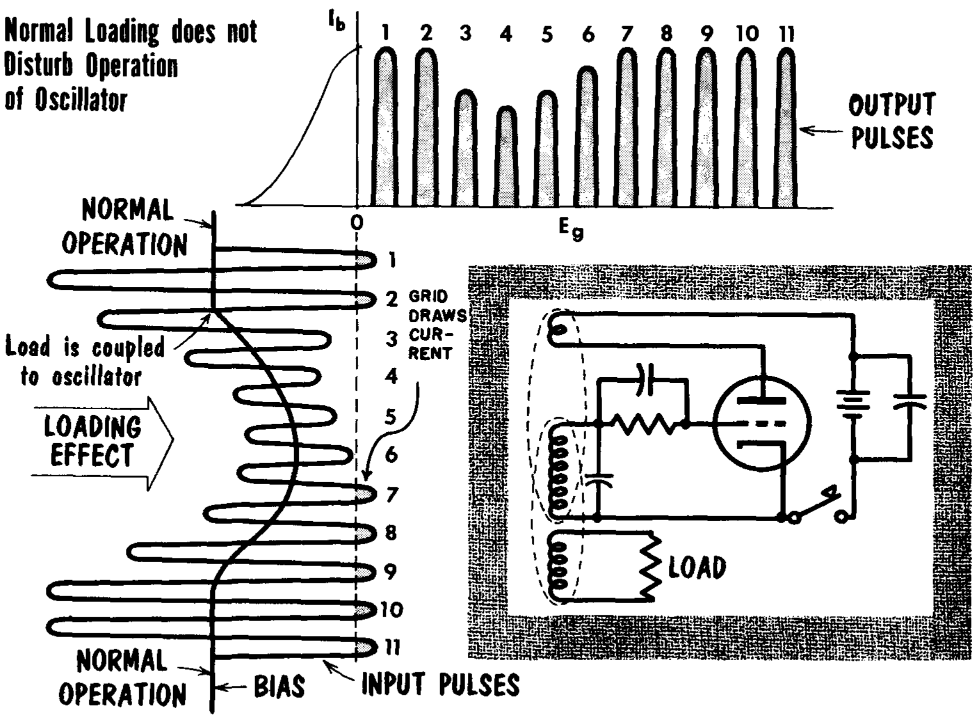
The output plate current signal is inductively coupled from the tickler (feedback) coil to the tank circuit, in phase, to maintain oscillations in the tank circuit. If the feedback of the tickler coil is not in phase, the circuit will not oscillate. The amount of feedback required is small, only enough to make up for the tank circuit losses. It is analogous to the pendulum of a clock; once swinging, it needs only a small push to make up the losses due to friction. With only a slight "push" required to make up for circuit losses, the output signal applied to the grid need not be the complete sine wave. To provide this slight push a class-C amplifier is used.

Oscillator Characteristics

The oscillator is self-starting. When voltage is first applied, there is no grid bias, causing heavy plate current flow. The large pulse of current flows through the tickler coil, building a strong field about it. This induces a voltage in the tank circuit, which starts to oscillate. The procedure is called shock excitation. The signal swing at the grid circuit causes the grid to draw current on the positive alternations, building up to the required bias.

The tank circuit may be coupled to another circuit. The coupling could cause the tank circuit to become "loaded down", reducing the amplitude of the oscillations. The signal swing being reduced, the tips of the positive alternations may not drive the grid voltage positive. There is then no grid current flow. The capacitor discharges and for a short time maintains the proper

The OSCILLATOR IS SELF REGULATING

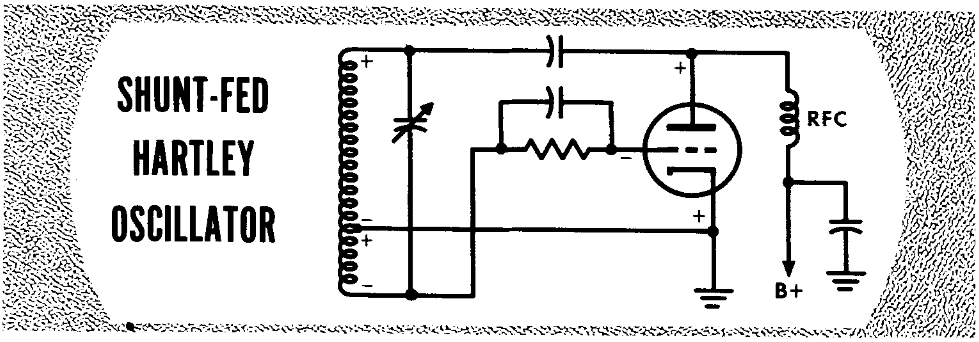


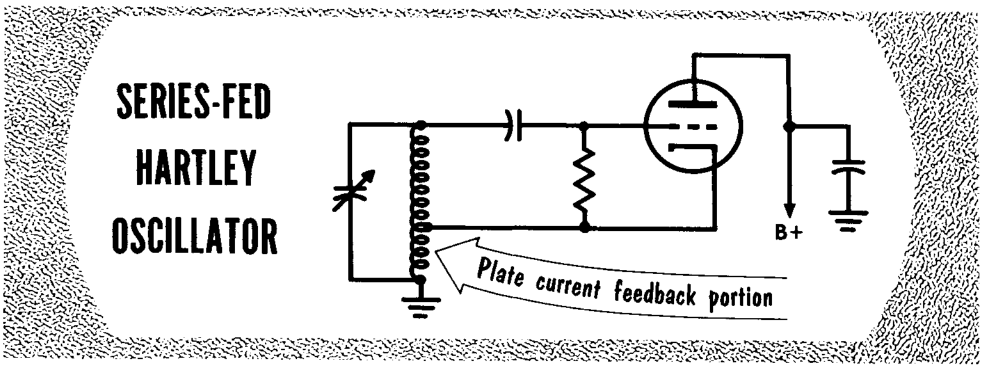
bias. Soon the capacitor discharges to a point where the grid-bias voltage is greatly reduced. After a few cycles the bias will have gone sufficiently positive-going to cause the tips of the alternations to drive the grid positive, causing grid current to flow again. The voltage drop across the grid-leak resistor again develops the correct bias. At the same time the positive grid causes increased plate-current pulses to flow, maintaining the feedback. The end result is to have the oscillator again running with its original values.

Hartley Oscillator

One of the most popular oscillators is the Hartley, named after its inventor. It uses a single tapped coil and capacitor in the tank circuit.

Taking one alternation of the circuit, the top, or plate end of the coil is positive, the bottom, or grid end is negative. The cathode is placed at the tap on the coil. To vary the feedback the tap can be moved. This controls the value



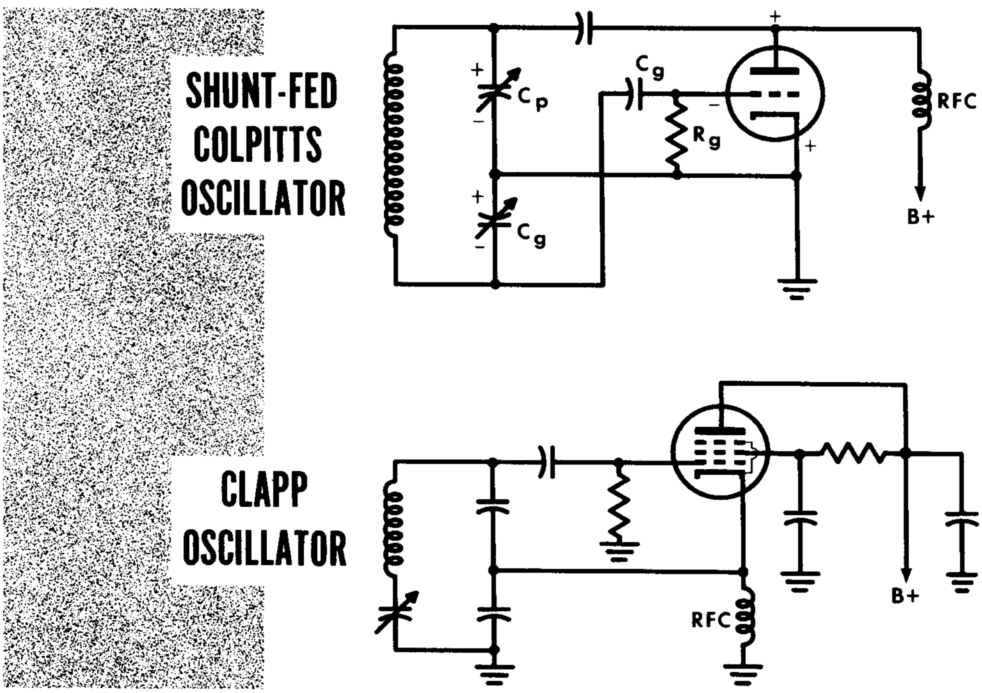


of feedback signal applied between cathode and grid. With the grid signal negative with respect to the cathode, the output pulse at the plate is 180° out of phase, or positive. Coupling the output pulse to the positive end of the coil provides the proper in-phase feedback. For the following alternation the top of the tank coil is negative; the bottom positive. The grid has a positive signal with respect to the cathode. The pulse at the plate is 180° out of phase, or negative-going. This is coupled to the negative end of the coil, again giving the proper in-phase feedback.

It is desirable to free the plate of the plate load and feedback task, using it merely as an attracting force for the electron stream. In the circuit shown, feedback is accomplished by plate current flow through that portion of the coil that is in the cathode circuit.

Colpitts and Clapp Oscillators

The most popular type of capacitive-coupled oscillator is the Colpitts, also named after its inventor. To prevent plate current from flowing through the tank circuit, it is shunt-fed. Its operation is similar to that of the Hartley oscillator. Instead of a tapped coil, there is the equivalent in two variable capacitors forming a voltage divider for the signal voltage. The voltage developed across the reactance of the grid circuit capacitor C_g, is applied between the grid and cathode. As shown, the grid is negative in respect to the cathode. The signal at the plate is a positive-going pulse which provides the in-phase feedback to the positive end of the coil. To vary the value of feedback the relation of the voltage division between the two capacitors must be



varied. To vary the frequency of the tank circuit, both capacitors should be varied. Quite often the frequency of the tank circuit is varied by using a slug-tuned coil to change its value of inductance.

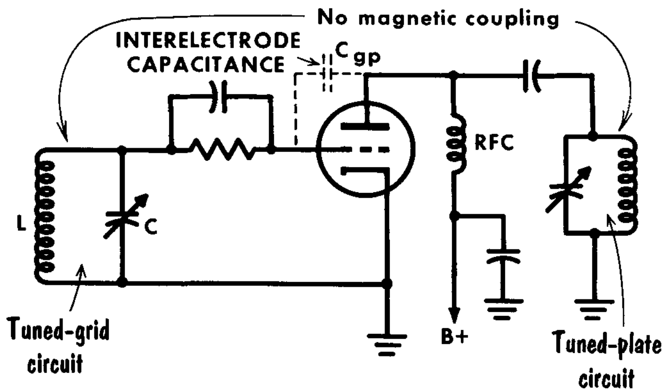
The Clapp oscillator closely resembles the Colpitts oscillator. The major differences are the use of a pentode, with the plate bypassed, placing it at a-c ground potential. With no plate load, the plate voltage is relatively constant, where the plate voltage of a triode varies, in turn varying the transconductance of the tubes. Another difference is the use of a series-tuned resonant circuit. This allows the use of fixed values of capacitors for the voltage divider. The Clapp oscillator provides stability of a high order.

The Tuned-Plate Tuned-Grid Oscillator

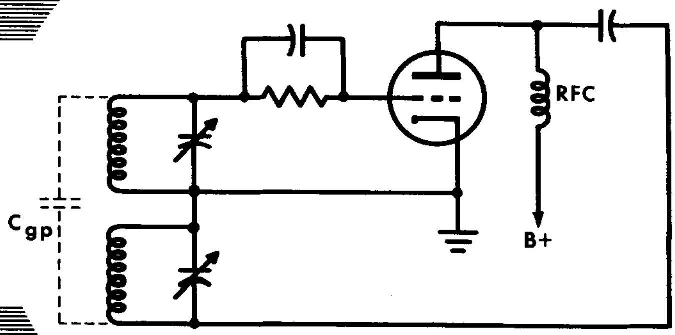
The tuned-plate tuned-grid (TPTG) oscillator uses a tuned circuit in both the plate and the grid circuits; the circuits are not inductively coupled. This type of oscillator may be used over a wide range of frequencies. However, because of reduced feedback between plate and grid at low frequencies, the TPTG oscillator is not particularly satisfactory at low frequencies. The feedback necessary to maintain oscillations is coupled from the plate circuit to the grid circuit by means of the interelectrode capacitance between plate and grid. Conventional grid-leak bias is used, with the r-f signal bypassed around the power supply.

In general, the frequency of oscillation is determined by the tuned L-C circuit having the higher Q. In most cases this is the grid-tank circuit. However, the natural oscillation frequency of the grid tank is lowered by the presence of the large effective input capacitance in parallel with the tank capacitance. To compensate for this, the grid-tank circuit must be tuned to a frequency slightly higher than called for, by reducing the capacitance of C. The plate-tank circuit must be tuned below the operating frequency to obtain the proper phase relations for oscillation. When redrawn, the similarity of the TPTG to the Hartley oscillator becomes apparent.

TUNED-PLATE TUNED-GRID OSCILLATOR

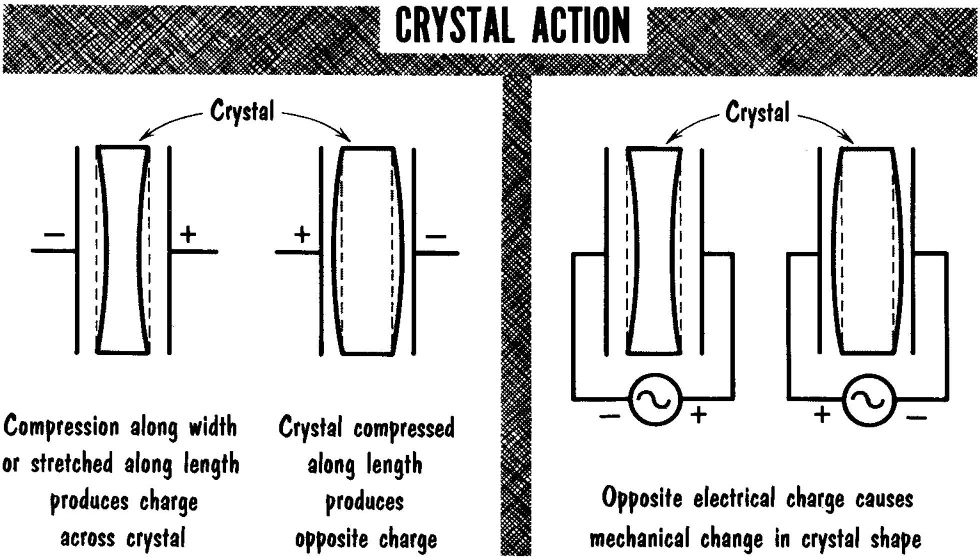


The oscillator redrawn to show the similarity to a Hartley oscillator



Action of Crystals

Types of crystals occasionally abbreviated as "xtal" exhibit the piezoelectric effect. The word piezo is a Greek word for pressure, or pressing. Piezoelectric then is "electricity from pressure". This describes the action of a quartz crystal. When squeezed, or compressed, a difference of potential is developed between its two faces. Increased pressure results in increased



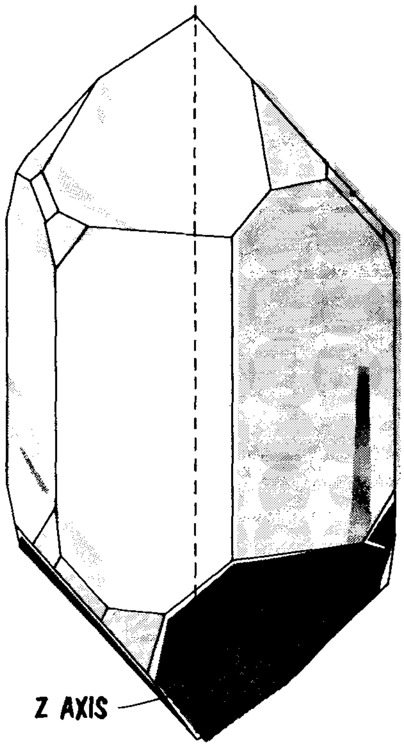
potential difference. Compressing a piece of quartz crystal in its width (A) causes it to develop a potential with each side having an opposite polarity. When the crystal is squeezed so that it bulges outward (B), the charge across each face reverses polarity.

In addition to generating a voltage, application of an alternating voltage to the two sides of the crystal causes it to vibrate mechanically. These vibrations become very strong when the frequency of the applied voltage is the same as the mechanical resonant frequency of the crystal (determined by its size and shape). These strong vibrations at the resonant frequency in turn generate a strong alternating voltage at the same resonant frequency. When the applied alternating voltage is sufficient to overcome the mechanical losses of the vibrating crystal, the generated alternating voltage of the crystal appears as a constant-frequency signal voltage that can be used as an L-C tank in an oscillator circuit.

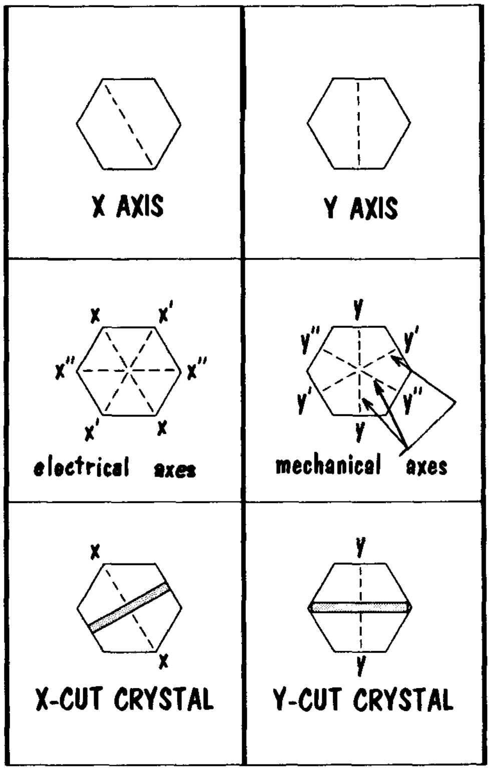
Many crystals exhibit the piezoelectric effect, but three particular types have been found to be most useful; Rochelle salt, tourmaline, and quartz. Rochelle salt is the most active crystal, generating the greatest amount of voltage for a given mechanical strain. However, quartz crystal is the most common type used for crystal oscillator circuits. It is inexpensive, rugged, stable, and can withstand reasonably high temperatures.

Crystal Cuts

HEXAGONAL CRYSTAL



QUARTZ CRYSTAL AXES AND CUTS



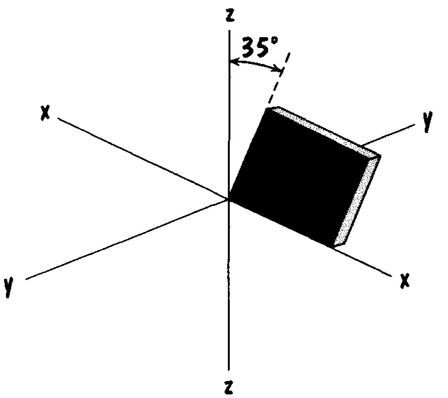
Natural crystals are found in various shapes, but those used for electronics are six-sided, or hexagonal. The crystal is divided and identified by various imaginary axes running through it. The Z axis, sometimes referred to as the optical axis, runs from end-to-end of the crystal. Strain or stress applied along the Z axis causes no piezo effect. Drawing a line from one of the six corners to an opposite corner produces an X axis. The three combinations of corners are labeled X, X', and X". The X axes are called the electrical axes because they provide the greatest piezoelectric effect. Joining opposite faces produces Y axes.

Crystal wafers are cut from the natural crystal, called the mother stone, or crystal. The thickness of the wafer, and the method in which it is cut or oriented to the mother stone will determine its frequency and temperature characteristics. The thinner the crystal wafer, the higher the frequency at which it can vibrate. A crystal wafer cut so as to have its two flat sides, or faces, facing an X axis, is called an X-cut crystal. An X-cut crystal is sliced from the mother crystal at a Y-axis angle. A crystal wafer having its two faces perpendicular to a Y axis is called a Y-cut crystal. Y-cut crystals are sliced at an X-axis angle.

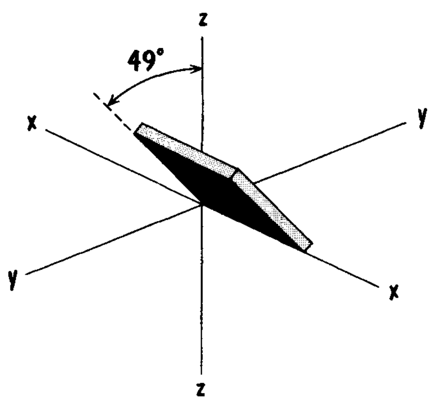
Temperature Coefficient

To obtain improved characteristics a wafer may be cut from the mother crystal on axes that are neither X, Y, or Z. The wafer may be cut to appear similar to a Y-cut crystal, but actually be tilted about the X axis at a 35° clockwise angle. This changes the notation of the crystal to an AT-cut. When the wafer is similar to a Y-cut crystal but is rotated to a 49° counterclockwise angle, it is called a BT-cut.

A vital characteristic of a wafer crystal is its temperature coefficient. This refers to the change in frequency that occurs when the temperature surrounding the crystal changes. An increase in temperature causes the crystal to expand, changing its resonant frequency (the crystal "drifts").



AT-CUT CRYSTAL



BT-CUT CRYSTAL

The variations in frequency are expressed in p/m (parts per million), sometimes written as ppm, or cycles per mc, for an increase in temperature of 1°C. The combined notation reads p/m°C. The change in frequency, or drift, may be upward to a higher frequency for a positive or plus coefficient, or it may be downward to a lower frequency for a negative or minus coefficient. A crystal wafer having no frequency drift with changes in temperature is rated as having a zero temperature coefficient.

Some crystals have poor temperature coefficients. An example might be a rating for a crystal of -20 cycles per mc per degree centigrade, abbreviated -20p/m°C. This indicates that the crystal will provide an output resonant frequency 20 cycles lower per mc for each °C increase in temperature. The exact value of drift varies with the type of cut and the thickness of the crystal. The AT-cut, and BT-cut wafer crystals have better temperature coefficients than the X-cut, and Y-cut crystals. An AT-cut crystal may have a rating of 0 p/m°C at 45°C. This means that keeping the temperature of the crystal close to 45°C will keep the output frequency constant. When operated at 85°C the same crystal may drift as high as +20p/m°C.

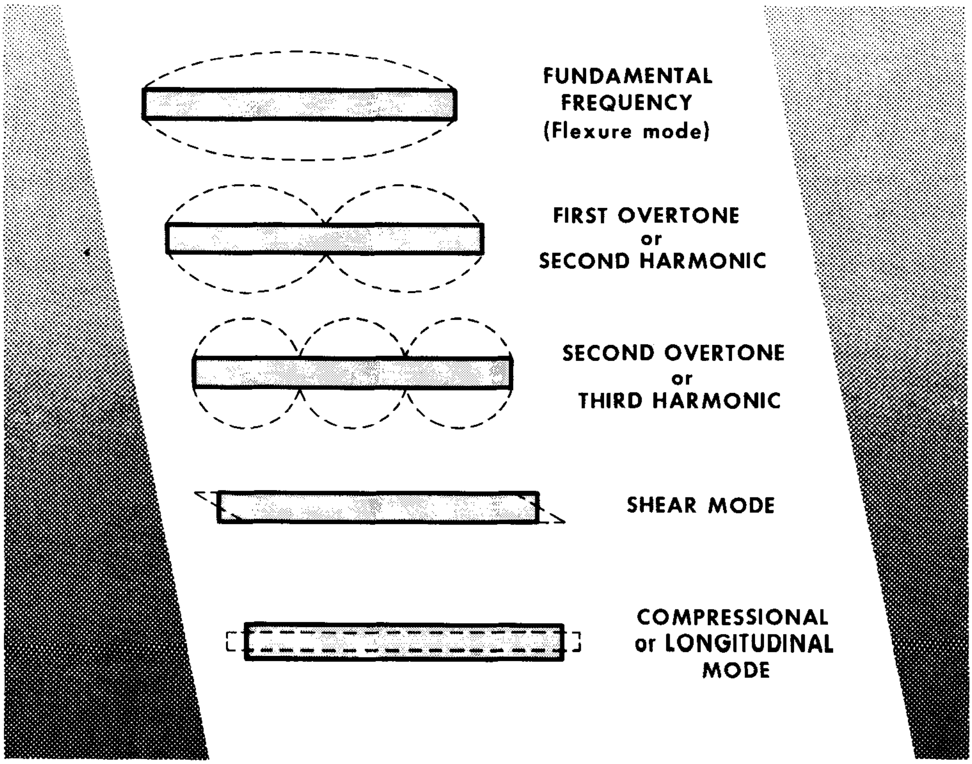
Oscillator Harmonics

When the crystal wafer has applied to it an a-c signal of the same frequency to which the crystal is mechanically resonant, the crystal develops strong vibrations. If not controlled, the vibrations may become excessive and fracture the crystal. When the entire crystal vibrates, the frequency at which it vibrates is called the fundamental frequency.

The same size crystal may vibrate in two motions, or wavelengths of the fundamental. This provides a frequency that is two times the original fundamental frequency, called the first overtone, or the second harmonic. When the same crystal vibrates in three wavelengths of the fundamentals, the second overtone, or third harmonic, is generated.

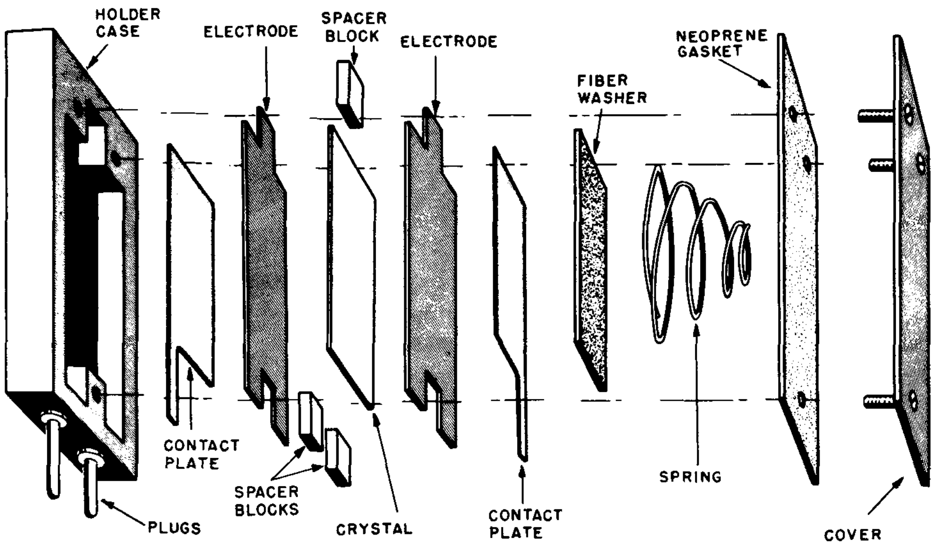
The physical direction in which the crystal wafer vibrates, called mode, may vary. When the entire crystal wafer vibrates at its fundamental frequency it is oscillating in a flexure mode. When vibrating so as to have each face move in an opposite direction, the crystal is oscillating in a shear mode. When vibrating so as to have the two faces compress and expand, varying the thickness of the crystal, it is oscillating in a compressional mode, sometimes called longitudinal mode.

VARIOUS CRYSTAL MODES and OVERTONES



Crystal Holders

CONSTRUCTION of "PRESSURE SANDWICH" TYPE MOUNT



The crystal wafer is carefully ground to exact thickness for resonance at the desired frequency. The crystal must be kept perfectly clean and placed in a special crystal holder. A basic holder consists of two metal electrodes to make contact with the faces of the crystal, with an air gap to minimize damping of the vibrations. The clamp holder is identical to the basic type, except that the crystal is firmly clamped between the two electrodes. The symbol for a crystal holder closely represents its physical construction.

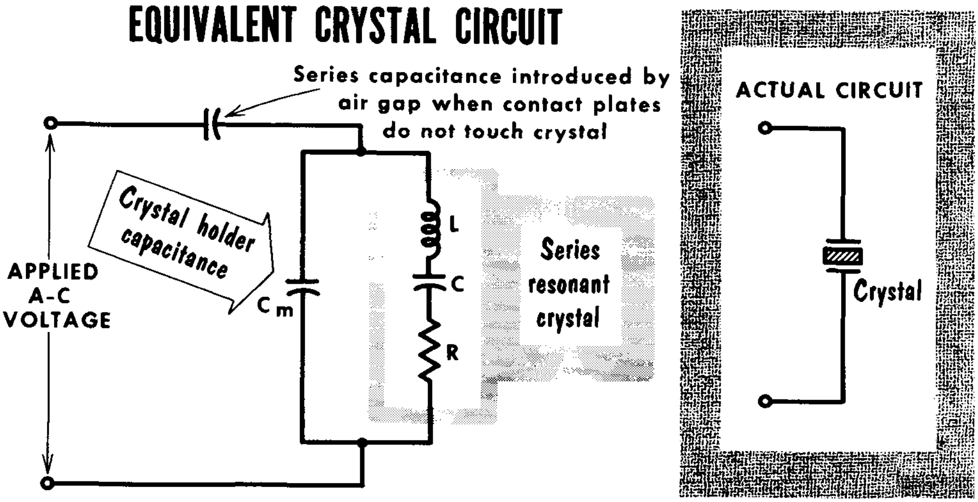
The physical construction of the crystal holder varies widely, some taking the shape of a "pressure sandwich" in which the crystal is clamped in a spring-mounted sandwich and then placed in a holder of the type shown. Other holders may clamp the crystal between tuned lengths of wire, or clamp the crystal at its edges.

The holder may take other forms such as that shown enclosed in a vacuum, which mounts in a standard octal socket. Other holders use various sockets depending upon their pin dimensions; most crystal holders will fit in standard tube sockets.

A special crystal oven can be used to maintain the crystal at a near-constant temperature to prevent oscillator drift. The crystal oven is designed to keep the temperature of the crystal at a constant higher value than the surrounding or ambient temperatures. This prevents ambient temperature changes from affecting the oven-controlled crystal. Operating the crystal at a higher temperature is of little consequence; more important is that the oven temperature be held constant.

Crystal Characteristics

The crystal wafer, in its crystal holder, represents a circuit component. When shown as a symbol in a schematic diagram it appears merely as a slab of crystal between two electrodes. As a circuit component it appears vastly different, having the properties of a resonant tank circuit, with appropriate



values of L, C, and R. The capacitance of the electrodes, or crystal mounting device is C_m. The series resonant action of the crystal is represented by L, C, and R. R represents the electrical equivalent of the mechanical friction present when the crystal is vibrating; C represents the electrical equivalent of the mechanical compliance (stiffness) of the crystal; L represents the electrical equivalent of the vibrating mass of the crystal. L and C of the crystal determine its series-resonant value.

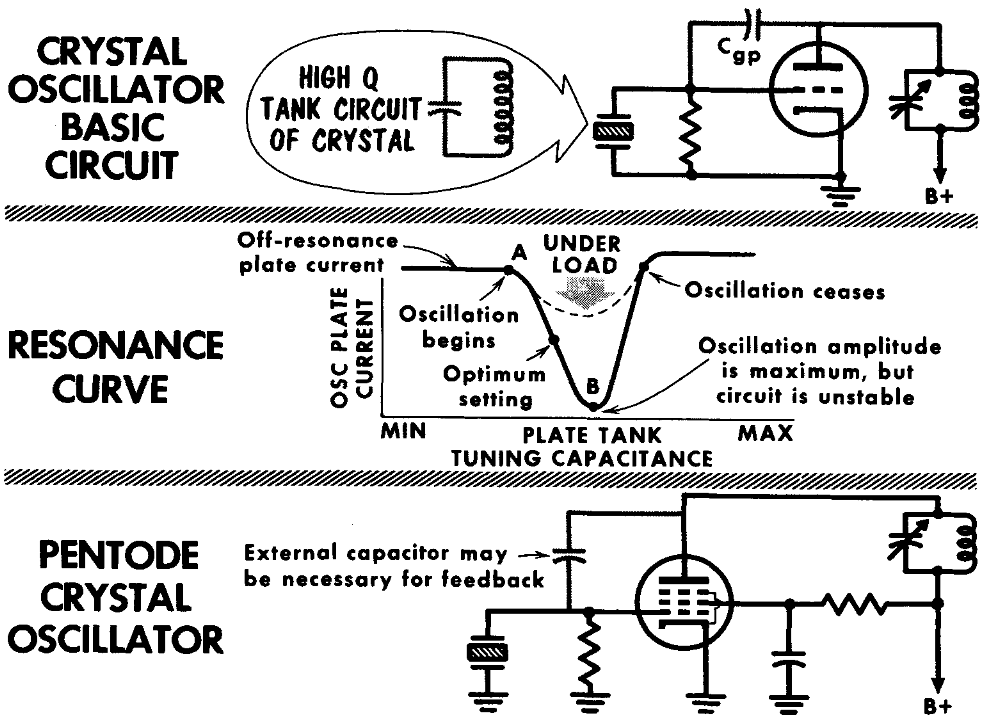
The capacitance of C_m is high, and in parallel with the series-resonant circuit of the crystal, forming a parallel-resonant circuit. The series-resonant circuit of the crystal is lower in frequency than the parallel-resonant circuit. When a series-resonant circuit is above resonance, the increased reactance of L, and decreased reactance of C, causes it to appear as an inductive circuit. The parallel-resonant circuit of the combined crystal holder capacitance, and the crystal, is slightly higher in frequency than that of the series-resonant crystal itself. At the higher resonant frequency of the parallel-resonant circuit, the crystal appears as an inductive circuit.

Parallel resonance occurs when the X_C of C_m is equal to the X_L of the series-resonant circuit of the crystal. The actual values of resonant frequencies of both the series-resonant circuit of the crystal and the parallel-resonant circuit of the crystal and holder, is fairly close. This results in a parallel-resonant circuit of extraordinarily high Q. A normal high-Q L-C tank circuit might have a Q of 100; the Q of a crystal-tank circuit could well be 30,000.

Basic Crystal Oscillator

Note the resemblance of the crystal oscillator to the TPTG type. The crystal replaces the grid tank, and feedback is through grid-plate capacitance. The high-Q grid circuit requires critical plate-tank circuit adjustment for correct oscillator operation. As the tuning capacitance in the plate circuit varies from minimum to maximum capacitance, the first signal of plate current reduction, A, indicates the start of oscillations. As capacitance is increased, the oscillations grow stronger, indicated by a continuous decrease of plate current, until B is reached. When capacitance is increased beyond B, oscillation ceases, and plate current immediately rises to maximum.

Optimum operation occurs between points A and B. Operating too close to these points causes critical conditions where small changes in circuit constants may stop oscillation. When the circuit is loaded (energy is coupled to another circuit), the plate-current dip is not as great. However, the dip is

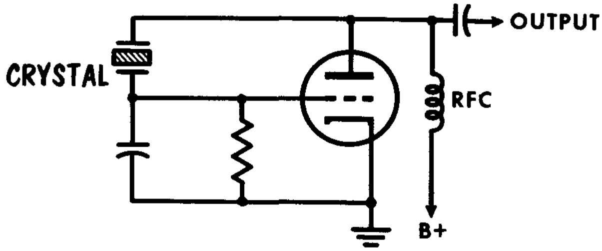


still there, and careful tuning is required to place the plate current value between A and B. In addition to monitoring plate current, the grid current or bias voltage developed across the grid resistor may also be monitored. Grid current values are the opposite of plate current values – as the plate current dips, the grid current rises. Use of a beam-power pentode increases output, but the crystal receives less feedback due to the reduced plate-to-grid interelectrode capacitance of a pentode. However, a small-value external capacitor may be placed from plate to grid for sufficient feedback.

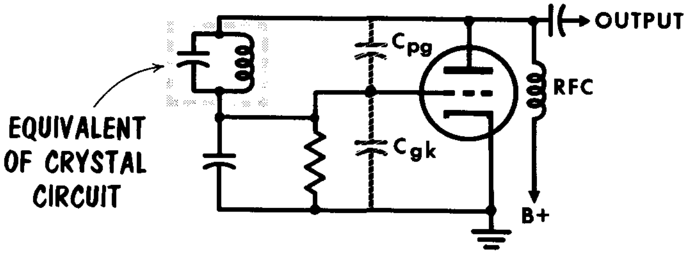
Pierce Oscillator

The Pierce-oscillator circuit provides high stability with a simple circuit that requires no L-C tank circuit. The lack of a tank circuit makes the oscillator convenient for use with numerous crystals that may be switched into position as desired. With the circuit redrawn, the crystal represented as a

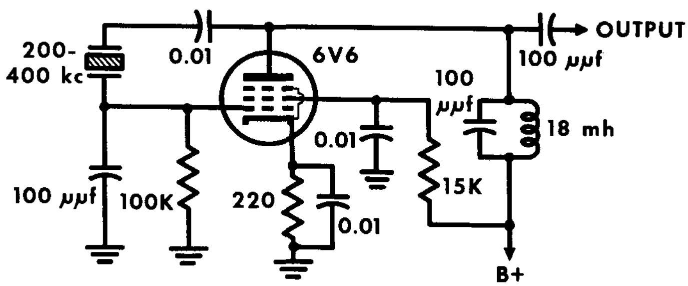
PIERCE OSCILLATOR BASIC CIRCUIT



Redrawn to show Similarity to the Colpitts Oscillator



Typical Pentode-Operated Pierce Oscillator



tank circuit, and the interelectrode capacitance as shown, the circuit is identical with the Colpitts oscillator.

The use of a pentode for a Pierce oscillator is quite common. The circuit is identical with regard to placement of the crystal. The additional components (aside from the screen-bypass capacitor and voltage-dropping resistor) are the cathode-bias resistor, cathode-bypass capacitor, and the coupling capacitor between the plate and crystal holder. The small bias developed in the cathode circuit is merely a safeguard. With the crystal removed the tube will have a small amount of bias to keep the plate current from reaching an excessive value. The capacitor between the plate and crystal holder removes B plus from the crystal holder and crystal.

The Crystal-Overtone Oscillator

Sometimes the desired frequency of operation is so high that a crystal ground so that its fundamental frequency is the desired frequency, would be so thin as to be easily fractured. An alternative method of achieving crystal control of the high frequency is to use an overtone crystal. An overtone crystal is a special cut that affords more than the normal output at a particular overtone or harmonic frequency. It may be labeled 33.3 mc and be used to generate this frequency although its mechanical resonant or fundamental frequency is 11.1 mc. Thus the crystal is operating at its third harmonic.

The Butler overtone oscillator is a popular circuit for an overtone oscillator. It is simple and has high frequency stability. Tube V1 is a cathode follower, while tube V2 is a grounded-grid amplifier. The cathodes of the two tubes are coupled to each other via the crystal, which, at its overtone frequency, is series resonant and therefore presents very low impedance (resistance). The

In the BUTLER OVERTONE OSCILLATOR the OUTPUT FREQUENCY is an OVERTONE of the FUNDAMENTAL

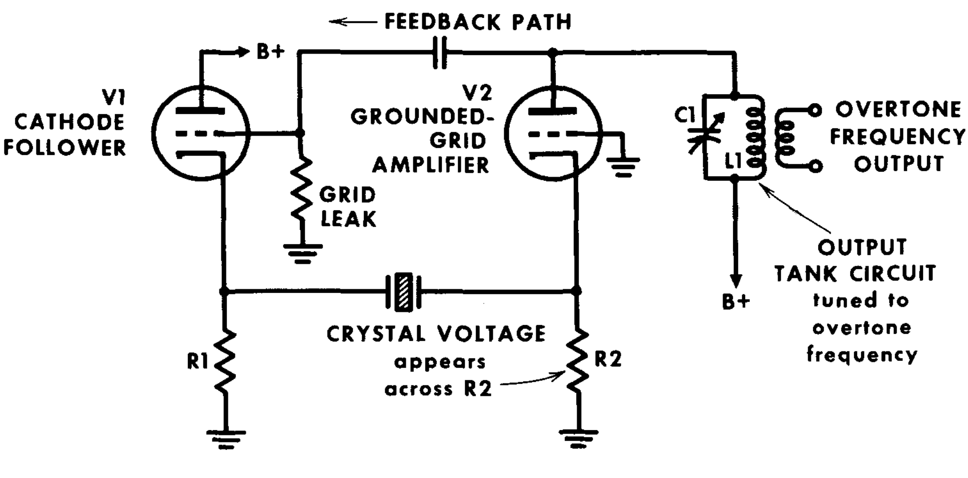
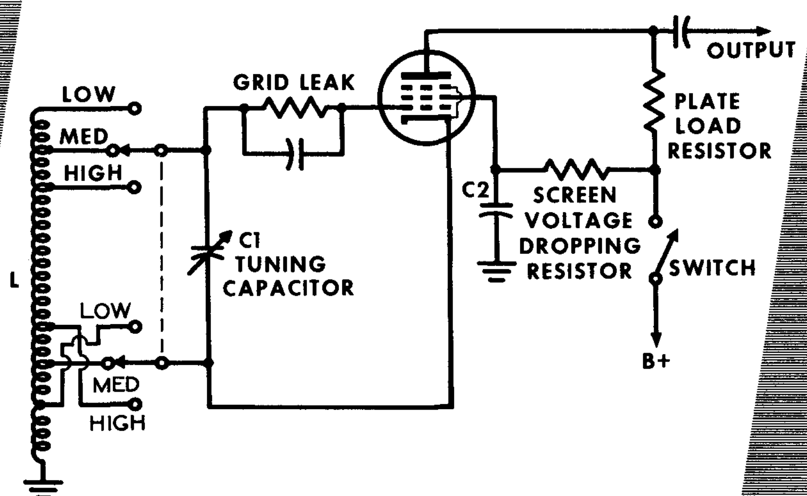


plate tank L1-C1 in the grounded-grid amplifier plate circuit is tuned to the overtone or oscillator output frequency. Note that the plate of the amplifier stage (V2) is coupled to the control grid of the cathode follower (V1). A pulse appearing in V2 when the operating voltage is first applied, develops an output voltage that is fed back to V1, there to cause a voltage to appear across the cathode resistor R1. This voltage starts the crystal vibrating. The resultant voltage developed by the crystal appears across R2. With the control grid of V2 connected to ground, the crystal voltage applied across R2 appears between the cathode and control grid of V2. An amplified version appears in the plate circuit of V2 and is fed back to V1. The action is repeated until sustained oscillations are generated.

The Variable-Frequency Oscillator (VFO)

Many transmitters are designed to function anywhere within one or more bands of frequencies. Amateur radio stations, also known as "ham" stations use such transmitters, as do some other services. To permit such operation the oscillator portion, as well as other parts, of the transmitter are made

A BASIC VARIABLE-FREQUENCY L-C ELECTRON-COUPLED OSCILLATOR



R-f connection between low end of L and screen grid is via ground and C2

Temperature compensating capacitors have been omitted.

tunable over bands of frequencies. An oscillator of this kind is known as a variable-frequency oscillator, or vfo.

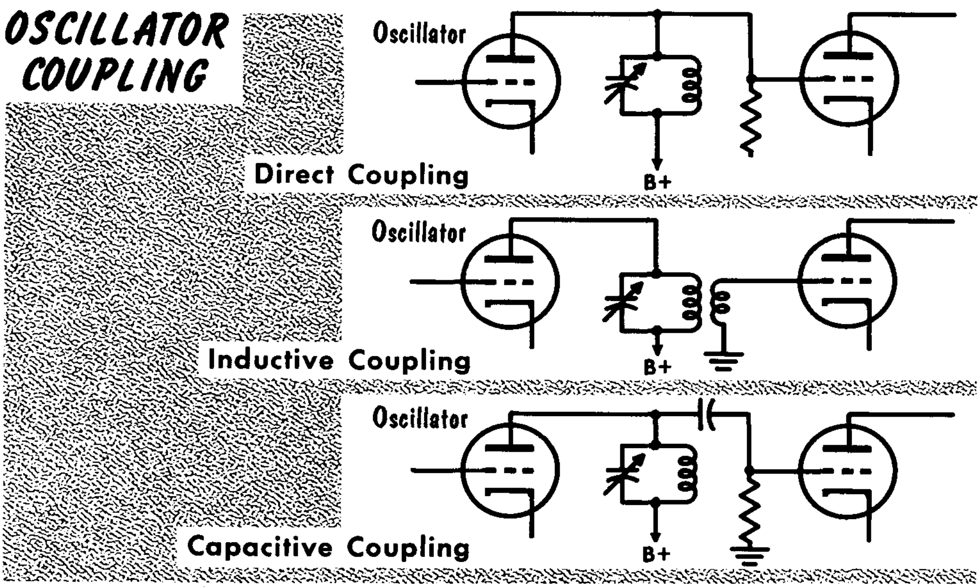
Variable-frequency oscillators are L-C oscillators. Most often they switch in different values or taps on a coil to select the desired band of frequencies. The variable C affords the choice of the exact desired frequency. When used in transmitters, variable-frequency oscillators are similar to the variable frequency heterodyning (local) oscillators used in superheterodyne receivers.

The number of frequency bands covered by a variable-frequency oscillator may be from one to as many as allowed for in the design. Usually the vfo is an integral part of the transmitter. Sometimes it is a separate device connected to the remainder of the transmitter. We show one type of circuit used for generating oscillations, but the variable-frequency oscillator may use any one of a number of other circuit configurations.

Oscillator Coupling and Loading

To use the oscillator circuits previously discussed the signal must be coupled to additional circuits. Although the output of an oscillator circuit could be coupled to an antenna, it most often is coupled to additional circuits. Shown here is direct coupling of the output tank circuit to the grid of the following stage. Direct coupling is rarely used; it provides no isolation between stages, and requires careful application of voltages. Inductive coupling has the output coupled to a secondary winding of a transformer. Quite often the secondary is also a tuned circuit for maximum signal coupling. Capacitive coupling will couple the full output signal to the succeeding stage.

The stage following derives its energy from the oscillator circuit. The removal of energy from the oscillator is termed loading. Oscillator circuits are to a degree self-regulating, and normally make up for circuit loading. Despite this, loading of the oscillator circuit can make itself felt in other



directions. Coupling to the output circuit is paralleling the output circuit; this changes the values of the output-tank circuit causing it to be mistuned, requiring realignment. Loading the oscillator output-tank circuit also reduces the Q of the circuit, reducing the feedback and efficiency of the oscillator. A low-Q tank circuit can readily have its output frequency shifted, since the tank circuit will easily resonate over a wide band of frequencies.

To keep adverse effects to a minimum it is best to load the oscillator output as little as possible. Minimum loading is often accomplished by the use of electron coupling. The electron-coupled oscillator separates the frequency-determining portion of the oscillator from the output-tank circuit, providing a minimum of loading.

Basically, the radio transmitter is a device that generates an electrical signal that can be fed to the antenna, and from there radiated into space.

Large transmitter tubes produce large amounts of heat which must be removed to prevent damage to the tubes and associated circuits. This heat is produced mainly at the plate, but also at any grid drawing current.

Heat removal can be accomplished by natural air currents, forced-air cooling, or water cooling. Blackening the plate also assists in heat removal. Grids drawing large currents are constructed with internal tubing through which water is circulated.

Materials with a high melting point, such as molybdenum, tantalum, and tungsten, are used in the construction of grids.

Typical plate materials are graphite, copper, molybdenum, nickel, tantalum, and tungsten. Nickel is used for low-power transmitter-tube plates, graphite and molybdenum for medium-power, tantalum for high-power, and tungsten for extremely high-power.

To produce oscillations, an electron-tube circuit must have the following characteristics:

1. A tuned circuit having the proper amounts of inductance and capacitance to oscillate at the desired frequency.
2. A tube capable of amplifying a signal at its control grid.
3. A means of providing the tuned circuit with sufficient regenerative energy to sustain oscillations.

The tuned-grid oscillator obtains regenerative feedback by coupling the plate circuit to the tuned-grid circuit.

The tuned-plate oscillator has its tuned circuit on the plate side. Regenerative feedback is obtained by coupling a part of the oscillation to the plate circuit.

In the tuned-plate tuned-grid oscillator, regenerative feedback occurs through the grid-to-plate capacitance of the tube.

There are two basic types of split-tank oscillators – the Hartley, and the Colpitts.

Certain types of crystals, such as quartz or tourmaline, can be used as tuned circuits in oscillators. Crystals are used to give frequency, precision, and stability.

REVIEW QUESTIONS

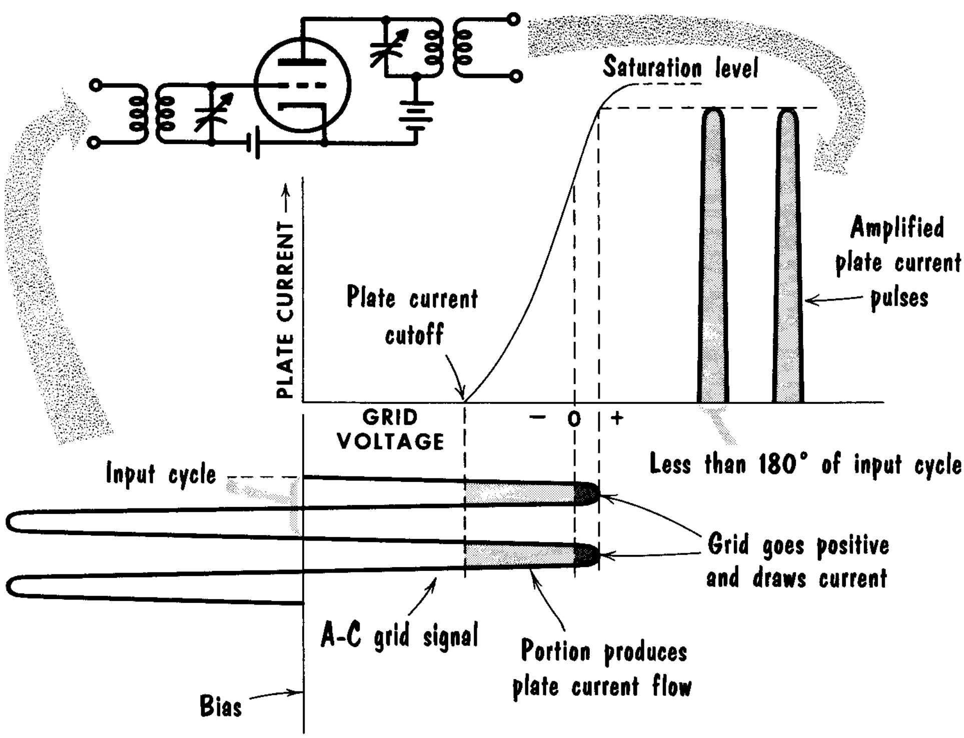
1. What is the basic function of a transmitter?
2. Basically, how do transmitting tubes differ from receiving tubes?
3. State two methods of removing heat from a transmitting tube?
4. What type of filament material is used in high-power transmitters?
5. How is energy coupled from the plate circuit to the grid circuit in the TPTG oscillator?
6. What are the proper conditions for producing regenerative feedback in the TPTG oscillator?
7. What is the function of a crystal as used in an oscillator?
8. What are the characteristics of an X-cut crystal?
9. What is meant by the "piezoelectric" effect?
10. What is meant by the temperature coefficient of a crystal?
11. What effect does the crystal mounting have on the equivalent circuit of the crystal?
12. How does the Hartley oscillator differ from the Colpitts oscillator?

The Class-C Amplifier

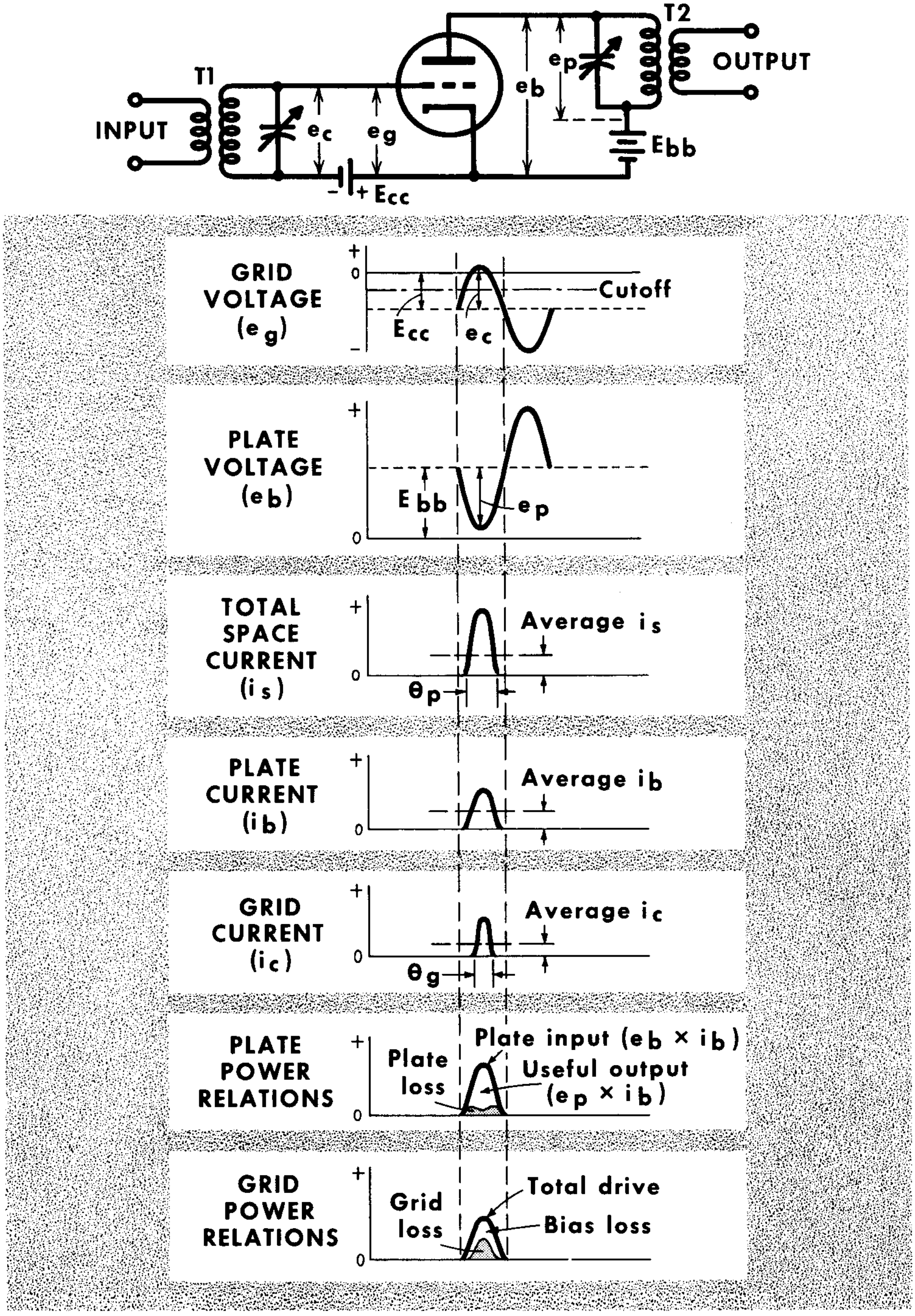
A class-C amplifier is biased well beyond (1.5 to 4 times) cutoff so that plate current flows for less than 180° of the input cycle. On the positive peaks of the grid-input signal the grid is generally driven positive, and as a result, draws current. In the absence of any input-signal voltage, the plate current is zero, and remains zero until the signal voltage has a positive amplitude greater than the applied grid bias. When the input signal is applied and grid current flows, the rise in plate current takes place along the linear portion of the characteristic curve – ideally up to, but not exceeding, the saturation level.

A feature of the class-C amplifier is that it affords higher plate circuit efficiency than any other class of amplifier. This is the efficiency with which the d-c power supplied to the plate circuit is converted into amplified a-c energy. The class-C amplifier (biased beyond cutoff) consumes power only during a portion of the input-signal period. The efficiency of a class-C amplifier may be as high as 80%. This amplifier cannot be used to reproduce variations in the waveform of the driving signal because, regardless of the input-signal voltage, the plate current appears as individual pulses. When amplifying sine-wave signals, the high distortion introduced by the class-C amplifier is overcome by the flywheel effect of the tuned-plate circuit.

CLASS-C AMPLIFICATION



CURRENT and VOLTAGE in CLASS-C AMPLIFIERS



Class-C Amplifiers (Contd.)

A typical class-C amplifier is shown, with the input signal supplied through tuned transformer T1. The output is developed by the r-f signal appearing across the parallel resonant circuit T2. The various voltage and current relationships in the circuit are shown beneath the circuit. The signal voltage e_c is developed across the tuned circuit T1.

The class-C amplifier operates with grid bias much greater than cutoff. Therefore, the grid excitation voltage causes plate current to flow during only part of the cycle. During the remainder of the cycle the voltage on the grid is below the cutoff value, the plate current i_b is zero, and the corresponding plate voltage e_b rises to its highest value, or E_{bb}. Since no plate current flows, the voltage drop across the plate-load impedance must be zero. The voltage drop across the load, therefore, is 180° out of phase with the grid voltage. The a-c components of the plate and grid voltages are sinusoidal because of the sharply-tuned resonant circuits.

Plate current flows when the grid voltage e_g rises above cutoff. The angle of flow of plate current is θ_p and is always less than half a cycle. Grid current flows during the angle θ_g when the grid voltage e_g becomes positive. The sum of these two currents, $i_b + i_c$, is the space current, i_s, and represents the total current leaving the cathode. The angle of grid current flow depends on the ratio of the grid bias to the peak signal amplitude. This is equivalent to saying that, in a particular amplifier, the value of the grid bias chosen determines the angle of plate current flow for a given input signal. Short angles of flow give high efficiency and low power output, whereas large angles give low efficiency and higher power output.

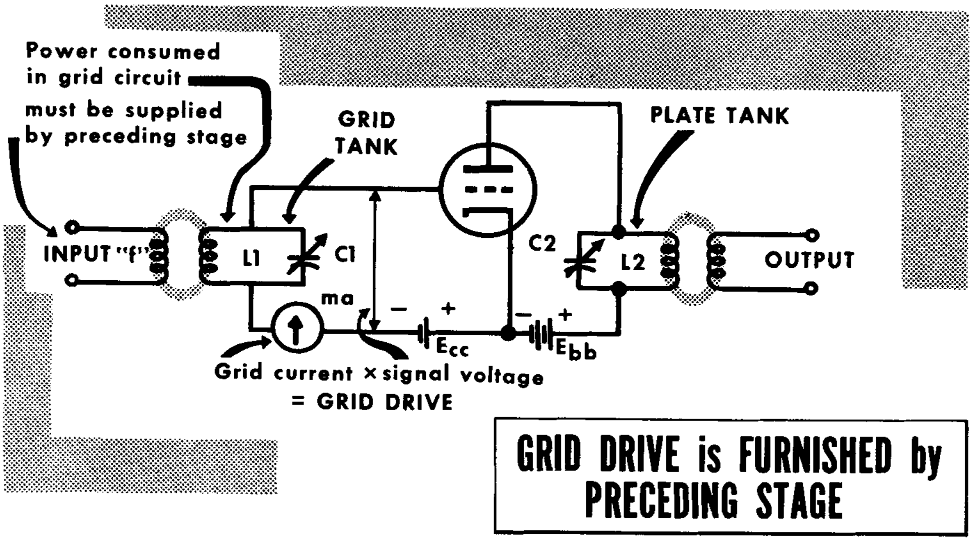
At any moment the total power input to the plate is the product of the total voltage e_b supplied to the plate, and the instantaneous plate current i_b. The power output is equal to the product of the load voltage and the plate current. The power loss at the plate is the difference between the input power and the output power. The efficiency of a class-C amplifier is the ratio in percent of the output to input power, and is usually between 60% and 80%. This high efficiency is possible because the plate current flows only when most of the voltage drop is across the output circuit. Therefore, only a small part of the supply voltage is wasted as a voltage drop between the plate and cathode of the tube.

Since the grid of the tube swings positive and draws current during part of the cycle, power is absorbed from the excitation circuit, which is the product of the exciting voltage e_c and the grid current i_c. Some of this power is lost at the grid, and the remainder is dissipated in the bias battery. If grid-leak bias is used, the remainder is dissipated as heat in the grid-leak resistor.

Special emphasis was placed on the action of the class-C amplifier because it represents an extremely important transmitting circuit. We will discuss it further in its numerous applications of voltage amplification and power amplification, together with methods of tuning and circuit variations.

The Class-C Tuned R-F Amplifier (The Grid Tank and Grid Drive)

The class-C amplifier as used in transmitters is a tuned r-f amplifier. A distinguishing feature is the variable-tuned parallel-resonant L-C circuit in the plate circuit. This tuned circuit is generally referred to as the plate tank. The control-grid circuit of the same amplifier stage may or may not contain a similar variable-tuned L-C circuit. When it is present it is called the grid tank. For purposes of explanation let us consider a class-C pentode tuned r-f amplifier with grid and plate tanks – L1-C1 in the grid circuit, and



L2-C2 in the plate circuit. The different methods of procuring the negative bias are discussed later.

Assume a signal voltage of a single frequency "f" coupled to the grid tank. The magnitude of the signal voltage delivered to the control grid will depend on the tuning of L1-C1. When the grid tank is tuned to frequency "f", maximum signal voltage is delivered to the input circuit of the tube. If the level of the signal derived from L1-C1 is sufficient to override the applied negative grid bias, the positive peaks of the signal will drive the control grid positive and cause grid current to flow. The current will be maximum when the grid tank is tuned to resonance with the incoming signal, and will drop off rapidly each side of resonance. If we measured the signal voltage delivered to the control grid, and multiplied this voltage by the grid current, the product would be a certain amount of electrical power which is consumed in the grid circuit. This power is known as the grid drive, and is expressed in watts.

Tetrode- and pentode-type amplifiers require less grid drive than triode-type amplifiers. In all cases, however, class-C amplifiers bear a rating that states the amount of signal power (grid drive) that must be delivered to the control grid to derive maximum power from the tube. In the absence of adequate grid drive to override the applied bias, the tube will be inoperative.

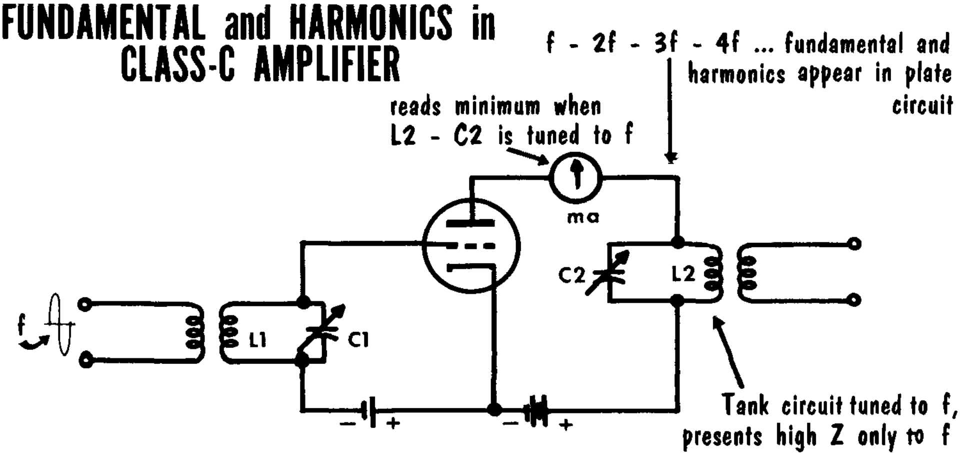
Class-C Amplifiers – The Plate Circuit

When the voltage between control grid and cathode is below cutoff, plate current flows. For any given value of applied plate voltage, the plate current is determined by the fixed grid bias, the amount of grid drive, and the state of resonance of the plate tank L2-C2. The plate tank is a parallel-resonant circuit which serves as the load for the plate circuit of the tube. As such, it presents maximum impedance at resonance.

Let us view the plate-cathode circuit inside the tube as a source of pulses of plate current, each of which consists of a number of frequencies. One of these is the fundamental frequency "f", corresponding to the frequency of the signal fed to the control grid of the tube. The others are even and odd harmonics of f, or 2f, 3f, etc. If we now visualize the plate tank tuned to the fundamental frequency of the plate current pulses, i.e., to f, then L2-C2 will present maximum impedance to the flow of this frequency component of the plate current, and theoretically zero impedance to the flow of the other frequency components of the plate current. The overall result is minimum plate current. A d-c milliammeter connected in series with the plate circuit would show minimum plate current. If the plate tank is detuned from this frequency, the impedance it presents to the flow of plate current is greatly reduced, and the plate current rises very sharply. In fact, in high-power amplifiers it can become so high as to damage the tube.

A reduction in plate current would occur if the plate tank were tuned to the second harmonic, rather than to the fundamental, except that now the decrease would not be as great as when L2-C2 was tuned to the fundamental. The reason is that the amplitude of the 2f component of the plate current is much less than the f component. Limiting the 2f component has a lesser effect on the total current than limiting the f component. The resonant plate tank would present maximum impedance to the flow of the 2f component, and theoretically zero impedance to the f component, and to frequency components higher than 2f. A corresponding action develops when the plate tank is tuned to 3f or to higher harmonics.

FUNDAMENTAL and HARMONICS in CLASS-C AMPLIFIER

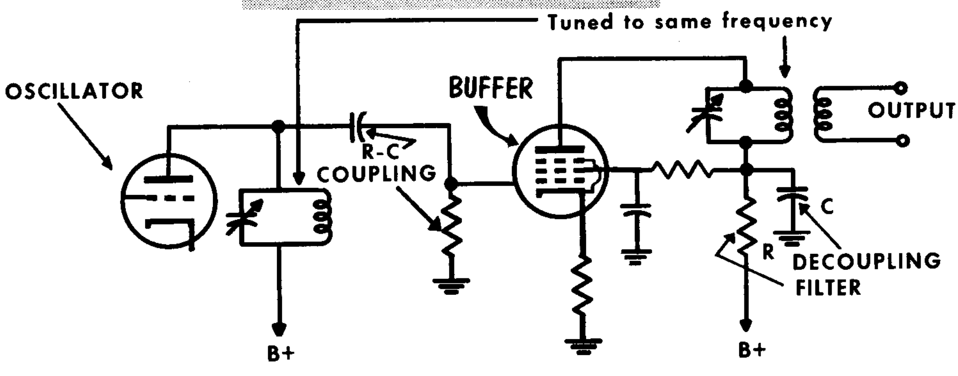


Buffer Amplifiers

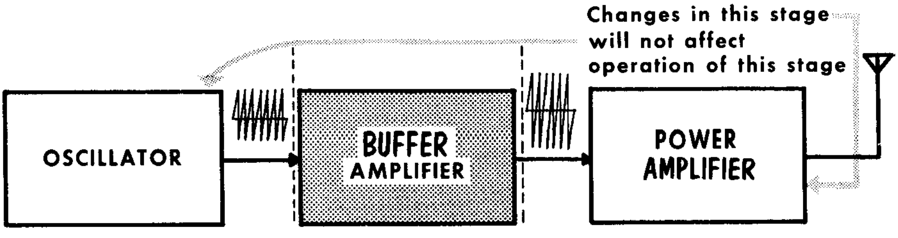
For increased power and isolation of the oscillator, a stage of amplification is placed between the oscillator and the antenna or following stages. This stage is called a buffer amplifier. Its basic function is to simply amplify, making no frequency changes. Isolating, and reducing the loading of the oscillator, improves its frequency stability. Any changes in the buffer output caused by loading from the succeeding stage is not reflected back to the oscillator; hence, buffering action.

The buffer amplifier is generally biased to operate class-A. It is designed so that the grid will not go positive and draw grid current. By doing this, the grid circuit of the buffer stage always presents a high-impedance load to the oscillator feeding it. This minimum loading produces high oscillator stability. Coupling from the oscillator to the buffer amplifier is usually of the R-C type. The output-plate circuit of the buffer has a tank circuit tuned to the oscillator frequency. The grid circuit does not use a tuned circuit, since it would then resemble a TPTG oscillator and may start oscillating. In the typical buffer circuit shown, the cathode-bias resistor is not bypassed thus providing degeneration. This acts both to keep the signal undistorted and to prevent any possible regeneration from the plate to grid circuit. In some low-power transmitters frequency multiplying is sometimes done in the buffer stage, and it may be operated class-C.

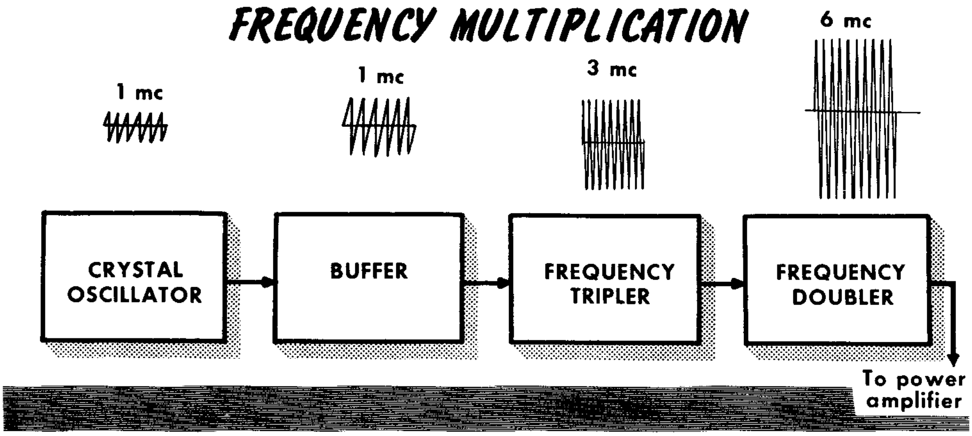
BUFFER AMPLIFIER

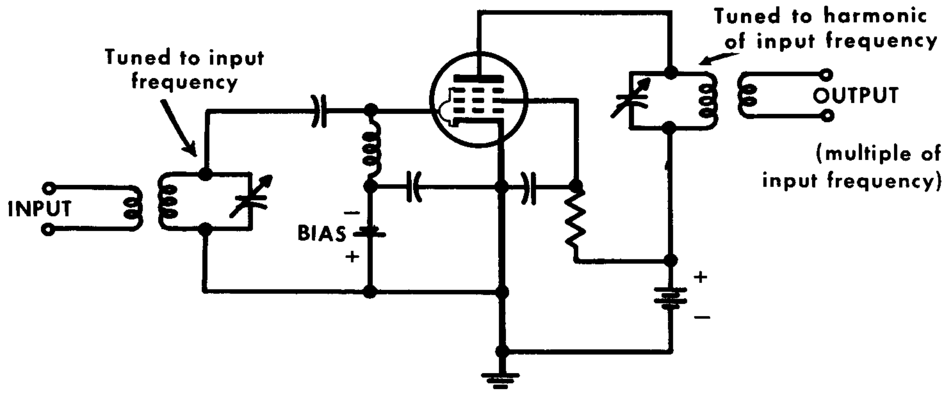


USING a BUFFER AMPLIFIER for ISOLATION



Frequency Multipliers





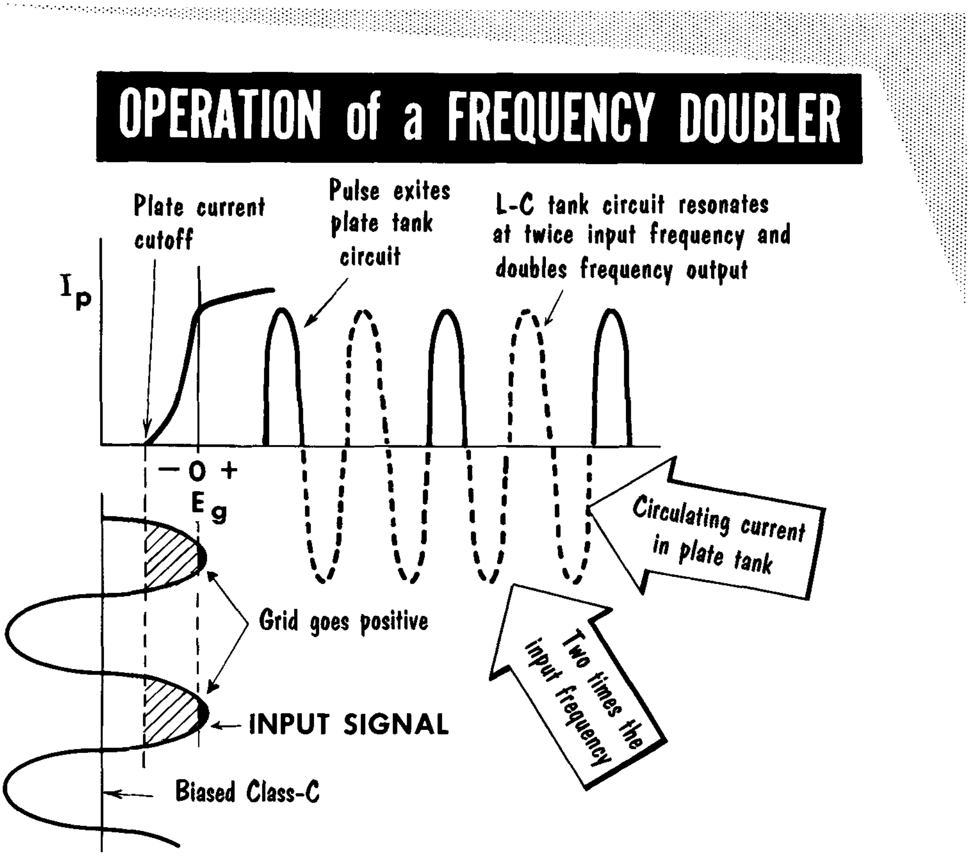
Frequency Multiplier

At the higher frequencies, oscillator frequency stability becomes difficult to maintain. Hence, it has been found best to keep the oscillator stage operating at a lower frequency, while arriving at the desired frequency by frequency multiplying. If the plate-tank circuit of the buffer amplifier is tuned to the second harmonic of the driving signal applied to the grid, the stage becomes a frequency doubler and the output voltage has a frequency equal to twice that of the input. Likewise, the buffer amplifier may become a tripler or a quadrupler. Where a greater amount of frequency multiplication is desired, two or more stages of frequency multiplication are operated in cascade.

The output of a frequency multiplier is a harmonic of the input signal. The harmonic output of a class-C amplifier can be controlled by controlling the value of the peak plate current. The harmonic output increases at first, as the width of the plate current pulse is decreased, but will then start to decrease as the pulse width is decreased still further.

Frequency-Multiplier Amplifiers

To act as a doubler, the plate-tank circuit must be tuned to twice the input frequency. As shown, following the plate-current pulse, the circulating current of the tank will carry the oscillations through two complete cycles before the next reinforcing plate-current pulse. With plate-current pulses applied at every other cycle, the output power available when operating as a doubler



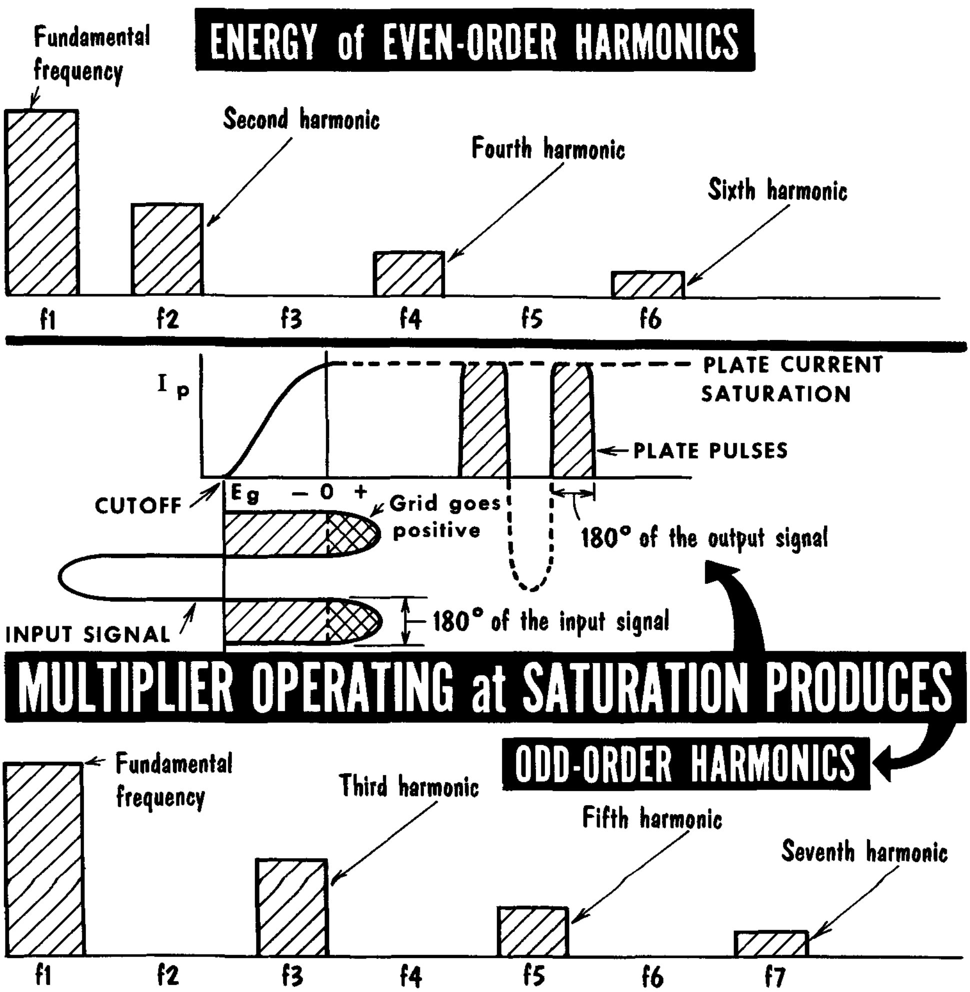
will be approximately one-half of that available when the amplifier is operated straight through.

The action of the circuit as a tripler and quadrupler is identical. The efficiency of a tripler will be less than that of a doubler, approximately one-third the output power available from a straight-through amplifier. The output of a quadrupler is similarly reduced in value to approximately one-fourth. Despite the reduced output power from the higher-order harmonic frequency multipliers, use of a high-Q output-tank circuit keeps the damping of plate-tank circuit oscillations to a minimum, and permits ample output power.

Frequency-Multiplier Amplifiers (Contd.)

When even-order harmonics are produced, the second harmonic contains approximately one-half the energy contained in the fundamental frequency, with the fourth and sixth harmonics proportionally reduced. Circuits designed to produce even-order harmonics will not generate odd-order harmonics.

Odd-order harmonics can be produced by application of sufficient bias and signal level to cause plate current saturation and "square" the input signal. Assuming plate current conduction for 180° of the input cycle, there will be output pulse energy present at various odd-order harmonics of the fundamental frequency. As shown, the third harmonic contains approximately one-half the energy contained in the fundamental frequency, with the fifth and seventh harmonics proportionally reduced. A circuit operating with the bias set as shown, provides an excellent frequency tripler.



Tuning Frequency Multipliers

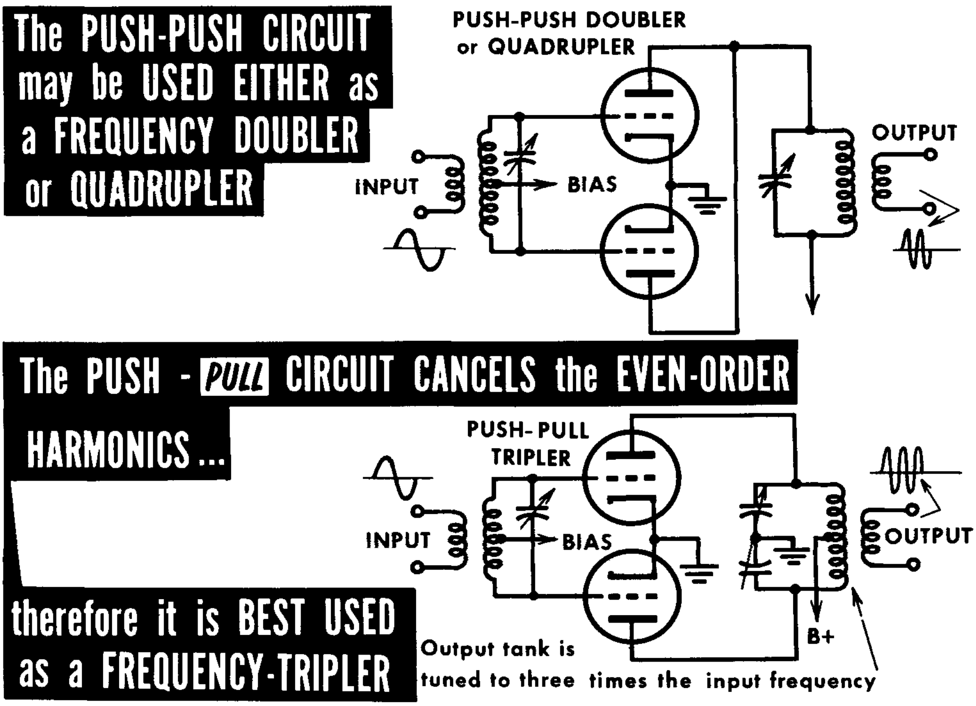


Plate current pulses in the frequency multiplier are usually of much shorter duration than in the conventional class-C amplifier. In this way the frequency content of the plate current pulses is enriched. The conduction angle in frequency doublers is about 110°-130°; in triplers 80-90°; and in quadruplers 60-70° of the input cycle. Tuning the frequency multiplier is like tuning the conventional class-C tuned r-f amplifier. Resonating the plate tank to a harmonic component of the plate current introduces a high impedance into the plate circuit at that frequency, thereby causing a fall (dip) in plate current. The deepest dip for a harmonic frequency occurs when the plate tank is resonant to the second harmonic of the fundamental. The dips become shallower as the order of the harmonic increases. Care must be taken when tuning the plate tank so as not to miss the point where the plate current falls. This is especially true if the multiplier is loaded by another stage.

Neutralization is not required in the triode-frequency multiplier because of the widely different frequencies in the input and output circuits. Two other circuits are occasionally used for frequency multiplication. They are the push-push and the push-pull class-C tuned amplifiers. The push-push circuit uses push-pull input and parallel connected output. The circuit is a predominantly even-order harmonic generator, hence is very useful as a doubler or quadrupler, especially the latter. The push-pull circuit eliminates even harmonics in the output, hence is a very efficient frequency tripler.

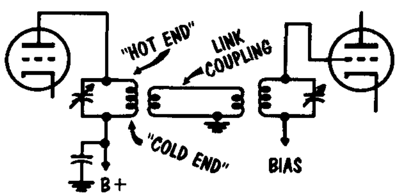
Interstage Coupling

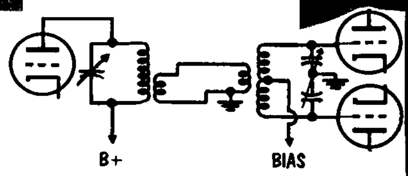
There are two main methods of coupling between stages of a transmitter – capacitive and inductive (transformer). Coupling is designed to provide the maximum transfer of energy with minimum loading. In capacitive coupling, also known as R-C coupling, the capacitor provides a low impedance path for the signal while simultaneously blocking the d-c plate voltage. To obtain a better impedance match the capacitor may be placed at various positions, or tapped down the coil. Inductive or transformer coupling consists of using the primary as the plate-tank coil, and the secondary as the grid coil for the next stage. The closeness of the coils determines the degree of coupling, and therefore the loading. The positions of the coils relative to each other also changes the degree of coupling. When parallel to each other, the coupling is maximum; when placed at right angles coupling is minimum. Coupling is unity or perfect when the coils are wound so that the wires are interwound. Then the tuning of the plate circuit also tunes the grid circuit.

In another form of inductive coupling, known as link coupling, a coil is connected at each end of the transmission line, and serves to link two tuned circuits. Long distances are permitted between the coils, since the transmission line impedance is low. The closeness of each coupling may be varied. The r-f signal developed across both the plate and the grid coils finds the top ends hot with respect to the bottom end which is cold (being at r-f ground). The best place to couple to the tuning coils is at the cold ends to prevent r-f voltage arcing. To keep developed harmonics to a minimum, one end of the link is grounded. This also reduces capacitance coupling between the tuned circuits.

In push-pull or balanced circuits, the centertap is the coil coupling point. Link coupling is also of advantage when coupling is required from an unbalanced to a balanced circuit.

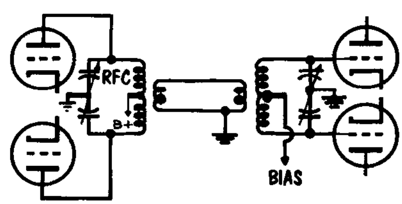
LINK COUPLING of TWO CIRCUITS





LINK COUPLING an UNBALANCED to a BALANCED CIRCUIT

LINK COUPLING TWO BALANCED CIRCUITS



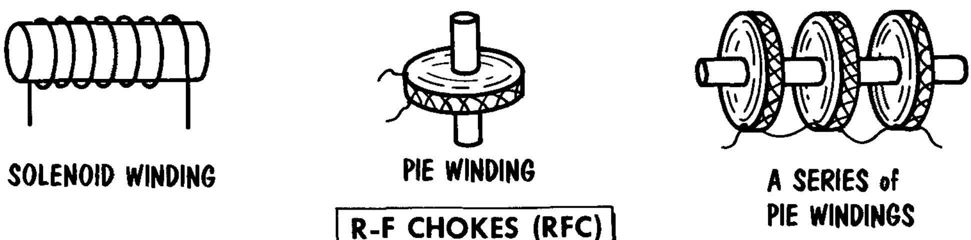
Interstage Coupling (Contd.)

Since the action of impedance coupling is so dependent upon the use of r-f chokes, it is worth discussing them separately.

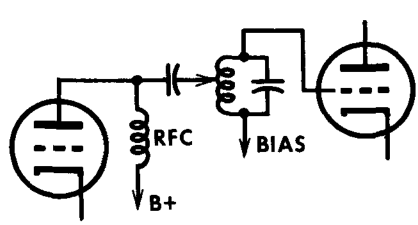
R-f chokes are a special form of inductance, usually consisting of many turns of fine wire wound either as a single coil, called a solenoid, or in a layer, called a pie winding, or a series of interconnected pie windings. They are specifically designed to have the least amount of possible stray capacitance between the windings. Every r-f choke has one frequency at which the value of inductance will resonate with the stray capacitance. This frequency should be much lower, or much higher than the frequency of the circuit in which it is placed. Some r-f chokes are designed to be used over a wide range of frequencies. Care must be taken with these chokes to see that nowhere in the range of frequencies used will there be a frequency that is resonant with the stray capacitance. Also, sufficient reactance must be offered at all frequencies. The higher the frequency, the higher the value of X_L with a given value of choke. This means that at high frequencies it is easy to obtain a large value of reactance, and therefore high-frequency r-f chokes usually contain only a small number of turns. The size of the wire used to wind the choke will determine how much current it can handle safely.

Using the r-f choke as a plate load, the entire signal is coupled through a capacitor to the tuned-input circuit. To vary the coupling the tap on the input coil is varied.

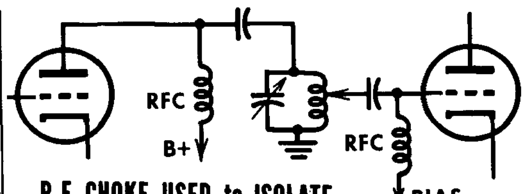
To vary the tap on the plate coil of a tank circuit, exposure to the d-c plate voltage becomes unavoidable. By using the r-f choke to provide a parallel path for the application of the B plus voltage, a parallel plate-feed circuit is obtained, freeing the tank circuit of the d-c plate voltage.



R-F CHOKES (RFC)



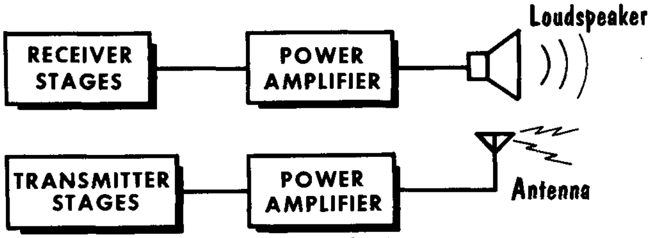
R-F CHOKE USED as a PLATE LOAD



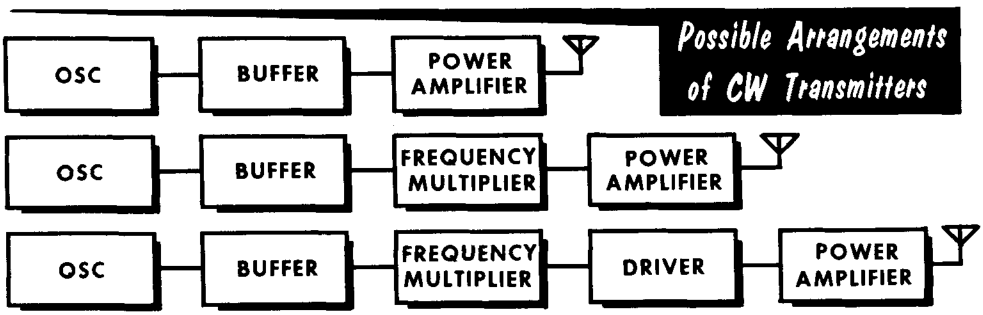
R-F CHOKE USED to ISOLATE the B+ FROM the TANK CIRCUIT, allowing safe handling of the tap

The Tuned R-F Amplifier and Driver Stage

The FINAL STAGE of a TRANSMITTER is the POWER AMPLIFIER



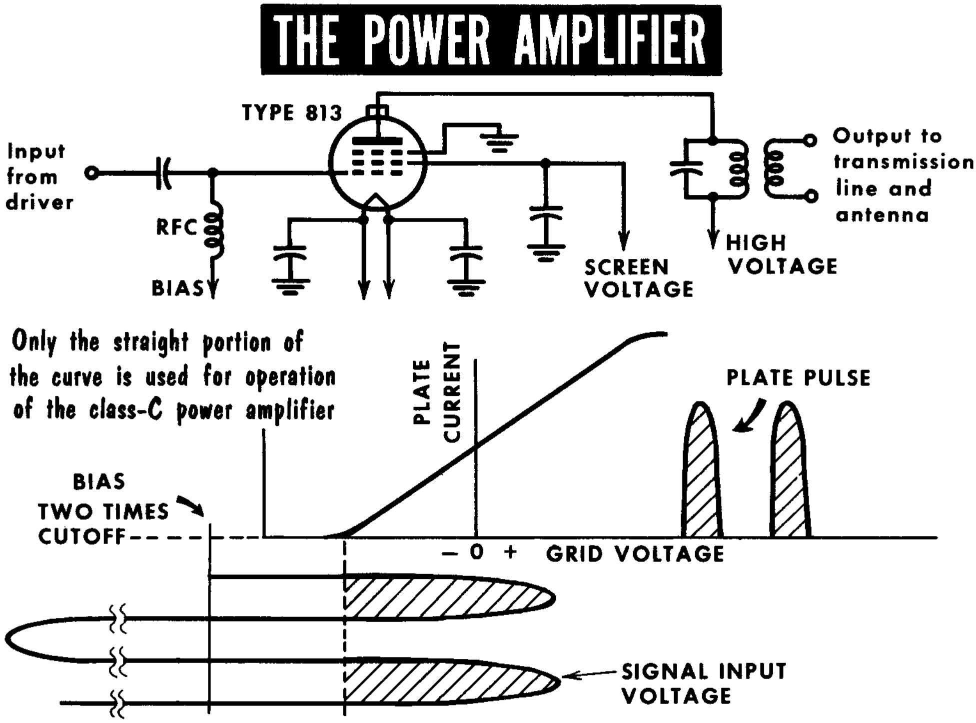
Just as the final stage of a receiver drives the loudspeaker, the final stage of a transmitter drives the antenna



The r-f power amplifier, sometimes referred to as the "final", is a class-C tuned amplifier. It is the last stage in the chain of amplification between the oscillator and the antenna. A transmitter may or may not include the buffer amplifier and frequency multipliers, but every transmitter, regardless of the kind of intelligence that it processes, contains an r-f power-amplifier stage. Sometimes the input signal required by the power amplifier is greater than that which might be obtained from a buffer stage or from a frequency multiplier. In this event a driver stage is used ahead of the power amplifier. The driver stage is generally a straight-through amplifier, i.e., the tuned circuit in its plate circuit is resonated to the frequency of the signal fed to the control grid of the same stage. In almost all respects the driver stage is a low-power version of the r-f power amplifier, differing from the latter mainly in the amplitude of the signal it processes, and in the lower values of grid bias and plate voltages used. While all transmitters contain a power amplifier stage, the amplifiers are not identical in all cases. They differ in the type of tube used to satisfy the output-signal power requirements, and in the operating voltages made necessary by the type of tube.

Several organizations of continuous wave (CW) transmitters are shown in block diagram form on this page. The relationship of the power amplifier to the rest of the circuit is readily apparent in each system. For ease of understanding, the keying method is omitted, as are the sources of operating voltages. They are dealt with later.

The Tuned R-F Power Amplifier

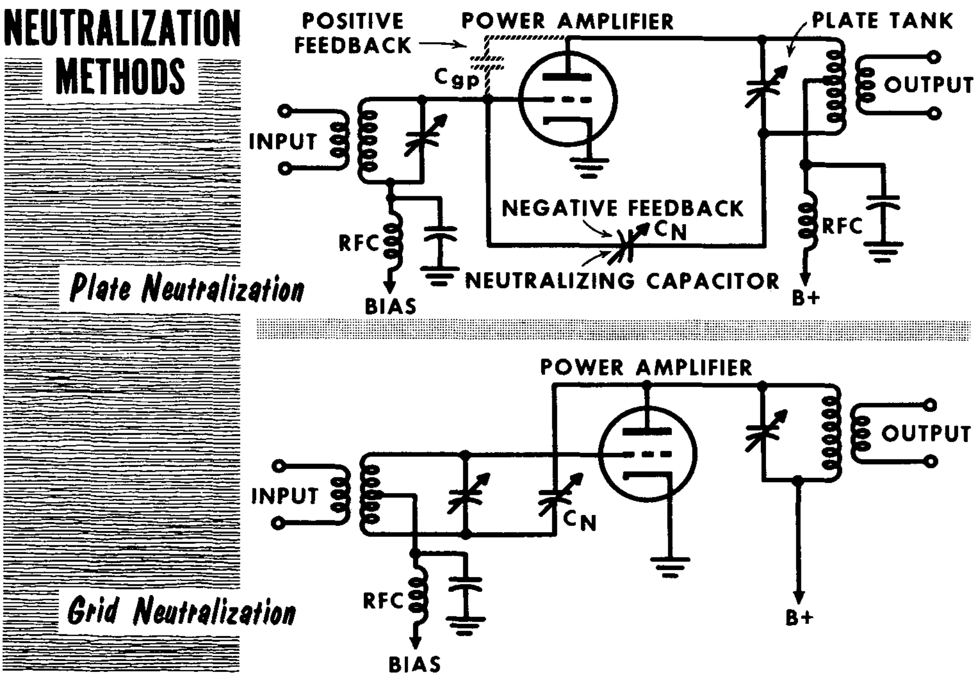


The basic operating conditions for the class-C tuned r-f amplifier have been described, but several qualifications pertinent to the power amplifier must be emphasized. First, is the requirement that the grid bias and plate voltage values be such that the changes in plate current for any given input-signal voltage occur along the straight portion only of the grid voltage-plate current characteristic. The plate current should rise from zero to maximum, but the maximum should not be beyond the saturation level, which as you will remember, was the maximum current condition in the class-C frequency multiplier. The abovementioned restriction establishes a limit on the peak amplitude of the input-signal voltage. Limiting the plate current in this way reduces the amplitudes of the harmonics in the plate current, thereby minimizing the possibility of delivering harmonic frequency signals to the antenna.

Another consideration is that the stage from which the power amplifier derives its input-signal voltage must be capable of supplying the necessary power consumed in the power-amplifier grid circuit (input signal voltage x grid current) during the operating cycle of the plate current. R-f power amplifier tubes bear input power ratings for full output, as for example, "8 watts driving power." The stage ahead of the power amplifier should be capable of delivering more than the stated minimum input power requirement of the amplifier stage. In this way adequate grid drive is available when the amplifier is delivering full output.

Neutralization

In circuits such as buffers, drivers, and power amplifiers where both the input and the output frequency are identical, an acute problem exists. The positive feedback through the grid-to-plate interelectrode capacitance may cause oscillations. The higher the frequency the more acute the problem becomes. In low-power tetrode and pentode circuits, the problem still exists



due to the high signal voltage. In spite of the low interelectrode capacitance, the small positive feedback may still be sufficient to start unwanted "parasitics", or oscillations. To prevent them, the simplest method, called neutralization, is to counter or "neutralize" with an equal amount of negative feedback.

One method, called plate neutralization, uses a neutralizing capacitor the purpose of which is to obtain the correct amount of negative feedback, and to apply it to the grid in the proper phase. For this purpose we connect the centertap of the plate r-f coil to ground. This makes the bottom end of the tank coil "hot" and of opposite polarity from the top of the coil, and suitable for application of a negative feedback voltage. Hence, the negative feedback is obtained from the end of the coil opposite to the plate.

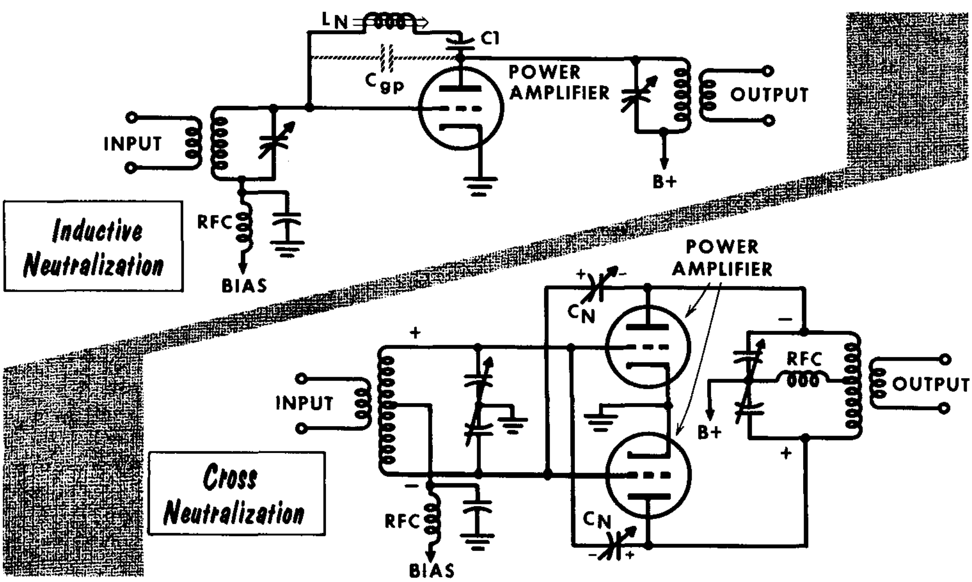
In another method, called grid neutralization, positive and negative feedback emanate from the same point. The d-c grid-bias voltage is applied to the centertap of the grid coil, instead of to the bottom of the grid coil which is now "hot", permitting the negative feedback voltage to be applied there.

Neutralization (Contd.)

For balanced circuits, cross neutralization is used. Portions of the output voltage of each plate are coupled to the grid of the opposite tube. Assume that at a given instant the signal voltage across the grid coil appears with the voltage shown, and which in turn places a voltage of opposite polarity on the output coil. The charge on the plate side of the neutralizing capacitors causes the opposite plate on the grid side to have an opposite charge. Hence, degenerative feedback is applied to each grid.

To adjust the negative feedback to the required value, the procedure is as follows: remove the plate voltage from the stage being neutralized. Apply the signal voltage to the grid and tune for resonance at the desired frequency. Use an r-f indicator which consists of a lamp in series with a turn of wire, inductively coupled to the plate-tank coil. If the lamp lights, an r-f voltage is present in the plate tank, and the circuit is not neutralized. The negative feedback is correctly adjusted when no r-f voltage is indicated in the plate-tank coil. After each adjustment, the grid and plate coils must be retuned, and the plate-tank coil rechecked for r-f voltage.

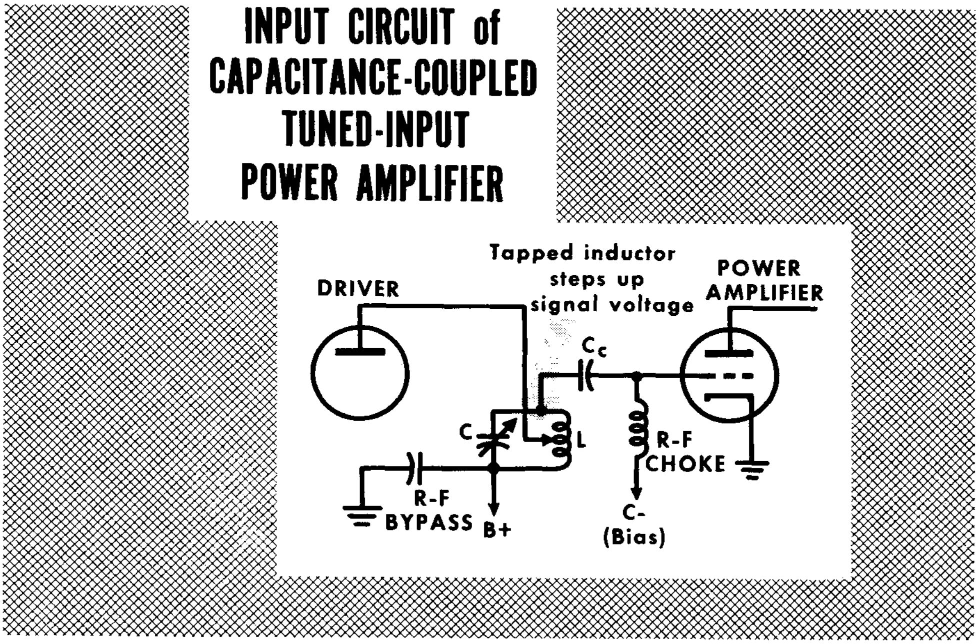
Another method, called inductive neutralization consists of canceling the interelectrode capacitive reactance by shunting it with an inductive reactance (coil), connected between grid and plate. The desired effect is obtained when the two reactances have equal values. Being of opposite signals they cancel each other. We have therefore a parallel-resonant circuit between plate and grid which at resonance offers maximum impedance, and blocks any energy transfer from plate to grid. The parallel-resonant circuit is at resonance for only one specific frequency, hence it must be retuned for any signal frequency change. C1 is a high value, low-reactance d-c blocking capacitor, and does not affect the parallel-resonant circuit.



Power-Amplifier Input Circuits

It is highly desirable to have as efficient a transfer of power from the driver stage to the power amplifier as possible. Therefore, the grid tank circuit must provide an impedance match between the grid input impedance of the amplifier and the plate output impedance of the driver stage. If a grounded-grid amplifier is used, similar considerations apply to the cathode tank circuit.

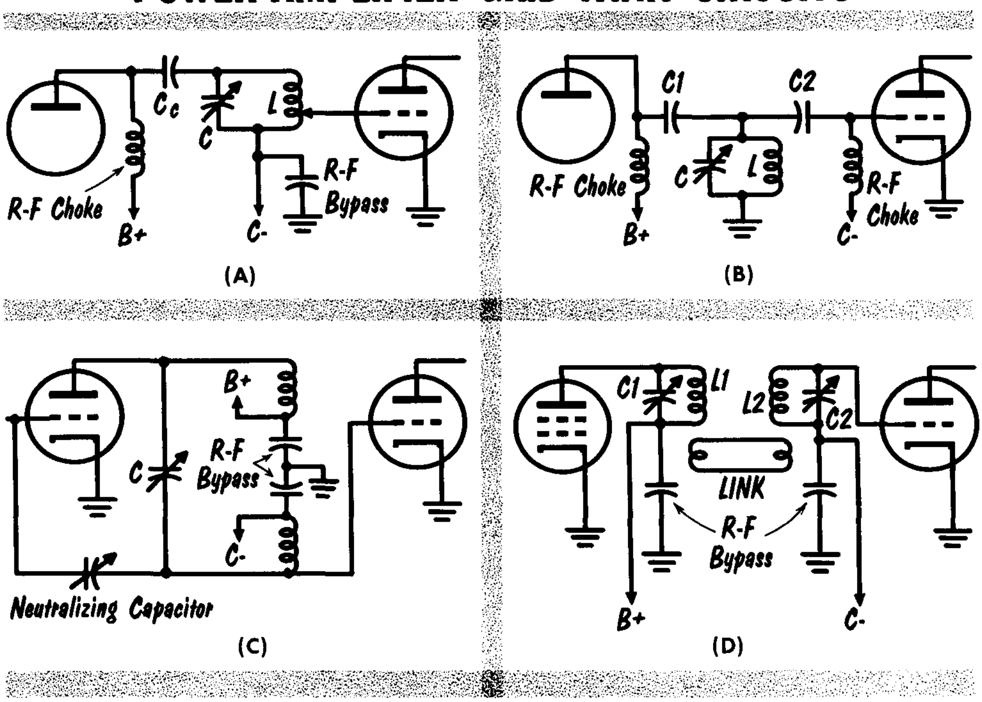
The impedance of a circuit normally is defined as the ratio of voltage to current. However, in the grid-circuit of a class-C amplifier, this ratio is far



from constant. When the grid voltage goes highly negative, no current is drawn at all; when it is positive, a great deal of current flows. Therefore, the impedance of the grid circuit varies over a range from an extremely high to an extremely low value through the operating cycle. If the input impedance of the grid circuit is too high, the heavy current demanded by the extreme grid swing cannot be drawn. As a result, actual grid voltage and consequent loss of peak efficiency are reduced in the operation of the amplifier. If the impedance of the grid tank circuit is too low, a great deal of power from the driver stage is required to operate it, and the losses in the inductor consume a considerable amount of the applied power. Generally, a compromise value is used which is approximately equal to the ratio of the driving power in watts divided by the square of the grid current. The choice of values for the components in the grid tank circuit is determined by this impedance. The result usually is satisfactory regulation of the grid voltage without excessive power loss.

Power-Amplifier Input Circuits (Contd.)

POWER AMPLIFIER GRID-TANK CIRCUITS

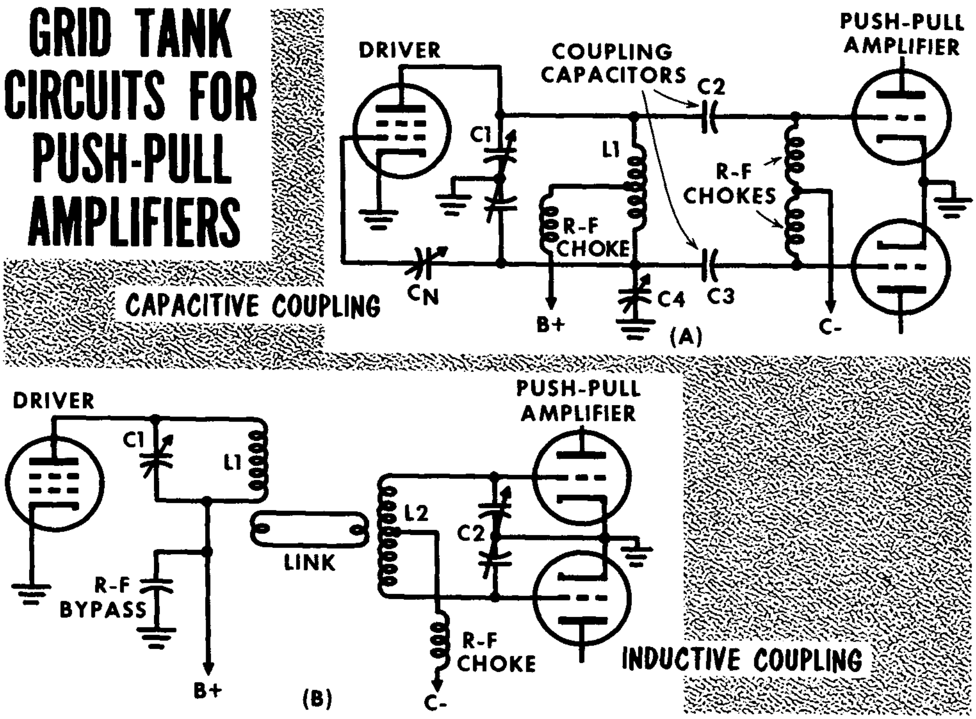


The circuit at A is much the same as the previous circuit except that the tuned circuit is now in the grid circuit of the power amplifier, and driver plate voltage is supplied through an r-f choke. This arrangement is used when the required impedance at the grid of the power amplifier is lower than the output impedance needed in the driver stage. Circuit B permits complete d-c isolation of the tuned circuit from the driver and amplifier stages. C1 and C2 block the d-c voltages, and at the same time couple the signal from driver plate to amplifier grid. There is no means for adjusting the impedance between grid and plate circuits.

Circuit C permits the driver to be neutralized to prevent oscillation. It also provides variable drive for the amplifier, and d-c isolation for the bias and high-voltage circuits without need for r-f chokes. Inductor L of the tuned circuit is split into two parts at the center, each of which is grounded separately with r-f bypass capacitors. Driver plate voltage thus cannot reach the amplifier grid. In circuit D the tuned-plate tank of the driver is inductively coupled through a low-impedance link to a tuned-grid circuit. The link inductance is small and therefore the impedance of the coupling circuit is low. This minimizes losses in the transmission of driving power, and provides great flexibility in matching impedances between the driver and power amplifier.

Input Circuits For Push-Pull Amplifiers

The input circuits for push-pull amplifiers are variations of those used in single-ended amplifiers. The capacitive coupling arrangement in A uses a tuned-plate circuit for the driver tube formed by L1 and C1. Because each half of the split coil is out of phase with the other half, each half can supply grid drive in push-pull directly through C2 and C3. A split-stator capacitor, C1, is used as the main tuning capacitor for the coupling arrangement. Two r-f chokes are used, one from each grid to the bias supply. C4 introduces a small amount of capacitance from the lower end of the tank circuit to ground to compensate for the plate-to-ground capacitance of the driver tube, which



appears across the upper half of the circuit. Driver plate voltage is applied through an r-f choke. Neutralization of the driver stage is provided by C.

An inductive coupling arrangement for the grid tank circuit is shown in B. The link circuit transfers energy from the driver tank, L1-C1; out-of-phase voltages to the push-pull grids of the amplifier are developed across C2.

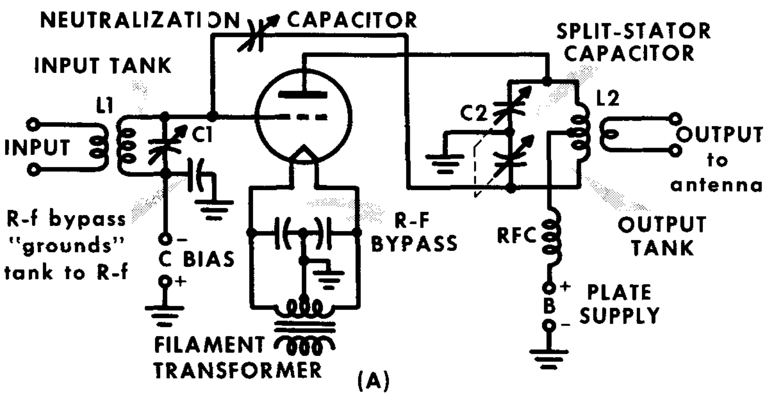
The choice of circuits depends on several considerations. Where the tuning capacitor is grounded directly, it must have twice the voltage rating of one grounded through a capacitor. In some instances, the use of an r-f choke for bias is undesirable. The link-coupled circuits permit the driver to be located at some distance from the amplifier and connected to it through a low-impedance transmission line.

Single-Ended Power Amplifiers

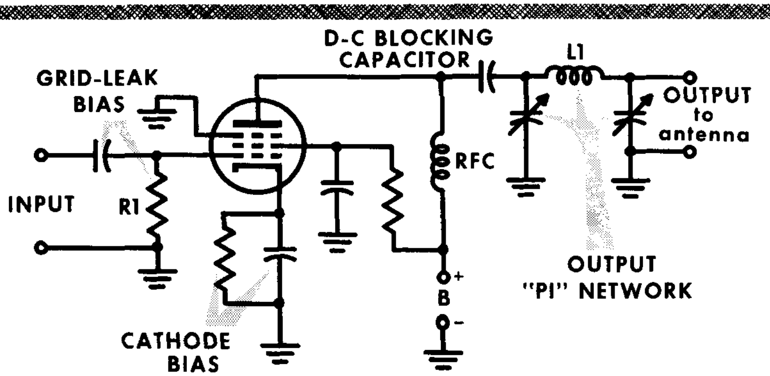
A power amplifier that uses one tube is called a single-ended circuit. The circuit is usually the same whether operated class-B or -C. The difference is in bias and excitation voltages and in component values. The single-ended circuit generates more harmonic output which sometimes is a disadvantage. It is however, the simplest circuit to use over a wide range of frequencies. Several types of tank circuits provide a wide range of impedance matching. These circuits are necessary where a variety of different antennas must be used. In A, the plate tank is composed of a variable capacitor with a single rotor and two stators, and L2. The capacitor is called a split-stator, and is used to obtain the out-of-phase voltage that is fed back for neutralization. The output tank is coupled to the antenna by a link coil around L2. This circuit is popular with high-power triodes.

In circuit B, grid-leak bias is used. Cathode bias also is provided as a protective measure in case excitation fails, which would result in a loss of grid bias if R1 alone were used. The tank circuit is a special impedance-matching network known as a pi network. It can be used with almost any length antenna.

SINGLE-ENDED POWER AMPLIFIER CIRCUITS



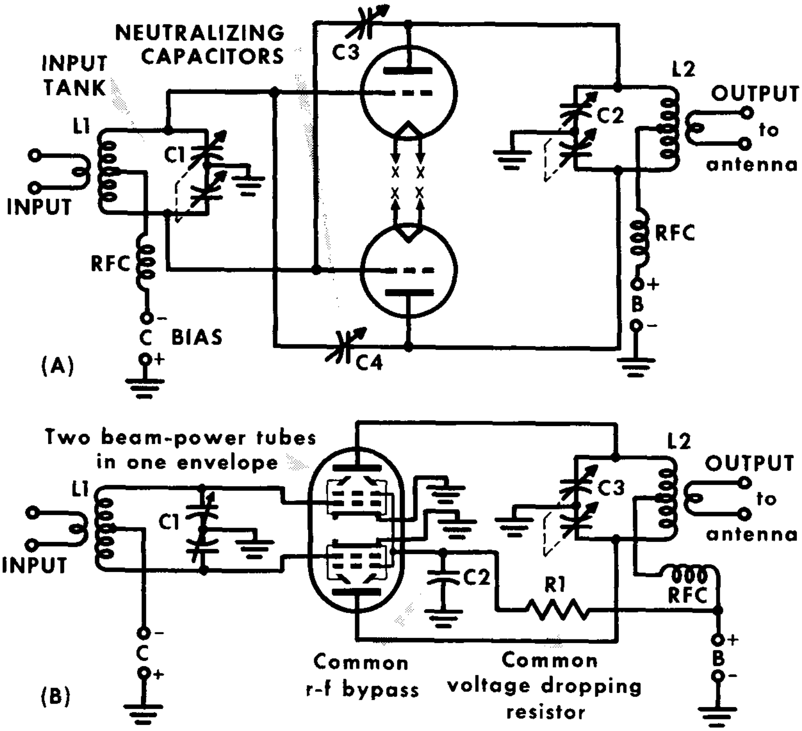
(A)



(B)

Push-Pull Power Amplifiers

PUSH-PULL POWER AMPLIFIER CIRCUITS



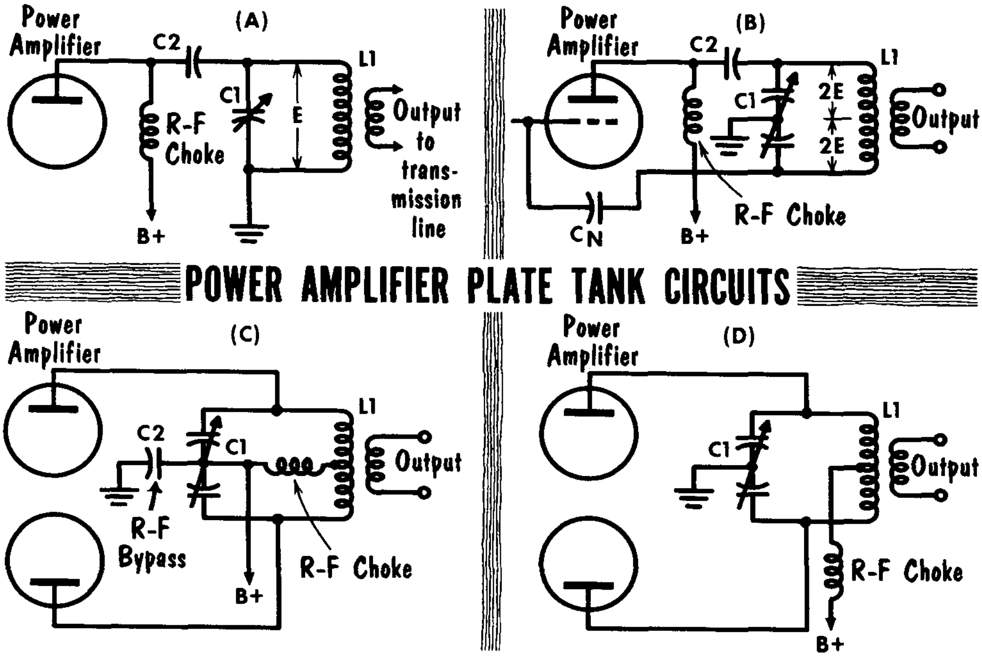
Since a push-pull circuit uses two identical tubes, advantage is sometimes taken of this, and the two tubes are combined in one glass envelope. This is especially useful at very high frequencies where the length of connecting leads becomes important, and must be kept short. Push-pull circuits have the advantage of lower harmonic output, ease of neutralization, and greater output for a given amount of excitation. Such a circuit requires only as much excitation as one of its tubes does when run by itself in a single-ended connection. This is because both halves of the grid-voltage cycle are used alternately by each tube. The output is recombined in the plate circuit.

In circuit A, the input tank circuit provides equal and opposite voltages to the grids of the two triodes. The r-f choke provides a means of inserting grid bias from a fixed bias supply. The plate tank is made with split-stator capacitor C2, and recombines the output from the two triodes. Since opposite sides of a tank are 180° out of phase, neutralizing voltages may be obtained simply by feeding back some of the energy at these points to the opposite grids. In circuit B the grid tank L1-C1 provides drive for the tube. Grid bias is furnished from a fixed supply. In the tube, the screens of both sections are connected, and C2 serves as a bypass for them. R1 provides screen voltage by dropping the plate voltage because of the current that passes through it. The output tank circuit is the same as in A. The r-f choke permits feeding plate voltage to both halves of the tube.

Power-Amplifier Output Circuits

A large variety of practical circuits has been devised which present the proper load impedance to the power amplifier when connected to the transmission line or antenna. The simple parallel-resonant tuned circuit in A is frequently used for single-ended tetrode amplifiers. The circuit is shunt-fed. Tank circuit C1-L1 is coupled to the plate by C2. The advantage of this circuit lies in the removal of all d-c voltages from the tuning capacitor. This means a lower value of total voltage across this component with corresponding smaller size. A major shock hazard from contact with an exposed portion of the tank circuit is removed, but possibility of a bad r-f burn always exists.

Although circuit A is satisfactory when used with tetrodes and grounded-grid triodes, it provides no means of neutralizing an ordinary grounded cathode

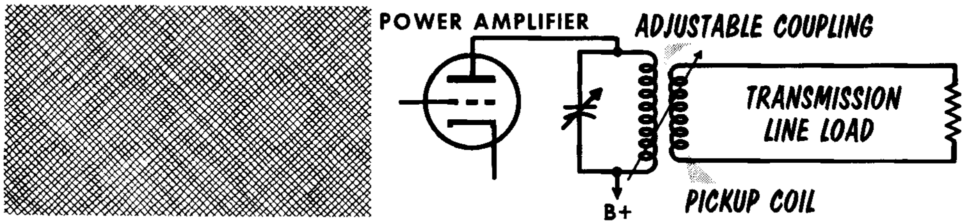


triode. In B, C1 has a split-stator with the rotor directly grounded. This permits an out-of-phase voltage to be taken from the lower end of the coil and returned to the input through neutralizing capacitor C_N. The split-stator capacitor effectively divides the circuit in two parts, and an r-f peak of twice the d-c plate voltage can appear across each. This requires a physically large capacitor. The circuit in C is the push-pull counterpart of the simple resonant tank. A split-stator capacitor is used with the push-pull version, and the rotor is grounded for r-f through C2. To reduce the voltage across each half of C1, the plate voltage sometimes is connected to the rotor. The shock hazard introduced by this can be avoided by grounding the rotor of C1 directly (D), and applying the plate voltage through a choke.

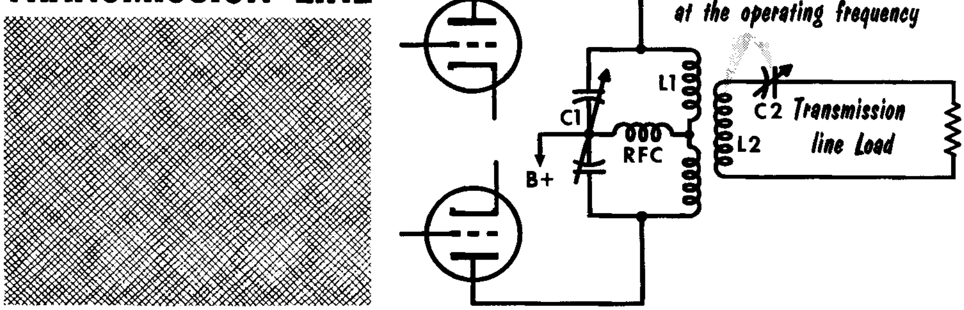
Output Coupling

The output of the final stage, or power output stage, has to be coupled to the transmission line as efficiently as possible to ensure maximum output to the antenna. A transmission line, to be discussed later, can be assumed for present purposes to appear as a pure resistive load, the value of which is determined by the characteristics of the type of transmission line used. To

COUPLING the OUTPUT to the TRANSMISSION LINE



USING a TUNED CIRCUIT for COUPLING to the TRANSMISSION LINE



couple the output to the transmission line, a pickup coil is used. The coupling between the output plate-tank coil and the pickup coil is made adjustable to permit changes in loading.

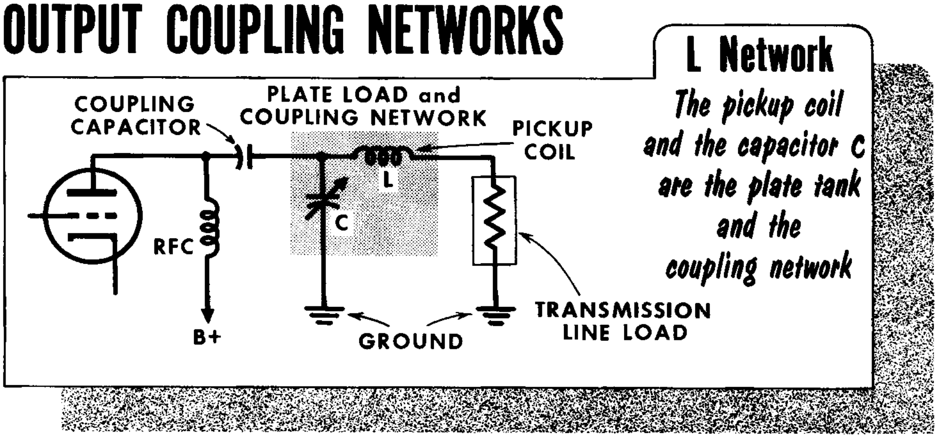
The pickup coil acting as the source for the load represented by the transmission line should match the impedance of the transmission line for maximum transfer of power. This is not a simple thing to accomplish; a mismatch will often occur, which when reflected back to the output plate-tank circuit will cause it to be detuned. To overcome this obstacle a capacitor is placed in series with the pickup coil, providing a tuned circuit. Additionally the tuned circuit provides increased selectivity of the output frequencies, making it helpful in suppression of harmonics. The resonant circuit formed by the pickup coil and capacitor should not have too high a Q. If the output becomes too selective it will have to be retuned for each small change in output frequency.

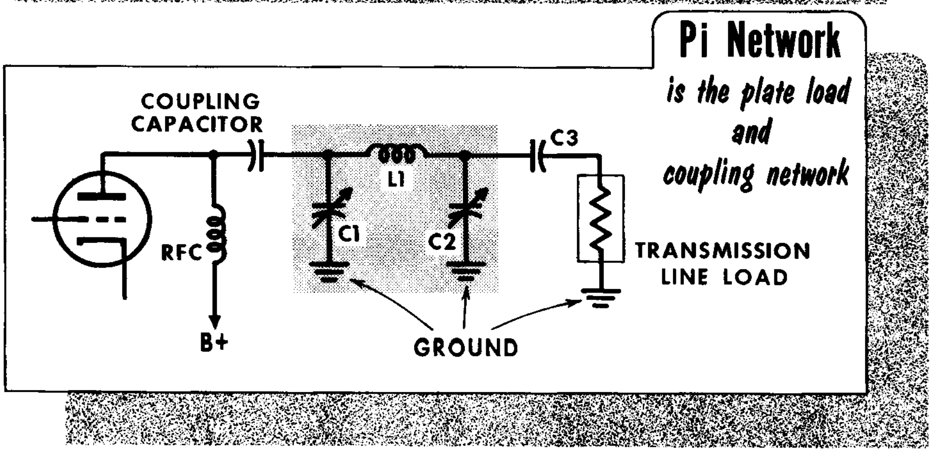
Output Coupling (Contd.)

In place of the output tank circuit and transmission line tuned circuit, a network may be used which will act as both the tank circuit and matching circuit to the transmission line impedance. An L network uses an r-f choke for the plate load and a capacitor to couple the output to the tuning capacitor and inductance forming the L network. This places the transmission line impedance in series with the L and C of the tuned output circuit.

One of the most frequently encountered variable matching networks for the output of an r-f amplifier is the pi network. The plate voltage fed through the r-f choke is prevented from reaching the antenna by blocking capacitor C3. The pi network of C1, L1, and C2 is capable of matching a wide range of impedances, and operates as a voltage divider. The combination of L1 and C2 forms the divider circuit which develops higher or lower voltages at the output terminal. C1 then tunes the combination of C2 and L1 to resonance at the operating frequency. Depending on the relative values of C1 and C2, a voltage much lower than the a-c plate voltage can be developed. Thus, this circuit can match an extremely wide range of impedances.

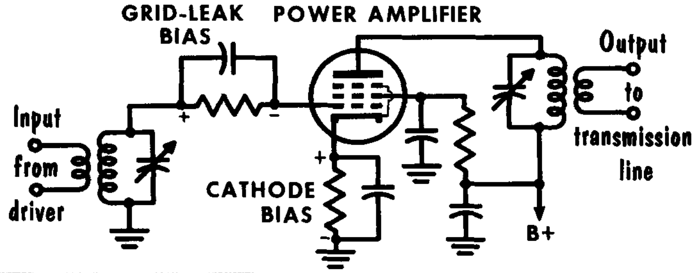
OUTPUT COUPLING NETWORKS



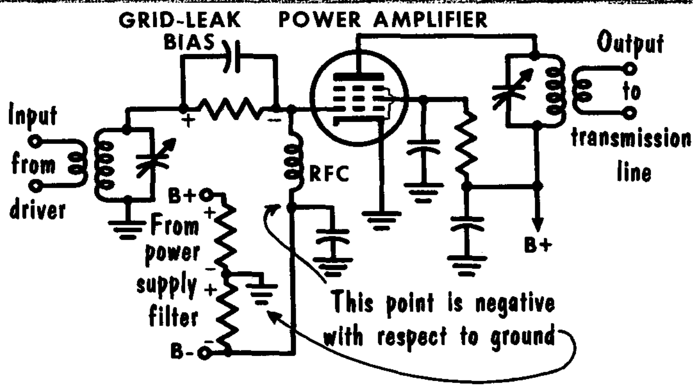


Bias Voltages

COMBINING GRID-LEAK and CATHODE BIAS



COMBINING GRID-LEAK and POWER SUPPLY BIAS



The proper operation of an amplifier depends to a large degree upon the correct bias voltages. A tube that operates at or beyond cutoff, such as in class-B and -C operations, must use some form of external bias. In class-A operation, the continuous flow of plate and screen current through a cathode resistor can be used to develop self-bias. Most transmitter circuits operate class-C; therefore some form of bias is required.

Grid-leak bias develops a bias voltage due to the presence of a signal. Driving the grid positive draws grid current, develops a voltage drop across the grid resistor, and charges the capacitor. After several cycles, the charge on the grid levels off at the correct bias value. Loss of grid drive, however, causes loss of grid bias. In addition to accidental loss of grid drive, a condition exists when CW code is transmitted for short durations. Between "dots and dashes" the lack of signal is as though no signal were applied to the grid, causing a loss of grid-leak bias. With high-power tubes, the lack of bias may cause excessive plate current flow, and quickly ruin the tube.

A typical circuit shown, which includes both grid-leak and cathode bias, prevents excessive plate current flow. When no signal is applied to the grid, the heavy plate current flow through the cathode resistor develops a large voltage drop. This is sufficient self-bias to quickly reduce the plate current and keep it within safe limits. Another circuit shown, makes use of a minimum value of bias voltage from an external source, such as the tap of a voltage divider or a separate bias supply. When the signal is applied, the grid drive develops its own grid-leak bias, in addition to the external bias.

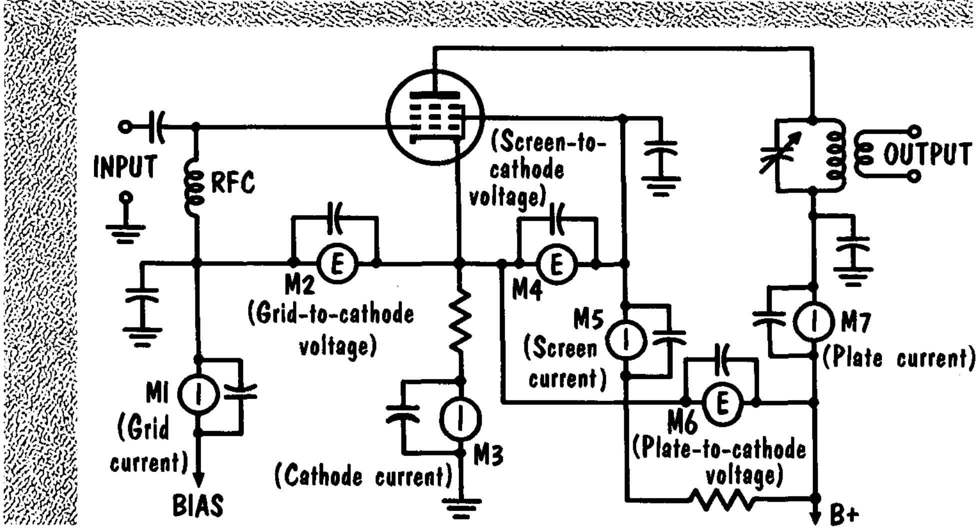
RUNNING-TIME METER

Metering performs a vital job in maintaining a transmitter. Metering the many voltages and currents ensures high performance, and indicates present or possible future troubles. In high-power, high-cost commercial transmitters, individual meter movements are generally used for each reading. In less complex transmitters, a single meter is used to monitor many readings by switching it into different circuits. When metering in r-f circuits a bypass capacitor is always placed directly across the meter terminals.

To measure the length of operation time of a transmitter, a "running time" meter is used. The running-time, or elapsed-time meter is essentially an electric clock that reads in hours and tenths of an hour as shown. The running-time meter is usually connected across the power transformer primary winding to measure the time duration that ac is applied to the transmitter. Occasionally an a-c meter is used to measure the filament voltage of a high-power tube, but mostly d-c voltages and currents are measured.

In the typical placement of individual meters shown, meter M1 indicates grid current flow. To measure the voltage from grid to cathode, meter M2 is used. The current meter M3, in the cathode circuit measures the combined control grid, screen grid, and plate currents. At times, a cathode-current meter is used in place of a plate-current meter. This eliminates the placing of a meter where high voltage is present. To obtain the plate current value, the control grid and the screen grid currents must be deducted from the cathode-current reading.

TYPICAL PLACEMENT of METERS in a TRANSMITTER CIRCUIT

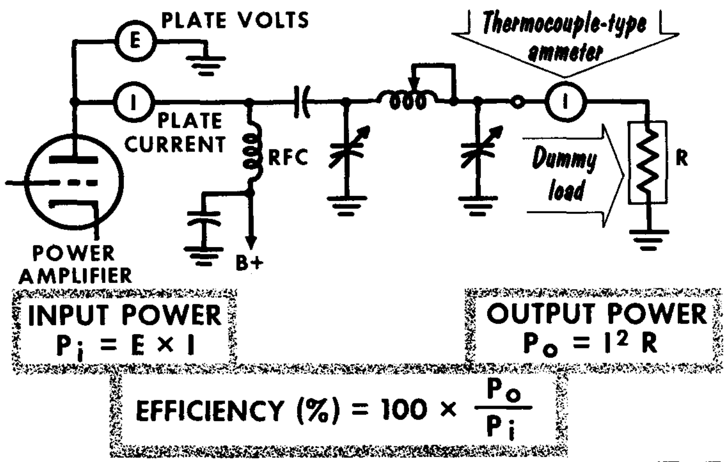


Output and Input Power

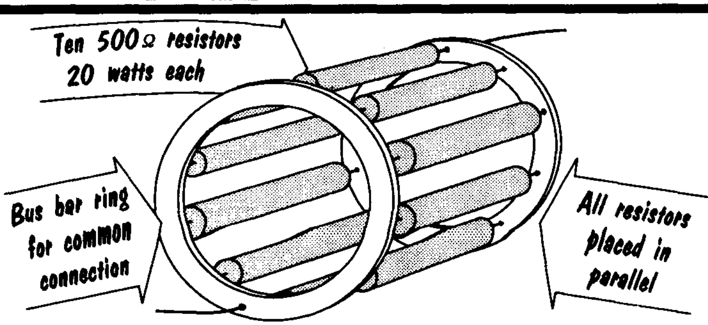
A transmitter is most often rated by its output power. To measure it, the transmission line is replaced with a dummy load that is a pure resistance of the same value as that of the transmission line it replaces. The dummy load's resistance must be capable of handling the dissipated heat. Many resistors may be placed in parallel; for example, placing ten 500-ohm resistors rated at 20 watts in parallel, provides a dummy load of 50 ohms, capable of dissipating 200 watts. In another method, the dummy load is immersed in oil and permits a still higher power dissipation. The true output power indication is obtained by measuring the actual watts of power being dissipated as heat in a resistor. A more accurate method of measuring heat dissipation uses a heat-measuring instrument called a calorimeter.

Most often, however, a thermocouple-type ammeter is inserted in series with the dummy load, as shown. The current indicated on the ammeter, and the resistance offered by the load gives us the output power with the formula $P = I^2R$, where I is the measured current, and R the resistance of the load. The input-power rating of the final transmitter stage is obtained from the plate voltage (E) and plate current (I) values with the formula $P = E \times I$. Knowing both the input and the output powers of a final stage, its efficiency may readily be calculated.

MEASURING OUTPUT and INPUT POWER



DUMMY LOAD SIMULATING a 50-OHM LOAD CAPABLE of DISSIPATING 200 WATTS

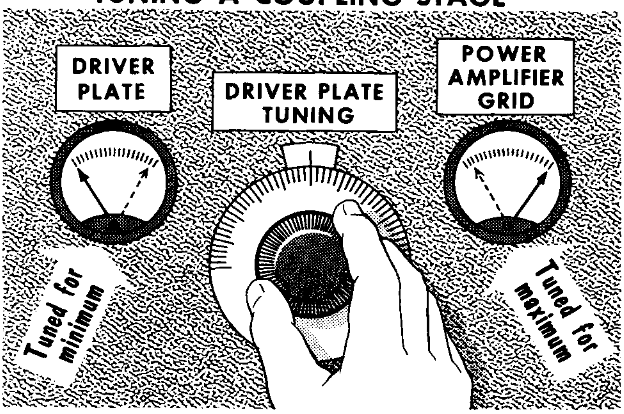


Tuning Procedure

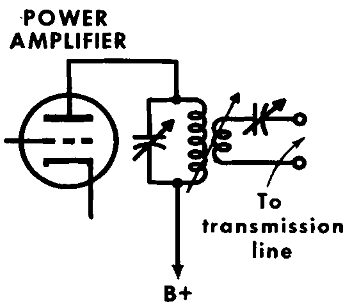
Assuming a correctly operating transmitter, the task of tuning it to the desired frequency can be difficult or easy, depending upon the skill of the operator. If the circuits are designed for it, it is best to start by removing plate voltage from all stages following the oscillator. The procedure generally is to tune the grid circuit first, then apply plate voltage and tune the plate circuit.

Tuning a driver output stage is shown. The plate circuit is tuned for a dip at resonance as indicated by a dip in the plate-current meter. At the same time

TUNING A COUPLING STAGE



TUNING THE POWER OUTPUT STAGE



the grid-current reading of the power-amplifier stage indicates maximum at the same point. This is true of a capacitor-coupled circuit; if the grid circuit of the following stage is a tuned circuit, it will have to be tuned separately.

When plate and screen voltages are applied and the plate circuit tuned, the grid-current reading may become slightly lower. If the grid current has reduced to a value below that required, the coupling to the preceding stage should be increased to increase the grid drive. With the correct value of grid drive, and the grid circuit tuned, the plate circuit is then tuned for a dip at resonance. This procedure is followed for all stages except the final.

When tuning the power-output stage using a coupling circuit such as that shown, the following procedure is typical. The grid circuit should be tuned for maximum grid drive. The pickup coil of the transmission line is coupled approximately half way. With plate voltage applied, the tank circuit is then tuned for a dip at the resonant frequency. The tuning capacitor in series with the pickup coil is then adjusted for a maximum reading of plate current, indicating that the transmission pickup loop is tuned and drawing more energy from the output-tank circuit. This slightly detunes the output-tank circuit. This procedure is then repeated, the plate-tank capacitor retuned for a dip, and the pickup-coil tuning capacitor tuned for maximum, until the plate-current reading is set to the value recommended for the power-output tube being used.

Tuning Procedure (Contd.)

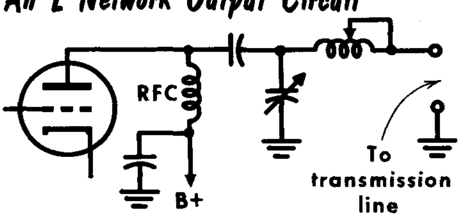
When tuning an L network such as that shown, the grid circuit, as in other circuits, should first be tuned for maximum grid drive. When tuning the output circuit it is possible to have a shallow plate-current dip, making it difficult to find. For a quick approximation the transmission line should if possible be temporarily shorted. With no load the plate-current dip at resonance will then be sharp and readily noticed.

With the short-circuit removed from the transmission line the value of plate current should then be noted. If the plate current is less than the recommended value the tap on the plate coil is moved to decrease the value of inductance. If the plate-current value noted is too high, the plate-coil tap should be moved to increase the value of inductance. Following any resetting of the plate-coil tap, the tuning capacitor should be readjusted for minimum plate current, and the grid circuit retuned for maximum grid drive.

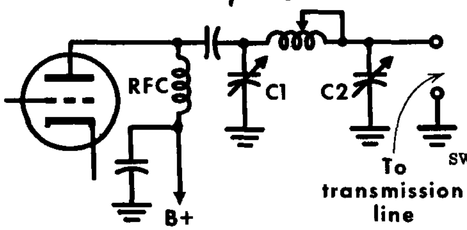
To tune a pi network output the grid circuit is tuned for maximum grid drive. C2 is set for maximum capacitance, and the coil is set for maximum inductance. C1 is then tuned for a dip in plate current. If the plate-current dip is so shallow as to be difficult to locate, the tap on the coil should be moved to lower the value of inductance by approximately one third. When the dip is located, the value of plate current should be noted. If the plate current is lower than the recommended value, reduce the capacitance of C2 and retune C1. If the recommended value of plate current is reached with C2 very close to minimum capacitance, a small amount of inductance should be added by shifting the tap on the plate coil.

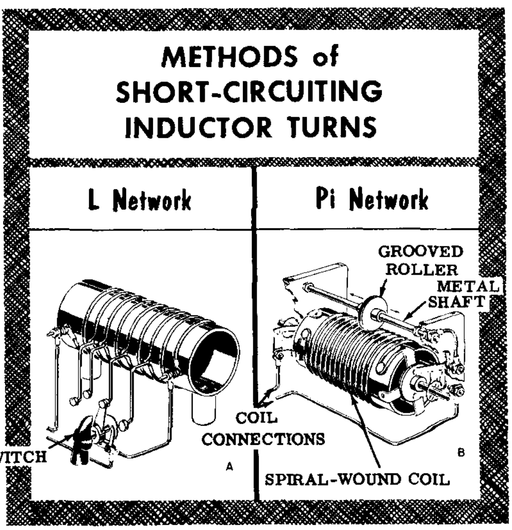
In operation, a small change of frequency may require that the pi network be quickly retuned. This is done by tuning C1 for a plate-current dip. C2 is then set for the correct value of plate current, causing a slight detuning of C1. C1 is then retuned for a plate-current dip.

An L Network Output Circuit

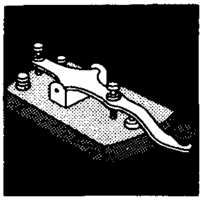


A Pi Network Output Circuit





Keying

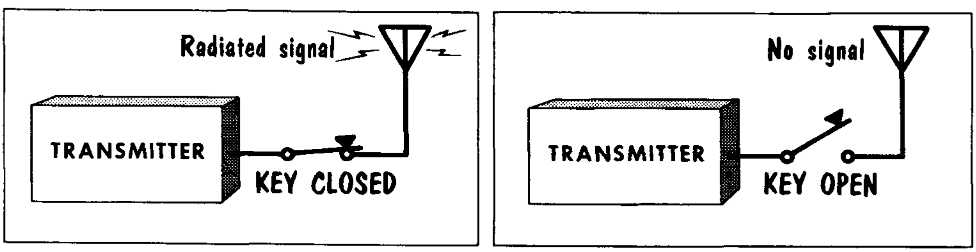


A TYPICAL KEY

The output signal of a transmitter using the various stages just described – oscillator, buffer, frequency multipliers and power output – would be continuous-wave type-AO transmission. The output signal radiated from the antenna is called a carrier. To have the carrier convey information it is alternately turned on and off in the form of dots and dashes to transmit code characters. This type of interrupted transmission is called telegraphy, or type-A1 emission. To achieve telegraphy with a CW transmitter the ability to quickly and easily turn the carrier on and off must be placed somewhere in its circuitry. The instrument most often used to turn the transmitter signal on and off is the familiar telegraph key. Pressing the key down closes the contacts and turns the carrier on. By releasing the pressure upon the key a spring will return it to open, turning the carrier off. This is known as keying the carrier.

At first glance keying a transmitter appears simple, merely turning any stage on and off will cause the carrier to be on or off. Unfortunately it is not this simple; keying the transmitter requires careful thought and attention as to what stage should be keyed. In addition, keying a circuit brings with it attendant troubles. A typical example is a backwave. This consists of some carrier signal being transmitted when the transmitter output should be zero. The correct carrier signal is one that is fully on when the key is down, and zero when the key is up. Backwave caused by not having the transmitter completely off when the key is up will cause difficult and unpleasant reception for the receiving operator. When the carrier signal is turned on and off too

KEYING TURNS the CARRIER WAVE ON and OFF

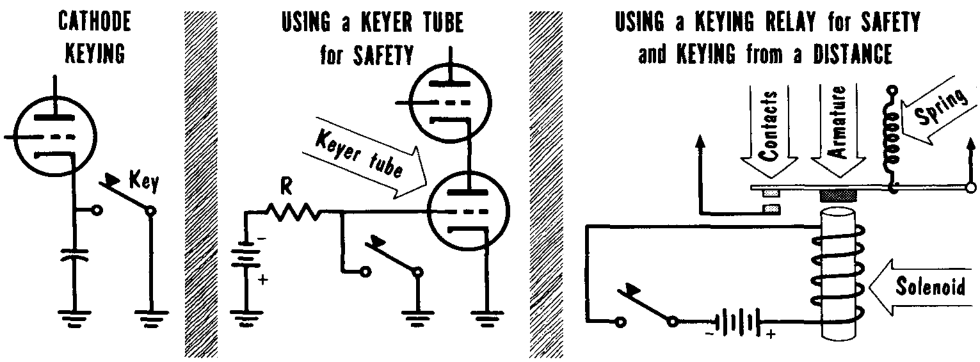


sharply the rapid change in power level produces sidebands which are heard as clicks, called key clicks. The sharper the rise and fall of the carrier the wider the sidebands causing the key clicks, increasing their amplitude and possibly interfering with close-by signals. When the stage that is keyed is the oscillator, or a stage close to the oscillator, there is the possibility of a shift in frequency. This shift in frequency caused by keying produces a chirp when the key is pressed down, and another chirp when released. A change in oscillator frequency of only 25 cycles will cause sufficient chirp to make copying the signal unpleasant. A chirp of 200 cycles is enough to make copying extremely difficult.

Keying (Contd.)

The most frequent cause of backwave is keying in such a manner that the grid bias is not reduced beyond cutoff, permitting r-f energy to leak through and be amplified in succeeding stages. To eliminate this it is possible to key the oscillator stage directly. This, however, may result in chirping. Another method of combating backwave is to key the cathode circuit. The up or open key position opens both the plate circuit and grid return, blocking the grid and providing no plate current. Hence, no r-f signal is passed to the succeeding stage. Another possible cause of backwave is the keying of a circuit following the oscillator. This permits the oscillator to run continuously. If energy from the continuous-running oscillator leaks through to the antenna, a backwave may be generated. Careful shielding of the oscillator stage prevents this.

In low-power stages the voltage developed across the key may be safe; in high-power stages dangerous. To remove this danger an electronic switch can be substituted for the key, with the key controlling the switch. The circuit here shows vacuum-tube keying using a vacuum tube as an electronic switch. With the key in the up position the bias to the grid of the keyer tube keeps it cut off. The high-resistance open-keyer tube isolates the cathode of

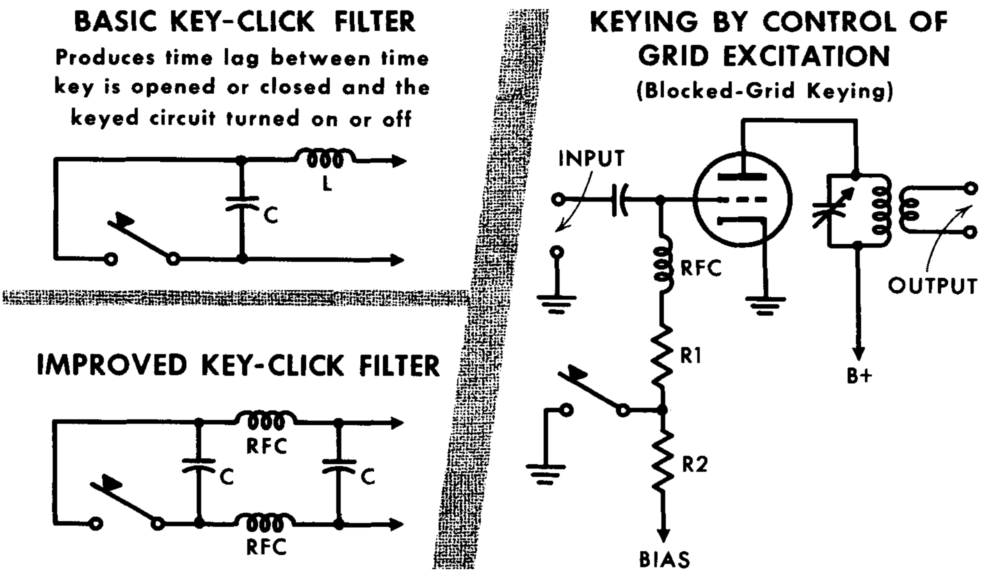


the tube used in the transmitter stage. With the key down the bias to the grid of the keyer is removed; the bias voltage is developed across resistor R. With no bias the keyer tube conducts heavily, providing low plate resistance that essentially places the cathode of the tube used in the transmitter stage at ground potential. The value of plate resistance of the keyer tube acts to place the plate of the keyer tube and the cathode of the transmitter tube at a slight positive potential; thus the keyer tube also acts as a cathode-bias resistor.

A method of keying any circuit where unsafe voltages may be present, is that of using a keying relay. When the key is closed the solenoid is energized, pulling the armature to it, closing the relay. Opening the key releases the armature and the contacts to open the circuit. A keying relay is quite useful for keying at long distances from the transmitter. The voltage used to energize the solenoid can be run for long distances in place of the cathode circuit or other circuit being keyed.

Keying (Contd.)

Eliminating key clicks requires a circuit that will cause a small time lag between the time the key is opened or closed, and the keyed circuit turned off or on. To do this, key-click filters are placed between the telegraph key and the circuit it is keying. A typical key-click filter which provides a time lag is shown. With the key closed the inductance in series with the circuit being keyed provides a lag in current flow by its reactance. This causes the keyed



circuit to be turned on gradually. When the key is opened, the charge across C briefly continues the current flow, keeping the keyed circuit from being turned off too sharply.

Another type of key click is caused by arcing at the contacts of the key. To remove these, a filter such as that shown is placed close to the key.

The question of which stage or stages to key in a transmitter is an important one. Keying the oscillator stage, or too close to it, can, as previously mentioned, cause chirping. If the oscillator circuit is stable, it can be keyed. However, any difficulty caused by keying the oscillator will be magnified by the succeeding amplifier and frequency-multiplier stages. Occasionally the problem of keying is eased by keying more than one stage, or using the previously mentioned keyer circuits. To enable clean keying of any stage there are various basic circuits. These circuits fall into two main categories: the first type controls the excitation, the second controls the voltage to the stage being keyed. A basic circuit for control of grid excitation is shown. With the key in the up position sufficient negative bias is applied to the grid of the tube to ensure its being cut off. With the key down, the bias voltage develops across R2 and is no longer applied to the grid of the tube. Resistor R1 acts to develop grid-leak bias.

Keying (Contd.)

Keying by controlling voltages to the transmitter stage can be accomplished by as many means as there are tube electrodes. In a typical pentode with the suppressor grid internally connected to the cathode, keying can be accomplished at the cathode, screen grid, or plate. Cathode keying has been previously mentioned. Screen-grid keying can be accomplished to a degree by merely keying the screen-grid voltage from its normal value to zero. With zero screen voltage, however, there is still ample plate current flowing, providing backwave. To eliminate this backwave it is necessary to not only remove the positive screen voltage, but also to apply a negative screen-grid voltage to halt all plate current flow. A basic screen-grid keying circuit is shown. With the key up, a negative voltage is applied to the screen grid, effectively cutting off all plate-current flow. With the key down, a voltage is applied to the screen grid permitting normal operation of the stage.

Plate keying is usually done in an early stage of the transmitter where plate voltage values are not too high. The keying can take place in either the posi-

Keying by Control of the Screen-Grid Voltage

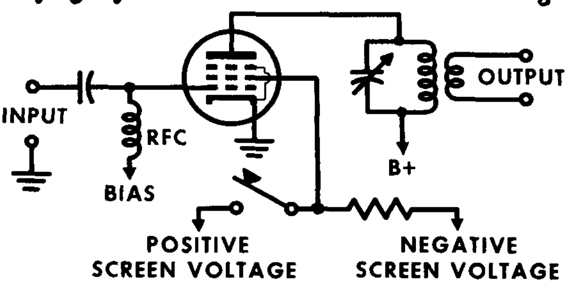
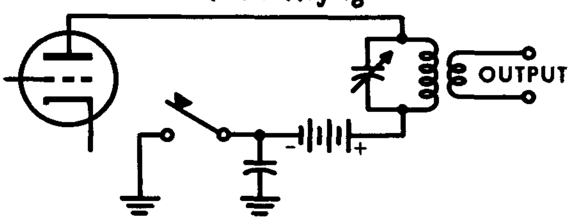
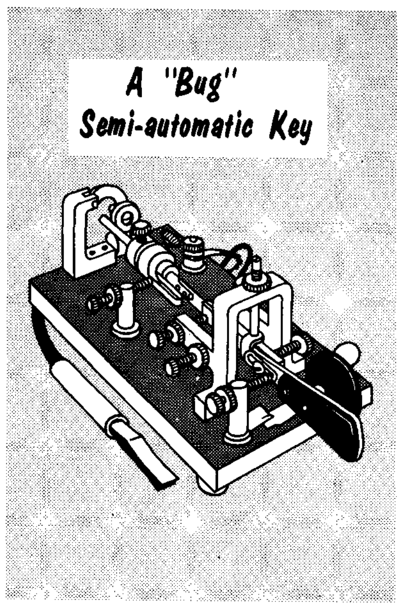


Plate Keying





A "Bug" Semi-automatic Key

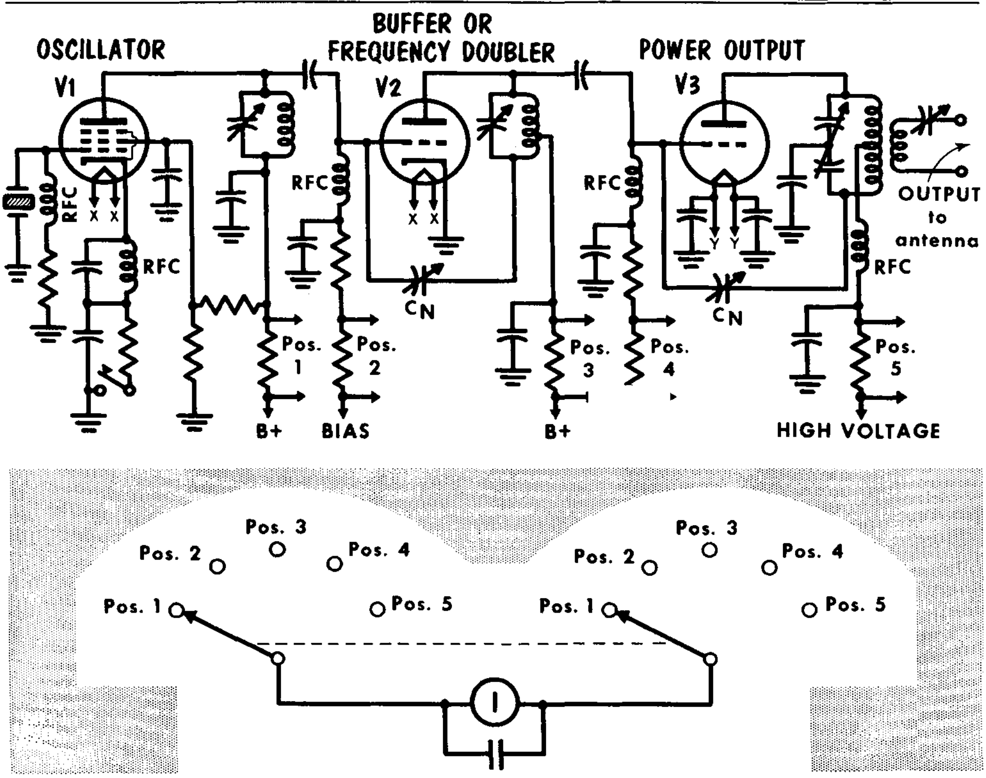
tive line of the circuit or in the negative line as shown. Most often it is done in the negative line since this permits one side of the key to be at ground potential.

In addition to the usual hand-operated telegraph key there are semi-automatic telegraph keys, one of which is shown. Popularly known as a bug, the semi-automatic key will make continuous dots when the paddle is pressed to the right. These dots are made by a spring action that causes the contacts to open and close at a rate set by an adjustable weight.

Basic CW Transmitter

A typical basic low-power transmitter such as might be found in some ship-to-shore installation is shown. The oscillator being crystal-controlled is quite stable, and is therefore the stage that is keyed. The components in the cathode circuit of the oscillator form a key-click filter. The screen-grid voltage of V1 is developed by taking the B-plus applied to the plate and applying it to a voltage divider. The output of the crystal oscillator is capacitor-coupled to the grid of V2 which can be used as either a buffer or frequency doubler. Being a triode, when operated as a buffer with the output frequency the same as the input frequency, neutralization is necessary. The power-output stage is also neutralized.

BASIC CW TRANSMITTER



Metering is accomplished with a single meter switched in to read oscillator-plate and screen-grid current, buffer-grid current, buffer-plate current, the final grid current and the final plate current. The oscillator screen-grid current being small in value, usually about 2 or 3 ma, it does not interfere with the reading of a plate-current dip in the oscillator plate-tank circuit.

Class-C amplifiers are used to amplify the output of the oscillator because of their efficiency. If triode tubes are used, they must be neutralized to prevent undesirable self-oscillation.

Buffers are used to prevent a power stage from loading down the stages preceding it. For example, a buffer is inserted between an oscillator and a power amplifier to prevent the power stage from changing the operating frequency of the oscillator.

A frequency multiplier is used to raise the output frequency of a crystal oscillator. It operates at some multiple of the oscillator frequency. It can be a frequency doubler, tripler, or quadrupler.

Power amplifiers are used in the output stages of transmitters to strengthen the modulated carrier for transmission. Usually, they are operated class-C. Class-B can also be used in single-ended operation.

A class-C amplifier operates with its grid-bias several times cutoff, thus plate current flows for less than 180° of the input cycle.

When the grid circuit of an amplifier draws current, the power must be supplied by the previous stage.

The power required by the grid circuit of a tube is called the grid drive.

The cancellation of feedback voltage from the plate to grid of a tube is called <u>neutralization.</u>

The maximum amount of r-f energy is transferred from one stage to another when the interstage-coupling system includes a tuned circuit which accepts r-f oscillations at the desired frequencies, and rejects r-f at all other frequencies.

Frequency multipliers are usually r-f amplifiers with input and output circuits tuned to different frequencies harmonically related to each other.

The output circuit of the power amplifier must be properly matched to the input impedance of the transmission line.

The operation of a class-C amplifier depends upon the angle of plate-current flow which in turn depends on the grid bias and the amplitude of the grid-driving voltage.

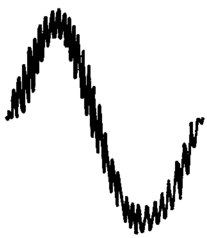
Keying circuits provide a means of periodically interrupting the transmission of r-f energy.

REVIEW QUESTIONS

1. What is meant by class-C amplification ?
2. Explain the term "grid drive".
3. What is the function of a buffer stage?
4. Explain briefly the operation of a frequency doubler.
5. Explain the need of neutralization.
6. What is the advantage of a push-pull power amplifier over a single-end type?
7. What is the function of a power-amplifier output-coupling network?
8. What are parasitic oscillations?
9. What is meant by keying?
10. Explain the operation of grid-bias keying.
11. What is link coupling?
12. What is cross neutralization?

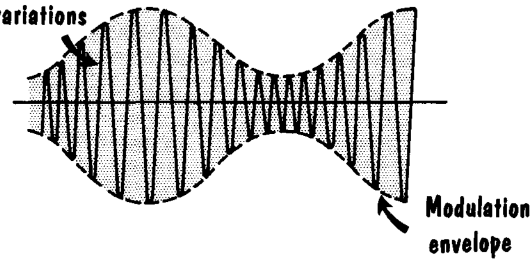
AM Transmission

RF and AF COMBINED IN LINEAR CIRCUIT



AUDIO WAVEFORM WITH R-F VARIATIONS

RF and AF COMBINED IN NONLINEAR CIRCUIT



AMPLITUDE-MODULATED R-F

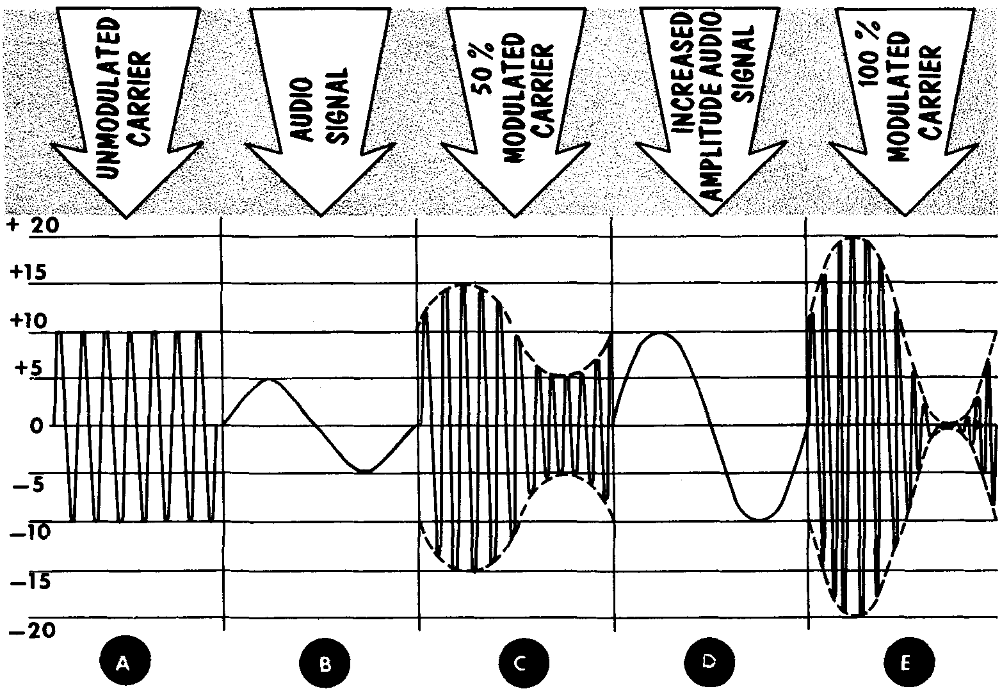
The word modulate as defined by a dictionary is "to vary in tone, inflection, pitch, or other quality of sound." In radio transmission we modulate or vary the r-f carrier. The unmodulated carrier is the r-f signal broadcast from a transmitter. It is called a carrier because it "carries" some form of intelligence. To have speech (or music) superimposed on the r-f carrier signal is to modulate the signal. There are several methods of modulating the carrier; the most popular, and the one we will discuss first, is that of varying the amplitude of the carrier. This is amplitude modulation, abbreviated as AM. The intelligence with which we will modulate the carrier will be an a-f sine wave. Actual speech and music frequencies are far more complex, but a sine wave permits a ready explanation of modulation which is identical to that caused by complex waveforms. To combine the carrier with the audio signal they are mixed together in special circuits discussed later. An important point is that the mixing can not be done with a linear device such as a pure resistance. When both the carrier and audio signals are applied to a pure resistance, they combine to have the audio signal vary about the carrier. The audio does not change the carrier amplitude; it merely swings the carrier up and down at the audio rate.

To have the audio signal vary the amplitude of the carrier, both must be combined in a nonlinear mixing circuit. The result is a combination in which the audio signal directly varies the amplitude of the carrier. The positive alternation of the audio signal adds to both the positive and negative alternations of the carrier; the negative alternation subtracts from both the positive and negative alternations of the carrier. This results in the carrier having two outlines of the audio signal. The upper outline is a duplicate of the original audio signal. The lower outline is also a duplicate, but 180° out of phase. Placing a dashed outline (called the envelope) on the modulated carrier shows up more clearly the outline of the audio signal.

AM Transmission (Contd.)

The carrier amplitude is important; but more so, is the audio signal's amplitude compared to the carrier's amplitude. The amount or degree of modulation with which the audio modulates the carrier is rated in percent.

In A, we show an unmodulated carrier with a peak amplitude level of ± 10 volts. In B, we have a modulating audio signal of ± 5 volts peak. C shows the resulting modulated carrier. Notice that where the +5-volt alternations modulate the carrier, the carrier increases in the same amount in both the positive and negative directions; similarly, where the -5-volt audio alternations modulate the carrier, the carrier decreases in the same amount in both directions. As a net result, the positive and negative peaks of the modulated carrier increase and decrease in value by 50%, or one-half of its normal value. Hence, C shows a 50% modulated-carrier signal. We now increase the audio signal to ± 10 volts, as shown in D. The modulated carrier is shown in E. Notice that the carrier's amplitude is doubled on the audio's positive peaks, and reduced to zero at the audio's negative peaks. Hence the modulated carrier increases and decreases in value by a full 100%. Therefore, the carrier is said to be 100% modulated.



MODULATION PERCENTAGE VARIES WITH STRENGTH OF AUDIO SIGNAL

Percentage of Modulation

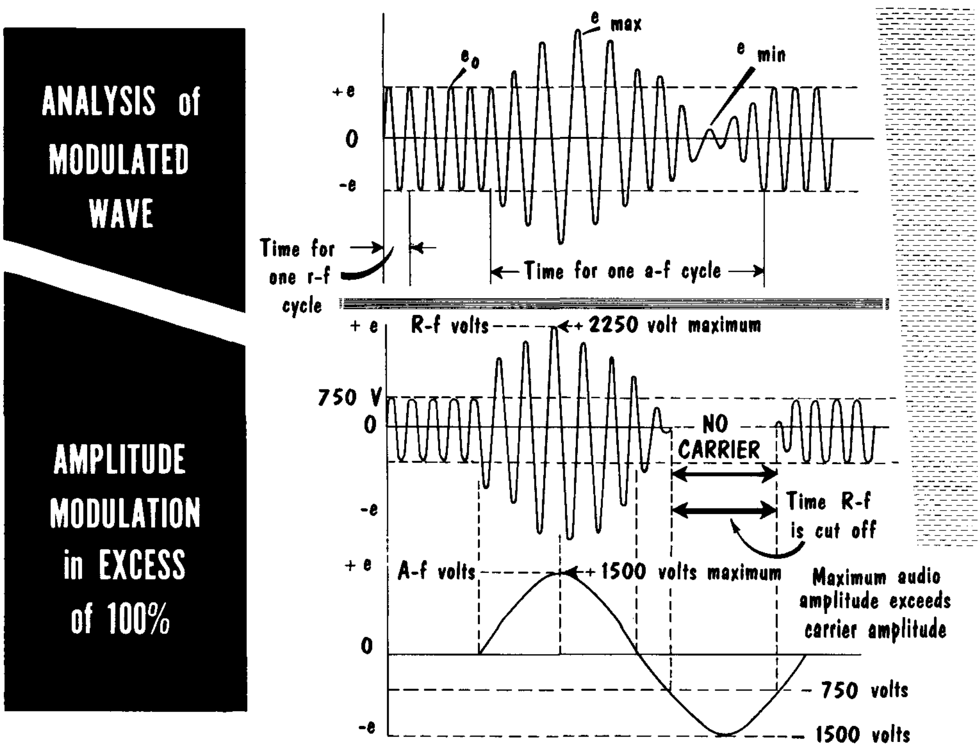
The degree of modulation in an AM wave is expressed by the percentage of maximum deviation from the normal amplitude of the r-f carrier. The effect of such a modulated wave, as measured by receiver response, is proportional to the degree, or percentage, of modulation.

The percentage of modulation may be determined by the equation:

$$\text{Percentage of modulation} = \frac{e_{max} - e_{min}}{2e_o} \times 100$$

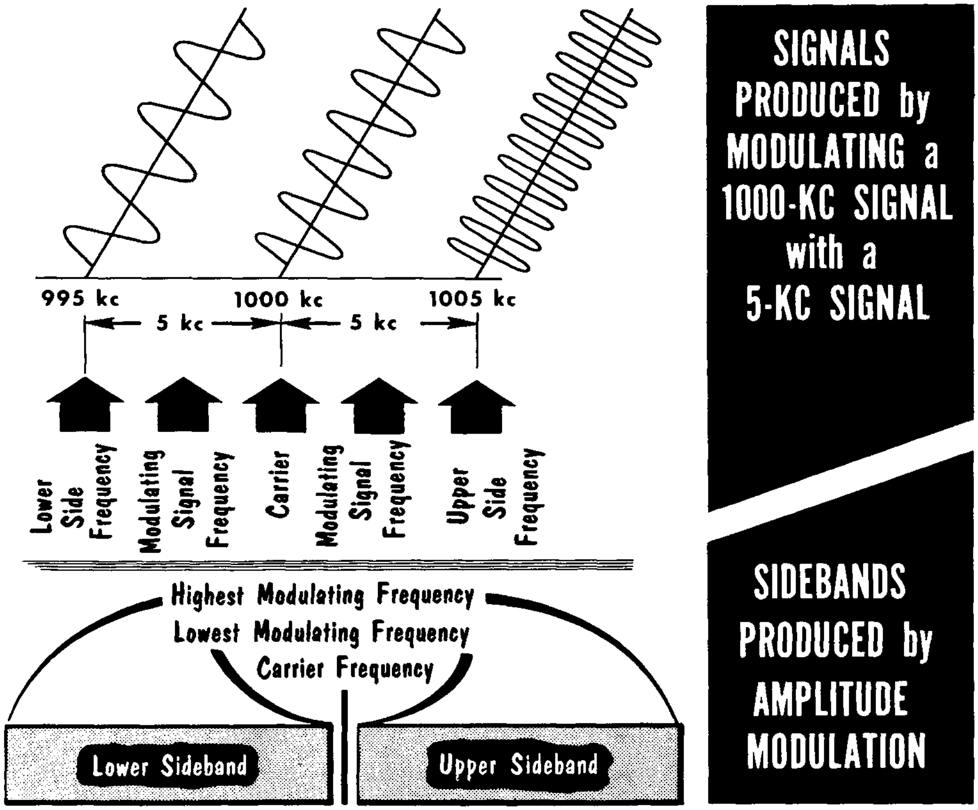
where e_{max} is the maximum instantaneous value of the r-f voltage, e_{min} the minimum instantaneous value of the r-f voltage, and e_o the maximum instantaneous value of the r-f voltage in the absence of modulation.

It is important that the amplitude be varied as much as possible, because the output of a detector in a radio receiver varies with the amplitude variations of the received signal. Thus a comparatively low-powered, but well-modulated, transmitter often produces a stronger signal at a given point than does a much higher-powered, but poorly modulated, transmitter located the same distance from the receiver. If modulation exceeds 100% there is an interval during the audio cycle when the carrier is removed completely from the air.



Sidebands

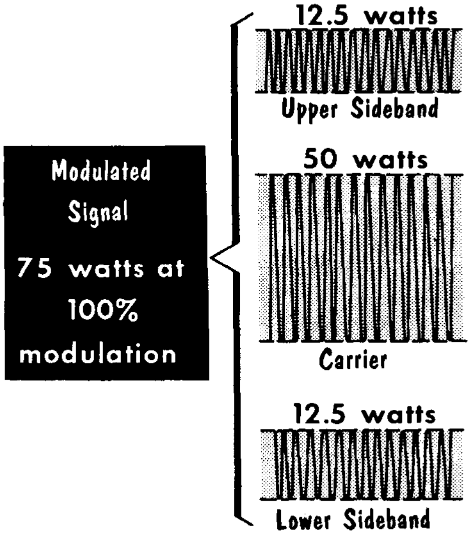
When an r-f carrier is modulated by a single audio note, two additional frequencies are produced. One is the upper frequency, which equals the sum of the frequency of the r-f carrier and the frequency of the audio note. The other frequency is the lower one, which equals the difference between the frequency of the r-f carrier and the audio note. The one higher than the carrier



frequency is the upper side frequency; the one lower is the lower side frequency. When the modulating signal is made up of complex tones, as in music, each individual frequency component of the modulating signal produces its own upper and lower side frequencies. These side frequencies occupy a band of frequencies lying between the carrier frequency, plus and minus the lowest modulating frequency, and the carrier frequency plus and minus the highest modulating frequency. The bands of frequencies containing the side frequencies are called sidebands. The sideband containing the sum of the carrier and the modulating frequencies is known as the upper sideband; the band containing the difference is known as the lower sideband. The space which a carrier and its associated sidebands occupy in a frequency spectrum is called a channel. The width of the channel (or bandwidth) is equal to twice the highest modulating frequency.

Power Distribution in an AM Wave

POWER DISTRIBUTION in an AM WAVE



$$P_{mod} = \left(1 + \frac{m^2}{2}\right) \times P_{car}$$

where

P_{mod} is the total power in the modulated wave

m is the degree of modulation

P_{car} is the power in the carrier

$$P_{mod} = \left(1 + \frac{1^2}{2}\right) \times 50 = \text{75 watts}$$

The power in an amplitude-modulated wave is divided between the carrier and the sidebands. The carrier power is constant (except in cases of over-modulation), and so the sideband power is the difference between the carrier power and the total power in the modulated wave. When a carrier is modulated by a single sinusoidal tone, the total power output is found from the formula shown. Assuming that a 50-watt carrier is modulated 100%, the power in the signal is 75 watts. Of this total, 50 watts are in the carrier and 25 watts are in the sidebands. The percentage of sideband power, 25/75 times 100%, equals 33.3%. Of the 25 watts of sideband power, there are 12.5 watts in each sideband, and the power content of each is therefore 16.6% of the total power output with 100% modulation.

The available sideband power takes a marked drop when the average percentage of modulation is well below 100%. This is shown by modulating the carrier only 50%, when the power in the carrier is 50 watts.

$$P\ mod = \left[1 + \frac{(0.5)^2}{2}\right] \times 50 = 56.25 \text{ watts}$$

The total modulated power is now 56.25 watts. Since 50 watts exist in the carrier, only 6.25 watts of power remain in the sidebands. Since 6.25 watts are one-fourth the value obtainable with 100% modulation, we see that reducing the modulation to 50% causes a 75% reduction in the available sideband power. Since all the intelligence being transmitted is contained in the sidebands, the desirability of a high percentage of modulation becomes evident.

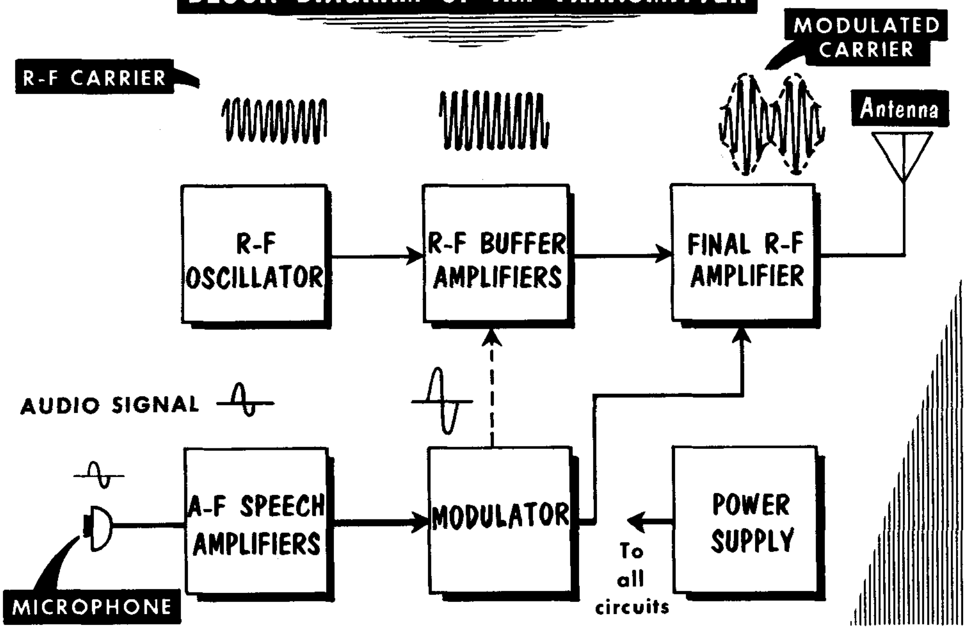
The Basic Amplitude-Modulated Transmitter

Referring to our diagram, we see that a feeble voice signal entering a microphone is amplified by several a-f speech amplifiers and also by the a-f modulator. The r-f oscillator produces the r-f carrier wave which is amplified by the r-f buffer amplifiers. The outputs of the a-f modulator and r-f buffer amplifiers are mixed in the final r-f amplifier to produce the modulated carrier wave.

Essentially, the r-f section of an amplitude-modulated transmitter consists of an r-f oscillator and several r-f amplifiers. In many cases, buffer amplifiers are used between the oscillator and the r-f amplifiers. As mentioned, buffer amplifiers isolate the oscillator from the following stages to minimize changes in oscillator frequency with changes in loading. Frequency multipliers are used to raise the oscillator frequency of the transmitter to the desired carrier frequency. It is desirable to have the oscillator operate at a comparatively low frequency for reasons of stability. Intermediate r-f amplifiers may be used to increase the driving power of the final r-f amplifier. The stage that the modulator feeds is known as the modulated r-f amplifier.

The modulator section is essentially an audio amplifier. The speech amplifiers are of the voltage type, with the modulator being basically a power amplifier. The modulator delivers the required amount of undistorted audio power to the modulated r-f amplifier. It may be operated class-AB, or class-B. To avoid distortion under these classes of operation, a push-pull stage must be used.

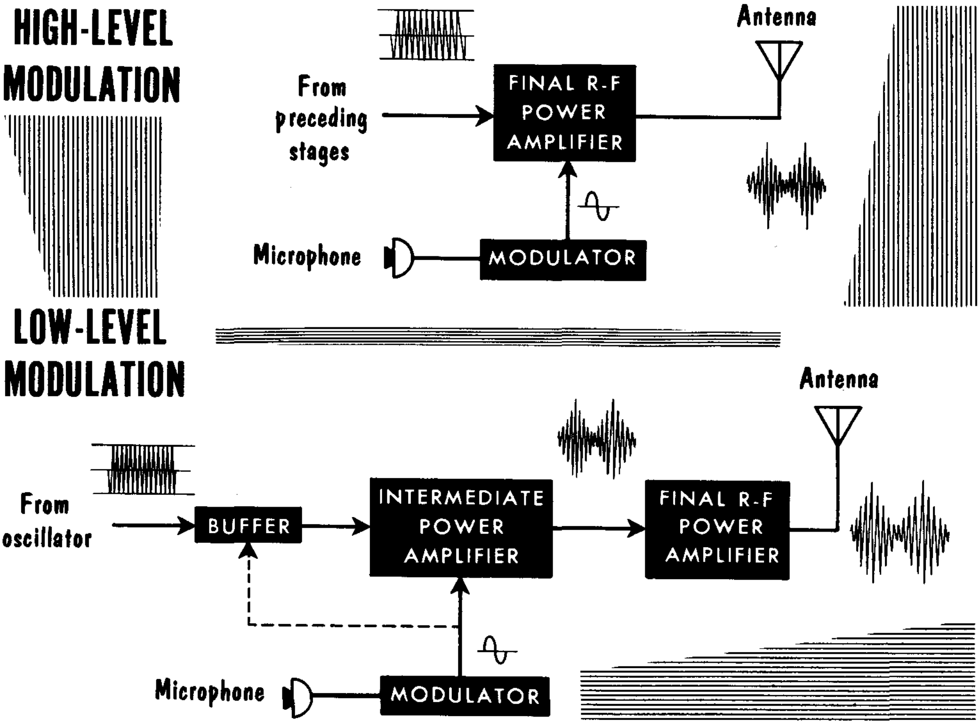
BLOCK DIAGRAM OF AM TRANSMITTER



Modulation Levels

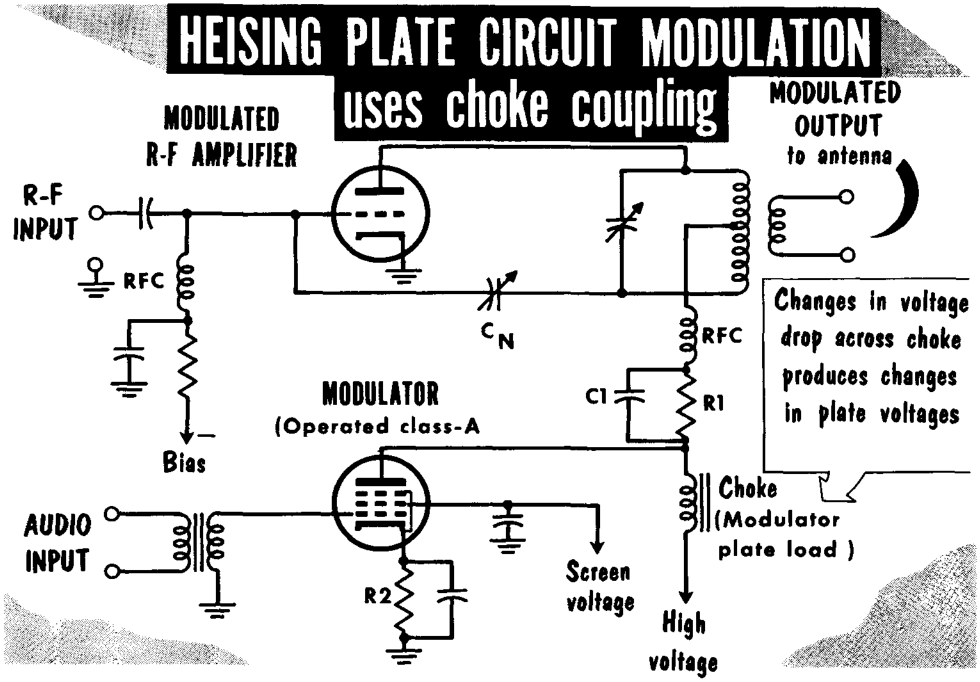
In high-level modulation the modulating voltage is applied to the final r-f amplifier. The stages preceding it need not be perfectly linear. Therefore they may be operated class-C with operating potentials adjusted for the desired circuit efficiency and gain. The final stage is always operated class-C. The overall efficiency of such a transmitter is high. A disadvantage of high-level modulation is that comparatively high audio power is needed, and several stages of voltage and power amplification may be required in the speech amplifier and modulator circuits.

In low-level modulation, modulation takes place in a buffer or intermediate power-amplifier stage, and modulating voltage is applied to a stage preceding the final amplifier. The r-f amplifiers which follow the modulated stage



must be operated so that their a-c output voltages are amplified, undistorted replicas of the modulated r-f voltages applied to their grids. Since little a-f power is required to modulate the carrier fully, the a-f section of the transmitter can be made comparatively simple. A disadvantage of this system of modulation is that the modulated stage must be followed by linear r-f amplifiers. Since lower efficiency is usually obtained from linear amplifiers, efficiency of a low-level modulated transmitter is low as compared with that of a high-level modulated transmitter using the same type of tubes and identical d-c operating voltages.

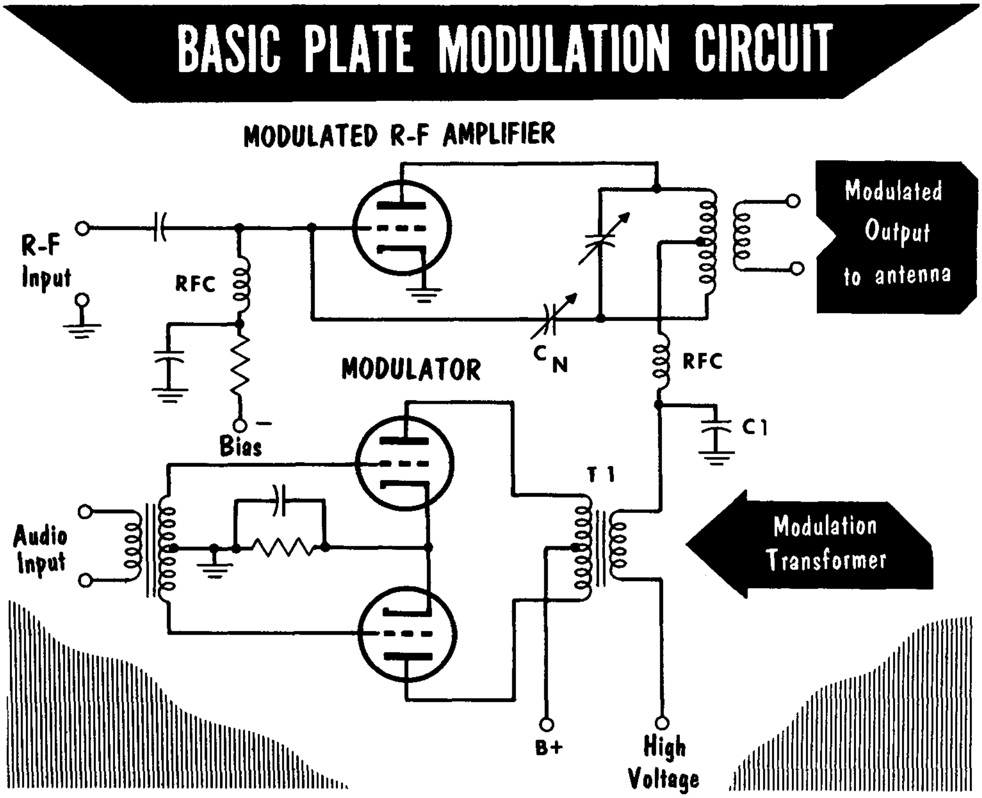
Heising Modulation



A method that uses a choke as modulator plate load is called Heising modulation, named after its inventor. In the basic circuit shown, the modulator tube is operated class-A. The choke used as modulator plate load provides maximum impedance at audio frequencies. Plate voltages for the modulator and the r-f amplifier are supplied from a single source. If, due to an applied signal input, the modulator grid voltage swings in a positive direction, the modulator plate current increases. Hence, the plate voltage decreases. Because of the manner of circuit connection, the r-f amplifier's plate voltage is also reduced. If the modulator grid voltage swings in a negative direction, the modulator plate current decreases, and the modulator and r-f amplifier plate voltages increase. The modulator input signal thus controls the r-f amplifier plate voltage, and causes it to vary in step with the modulating signal.

We can achieve 100% modulation only if the modulator plate-voltage swing equals the r-f amplifier plate voltage. Thus, with a positive-going modulating signal the r-f amplifier plate voltage is doubled; a negative-going plate voltage reduces the r-f amplifier plate voltage to zero. The modulator plate voltage must therefore swing to zero. Any attempt to approach this condition would result in a severely distorted signal. To overcome this difficulty, the modulator plate voltage is increased to a value half that of the r-f amplifier by placing R1 in series with the r-f amplifier plate. Bypass capacitor C1 prevents R1 from acting as a plate load. R2 protects the grid circuit from drawing excess plate current when the modulator input signal drives the modulator grid positive.

Transformer Coupling (Plate Modulation)



Another, and more popular method of plate-circuit modulation using transformer coupling is better known as plate modulation. The modulator output is coupled through the secondary of T1 which is connected in series with the plate circuit of the r-f amplifier. The audio signal developed across the secondary is in series with the r-f amplifier d-c plate voltage. Therefore, the audio signal adds to or subtracts from the d-c plate voltage, thus modulating the r-f amplifier. C1 is a bypass capacitor that presents a low-impedance path to ground for any r-f, and a high reactance to a-f.

T1 requires special design. The primary must match the modulator plate impedance. The secondary impedance must permit a voltage swing that can provide 100% modulation. The transformer action permits 100% modulation without a voltage-dropping resistor such as is used in Heising modulation. Transformer coupling also lends itself to the use of high-efficiency circuits of class-AB and -B modulators. When a tetrode or a pentode is used as an r-f amplifier, the screen grid acts as the main attraction for the electron stream, the plate being merely the collector of electrons. Since modulation of the plate only would be ineffective, the screen grid and the plate circuit are modulated. The screen receives its d-c voltage through R1, but the same modulated voltage is applied to the plate as well as to the screen.

Control-Grid Modulation

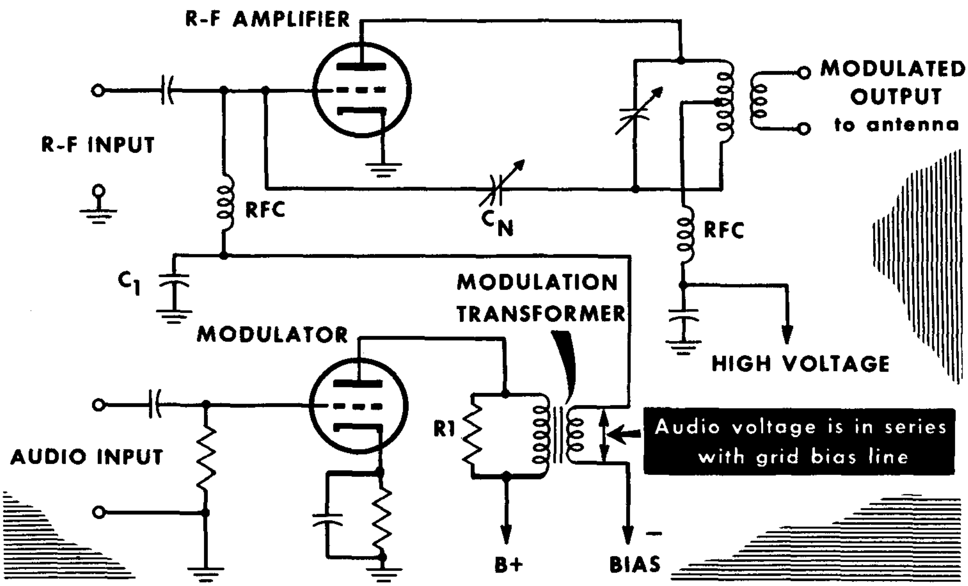
Economy and space requirements often dictate the use of modulating systems that require less power than is used in plate modulation. For example, in plate modulation, a 200-watt transmitter requires a 100-watt modulator for 100% modulation. By modulating the r-f amplifier through the control grid circuit much less audio power is required; the result is a saving in space and weight which is of great advantage in aircraft and mobile applications.

In the control-grid modulated r-f amplifier circuit shown, the audio modulating signal voltage is placed in series with the control-grid bias voltage. The modulating signal thus adds or subtracts from the grid-bias voltage. Distortion is kept to a minimum by keeping the load imposed on the modulator as steady as possible. Therefore, the modulator is often designed to provide twice the power necessary. The excess power is dissipated in resistor R1, placed across the modulation transformer primary. At the same time, R1 provides a more constant modulator load. An r-f choke and a bypass capacitor C1, prevent the r-f from reaching the modulator circuit. C1 is carefully chosen to bypass the r-f, yet offer a high impedance to the modulating a-f.

Only a small amount of audio power is required to modulate the control grid. Grid modulation, however, operates at reduced efficiency, and reduced power output, and there is difficulty in obtaining distortion-free output. The power output of the modulator can be small since it is only necessary to vary the negative grid bias slightly.

CONTROL GRID (GRID BIAS) MODULATION

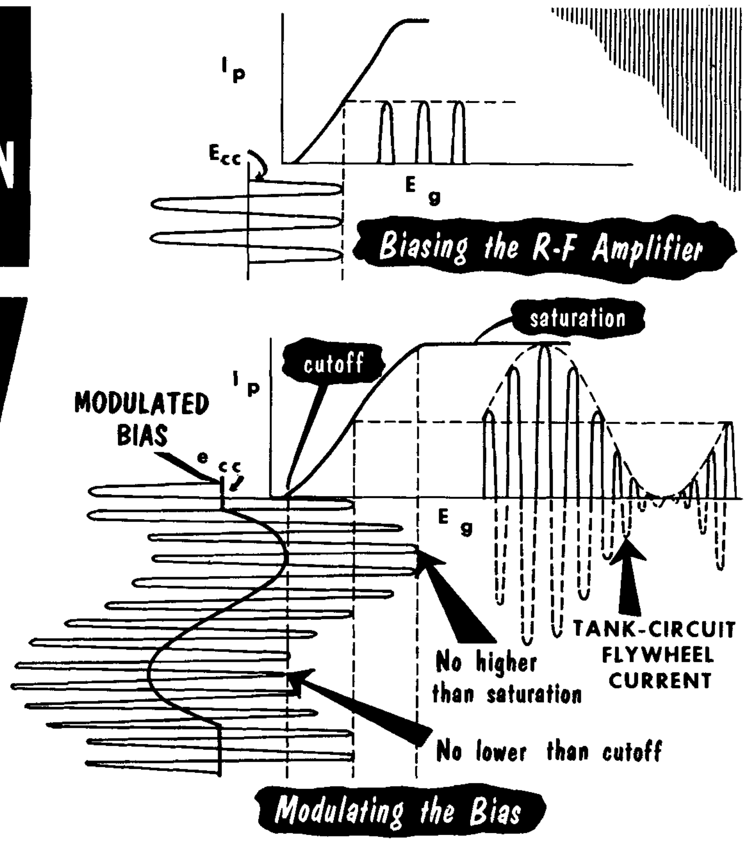
Comparatively little Audio Power is Required for Modulation



Control-Grid Modulation (Contd.)

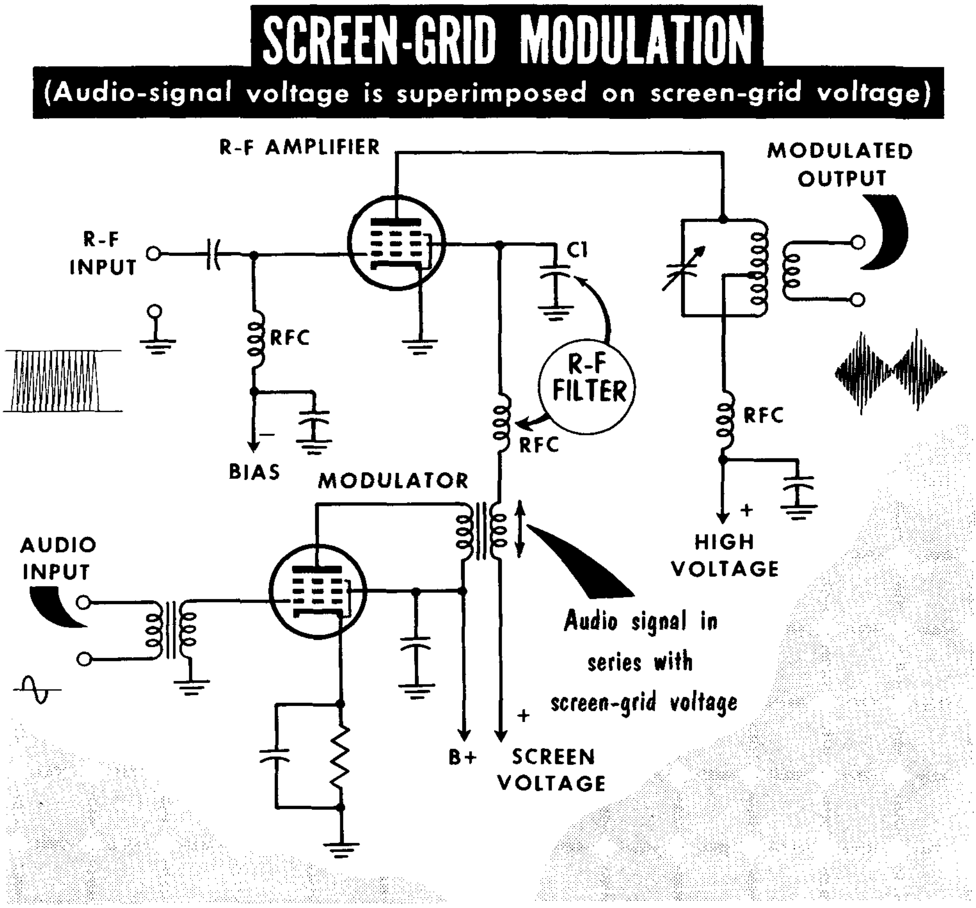
In the control-grid modulated r-f amplifier, the control-grid d-c bias is carefully adjusted so that the peaks of the carrier do not extend past the linear portion of the grid voltage – plate current curve, as shown. Unless the tube is operated on the linear portion of the curve, distortion results. The modulating signal superimposed upon the carrier is limited to a value which causes the carrier signal peaks to go no lower than cutoff, and no higher than saturation. Since the d-c grid bias is limited, the r-f amplifier does not operate at peak efficiency. With no modulating signal applied, the plate-current pulses are only one-half the maximum power available. As a result, the plate efficiency equals approximately 30%. When the modulating signal is applied, as shown, at the maximum positive peak, the plate-current

GRID-BIAS MODULATION



pulses, for an instant, provide an efficiency of 75%. Hence, the plate efficiency of a grid-modulated stage varies from 30% (no modulation) to 75% (at 100% modulation). The increased efficiency provides the power in the side-bands. The required modulator power is very small; usually a voltage-amplifier stage is sufficient. Should the r-f amplifier draw a grid current, the power required of the modulator is still small. For example, a 1000-watt r-f amplifier can be modulated with 25 watts of audio power.

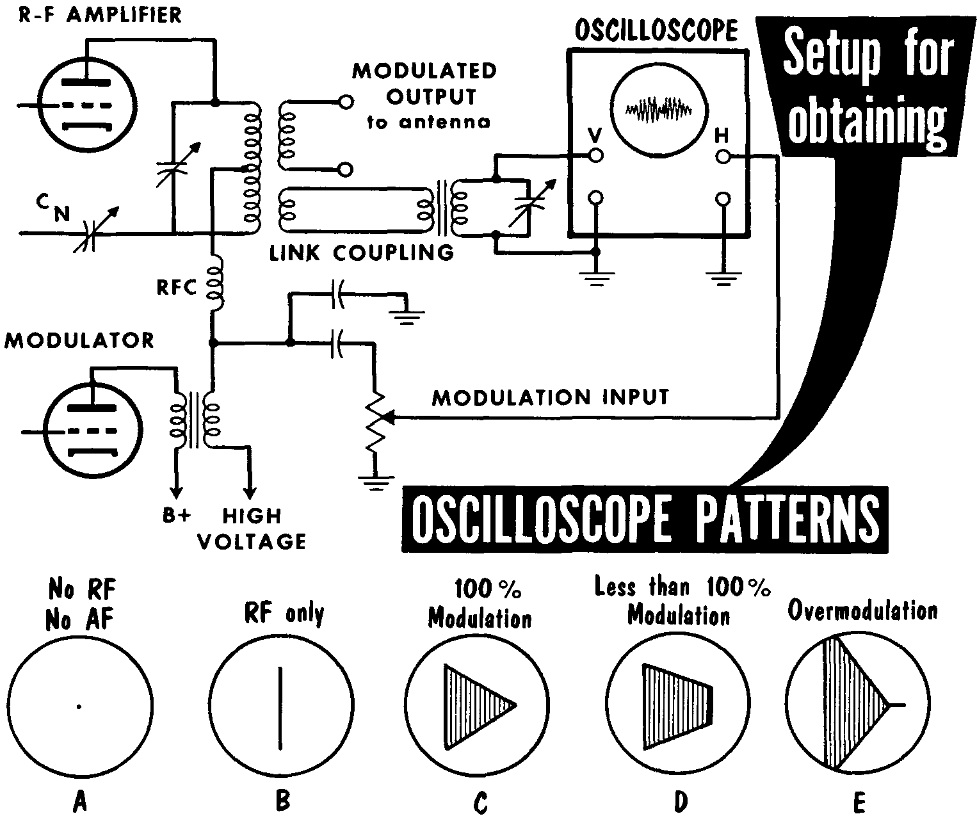
Screen-Grid Modulation



Screen-grid modulation also requires very little audio power. The modulating voltage and the screen-grid voltage are placed in series, as shown. Hence the modulating signal is superimposed upon the screen-grid voltage. The impedance offered by the screen-grid circuit varies, resulting in nonlinear modulation. For voice communication, the amount of nonlinearity is tolerable. Special degenerative feedback circuits are sometimes employed to improve the linearity. A problem exists in that with zero screen-grid current, plate current still flows. To obtain full modulation on negative peaks, the screen-grid voltage must often be driven to a negative value to obtain complete plate-current cutoff. In this circuit, the a-f power required for modulation is approximately one-fourth the power input to the screen under CW conditions. The peak audio voltage is approximately equal to the d-c screen voltage which is adjusted to one-half the value used for CW operation.

Screen-grid protection is obtained with a clamp tube which maintains a near-constant screen-grid voltage. The audio signal applied to the clamp tube modulates the screen-grid voltage.

Modulation Checking



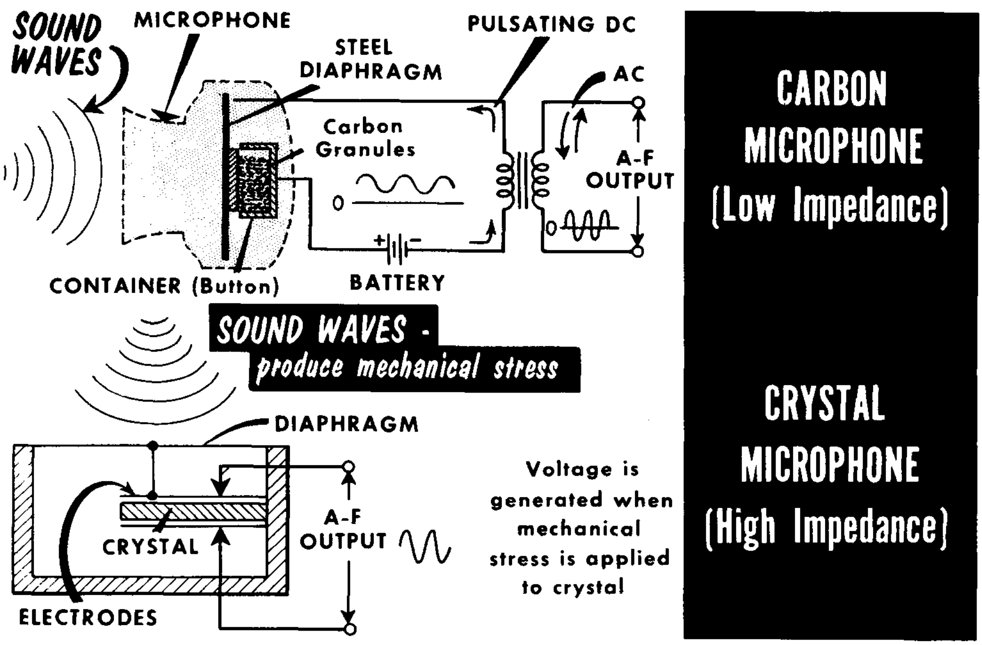
An oscilloscope can be used to present a visual picture of the modulated transmitter output. To obtain the popularly used "trapezoidal" pattern, the vertical plates of the oscilloscope are coupled to the output-tank circuit, and the horizontal plates are capacitively coupled to the modulating signal source, using a voltage divider, for adjusting the size of the display pattern. The input circuit to the vertical plates of the scope is tuned to eliminate unwanted harmonics.

In the absence of both excitation and modulation, the undeflected spot appears in the center of the screen (A). When excitation is applied, the spot swings upward and downward in step with the carrier-voltage variations, producing a vertical line on the scope screen (B). The length of the line represents the peak-to-peak amplitude of the carrier. When modulation is applied, the trapezoid pattern (C) is produced. This represents 100% modulation of a correctly adjusted transmitter. Diagrams (D) and (E) show a lower modulation percentage, and overmodulation, respectively. When the carrier is modulated 100%, the wide end of the pattern will be just twice the height of the carrier amplitude line of (B), while the narrow end will come to a point (C).

Microphones – Carbon and Crystal

The function of a microphone is to convert the variations in air pressure produced by the human voice or a musical instrument into an electrical voltage or current of the same frequency and corresponding amplitude. Microphones are rated in terms of frequency response, sensitivity, directivity, and impedance. A good broadcast-type microphone should have a frequency response of from 30 to 10,000 cycles, or higher. For general communications work, a frequency response of 75 to 4500 cycles would prove satisfactory.

Probably the most commonly used microphone is the carbon type. Sound waves striking the diaphragm set up vibrations which vary the pressure on the button, and thus vary the pressure on the carbon granules. This varies the d-c resistance of the carbon-granule pile. Since the granules are in series with a battery and the primary of a microphone transformer, the



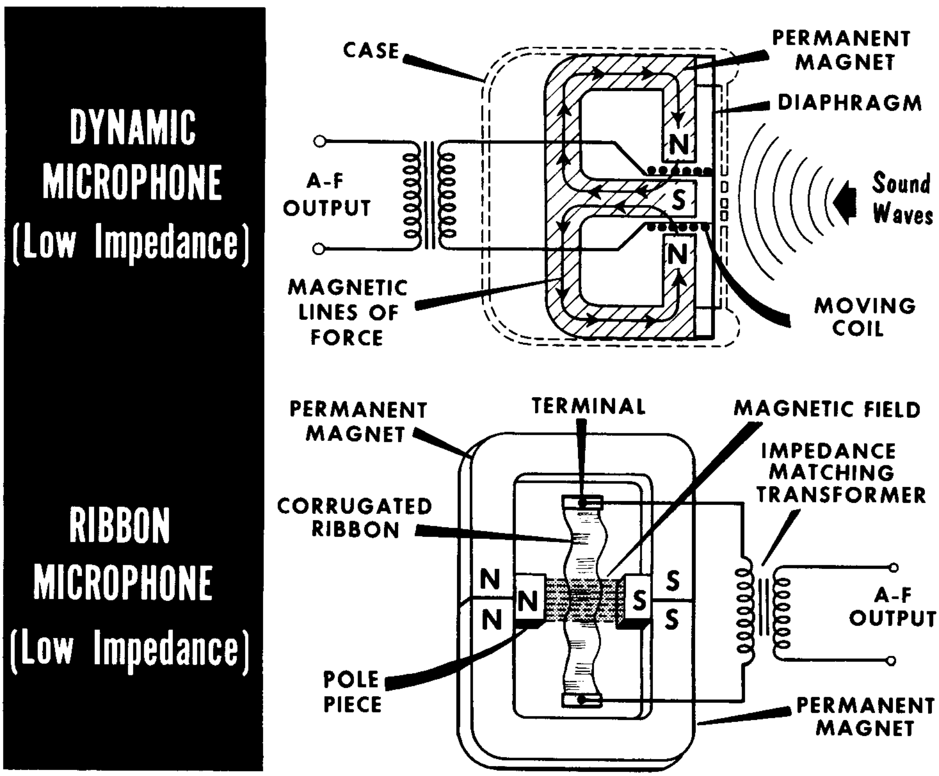
changing resistance produces a corresponding change in the circuit current. The resulting pulsating dc in the primary produces an alternating voltage in the secondary. The transformer is used to step up the voltage as well as match the impedance of the low-impedance microphone with the high-impedance grid circuit. Microphone current may be from 10 to 100 ma, and the resistance may vary from 50 to 90 ohms.

The crystal microphone uses a property of certain crystals such as Rochelle salts, known as the piezoelectric effect. The bending of the crystal, resulting from the pressure of the sound wave, produces an emf across the faces of the crystal. This emf is applied to the input of an amplifier.

Microphones: Dynamic and Ribbon

A dynamic or moving-coil type microphone consists basically of a coil of fine wire fastened rigidly to the back of a diaphragm so that it is suspended in the field of a strong permanent magnet. When sound waves make the diaphragm vibrate, the coil moves, cutting the magnetic lines of force of the magnet at the audio rate. This induces in the coil a voltage that is the electrical equivalent of the sound waves. The dynamic microphone is very sensitive, lightweight, and requires no external voltage. It is rugged, and has an excellent frequency response (a broadcast type may have a response of from 40 to 15,000 cycles); typical impedances are 50 to 500 ohms, thus requiring an impedance-matching transformer for connection to an amplifier.

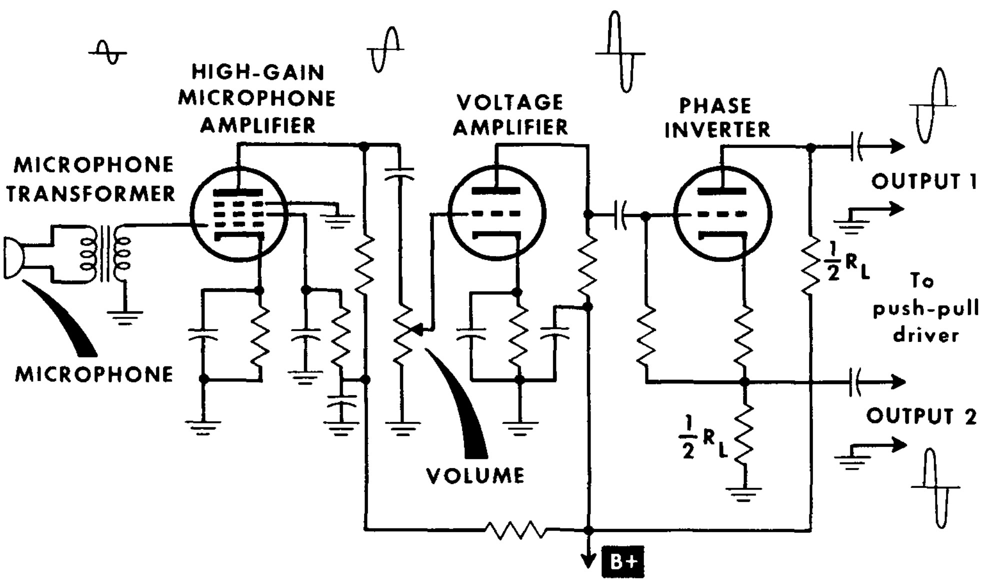
The ribbon or velocity microphone is a variation of the dynamic. It has a thin, lightweight, flexible, corrugated, metallic strip suspended between the poles of a permanent magnet. Sound waves make the strip vibrate in the magnetic field. It cuts the lines of force, and voltage is induced in it proportional to the frequency and strength of the sound waves. The velocity microphone has good frequency response, but responds only to those sounds originating directly in front of it. Since the ribbon resistance is so low (less than 1 ohm) this microphone usually has a built-in transformer which raises its impedance to 250-500 ohms. The voltage output is very low, and leads between microphone and amplifier must be shielded to avoid hum pickup.



Speech Amplifiers

The speech amplifier is used in a radiotelephone transmitter whenever the output of a microphone is lower than the signal voltage necessary to drive the modulator stage. Therefore, the speech amplifier is considered as all

"Voltage Amplifier" Portion of Speech Amplifier



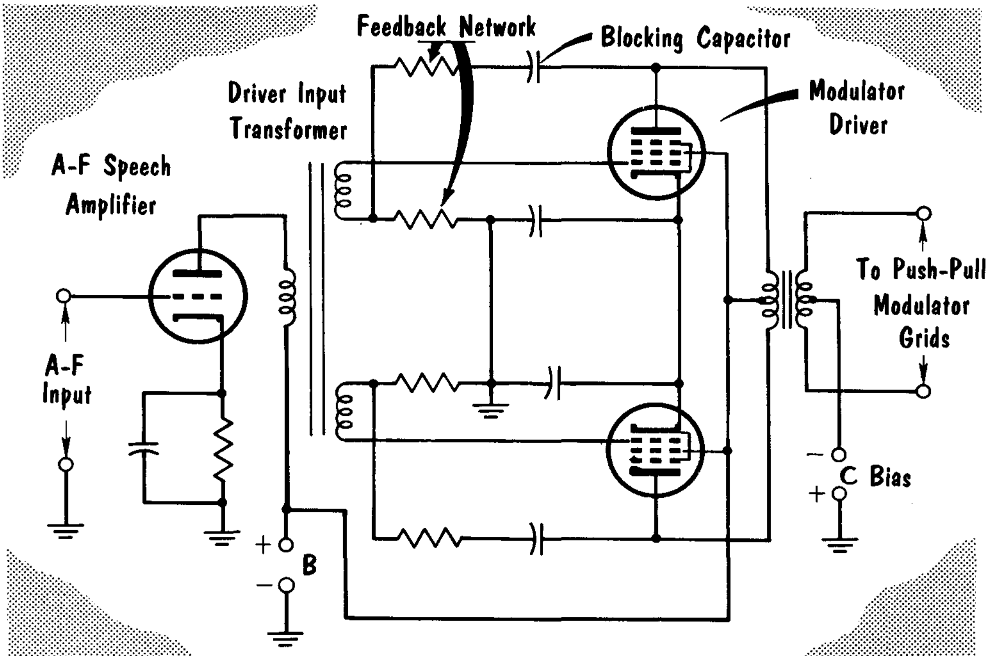
the a-f amplifier stages between the microphone and the input of the stage whose output actually modulates the r-f carrier. When a class-B modulator is used, the speech amplifier must include a power amplifier to supply power to the class-B amplifier grids. The power amplifier preceding the modulator is called the driver.

Audio amplifiers are designed to deliver either as much power or as much voltage as possible into a load impedance. When a large power output is desired, the stage can be operated class-A, -AB, or -B. When voltage gain is the prime requirement, amplifiers are usually operated class-A because of the low distortion. High voltage gain with low distortion is important in the design of low-level stages of a speech amplifier.

The speech amplifier must supply a peak voltage equal to the value of the d-c bias on the grid of the last class-A amplifier if it is a single-ended stage, and twice the d-c bias if the last stage is operated in push-pull. In actual practice, the speech amplifier must provide from 25% to 100% more voltage gain than needed to meet the requirements at the grid of the last a-f speech amplifier. The added gain compensates for various circuit losses.

Modulator Driver

The relationship between instantaneous grid voltage and grid current in class-B modulators is not linear, and the grids present a varying impedance to the driver over the a-f cycle. To avoid distortion in the modulator caused by the change in grid impedance, the modulator driver must supply a constant voltage to the modulator grids, regardless of the change in grid impedance. To meet these requirements a driver must have good voltage



TYPICAL MODULATOR DRIVER

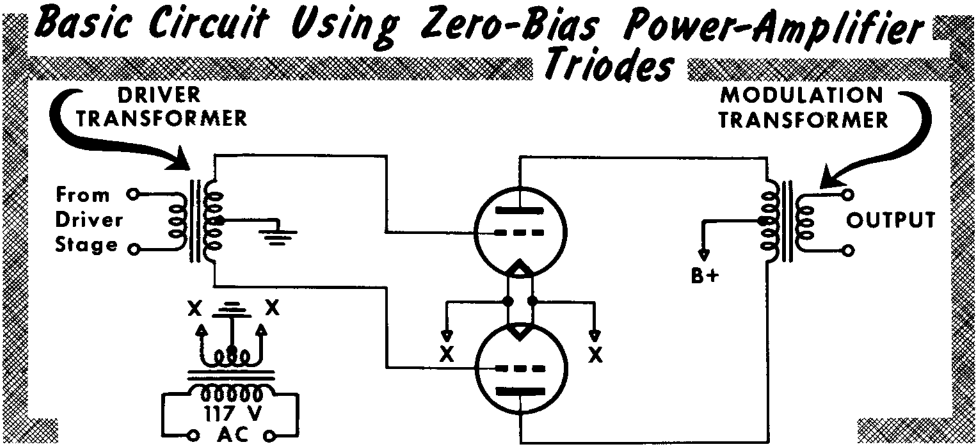
regulation. This is obtained by having a low internal resistance in the driver stage. In turn, this means that the modulator-driver tubes must have a low value of plate impedance. Low-mu triodes are best for this condition. Pentodes and tetrodes can be used if sufficient inverse feedback is used to lower the effective plate impedance.

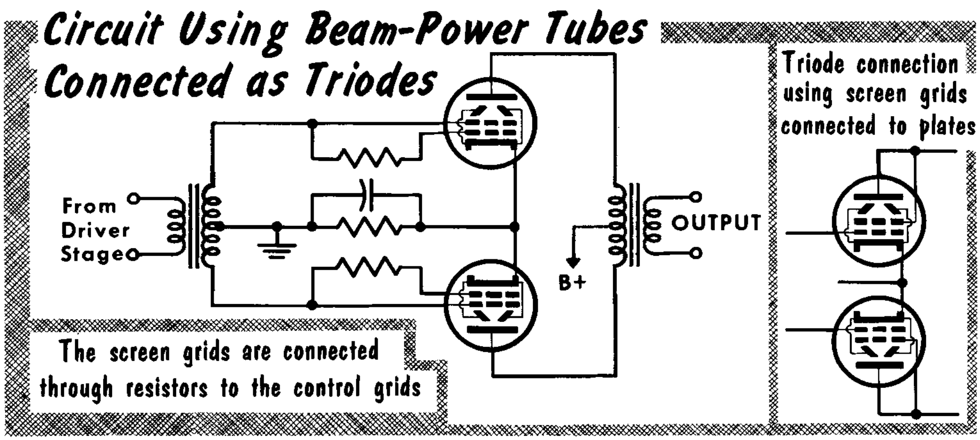
To obtain maximum power transfer from driver to modulator, we must match the higher driver plate impedance to the relatively low modulator input impedance. This is done by using a driver transformer having the highest possible voltage step-down between its primary and secondary. In this case, the plate resistance of the driver tubes, as seen by the modulator grids, is comparatively high. If high-mu triodes are used as class-B modulators, they can be operated with very little or no d-c biasing voltage on their grids. This reduces the variation in grid impedance over the a-f cycle, and gives the driver a more constant input-impedance load. Distortion is thus reduced. Tubes operated in this way are often called zero bias tubes.

Modulator Stages

To obtain sufficient power from a modulator stage, class-B operated push-pull amplifiers are most commonly used. Class-AB$_1$ and -AB$_2$ circuits are also found. The tubes in the modulator circuits are often of the same type as those used for r-f power-amplifier stages. The amount of audio power required determines the selection of a tube type for a specific application.

Plate modulation requires the modulating power to be somewhat higher than one-half the r-f amplifier power. The additional power covers any possible loss due to the coupling transformer, and readily permits 100% modulation.



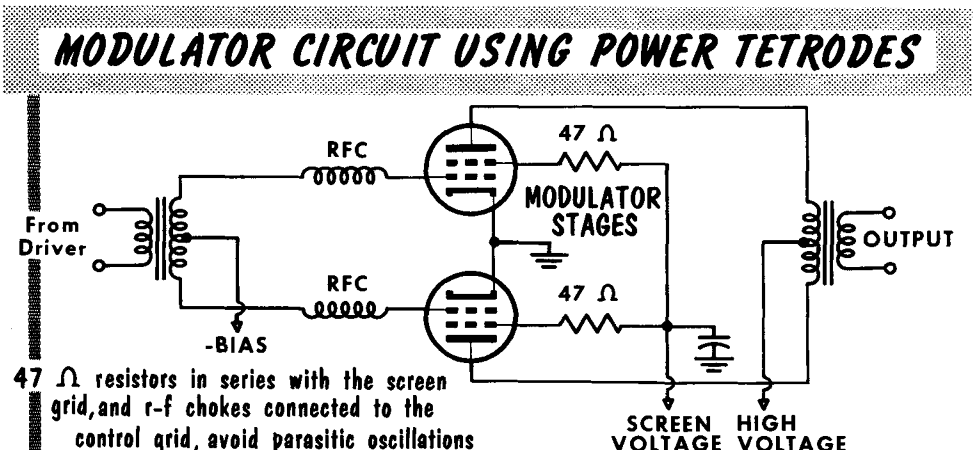


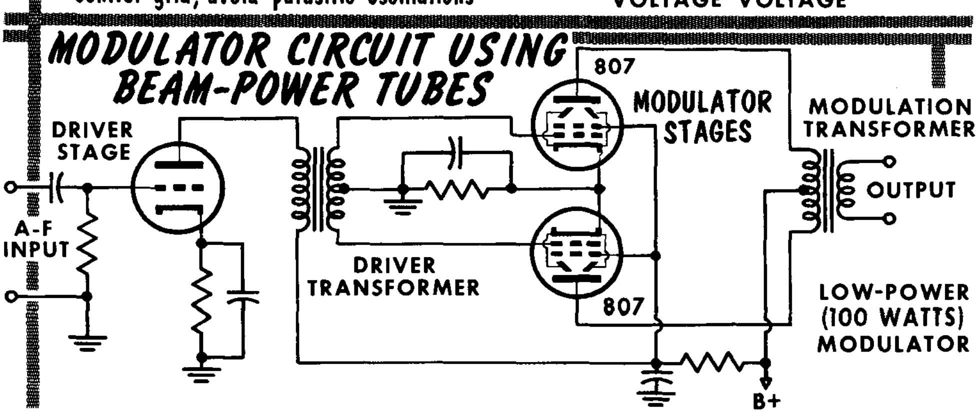
Special power-amplifier tubes are designed for class-B operation with zero grid bias. Then the grid current is so small that the tube may be considered as cut off. Specially designed power-amplifier triodes, tetrodes, and beam-power pentodes may be used. Beam-power pentodes and tetrodes may be used as high-mu triodes. The input signal is applied directly to the screen grids, which are connected through resistors to the control grids.

Modulator Stages (Contd.)

In the modulator circuit shown which uses a pair of power tetrodes, small resistors are connected in series with the screen grids; r-f chokes are connected to the control grids to avoid parasitic oscillations. In the low-power modulator circuit a voltage amplifier feeds a pair of beam-power pentodes. The modulation transformer is designed to match the plate-to-plate load impedance of the modulator tubes in the primary, and the load impedance of the r-f amplifier stage in the secondary. Commercial modulation transformers are available to match most of the popular tube combinations. For unusual tube combinations special transformers (universal) with numerous taps on both the primary and secondary permit impedance matching.

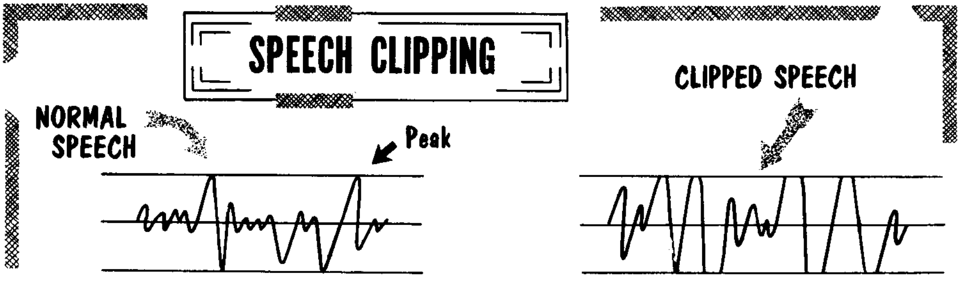
As an important special precaution, the modulation transformer secondary should always be under load. A dummy load of correct wattage and resistance is sufficient, for the r-f amplifier to be operating and drawing the correct amount of plate current. When no load is connected across the modulation transformer secondary, the primary impedance rises to a very high value, and develops unusually large audio-signal voltages across the primary which may break down the transformer insulation, causing a short-circuit.

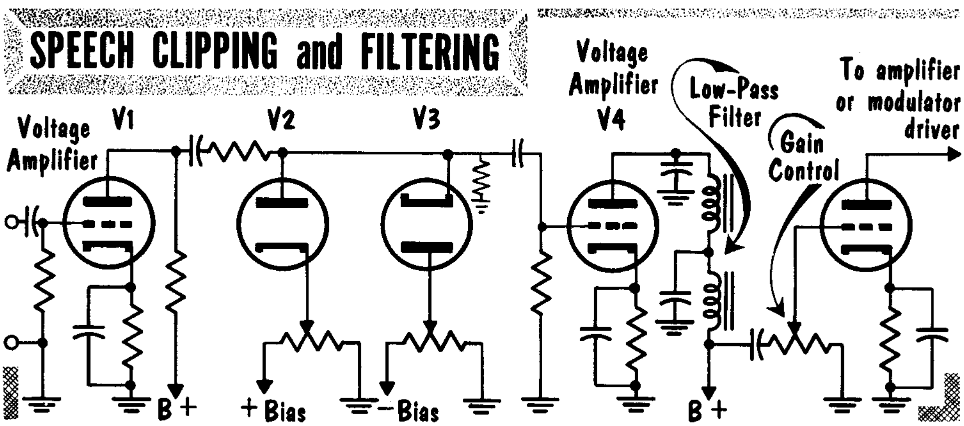




Omega Drafting & Art Services

Speech Clipping and Filtering



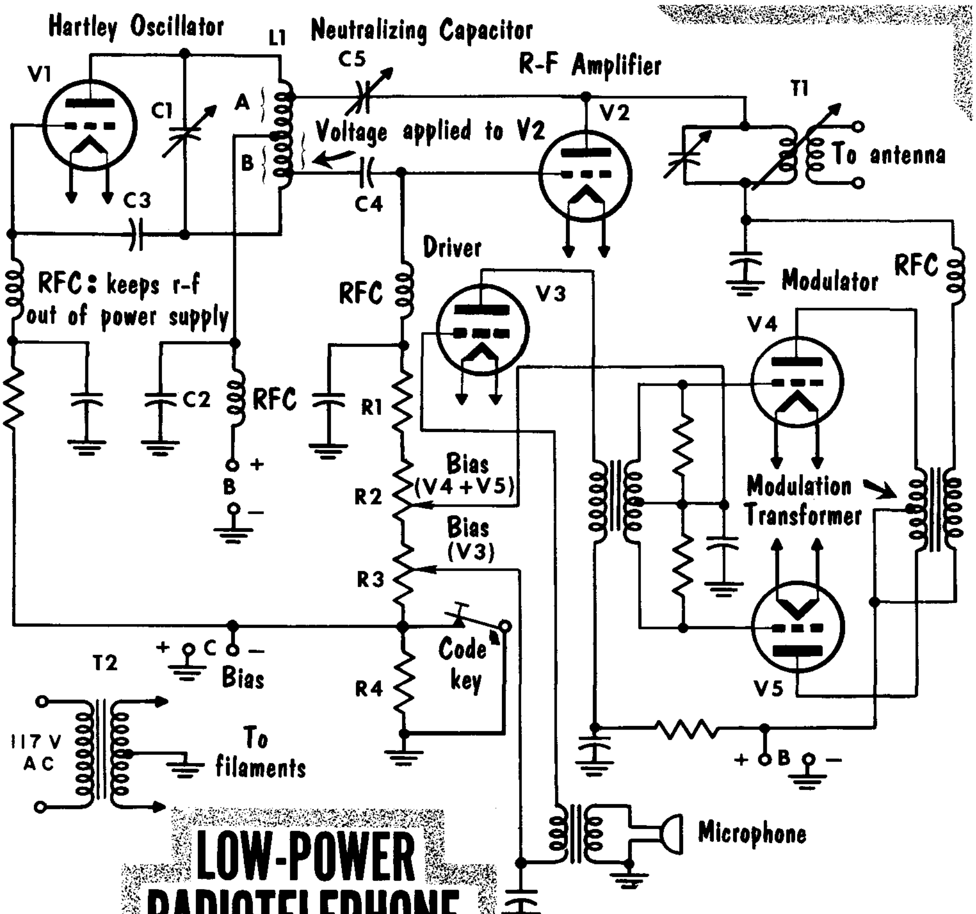


Complex speech waveforms contain much less average power than a sine wave of same amplitude. The high audio peaks that cause 100% modulation are few, and widely spaced. Increasing the amplitude of the complex waveform, and clipping the top and bottom of the peaks, prevents the amplitude exceeding a given value, and permits a power increase of the same speech waveform without causing overmodulation. There is a slight distortion of the speech waveform, but reduction in intelligibility is negligible. Clipping itself does not prevent overmodulation because clipping, like overmodulation, results in high-frequency sidebands called splatter. Therefore, a filter follows the clipper circuit and prevents frequencies of approximately 3000 cycles and up from passing through the following amplifiers.

The clipping level is usually adjusted. For example, we may clip a peak-to-peak signal of 10 volts to a level of 2 volts. The clipped 2-volt signal can then be amplified to 10 volts, resulting in a tremendous power increase. In the basic filter-clipper circuit shown, the amplifier output of V1 is applied to diodes V2 and V3. V2 is connected to conduct on the positive half of the audio signal. The bias control in the cathode of V2 sets the level at which the positive half of the signal is clipped or "clamped". V3 is connected to conduct on the negative half of the audio signal. The bias control in the plate of V3 clamps the negative half of the signal to the desired level. The clipped signal is then applied through amplifier V4 to an L-C low-pass filter, and then to the high-level amplifier circuits.

Basic CW-AM Transmitter

The transmitter shown uses a self-excited oscillator with L1-C1 making up the oscillator-tank circuit. Plate section A of L1 is between plate and filament, through C2, to ground and the centertap of T2; grid section B is between grid and filament, C3, C2, ground, and the centertap of T2. Coupling capacitor C4 is tapped on L1 to minimize oscillator loading. V2 bias is developed across R1-R2-R3 when key is closed. Oscillation in V2 is prevented by varying C5 to cancel out-of-phase voltages on grid. When the code key is open, negative voltage across R4 cuts off V2. V3 is the driver for class-B modulators V4 and V5. These stages are biased by the voltage developed by grid current flow in V2. R-f transformer T1 is the r-f modulated amplifier-tank coil and a coupling coil which goes to the antenna. Coupling is variable for correct load matching between amplifier and antenna.



LOW-POWER RADIOTELEPHONE-RADIOTELEGRAPH TRANSMITTER

Speech amplifiers are usually operated class-A to keep distortion at a minimum. High voltage gain, low circuit noise, and low hum pickup are desirable.

Three principal types of microphones are used with speech amplifiers: the carbon, the crystal, and the magnetic. The carbon microphone has the highest output and the lowest fidelity; the magnetic has the lowest output and the highest fidelity.

The modulator builds up the audio signal and modulates the r-f carrier with its output. It is usually operated class-B for relatively high power and fidelity.

If the audio-signal output of the modulator is injected into the final power amplifier, the process is called high-level modulation. If it is injected into some preceding stage, the process is called low-level modulation. The audio signal can be injected at any grid, or at the plate of the modulation stage.

In amplitude modulation, intelligence is superimposed on the carrier wave by causing the amplitude of the carrier to vary in accordance with the modulating signal.

The percentage of modulation is a measure of the depth or degree of modulation of a carrier wave. It is dependent upon the amplitude of the modulation envelope as compared with the amplitude of the carrier.

If the amplitude of the modulation envelope is twice as great as the amplitude of the carrier, then the modulated wave formed is said to be full, or 100% modulated.

When an r-f carrier is modulated, sidebands are produced on either side of the carrier frequency.

Sidebands are equal in frequency to the carrier, plus or minus the modulating frequency.

The power in an amplitude-modulating wave is divided between the carrier and the sidebands.

Plate modulation is most commonly used for high-level amplitude modulation.

In plate modulation the modulating voltage is impressed on the d-c supply voltage of one of the r-f amplifiers of the transmitter.

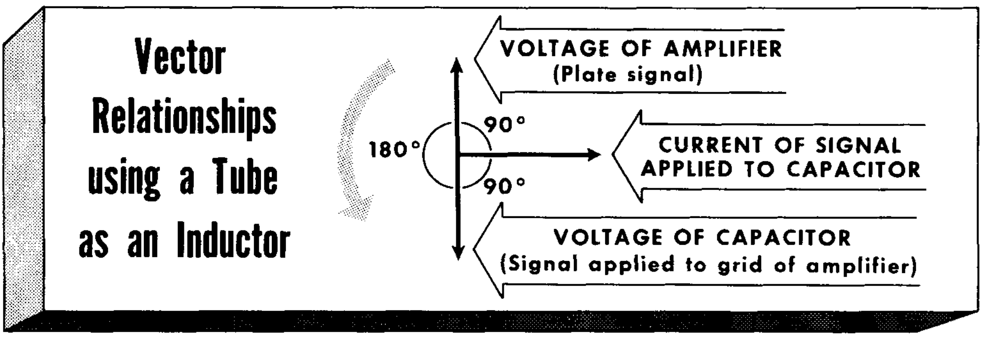
A speech amplifier is needed in a radiotelephone transmitter whenever the output of the microphone is lower than the signal voltage required to drive the modulator tube.

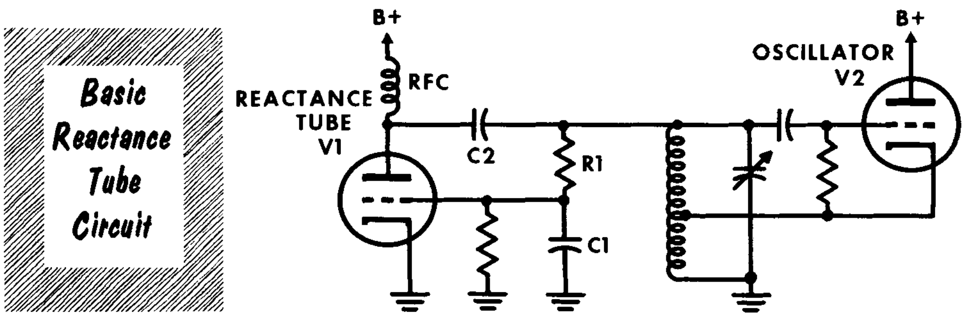
REVIEW QUESTIONS

1. What is meant by amplitude modulation?
2. What is meant by percentage of modulation?
3. What function do sidebands serve?
4. Explain Heising modulation.
5. What is meant by high-and low-level modulation?
6. Explain control-grid modulation.
7. Describe the operation of a carbon microphone.
8. Describe the operation of a dynamic microphone.
9. What circuits are included in the speech amplifier?
10. What is the function of the modulator drive stage?
11. What class of operation can be used in the modulator stage?
12. Describe a simple test procedure checking modulation.

Direct FM

The two most popular methods of developing an FM signal are by the direct and indirect methods. In the direct method an oscillator is modulated directly by the audio signal. In the indirect method, phase modulation is used to vary the carrier to create an FM signal. The simplest method of producing an FM signal is to have the modulating signal vary the value of L or C in an oscillator, causing the frequency to vary. This basic concept is illustrated by





having a capacitor microphone placed directly across the tuning capacitor of a Hartley oscillator. The oscillator frequency is determined by the total L and C in the tank circuit. When sound waves strike the movable plate of the microphone, the capacitance will increase and decrease in accordance with the sound wave. More capacitance means a lower frequency; less capacitance means a higher frequency. The result is a frequency-modulated wave.

A refined system for producing FM involves the use of a tube to vary the frequency of the tank circuit, rather than the microphone. We know that in a pure inductance E leads I by 90°; in a pure capacitance I leads E by 90°. A tube can act as an inductance by having plate voltage lead grid voltage by 90°; it can act as a capacitance by having plate voltage lag grid voltage by 90°. A reactance tube can be made to act as a capacitance or an inductance. How a tube can be made to act as an inductance is shown vectorially. A signal applied to a capacitor will have its voltage lag its current by 90°. The capacitor voltage is then applied to the amplifier, with the tube inverting the signal by 180°. Note that the plate signal leads the signal current by 90°.

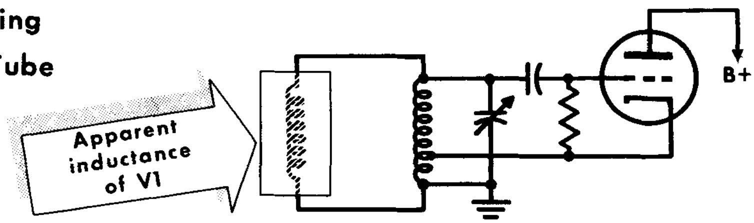
Direct FM (Contd.)

In the basic reactance-tube circuit the oscillator tank is paralleled by R1 and C1, placing the tank-signal voltage across these components. The resistance of R1 is much higher than the reactance of C1. Therefore, the R1-C1 circuit acts as a resistance, and the current flowing in it is in phase with the voltage of the tank circuit. The signal voltage across C1 lags the signal current by 90°. The tube amplifies the capacitor-signal voltage. This appears at the plate 180° out of phase with the grid voltage, and places the plate voltage 90° ahead of the signal current flowing through R1-C1. The plate-voltage signal is coupled back to the tank circuit by C2, a high-value low-reactance capacitor that couples the output signal and isolates the plate voltage from the tank circuit. The net result is to have V1 act as though it were an additional parallel inductance across the tank circuit inductance.

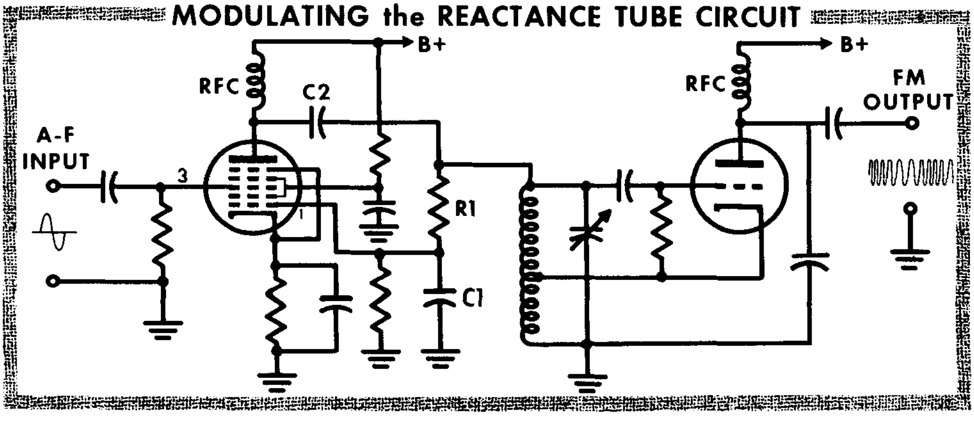
To put the reactance tube to work as a modulator, a pentagrid converter tube is used. The tank-circuit signal is applied to grid 1 with R1, C1, and C2, causing the tube to act as a reactance. The modulating audio signal is applied to grid 3 to modulate the reactance tube.

The variations in plate current caused by the audio signal determine to what degree the reactance tube will act as an inductor. These variations in turn cause the oscillator-tank circuit to change frequency accordingly. The frequency of the modulating signal results in the equivalent number of changes in the frequency of the oscillator. The amplitude of the modulating signal determines to what amount the reactance tube will vary, in turn determining to what extent the frequency of the oscillator will shift.

Result of Placing a Reactance Tube across the Tank Circuit



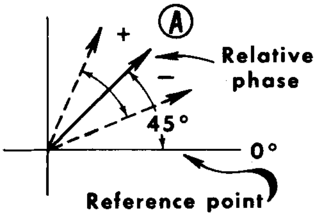
MODULATING the REACTANCE TUBE CIRCUIT



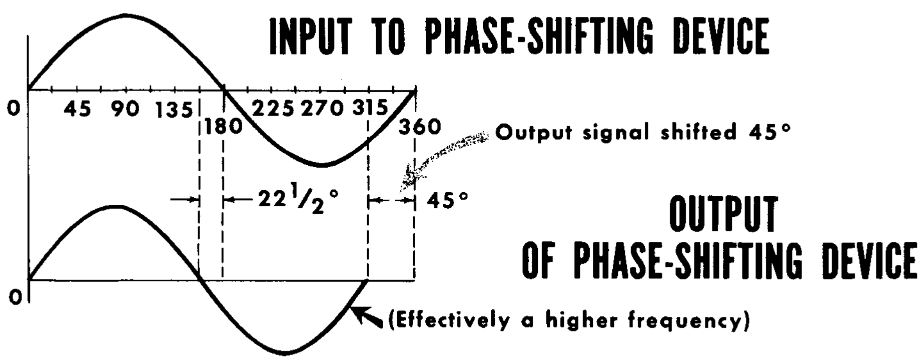
Indirect FM

The oscillator circuit used in the direct system of FM cannot be crystal controlled. This presents a problem in maintaining high-accuracy frequency control. The indirect method of FM permits the use of a crystal-controlled oscillator whose steady carrier output is phase modulated to produce FM.

To understand how phase modulation can create frequency modulation we will use a phase-shifting device set for 45° phase shift. A sine-wave input signal of constant frequency and amplitude is applied to the input terminals; the output terminals will have the same signal shifted to lead (or lag) by an amount determined by the phase-shifting device. As shown, the phase shifter is set to cause a 45° lead in the output signal. The output signal completes its cycle 45° ahead of the input signal. Each succeeding cycle will see the output lead-



Ⓐ is the phase-modulated carrier of 45° relative phase, and is constant in length. Phase deviation results when the vector varies above and below its relative phase angle as indicated by the dotted vectors.



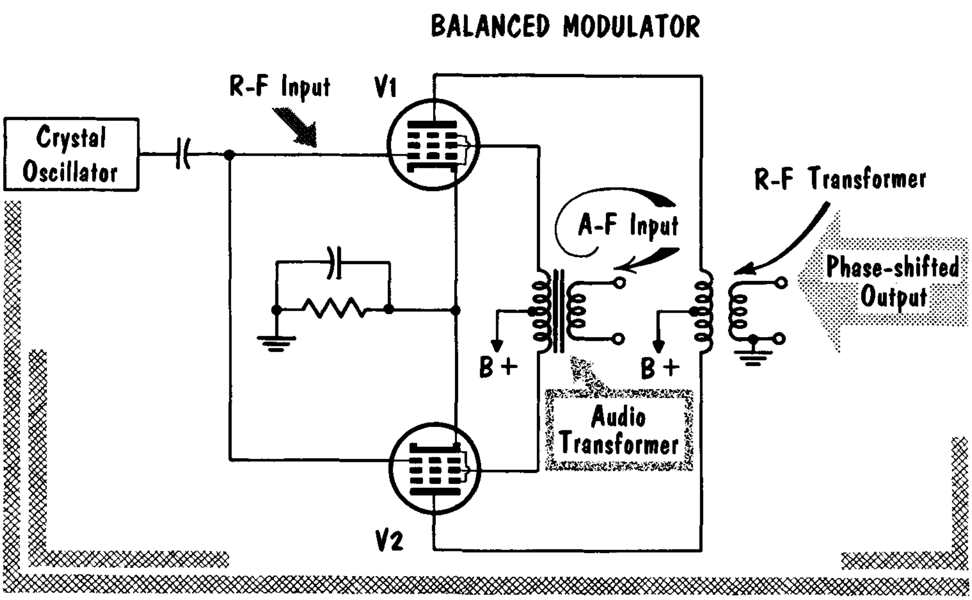
ing by 45° until 8 cycles have passed; the output signal will then lead by a full cycle, and the output frequency is thus higher than the input frequency. By the same process a signal that is caused to lag by 45° will after 8 cycles have the output lag the input by one cycle for a lower frequency.

The amplitude of the modulating signal will vary the degree of phase shift. The frequency of the modulating signal varies the frequency with which the phase shifting takes place. The resultant output of the phase-shifting device is a frequency-modulated signal.

The Armstrong Phase-Modulation Circuit

A circuit derived by Armstrong to achieve FM through PM uses the balanced-modulator stage. The circuit operates as follows: the output of a crystal oscillator is applied to the control grids of V1 and V2 in phase. The plate load for both tubes is a centertapped r-f transformer. The plate currents of each

THE ARMSTRONG PHASE-MODULATION CIRCUIT

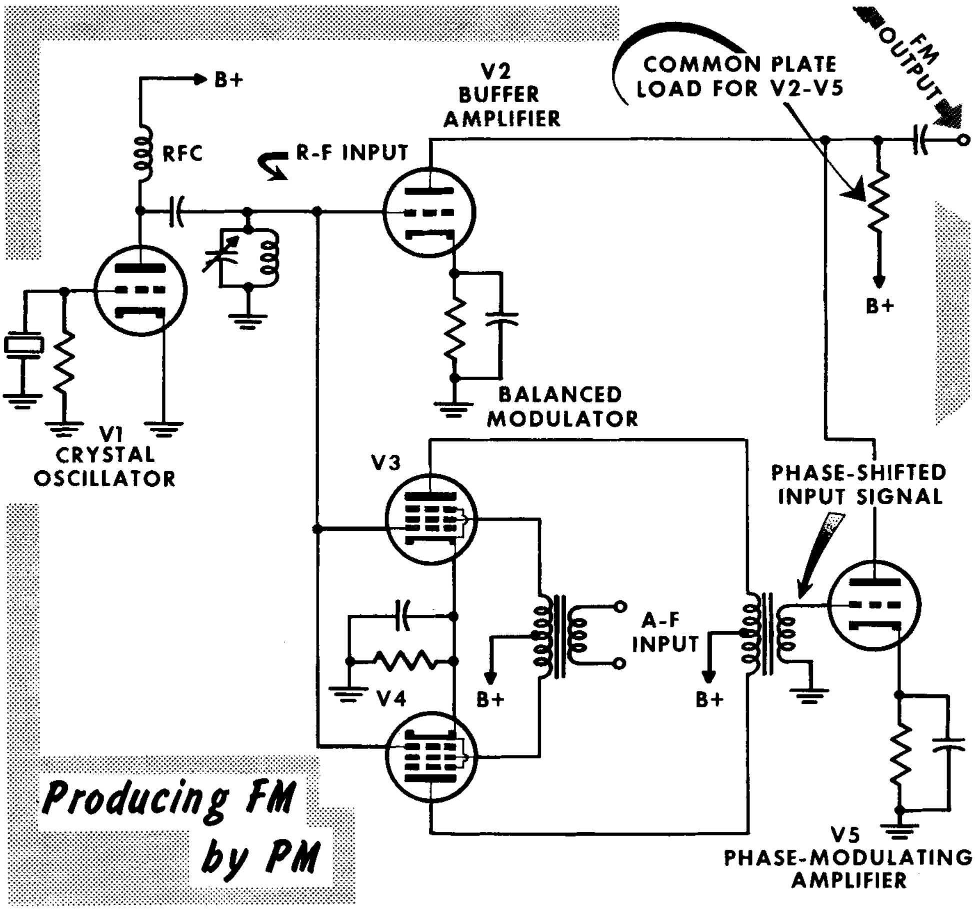


tube are 180° out of phase, and flow through the r-f output transformer. Therefore, they cancel, and no voltage is induced in the secondary. With no signal applied to the audio transformer, each screen grid has the same voltage; the same current flows in the plate circuit of both tubes. An a-f modulating signal applied to the audio transformer results in one end of the secondary being more, or less positive than the other. The resulting unbalance in the secondary causes the screen voltage of one tube to be higher, and the screen voltage of the other tube to be lower. Consequently the plate current of one tube increases, and that of the other tube decreases. The plate currents flowing in the r-f transformer primary are no longer equal but opposite in phase, and a voltage is now induced in the secondary.

The phase-shifting action of the tubes and the r-f transformer cause the r-f transformer output voltage to lead or lag the input voltage of the crystal oscillator. The phase shift leads or lags depending upon whether the plate current of V1 or V2 is increased. The amount of phase shift is determined by the modulating signal amplitude.

Phase Modulation (PM)

In the complete indirect FM circuit shown, the phase-shifted output of the r-f transformer secondary is applied to the amplifier tube V5. The crystal-oscillator output is applied to the balanced modulator and to the buffer amplifier V2. Tubes V2 and V5 use a common plate load. The phase-shifted output of



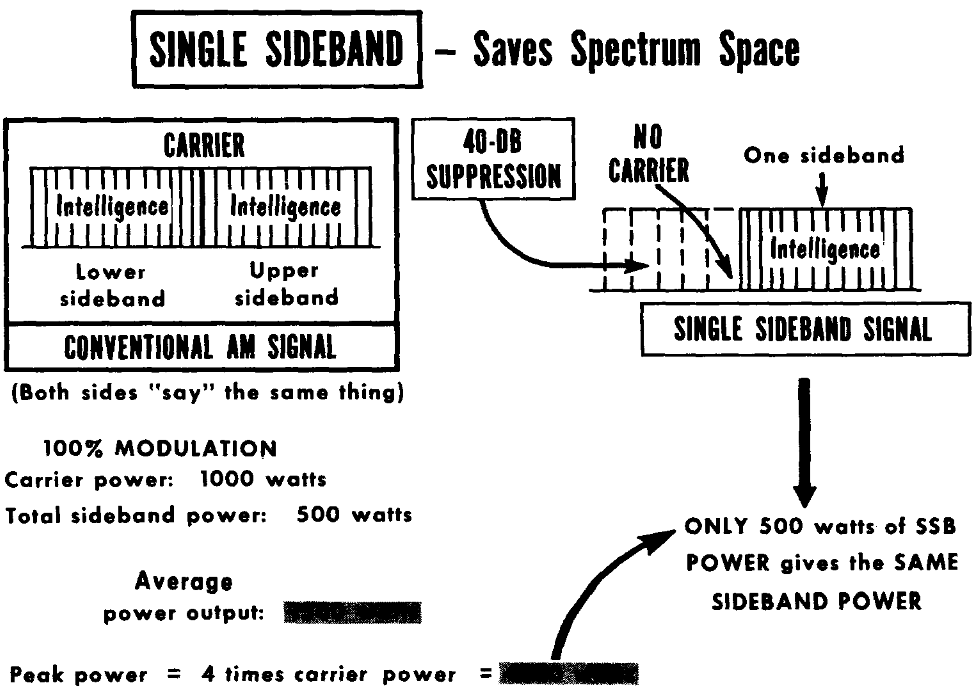
V5 phase modulates the output of V2. The resultant phase-modulated output of both tubes is changing in frequency to provide an FM output.

The degree of frequency swing of either the direct or indirect FM systems must be carefully controlled because the following stages consist of frequency-multiplier circuits that multiply both the center (zero-signal) frequency and the frequency deviation. The transmitter circuits used for FM transmitters are closely identical with those of AM transmitters. Frequency-multiplier circuits must use broader tuned circuits that are not too selective since the bandwidth used in FM is wider than that used in AM.

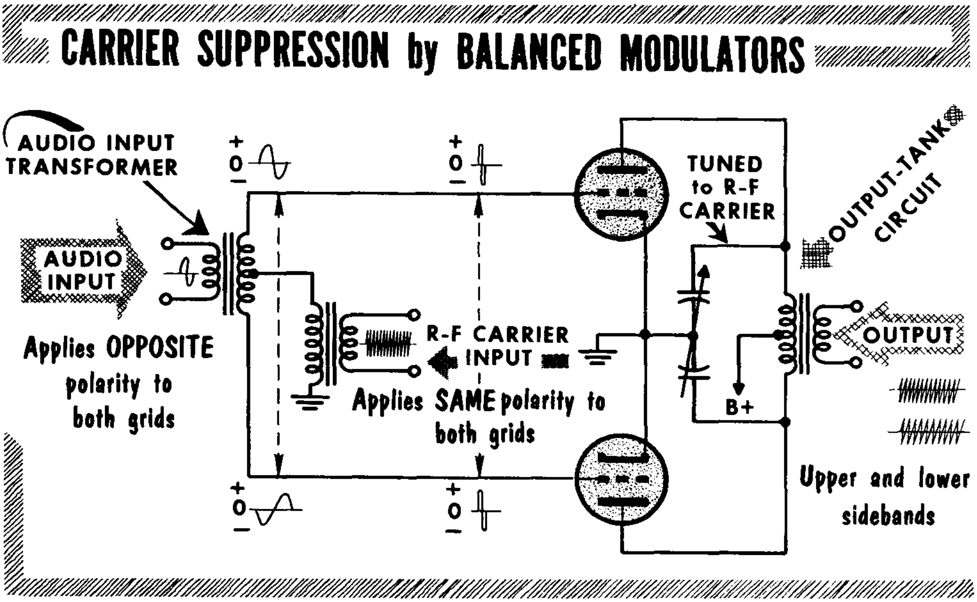
Basics of Single-Sideband Transmission

Single sideband (SSB) is a method of radio transmission in which one sideband and the carrier are suppressed; only one intelligence-bearing sideband is transmitted. Conventional AM signals consist of a carrier and two sidebands. The modulating signal determines the sideband frequency and amplitude. The upper and lower sidebands are mirror images of each other, hence contain duplicate information. At the receiver's detector, only one sideband is used; the carrier contains no information. Removing the carrier and one sideband leaves a single sideband, and does not alter the transmitter's intelligence-carrying ability because the entire modulating signal is contained in one sideband. By transmitting only a single sideband the power of the transmitter can be concentrated, and provides increased signal strength. Another way of expressing it is to say that a SSB of the same strength as an AM signal can be achieved with a lower-powered transmitter.

The SSB signal requires only one half of the r-f spectrum space taken by an AM signal. This relieves crowded r-f spectrum space, thus permitting more radio stations to go on the air. Since selective fading affects the sidebands in an AM signal, with only one sideband transmitted, intelligibility may be reduced. The stability requirements of a SSB transmitter are high. In an AM receiver the sidebands and the carrier are received together with no phase or frequency difference. In a SSB receiver a carrier must be provided for detection. A transmitter frequency shift is quickly noticed in the received signal. The SSB receiver too, must have a high order of stability, and include especially designed detector circuits.



Carrier Suppression by Balanced Modulators



There are many methods of accomplishing SSB transmission, all of which require carrier suppression. To remove the carrier while still leaving both sidebands is the job of the balanced modulator.

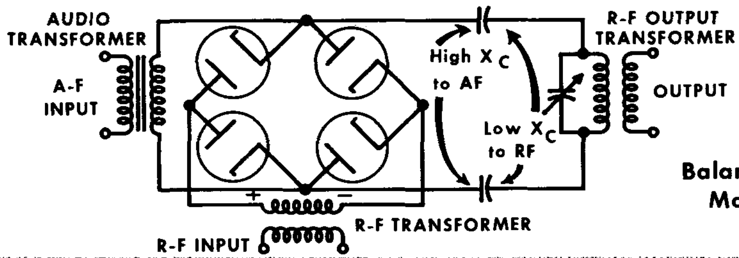
A balanced modulator using nonlinear vacuum tube amplifiers in a push-pull circuit is shown. Linear amplifiers will not mix frequencies; they will act only as amplifiers. The r-f carrier plate currents flow in equal but opposite directions to cancel each other. Thus there is no r-f carrier signal output. With the audio-modulating signal applied, the circuit acts as a conventional push-pull amplifier. The plate circuit, however, is tuned to the carrier. As a result, there is no audio output in the secondary of the plate circuit r-f transformer. Thus far, with either an r-f signal or audio signal applied, there is no output of the balanced modulator.

With both an a-f and r-f signal applied simultaneously, the circuit acts as follows: assuming that the polarity of the r-f carrier at the grid is negative at a given instant, the audio signal will have at any instant equal but opposite polarities at the secondary of the audio-input transformer. This places the modulating signal at one grid positive, and at one grid negative. The r-f and the a-f signals combine at the grid circuit. The grid having a positive a-f signal and negative r-f carrier signal will have a resultant cancellation. The grid, having a negative a-f signal and negative r-f carrier, will have the two signals add. The resultant mixing of both the r-f and a-f signals generates an output signal. The output frequency is the frequency of the carrier, plus or minus the modulating frequency. There is no carrier frequency itself, no audio frequency itself, only the two sidebands.

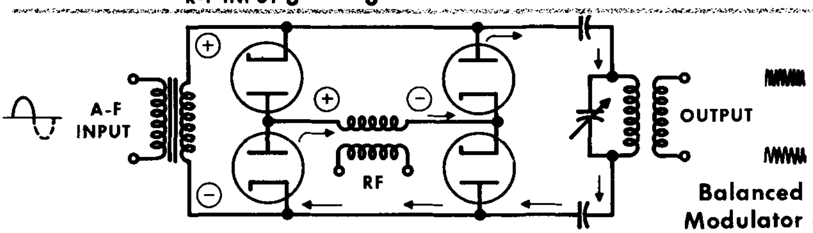
Carrier Suppression by Balanced-Bridge Modulators

A balanced modulator using a bridge rectifier circuit can provide a suppressed-carrier, double-sideband output. With only an r-f signal applied with the polarity shown, all diodes conduct, act as short-circuit, and prevent any r-f signal from getting to the output. With reversed polarity applied, the r-f signal places a positive polarity at the cathodes and a negative polarity at the plates of the diodes. The diodes do not conduct; they present an open circuit to the r-f input, and there is still no output signal. The audio modulation alone does not result in any output signal since the output-tank circuit is tuned to the r-f carrier frequency.

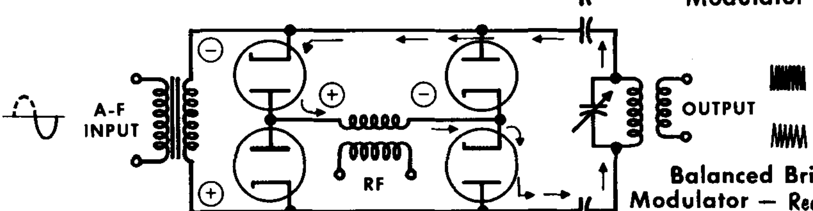
When a modulating signal and an r-f carrier signal are applied simultaneously, there is an output. The audio signal biases the diode, causing current to flow in one direction or another as determined by the polarity of the modulating signal. If the polarities of the audio and the r-f signals are as shown, the r-f carrier signal passes through the biased diodes, and provides an output signal. Reversing the polarity of the audio-signal voltage causes conduction in the opposite direction, and again provides an output signal. The modulating signal amplitude sets the diodes bias level, and modulates the r-f carrier. The output signal has no carrier, containing only sidebands. The output of a balanced modulator is a double sideband (DSB) with a suppressed carrier. A DSB signal is more efficient than a conventional AM signal because the power normally expended on the carrier can be applied to transmitting both sidebands.



Balanced Bridge Modulator



Balanced Bridge Modulator — Redrawn

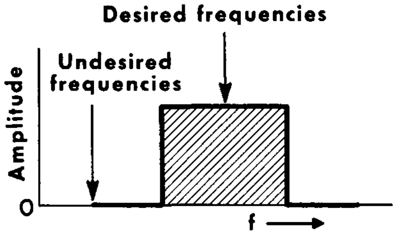


Balanced Bridge Modulator — Redrawn with opposite polarity a-f signal

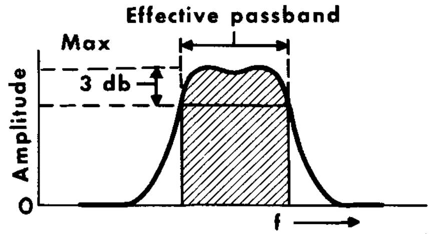
Sideband Suppression: Bandpass Filters

To obtain a SSB signal from the DSB output of the modulator, one of the sidebands is suppressed; either the lower or the upper. The two most popular methods for suppressing sidebands use filters and/or phasing circuits.

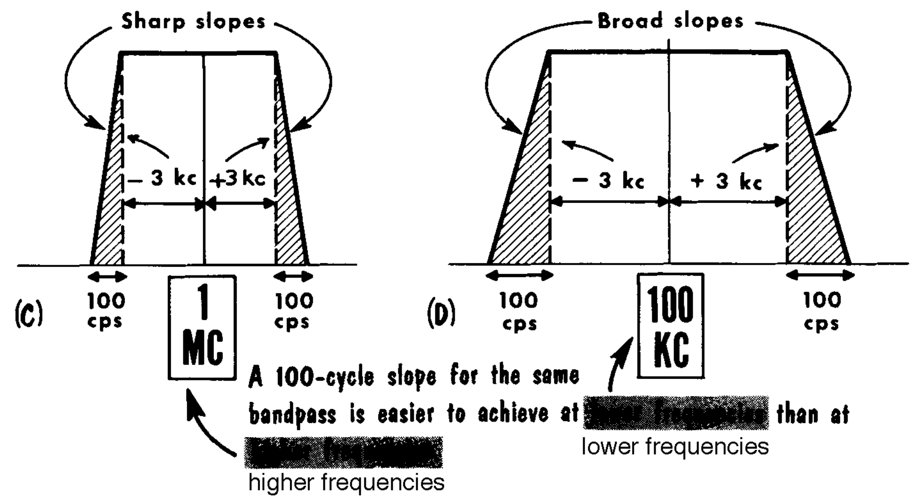
A bandpass filter passes a specific band of frequencies while rejecting all others. The requirement of such a filter is that the cutoff of the unwanted frequencies be as sharp as possible. A theoretical bandpass-filter frequency-response curve shown in A cuts off all frequencies exactly above and below



(A) IDEAL BANDPASS RESPONSE



(B) PRACTICAL BANDPASS RESPONSE

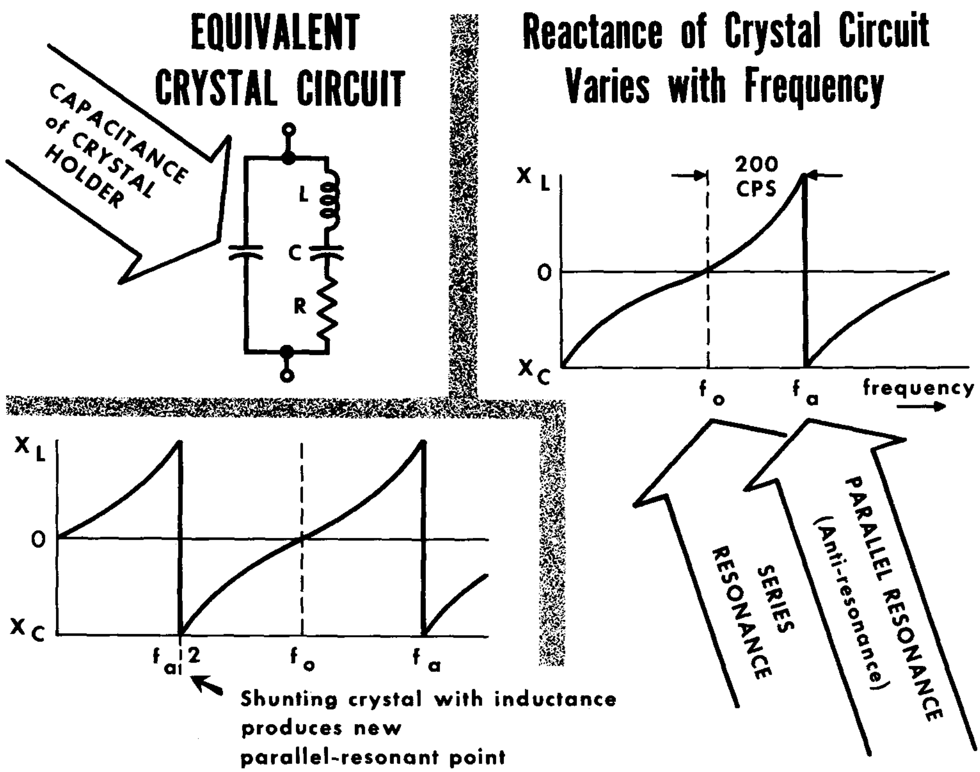


the desired passband. In practice, however, the sides or slopes of the pass-band response curve as shown in B, cannot be as steep as desired. Bandpass filters, however, may be designed with slopes steep enough to achieve the desired rejection of unwanted frequencies. The higher the frequencies, the greater the difficulties to achieve a steep bandpass slope. The bandpass filter shown in C is designed to operate at 1 mc, with a bandpass of ±3 kc and a tolerance of ±100 cps, and has extremely steep slopes. The same characteristic (D) of a bandpass filter designed to operate at 100 kc with a bandpass of ±3 kc, and a tolerance of 100 cps has reasonably less steep slopes.

Crystal Filters

Carefully designed, tuned L-C circuits can achieve the steep slopes required in sharp cutoff bandpass filters, but a more popular method is the use of crystals.

The crystal itself is basically a high-Q series-resonant circuit. The capacitance of the crystal-mounting device appears across the crystal to present a



parallel-resonant circuit. The series-resonant frequency of the crystal alone is lower than the parallel-resonant frequency of the crystal and holder. At series resonance the circuit is neither inductive nor capacitive.

Below series resonance, the crystal appears as a capacitive circuit. Above series resonance the crystal appears as an inductive circuit. The parallel-resonant frequency is called the antiresonant frequency and is labeled f_a. Between the series-resonant and parallel-resonant frequencies there is a bandwidth of approximately 200 cycles.

To increase the frequency spread between the series- and the parallel-resonant frequency, an inductor is shunted across the crystal. The shunted inductance causes the crystal to appear as a parallel-resonant circuit at a frequency below series resonance.

Crystal Filters (Contd.)

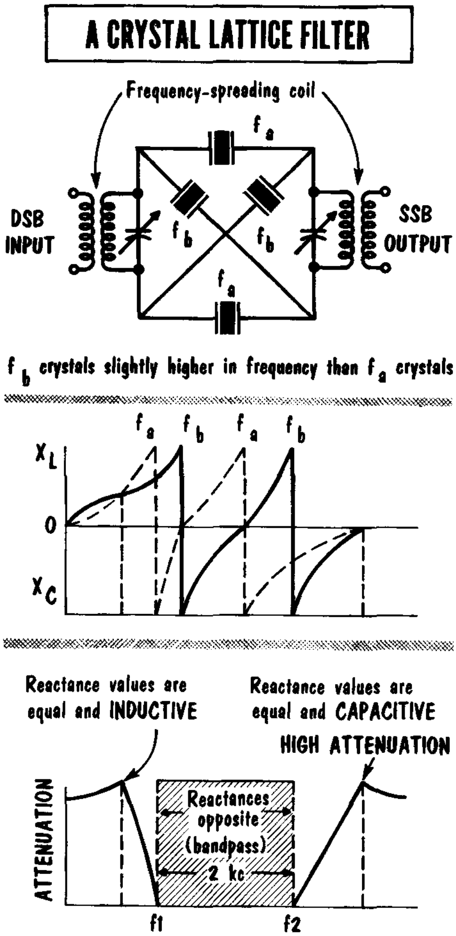
A crystal-filter circuit of the type shown is called a lattice type filter.

Two pairs of crystals are used; each pair may be tuned to a frequency approximately 1 kc apart. The input and output tuned circuits act as shunt inductors widening the frequency spread of the crystals. Two of the crystals, marked f_b, are placed in shunt with the input and output circuits. The other two crystals, marked f_a, are placed in series.

In the reactance curve of each set of crystals and the resultant bandpass characteristics shown, note that careful alignment is required to ensure that the series-resonant frequency of the series crystals (f_a) corresponds to the parallel-resonant frequency of the shunt crystals (f_b). The series-resonant frequency of the shunt crystals (f_b) must also correspond to the parallel-resonant frequency of the series crystals (f_a). Proper alignment is necessary to achieve the desired bandpass.

At those frequencies where the reactance of the crystal filter is predominately inductive or capacitive, the attenuation of the signal is high. At inbetween frequencies, approximately 1 kc above and below the resonant frequency, X_L and X_C are equal, or nearly so, thereby canceling and permitting these frequencies to pass. To achieve very steep slopes of the bandpass response curve the output of one filter may be applied to a second filter.

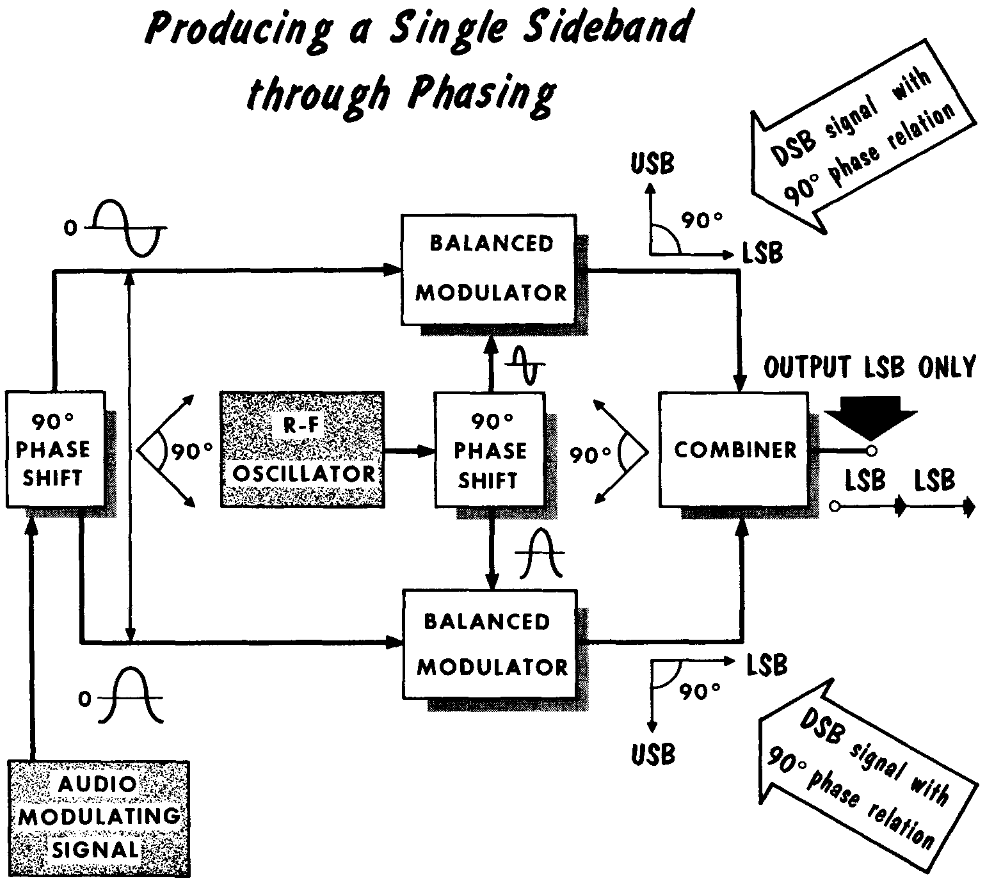
The price paid for the sharp selection of frequencies is a loss in power known as the insertion loss of the filter. This loss is the ratio of input power to output power of the filter.



Response curve of the individual pairs of crystals and the corresponding response curve of the filter

Phasing Method

Removing a sideband by phasing, is essentially done by applying out-of-phase signals to cancel each other. The modulating audio signal is applied to a phase-shifting network supplying two audio signals of equal amplitude, 90° out of phase, as shown by the accompanying vector. In addition, an r-f signal is also applied to a phase-shifting network to supply two r-f signals of equal amplitude that are 90° out of phase, as shown by the accompanying vector. The r-f and a-f signals are then applied to two balanced modulators.

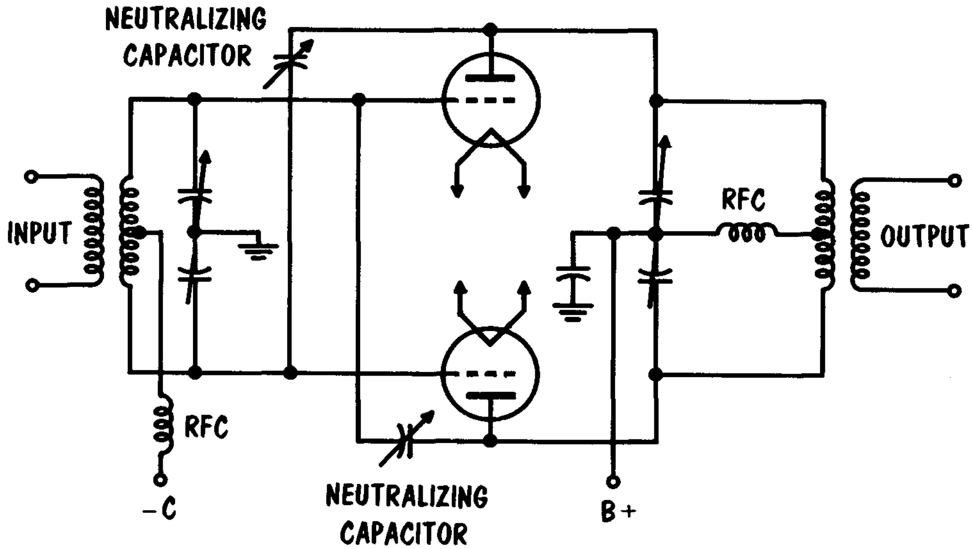


The output of the two modulator circuits consists of two DSB signals with the sidebands 90° out of phase with each other.

The two out-of-phase DSB signals are then placed in a combining circuit. The two upper sideband signals being 180° out of phase, cancel each other; the two lower sideband signals being in phase, add to produce a single sideband signal. To reverse the output to an upper sideband signal, the two out-of-phase audio signals going to the balanced modulators are reversed.

Linear Amplifiers

CLASS-B LINEAR AMPLIFIER for SSB AMPLIFICATION



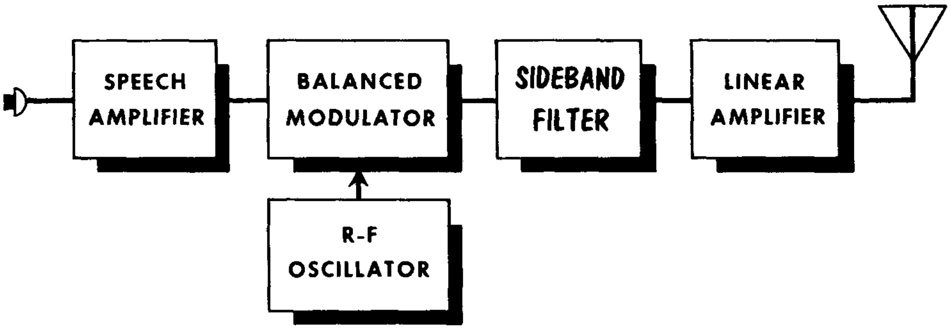
Having developed a SSB signal of low power it is necessary to amplify it to a value sufficient to provide the radiation energy required at the antenna. The best known linear amplifier is class-A; this, however, is costly, due to its low efficiency. The efficiency of a class-B linear amplifier is proportional to the applied signal voltage. A class-B linear amplifier operating with an unmodulated AM signal would only be about 35% efficient. A SSB signal is either present in its entirety, or completely absent. As a result, a SSB signal is essentially a constant-amplitude signal, permitting a class-B linear amplifier to operate at approximately 70% efficiency.

A nonlinear amplifier, even slightly nonlinear, is prone to develop harmonic-frequency signals of the applied signal. The harmonic frequencies and the fundamental frequencies can mix, to supply undesired frequencies. The undesired frequencies may well be in the suppressed sideband frequency. The result is poor SSB signal, with spurious frequencies present where the suppressed sideband is supposedly eliminated.

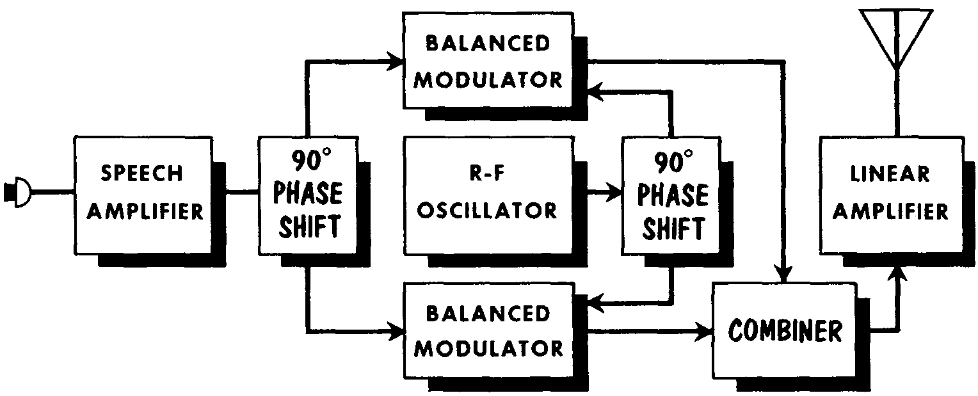
A class-A or -B linear amplifier resembles a typical class-C circuit, its main difference being in the value of grid bias. To operate class-B, the grid bias is set at cutoff; the tube amplifies only the positive-going portion of the signal. A highly-positive signal swing may cause the grid to go positive and draw grid current. When this happens, the grid circuit is no longer a high-impedance circuit. This power must come from the stage driving the linear amplifier. It is important that the driver stage be capable of supplying this power to the grid circuit. A well-designed driver will be capable of supplying at least three times the required power.

Basic SSB Transmitters

Special circuits are available to permit SSB reception with the least amount of difficulty. However, any good standard type commercial receiver with high stability, good band spread, and a beat-frequency oscillator, can be used to receive SSB. To receive a SSB signal, the prime consideration is that a carrier be reinserted to beat with the SSB signal for detection. In a commercial receiver this can be done with the bfo used for code reception. For proper detection the bfo signal level should be at least 10 times as strong as that of the received signal. Since the bfo level is fixed, and not as high as



BASIC SINGLE-SIDEBAND TRANSMITTER *using Filters*



BASIC SINGLE-SIDEBAND TRANSMITTER *using Phasing*

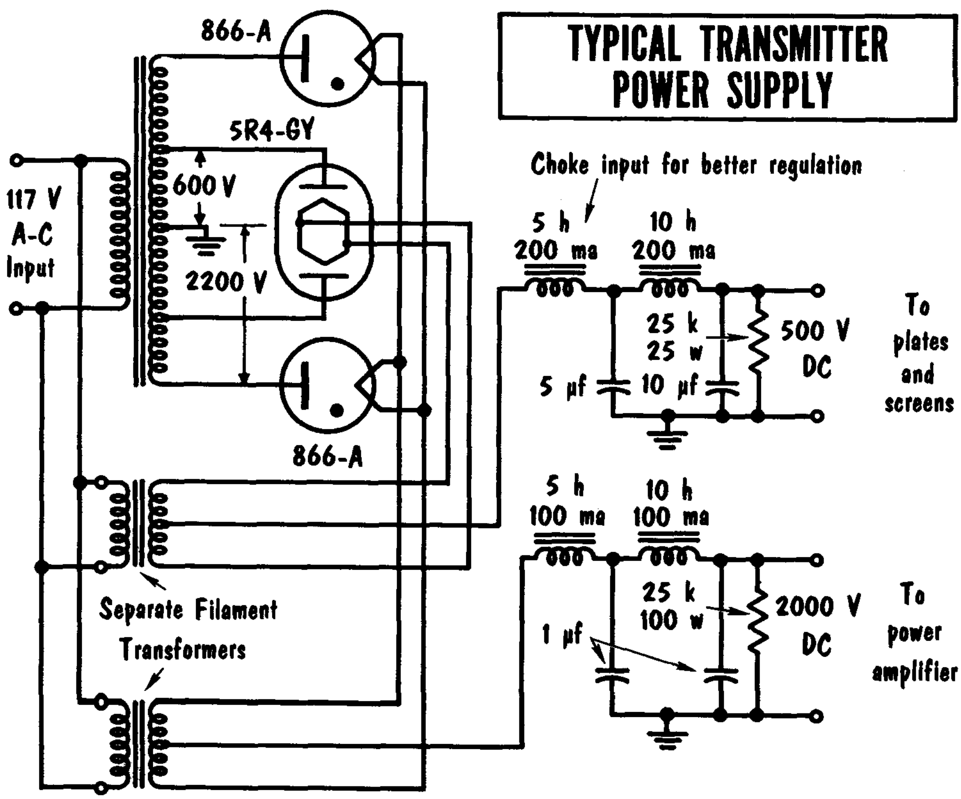
desired in the average receiver, the signal level can be reduced by lowering the r-f gain of the receiver.

Using the injected oscillator of the bfo as a reinserted carrier, it is important that the SSB signal be tuned in accurately to place the reinserted carrier at the correct operating point. Should the reinserted carrier beat with the SSB signal a few cycles too low, the speech will sound like Donald Duck. If too high, it will sound high pitched, making a bass sound like a tenor.

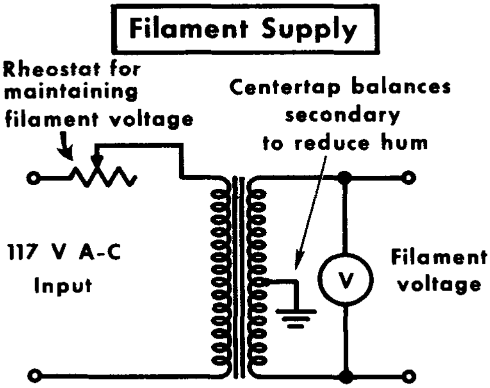
Transmitter Power Supplies

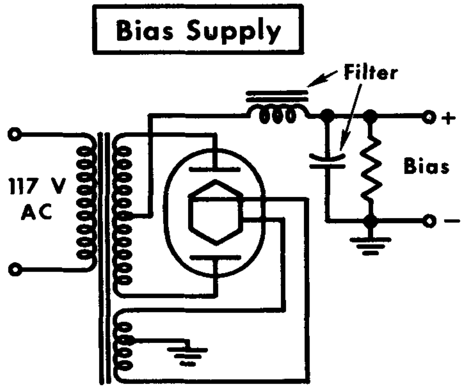
Power-supply circuits used in transmitters are basically the same as those used in receivers except for voltage and current considerations, and the many special types of controls. Half-wave, full-wave, and bridge-rectifier circuits are used. However, the rectifier tubes are much larger, being able to withstand many thousands of volts. Where large amounts of plate and screen currents are required, mercury-vapor tubes are used. These furnish high-power supply currents and produce small voltage drops across the rectifier tubes themselves. In transmitter power supply the power transformers are quite large, as are the filter capacitors, chokes, and insulators. All this reflects the high voltages and currents being consumed.

These power supplies must furnish certain basic voltages. A high voltage of from 500 to several thousand volts is required for the plate and screen of the r-f power amplifier, and the modulator stage. A low voltage of up to 500 volts or so, is required for the plates and screens of all the other tubes. Bias voltages of up to minus several hundred volts may be needed for the transmitter power stages. And, a high-current filament supply is generally required. In addition, a metering system is often included in the power supply for monitoring and adjustment, as well as for turning off certain voltages during warmup and tuning.



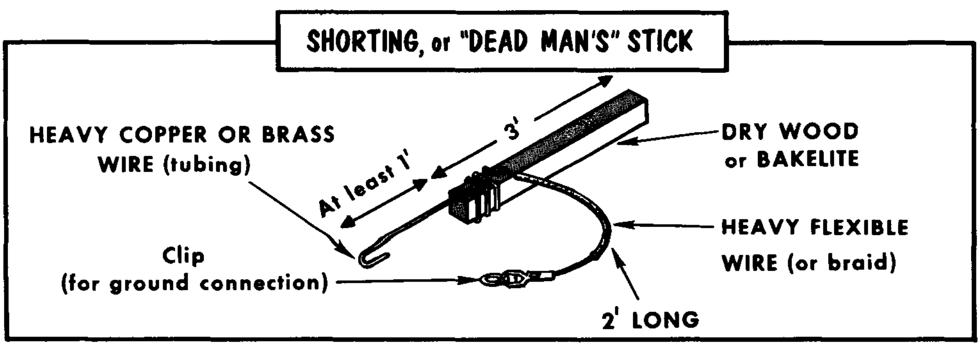
Transmitter Power Supplies (Contd.)



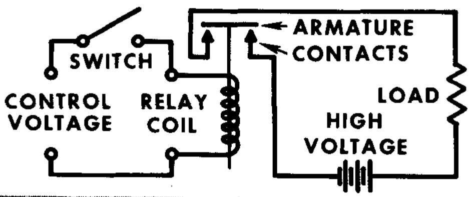


The filament and bias supply for transmitters are often completely separate units. The filament supply often has a means of adjusting and measuring, since a slight decrease in filament voltage can greatly affect the operation of a tube. The bias supply is often a conventional supply, except that it provides a negative voltage for the r-f and a-f power stages.

Most transmitters operate with plate voltages of from 250 volts to many thousands of volts. Contact with these voltages can cause serious shock or even death. It is necessary, therefore, to be especially careful when making transmitter adjustments. When the power is turned off in most transmitters, the bleeder and voltage-divider resistors discharge the filter capacitors. However, one or more of these resistors may open, and prevent the capacitors from discharging. If body contact is made with a charged capacitor, the capacitor can discharge through the body and cause shock, burns, or, as mentioned, death. Most transmitters with an output of 100 watts or more are equipped with switches, relays, or timing devices which open the primary circuits to high-voltage circuits whenever the transmitter doors are opened. However, any of these protective devices may be faulty, and then a "live" transmitter is even more dangerous than a charged capacitor. As a precaution, make sure there is no high voltage present, and discharge the high-voltage capacitors with a shorting stick.

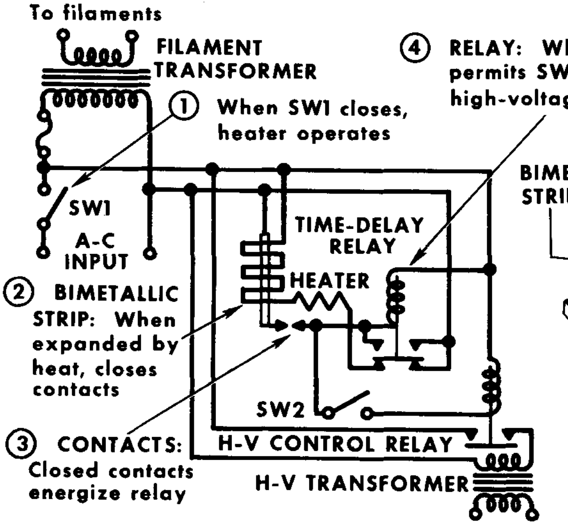


Transmitter Controls

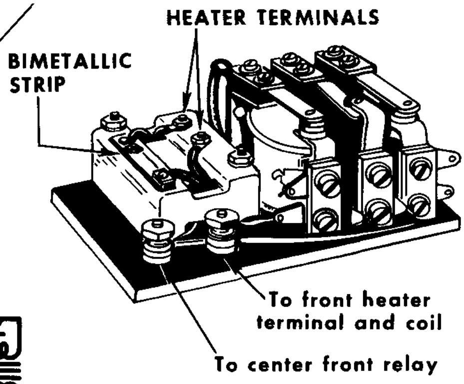


BASIC RELAY CIRCUIT

When switch is closed relay is energized, and load circuit is completed



Time-Delay Relay in High-Voltage Circuit



Time-Delay Relay

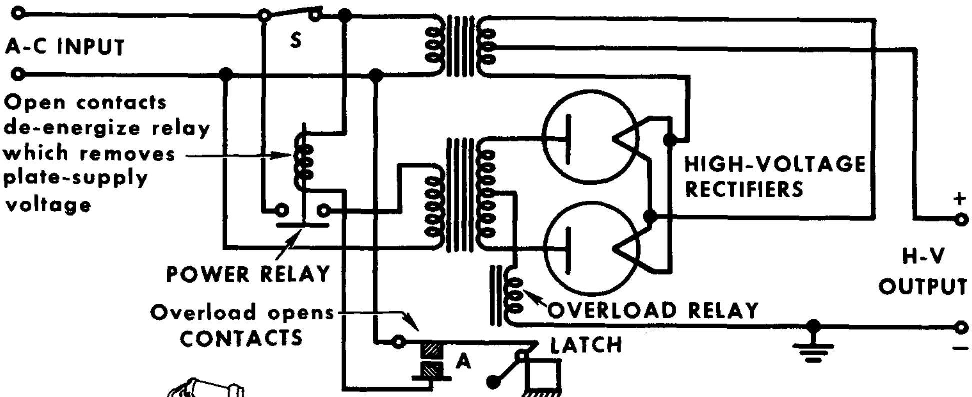
A number of devices are associated with power supplies to control and direct their functions, and to serve as protection to equipment and personnel. These controls take the form of key switches, circuit breakers, relays, panel-interlock switches, and time delays. Controls also operate signal devices that flash a light or sound an alarm when voltages and currents exceed a rated value. The common electrical fuse is a control device that opens the circuit when excessive current is drawn. These controls are connected at various points in a power supply.

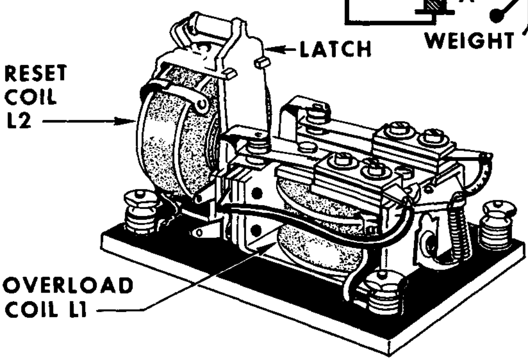
A relay is basically a switch operated by an electromagnet. It consists of a coil wound on an iron core and an armature that operates a set of contacts. The relay opens or closes an electrical circuit when the coil is energized. In the simple relay-circuit shown, when the switch is closed, the relay coil is energized, and the armature pulled down, closing the contacts. This, in turn, completes the high-voltage circuit. In the time-delay relay there is a definite time between operations. For instance, the relay provides a time delay between the heating of the filaments and the application of high voltage. A common form of time delay uses a bimetallic strip that bends as it is heated. A heater is mounted near the strip. When the strip expands from the heat, relay contacts are closed.

Transmitter Controls (Contd.)

An overload relay is a protective device that opens a circuit automatically when the current through it exceeds a certain established value. It is usually a d-c plunger relay in which the magnetic field of the relay coil acts upon a plunger. When the current becomes too great, the relay contacts open. A latch is usually included in the relay to keep the circuit open after the relay is energized, preventing the circuit from opening and closing repeatedly while the overload exists. The relay can have a mechanical release, or reset button, to restore it to operation after the overload has been removed. In

OVERLOAD RELAY CIRCUIT protects high-voltage system

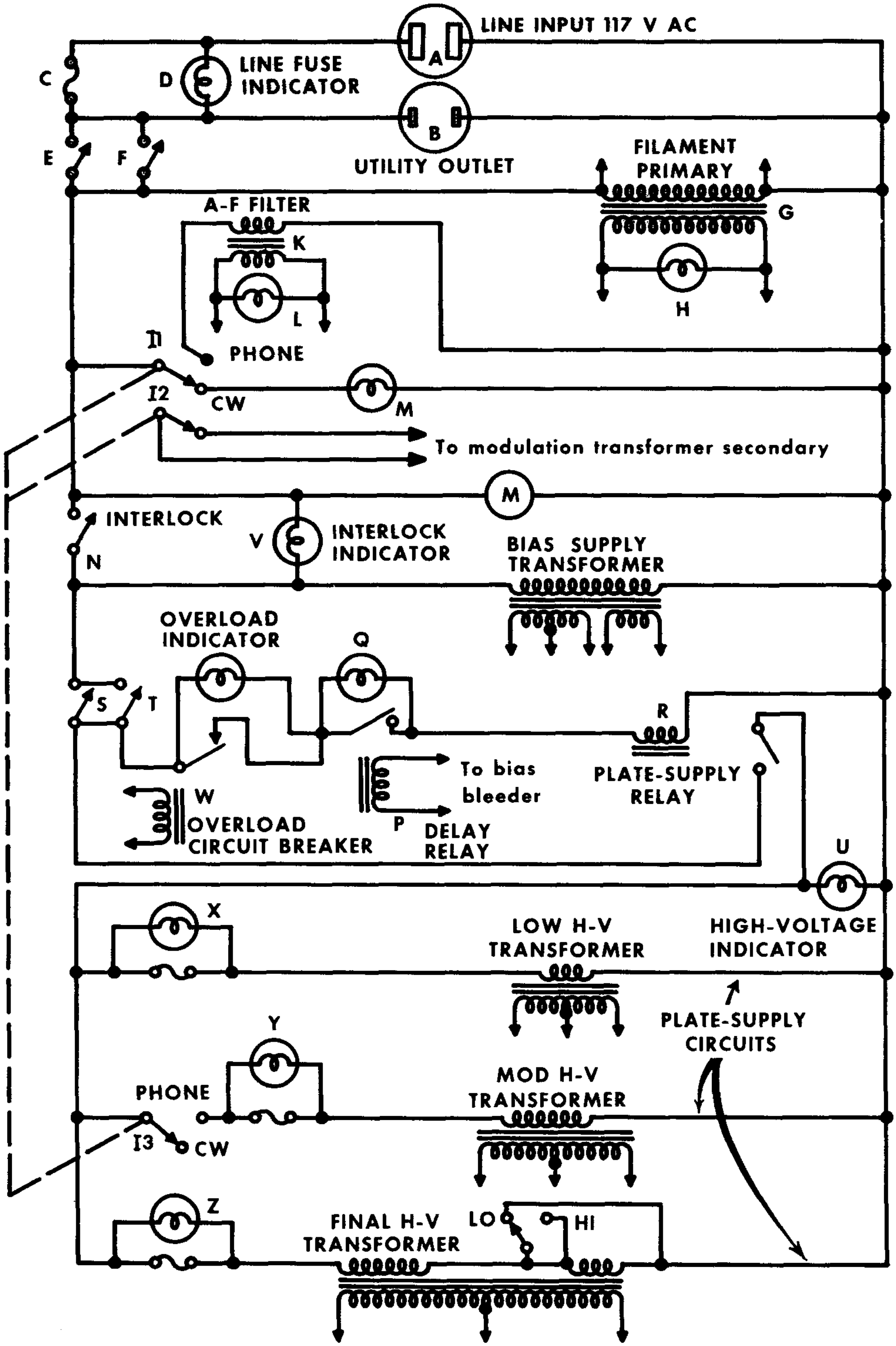




OVERLOAD RELAY with MAGNETIC RESET ACTION

the drawing, L1 is the overload coil, connected usually in the negative lead of the circuit being protected. L2 is the reset coil and, when energized, resets the mechanical latch after the relay has been energized.

In the relay circuit, the relay is connected in the low-potential side of the power supply to eliminate the need for high-voltage insulation. It is adjusted so that when the current reaches a predetermined point, the armature lifts, opening the contacts at A. This opens the circuit to the power-relay coil, de-energizing it and disconnecting the high-voltage transformer primary from the a-c input line. The weighted latch prevents the armature from falling and closing the contacts before the overload has been corrected.



TYPICAL CONTROL SYSTEM

Transmitter Control System

On the opposite page we see a complete transmitter control system, including line fuses, indicator lamps, and protective devices. At A, on top, is the basic 117-volt a-c power input. B represents a utility outlet that is not affected by subsequent switching. The input circuit is fused by C, with indicator lamp D lighting when C is open.

Filament switches E and F are connected in parallel. F energizes filament transformer G, and E supplies voltage to the remainder of the circuit, as required. Indicator lamp H, across the filament transformer secondary, lights when the filaments have been turned on. Ganged switch I1, I2, and I3, is the phone-CW switch. When it is in phone position, I1 closes the a-f filament transformer primary, I2 removes the short from the modulation transformer secondary, and I3 completes the circuit to the primary of the modulator high-voltage transformer. Voltage is applied to all filaments of the audio section through filament transformer K. At the same time, phone indicator lamp L lights. If the switch is in CW position, the filament primaries of the a-f section are open, and indicator lamp M lights. I2 short-circuits the secondary of the modulation transformer to prevent spurious oscillations in the modulation circuit on CW position. I3 does not affect any circuitry in the CW position.

Meter M is a line voltage indicator. N is a safety interlock switch in the circuit of the bias-supply transformer. Delay-relay P operates only after the bias supply is fully heated and delivering full output voltage. Operation of the relay shorts the bias indicator lamp Q, indicating that the plate-supply relay R is ready for operation. The time required for the operation of delay-relay P permits the bias supply to warm up enough so that it is safe to apply plate voltage to the r-f and a-f tubes.

With relay P closed, plate-supply relay R can be energized with the transmit-receive switch S (or T, an extension of the switch for remote operation). Relay R turns on all plate voltages, and lights the high-voltage indicator U. The transmitter is now ready for operation.

Indicator lamp V lights whenever the interlock switch is open. The interlock opens whenever the top or back of the transmitter has been removed, or has not been replaced securely. Whenever the interlock is open, all high voltages are removed from the transmitter, for the indicator lamp has enough reistance to prevent the bias power supply from operating.

W is an overload circuit breaker. Whenever the plate current to the final amplifier goes too high, exceeding the critical value for which the circuit breaker has been set, the circuit breaker operates. It opens the plate-supply relay circuit and causes all high-voltage stages to become inoperative. X, Y, and Z, are fuse-indicator lamps. They light whenever excess current opens the fuses associated with them.

The lo-hi switch in the primary of the final high-voltage transformer makes it possible to change from high to low power for tuning and testing purposes. For standby operation, only switch S (and/or T) has to be operated.

Varying the inductance or capacitance in an oscillator in accordance with an audio signal produces FM.

In phase modulation, the instantaneous phase of the signal is varied by the modulating signal.

A change in phase is equivalent to an instantaneous change in frequency.

In a phase-modulation system, the equivalent frequency deviation is proportional to the frequency of the modulating signal.

When the carrier frequency is varied directly, the process is called direct FM.

When the carrier is varied indirectly, the process is called indirect FM.

In single sideband transmission, the frequencies produced by the process of modulation on one side of the carrier are transmitted, and those on the other side are suppressed.

A crystal filter is a selective circuit using a filter that discriminates against all signals except those at the center frequency of the crystal.

A linear amplifier develops an output directly proportional in amplitude to that of the input signal.

In single sideband, since the carrier and one set of sidebands are not present in the output, a single sideband signal occupies less than half the spectrum space of a conventional AM signal.

In the filter method of sideband suppression, one sideband is removed by quartz-crystal filters or electromechanical filters.

In the phasing method of sideband suppression, the undesired sideband is removed by a process of phase shifting and balanced circuits.

The balanced modulator is desirable for single sideband applications because the balanced circuitry permits the carrier signal to be canceled out so that it does not appear in the output.

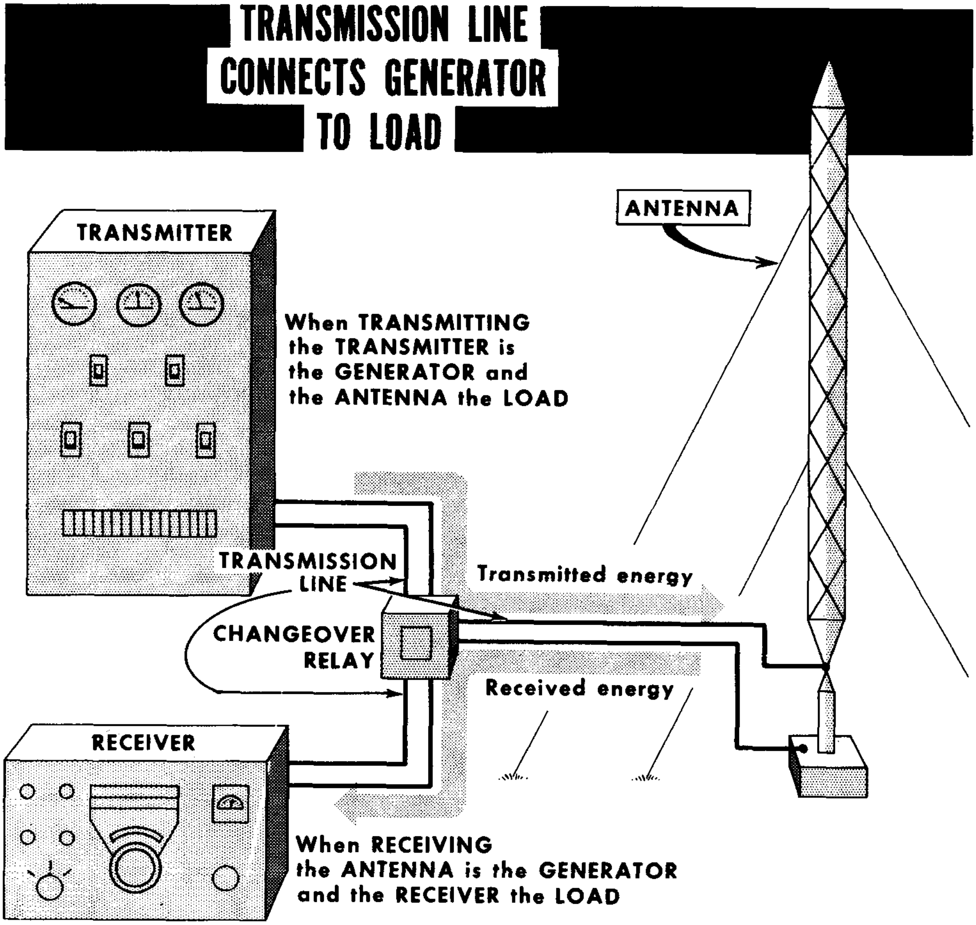
Transmitter power supplies differ little from conventional receiver supplies except in the components involved, which must handle much higher voltages and currents.

Transmitters use various types of relays to control the application of power to various circuits. All controls are designed for the protection of transmitter components, and for the safety of the operator.

REVIEW QUESTIONS

1. Explain the difference between direct and indirect FM.
2. Explain the difference between frequency and phase modulation.
3. Explain how FM is produced using a reactance-tube modulator.
4. How does a single sideband signal differ from a conventional AM signal?
5. How is the carrier suppressed by the balanced modulator?
6. Explain the filter method of sideband suppression.
7. Explain the phasing method of sideband suppression.
8. What is an advantage of single sideband transmission over double sideband transmission?
9. What is the function of a linear amplifier?
10. What is the basic difference between transmitter and receiver power supplies?
11. What is the function of a transmitter control system?
12. What methods are used to prevent a transmitting tube from overheating?

Transmission Lines



In most radio communication facilities the transmitting antenna is located at a distance from the transmitter. It is necessary, therefore, to transport the r-f energy from the transmitter to the antenna; similarly the signal energy intercepted by the receiving antenna must be fed to the radio receiver. A <u>transmission line</u> is used for this purpose.

Another way of describing the function of a transmission line is to say that it links a generator of electrical energy to its load. When transmitting, the <u>transmitter</u> is the <u>generator</u> and the <u>antenna</u> is the <u>load.</u> When receiving, the <u>antenna</u> is the <u>generator</u> and the <u>receiver</u> is the <u>load.</u>

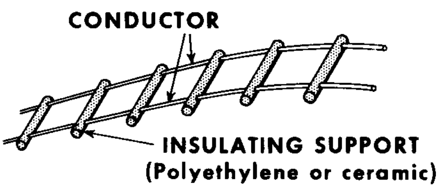
While it is true that any two wires can conduct electrical energy from one point to another, <u>any</u> pair of wires is not generally suitable for use as a transmission line. The commercial transmission line consists of two conductors arranged so that the device they form displays special electrical characteristics.

Types of Transmission Lines

There are many types of r-f transmission lines in general use. Of these, the three most commonly used are the open wire, coaxial cable, and twin or ribbon lead. Other types used include the twisted lamp cord, and single wire. In the coaxial type shown, the shield (outer conductor) is sometimes an aluminum tube with the separators made of rings of plastic, and the inside space is filled with an inert gas to prevent moisture formation.

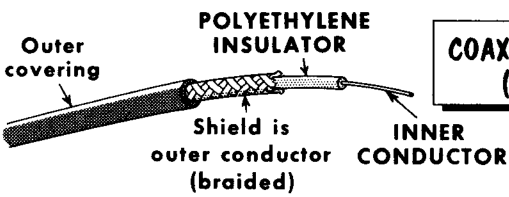
The physical characteristics of the transmission line influence the electrical characteristics, the amount of electrical power which each type of line can normally transport, the highest practical frequency of use, and electrical losses in the line. For example, the open-wire line is usable up to about 200 mc over a wide range of power levels, except perhaps for very high-powered equipment, and is the most efficient electrically, having the lowest losses. It is normally used when the feedpoint impedance of the antenna (discussed later) is several hundred ohms or more. The coaxial line is used over a much wider range of frequencies than the open-wire line, and handles a wide range of, and very high power levels. The twin-lead transmission line is also usable over a wide range of frequencies, but its power-handling capability is limited to perhaps several hundred watts input to the transmitter. Of these types of transmission lines the coaxial type is the most popular.

TYPES of TRANSMISSION LINES USED in RADIO COMMUNICATIONS



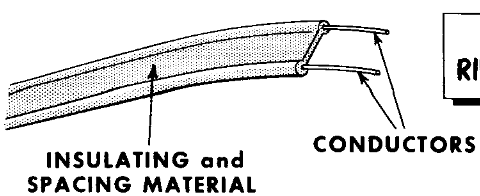
OPEN WIRE

Two uniformly-spaced parallel conductors, supported by spacers



COAXIAL CABLE (Coax)

Solid or multiple-strand inner conductor centered within outer shield



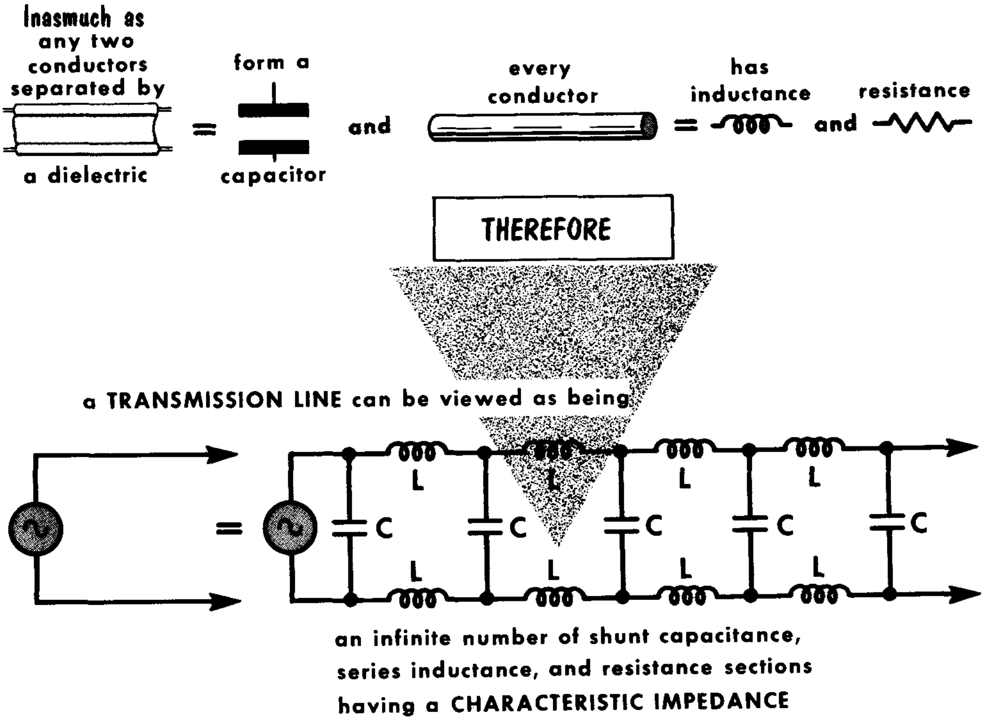
TWIN or RIBBON LEAD

Two uniformly-spaced parallel conductors encased in polyethylene

Inductance and Capacitance in Transmission Lines

When a d-c voltage or a low-frequency a-c voltage (say, 60 cps) is applied to two conductors of reasonable length which terminate in a resistive load, the voltage "sees" the total resistance of the circuit.

In an infinitely-long r-f transmission line, however, energy fed to it will travel down the line and be completely absorbed. We think of the two long



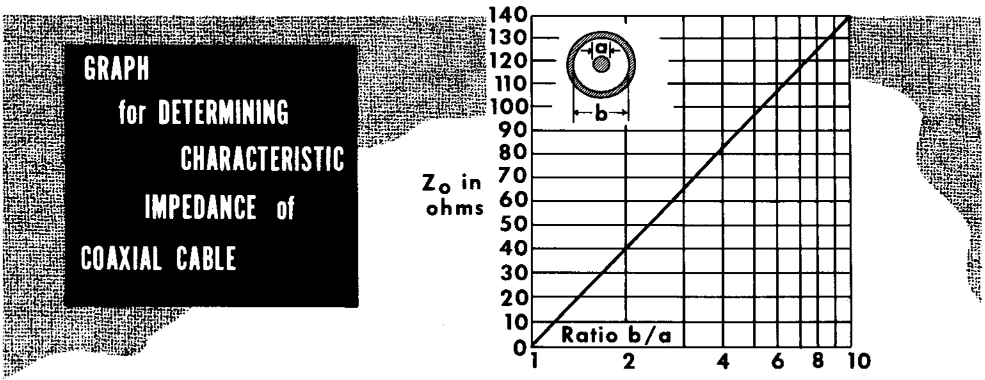
$$Z_o = \sqrt{\frac{L\ \text{(henries)}}{C\ \text{(farads)}}}$$

parallel conductors as forming a capacitance; the current-carrying conductors possess the property of inductance everywhere along the line. The same conductors also possess resistance, but this property is disregarded. Unlike the conventional circuit approach where capacitance and inductance are "lumped", capacitance and inductance in the infinitely-long transmission line are distributed throughout the length of the line. The distributed properties are apportioned into 1-foot lengths. The line can then be said to have, say, 30 micromicrofarads ($\mu\mu f$) per foot, and 0.15 microhenries (μh) per foot. An infinite number of sections of fixed C and L are assumed as making up the line, each section being identical.

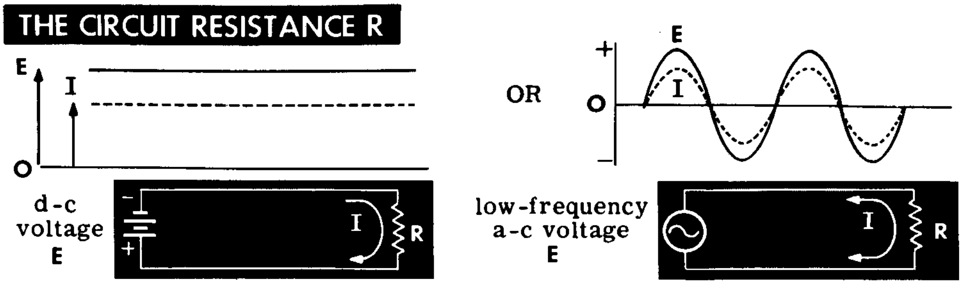
Characteristic Impedance of an Infinitely-Long Transmission Line

The shunt C and series L in each transmission line section comprise an impedance called the characteristic or surge impedance (Z_O) of the line. Z_O is purely resistive, and constant for any given line.

Your study of a-c electricity taught that Z varies with frequency. This is not so with the r-f transmission line; here, Z_O is a fixed quantity, whose ohmic value is determined by the physical construction of the line. Whereas



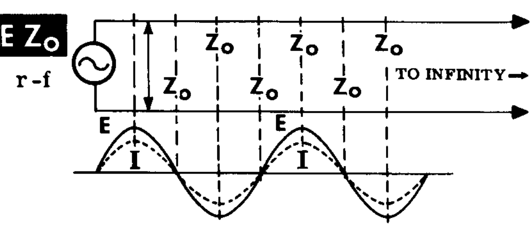
THE CIRCUIT RESISTANCE R



in a d-c circuit or in a low-frequency a-c circuit expresses the ratio between voltage E and current I in the circuit as a whole.

THE CHARACTERISTIC IMPEDANCE Z_O

of a theoretically infinitely long transmission line expresses the ratio between E and I at each point in the circuit, rather than in the circuit as a whole.



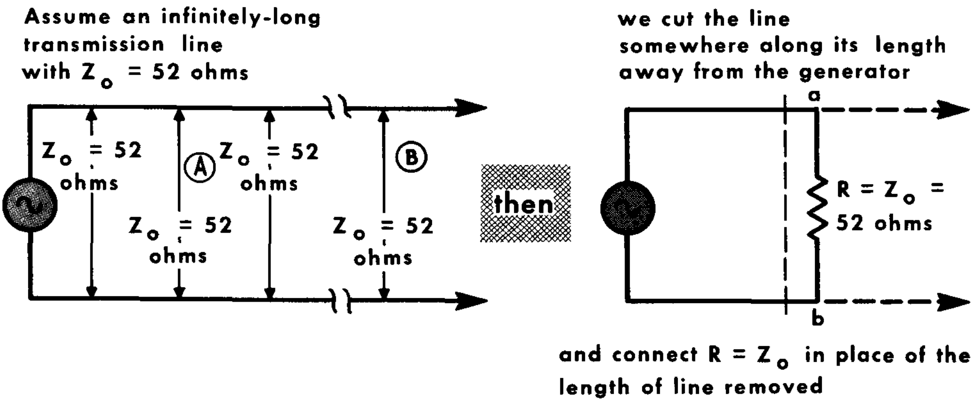
the voltage applied to a low-frequency circuit "sees" the total impedance Z of the circuit, the r-f voltage applied to the infinitely-long transmission line "sees" only the characteristic impedance Z_O at the input of the line. The current anywhere in the circuit is proportional to the instantaneous voltage at that particular point. If the applied voltage is 100 volts at one point, and the current is 1.92 ampere, $Z_O = 100/1.92 = 52$ ohms.

Terminating the Transmission Line

The infinitely-long r-f transmission line theoretically establishes the condition whereby a voltage applied to the line becomes energy which moves down the line never to return. When this happens, the current and voltage equivalent of the electromagnetic energy are in phase everywhere in the line. However, the infinitely-long line does not exist. Every practical transmission line is of finite length. How can such a line be made to perform like the infinitely-long line?

Imagine an infinitely long line having a Z_O of 52 ohms to be divided (but not separated) into two parts, A and B. A voltage is applied to this line. It travels down the line "seeing" 52 ohms everywhere along its path. Now imagine that part B of the line is cut off and replaced by a resistance equal to the Z_O of 52 ohms. If a voltage is applied to the line, the energy will advance down

A LOAD R = Z_o MAKES a FINITE LINE BEHAVE LIKE an INFINITELY-LONG LINE



The transmission line to the left of a-b will never "know" that section (B) was removed.

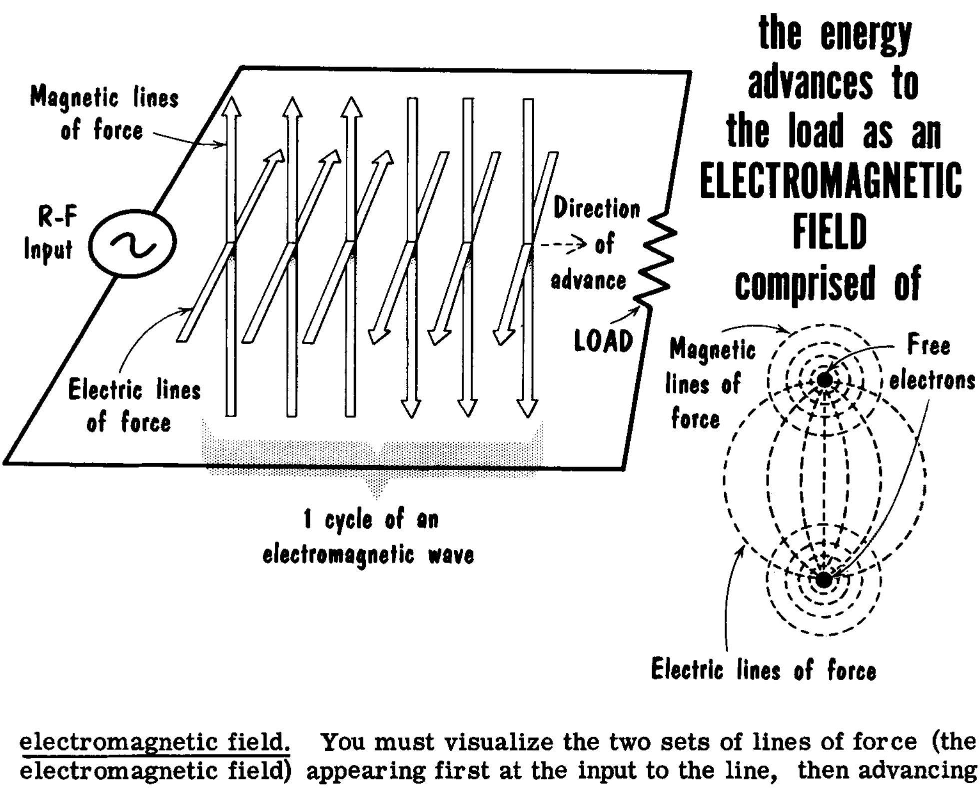
The load R = Z_o will absorb the energy exactly as the infinitely-long line would.

the line as before, not knowing how long the line is. Then it reaches the 52-ohm termination (the theoretical equivalent of part B). Just as the infinitely-long line absorbed all the energy traveling down the line, so the 52-ohm termination absorbs the energy when it reaches the load. In other words, if a transmission line of finite length is terminated in a resistance equal to its Z_O, the line will, regardless of length, demonstrate all the properties of the infinitely-long line. Such a line is called a terminated line. It transfers energy from the input to the load, with only slight loss of energy in the line. Whatever decrease in energy level (attenuation) does occur is the result of the R of the conductors as well as leakage in the dielectric that separates the two conductors. Substantially no losses are due to the characteristic impedance because it does not produce a power loss.

The R-F Transmission Line Transports Electromagnetic Energy

In the d-c circuit and the low-frequency a-c circuit energy is conveyed by the motion of charge (current). Referring back to the distributed L and C which make up the transmission line, applying an r-f voltage to the input causes current to flow into the line – into the first "section" of the line. One result is the creation of encircling magnetic lines of force around the free electrons which move as current in each conductor. The two fields combine in the space between the conductors, producing a composite field. At the same time, electric lines of force terminating on the moving charges in each of the conductors appear in the space between the conductors. So we have magnetic and electric lines of force, the two at right angles, occupying the space between the two conductors. The two fields combined are called an

IN the OPERATION of the R-F TRANSMISSION LINE



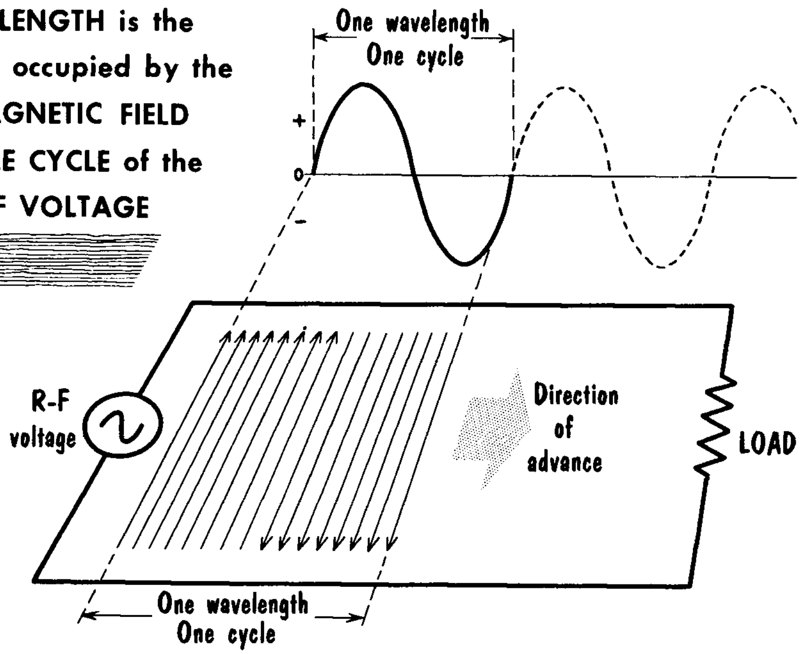
electromagnetic field. You must visualize the two sets of lines of force (the electromagnetic field) appearing first at the input to the line, then advancing down the line from one imaginary section to the other, carrying the energy with them. During each half cycle, the current in each of the conductors reverses its direction, as do the electric and magnetic lines of force. All the while the electromagnetic field they form advances away from the input of the line. Thus, current flow in a terminated transmission line can flow in opposite directions at many points, but the electromagnetic energy can be advancing in one direction only.

Electrical Wavelength

Let us take a look at the electromagnetic field generated by a 30-mc r-f voltage applied to the terminated transmission line. (For convenience, only the electric component of the field is shown.)

The first line of electric force appears with the first increment of input voltage and starts moving down the line, followed by the other lines of force cor-

ONE WAVELENGTH is the DIMENSION occupied by the ELECTROMAGNETIC FIELD for a SINGLE CYCLE of the APPLIED R-F VOLTAGE



responding to subsequent values of input r-f voltage. Following the positive alternation, the lines of force for the negative alternation appear and advance down the line. Each succeeding cycle produces the identical pattern of electric lines of force. Note that each set of lines of force corresponding to 1 cycle of the input voltage occupy the same lineal dimension. A measurement of 1 cycle for the 30-mc input signal shows this length to be 32.1 feet. This dimension is called 1 electrical wavelength, designated by the Greek symbol λ (lambda). Wavelength is related to both the electric and magnetic components of the field created by each cycle; the dimension is the same with both components present. A definition of 1 wavelength could be "the lineal dimension occupied by the electric lines of force created by all the amplitude variations of voltage in 1 cycle." The length of 1 electrical wavelength arises from a fixed relationship between velocity of the field (wave) in whatever medium in which it is traveling, and the frequency of the voltage. It is

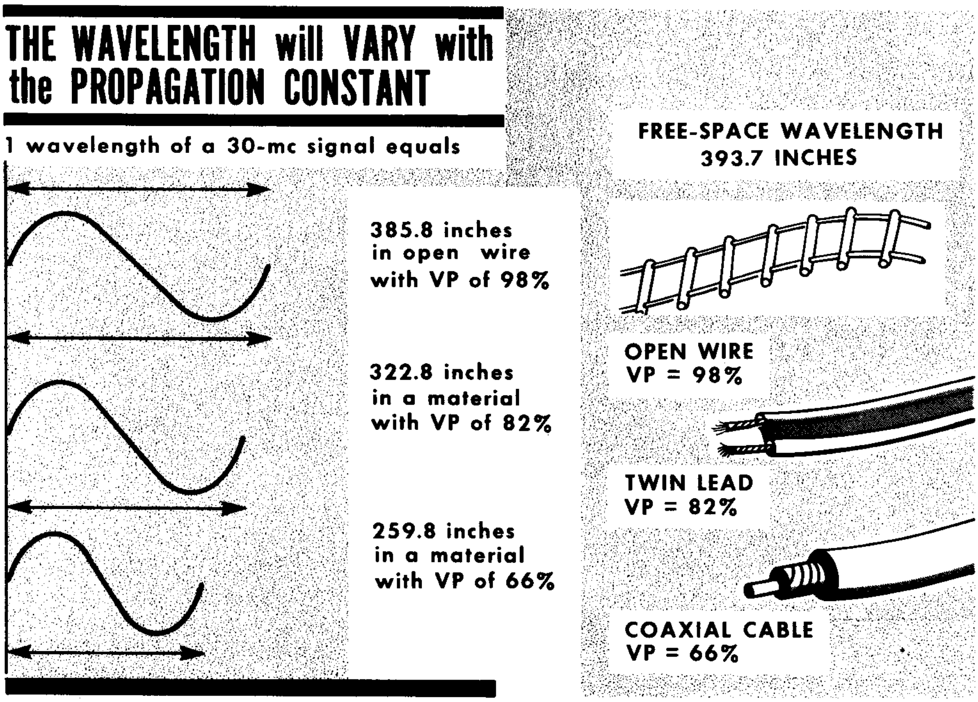
$$\text{wavelength } (\lambda) \text{ in meters} = \frac{\text{velocity of the energy in meters per second}}{\text{frequency in cycles per second}}$$

Electrical Wavelength and Velocity of Propagation

The reference for all velocities of electromagnetic radiations is the velocity of light in free space: 299,744,000 (usually stated as 300,000,000) meters per second, or 984 feet per μsec. Using the above velocity, the wavelength of a 30-mc wave in free space is

$$\text{one electrical wavelength (in meters)} = \frac{300,000,000}{30,000,000} = 10 \text{ meters}$$

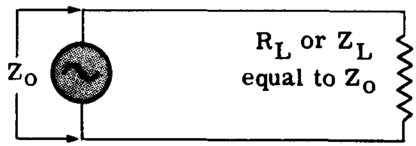
With 1 meter = 39.37 inches, 10 meters = 10 x 39.37 = 393.7 inches, or 32.8 feet. How was the previous figure of 32.1 feet derived? The answer is

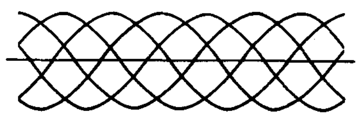


simply that the velocity of propagation (VP) of electromagnetic energy is most rapid in free space, and is slowed down when advancing in all other physical media. A transmission line is such a medium; the VP is lower in a transmission line than in free space. Different kinds of transmission lines bear different VP constants. When stated as a number, it is a percentage, with 100% being the velocity of light. Having determined that 1 wavelength of a 30-mc voltage in space is 32.8 feet, 98% of this length (in an open-wire line) is 32.8 x .98 = 32.1 feet. An electrical wavelength for 60 cps = 3100 miles. The lower the velocity the shorter the electrical wavelength for any given frequency. When a transmission line is spoken of as being so many wavelengths long, electrical wavelength is meant with the velocity of propagation already taken into account.

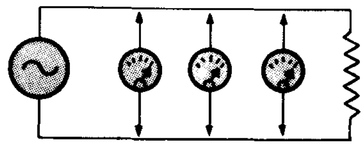
Transmission Lines Terminated in Characteristic Impedance

WHEN A PROPERLY TERMINATED LINE of any length is subject to a wave of energy, all points along it are subject to all values of voltage and current

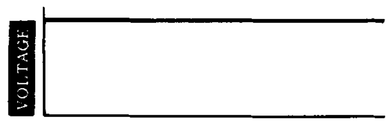




AND



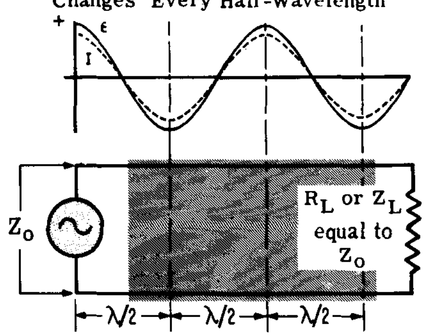
the line behaves as if a constant-amplitude voltage was applied to it.

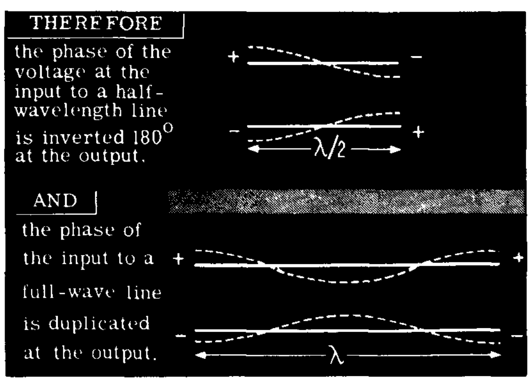


When an electromagnetic wave representative of energy between zero and maximum values advances along a transmission line terminated in a resistance equal to its Z_O, each point along the line from input to load is subject to like levels of energy from zero to maximum. Since this energy is the result of the application of a cyclic variation between zero and maximum of the input voltage, every point along the line is subject to voltage (and current) values between zero and maximum. Whether the line is long or short, no point will show any fixed value of voltage (or current) different from any other. In other words, there will be no points of maximum voltage or current (loops), or points of zero voltage or current (nodes), or any values in between. The line is flat.

Regardless of what may be the instantaneous polarity of the voltage (and current) at any point along the line, at a distance of 1 wavelength (λ) from this point, the polarity or phase is repeated. At a distance of a 1/2 λ however, the polarity or the phase are inverted. This is an important property of the 1/2-λ or full λ terminated line. A 1/2-λ line or odd multiple thereof is a polarity or a phase changer, while two 1/2-λ lines or even multiples of a 1/2-λ line become a polarity or phase repeater.

The Phase of Voltage and Current Changes Every Half-Wavelength

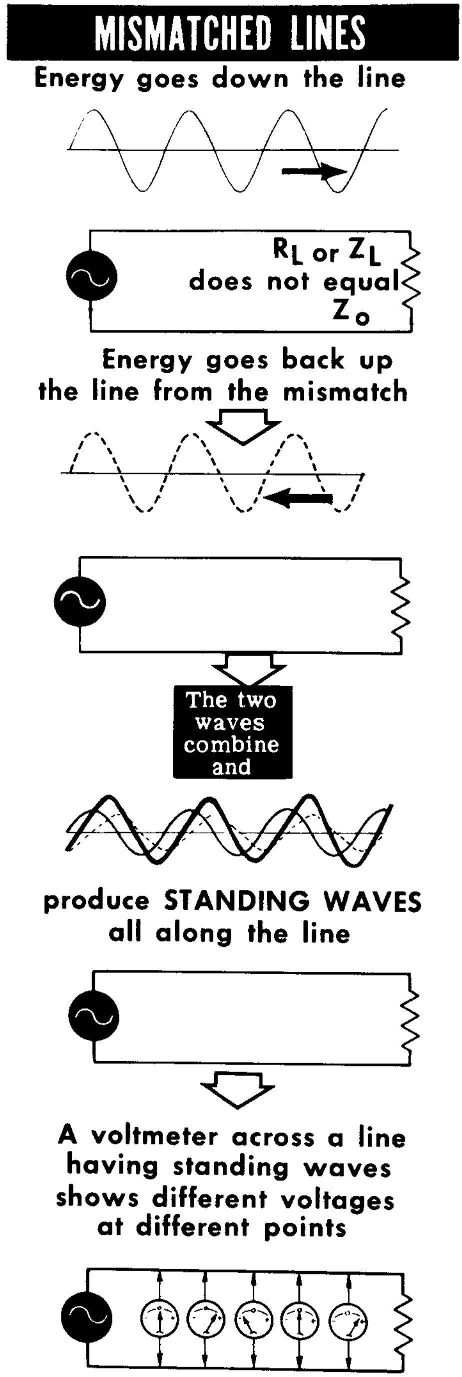




Transmission Line Not Terminated in Characteristic Impedance

It is conceivable that for any one of many possible reasons a transmission line may be terminated in a resistance not equal to its Z_O. If the difference in the ohmic values of the two is not too great, say 10 to 25%, it can be ignored in most instances. But if the difference is substantial, say 2 to 1, a new and important set of voltage and current conditions are created. Such a relationship between the ohmic values of the Z_O and the load is called an impedance mismatch. The greater the difference in the line and load impedances, the greater the mismatch.

Energy sent down the line toward the load finally reaches it, but is only partially absorbed. That portion not absorbed is reflected from the load, and travels back up the line to the generator. All the while energy is leaving the generator and advancing toward the load. So we have current and voltage traveling in two directions simultaneously in the line. Let us concern ourselves with only voltage, although reflection of current also occurs. The forward and reflected voltage waves combine and produce resultant sum voltage waves known as standing voltage waves. This means fixed points of maximum and zero voltage along the line. The useful energy absorbed by the load is the difference between the forward energy and the reflected energy. Hence, the greater the mismatch the less energy sent down the line becomes useful energy. When a transmission line is used to deliver electrical energy from one point to the other, correct matching of the load impedance to the transmission line impedance is vital. A badly mismatched line makes it difficult to properly "dip" the final amplifier in a transmitter.

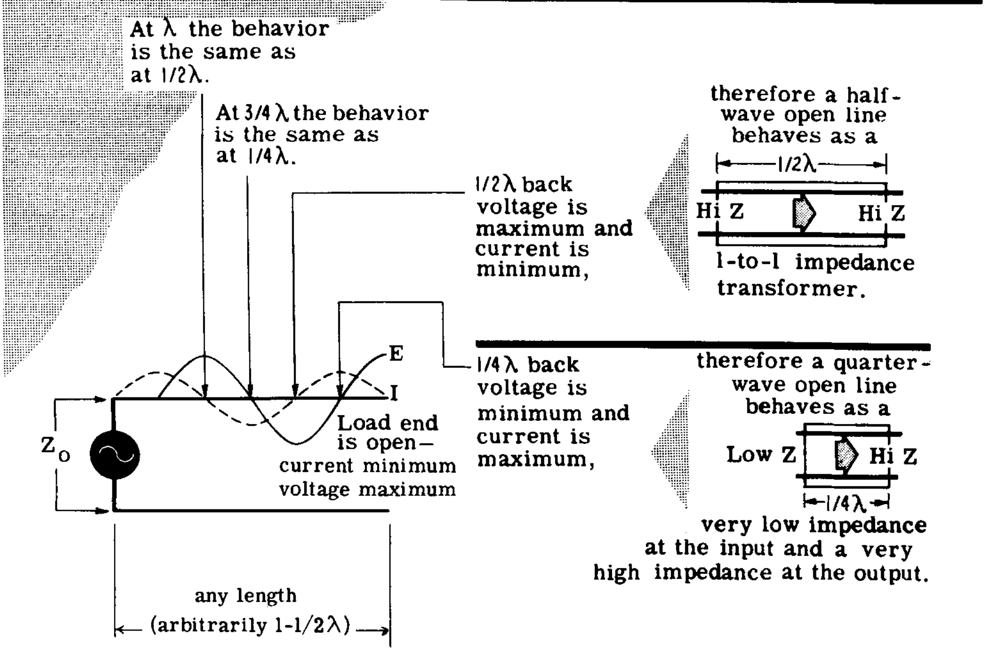


Resonant Transmission Lines (Open Circuited)

A resonant transmission line has standing waves of voltage and current. Imagine a transmission line open circuited at the load end. A mismatch exists, hence reflection of the applied r-f voltage and current occurs. The standing wave pattern of voltage and current distribution at the frequency for which the line is dimensioned is very useful.

Looking at the open end of the line, we say that the electrical condition of the open circuit is the equivalent of an infinite impedance; hence minimum current and maximum voltage. One quarter λ back from this point the standing wave of the current is maximum while the voltage is minimum. This indicates that the impedance is minimum, or a short-circuit exists. So, a 1/4-λ open line presents a minimum impedance at one end and a maximum imped-

CHARACTERISTICS OF OPEN-END LINES

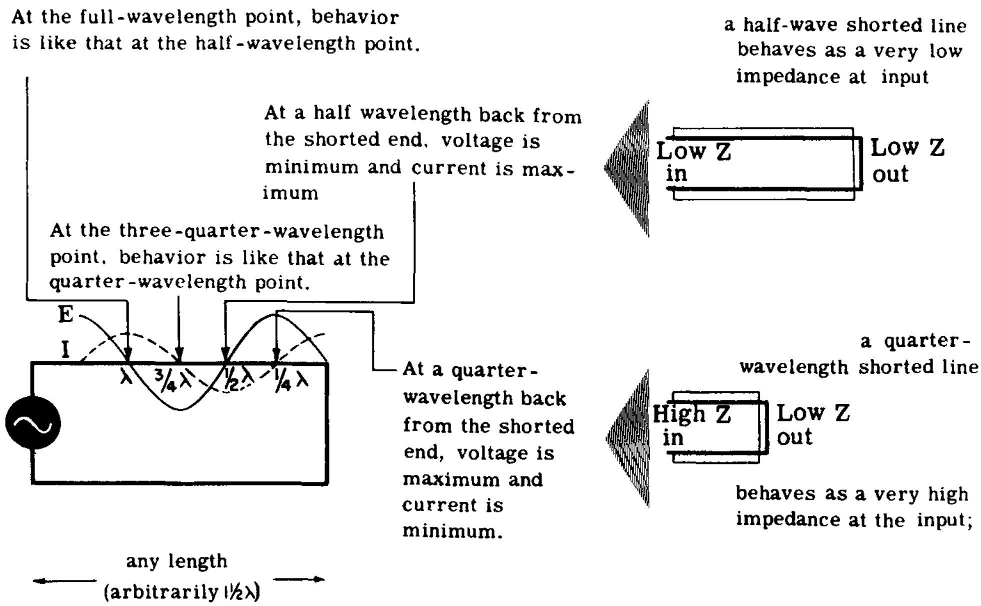


ance at the other, at the frequency for which the line is 1/4 λ. Either end can be the high-impedance end, which would make the other end the low-impedance end. Another 1/4 λ back along the line we note minimum current and maximum voltage; the same conditions as at the open load end. Thus, a 1/2-λ section of open line presents a high impedance at both ends. This is also true for whole number multiples (2, 3, 4, etc.,) of the 1/2-λ open line with 180° phase changes in voltage. Moving back another 1/4 λ to the 3/4 λ point, we note maximum current and minimum voltage again. Thus, odd multiples (3, 5, 7, etc.,) of the 1/4-λ open line duplicate the conditions of the 1/4-λ open line accompanied by 180° phase changes in the current.

Resonant Transmission Lines (Short-Circuited)

Imagine another extreme case of mismatch – the shorted transmission line. An r-f voltage is applied and reflection occurs at the point of mismatch. Standing waves of current and voltage appear on the line. The current at the shorted end is maximum while the voltage is minimum, a condition which develops when the impedance is minimum. Working back from the shorted end, we note changing minimum and maximum impedance values at points corresponding to 1/4 λ, 1/2 λ, 3/4 λ, and λ at the frequency of operation.

At 1/4 λ back from the shorted end is a maximum impedance point indicated by maximum voltage and minimum current. (If you compare the shorted 1/4-λ line with the open 1/4 λ discussed on the preceding page, the impedance conditions are exactly reversed.) At the 1/2 λ point the shorted line presents

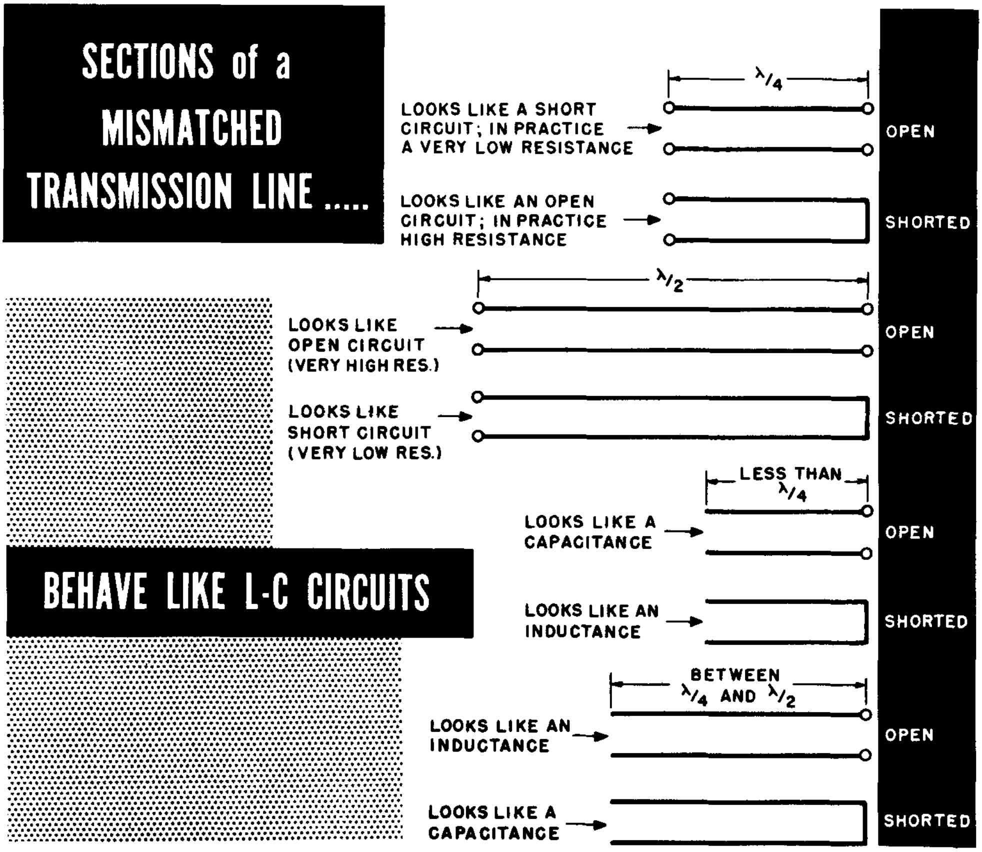


minimum impedance, a duplicate of the conditions at the shorted end. At a point 3/4 λ back from the shorted end the electrical condition is a repetition of that which exists 1/4 λ back from the shorted end. A full wavelength (λ) back, the impedance again is minimum, as at the 1/2 λ point.

Summarizing the electrical conditions described above, we can say that odd multiples (3, 5, 7, etc.,) of the 1/4-λ shorted line repeat the electrical conditions encountered at the 1/4 λ point, accompanied by 180° phase changes in the voltage. Also, shorted transmission lines containing whole number multiples (2, 3, 4, 5, etc.,) of 1/2 λ at the operating frequency duplicate the electrical conditions encountered in the 1/2-λ shorted line, accompanied by 180° phase changes in the current.

Properties of Transmission Line Sections

Although the main purpose of a transmission line is to deliver energy from one place to another, sections of an r-f transmission line can, under proper conditions, be used to simulate tuned circuits, circuit elements C and L, phase inverters, impedance transformers, and other similar devices. A

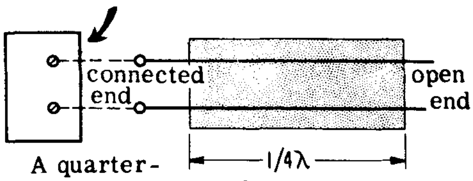


tuned circuit displays certain forms of behavior at and around the frequency to which it is tuned; so will a resonant transmission-line section when the physical length of the section corresponds to the desired part of an electrical wavelength allowing for the velocity of propagation of the line. The following equations are useful for dimensioning r-f transmission lines:

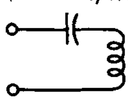
$$\underset{\text{(in inches)}}{1/2\,\lambda} = \frac{5904 \times VP}{f\text{ (mc)}} \qquad \underset{\text{(in inches)}}{1/4\,\lambda} = \frac{2952 \times VP}{f\text{ (mc)}} \qquad \underset{\text{(in inches)}}{1/8\,\lambda} = \frac{1476 \times VP}{f\text{ (mc)}}$$

The action attributed to the section of the line takes place at the "input" end. This is the end opposite to where the line is shorted or open circuited. The device connected to the input end of the line "sees" the electrical condition.

Typical Applications of Transmission-Line Sections



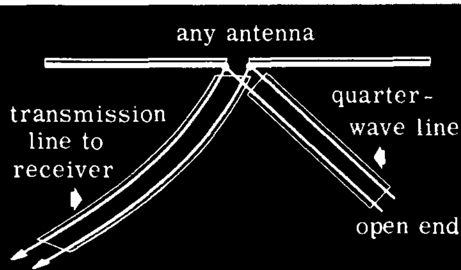
A quarter-wave line behaves like a

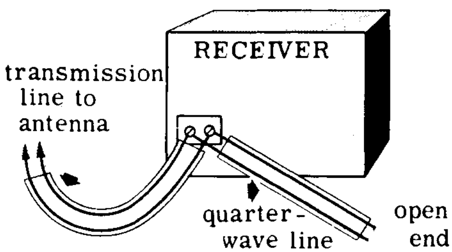


series resonant circuit.

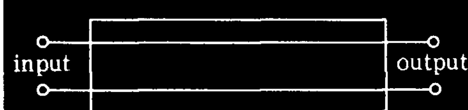
If a quarter-wavelength open line is connected across any source of signals, the source will "see" a very low impedance at the frequency for which the line was dimensioned.

A quarter-wave open line connected to an antenna will "short circuit" the antenna over a narrow band of frequencies around the frequency for which the line was dimensioned.



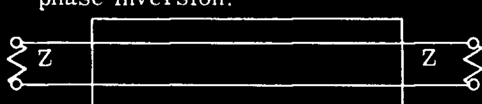


A quarter-wave open line cut for an interfering signal frequency and connected to the antenna terminals of a receiver will act as a short circuit (trap) over a narrow band of frequencies around the interfering frequency.



A half-wave open line acts as a 1-to-1 impedance transformer,

and a half-wave line terminated in a resistance or impedance Z will display the same impedance at the input, accompanied by a 180° phase inversion.

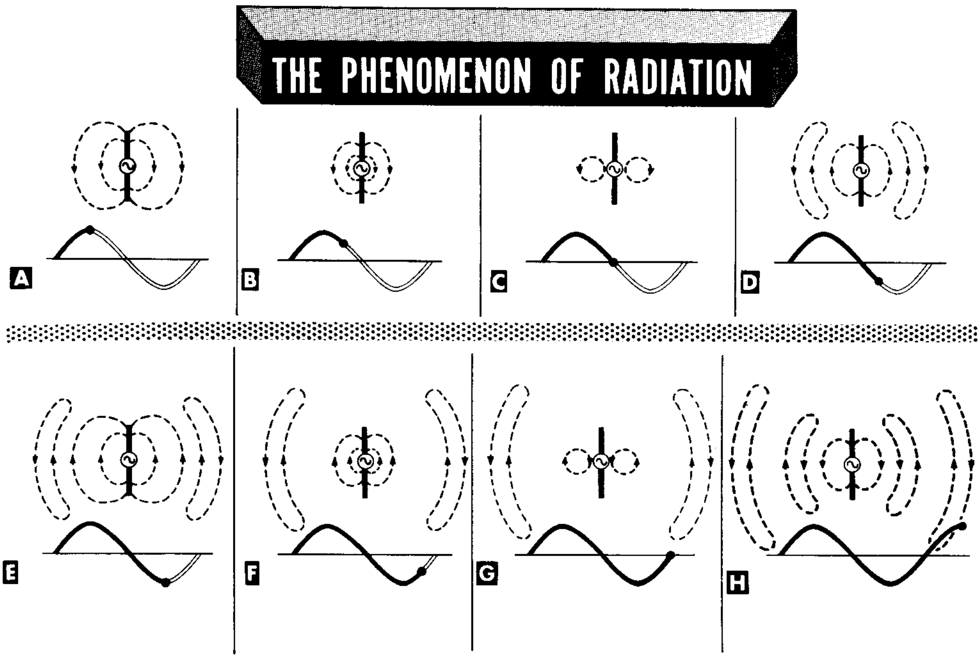


A few broad interpretations of the action of transmission-line sections are shown above. There are, of course, many others, several of which appear in the antenna section in this volume.

Radiation

The mechanism whereby a transmitting antenna liberates r-f energy delivered to it from a transmitter, is called radiation. For radiation to occur, changing current must flow in the antenna. The energy-converting system (antenna) must have capacitance with open rather than confined electric lines of force. The coaxial transmission line is an excellent example of a carrier of electromagnetic energy with confined electric and magnetic lines of force.

Although the action shown is very much simplified, and related to a particular type of antenna known as a dipole, the sequence of action is the same for all types. Only the electric loops of force are shown as a matter of convenience. Magnetic loops of force are also generated simultaneously, these being positioned at right angles to the electric loops of force. Lines of electric



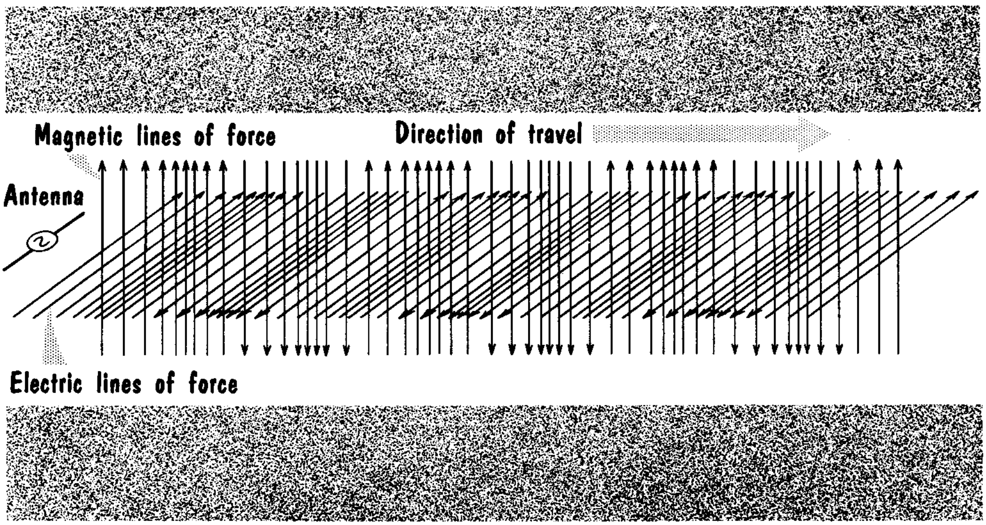
force having one direction of action build outward during the rise in voltage in an alternation, then tend to collapse back into the antenna when the voltage falls, but they cannot do so completely before the applied voltage changes its direction and generates new electric lines of force of a different direction. These too, expand outward, and in so doing push some of the previously collapsing electric lines of force away from the antenna, propagating them out into space. When the voltage starts to decrease, the new electric lines tend to collapse into the antenna, but cannot do so completely before the new cycle of voltage begins, hence these also are propagated into space. Thus the energy corresponding to the current and voltage fed to the antenna moves out into space as inseparably associated electric and magnetic loops of force.

Electric and Magnetic Fields in Free Space

The electromagnetic wave or radio wave radiated into free space consists of inseparably associated loops of electric and magnetic forces. Where they exist is called an electromagnetic field.

Unfortunately, there is no way of picturing energy; all we can do is to show a representation of where the energy is, and its direction of action. It is customary to show the electromagnetic field moving in free space in greatly

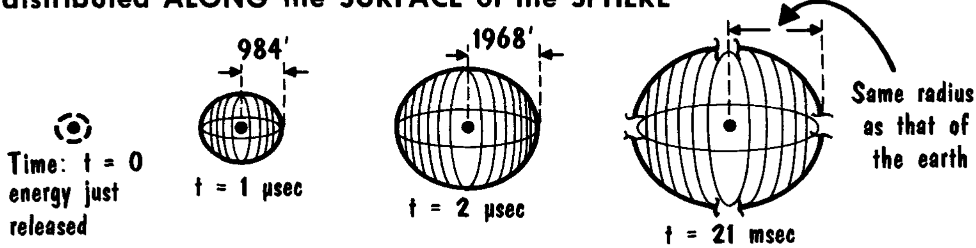
REPRESENTATION of the PROPAGATION of HORIZONTALLY-POLARIZED ELECTROMAGNETIC ENERGY



simplified fashion by imagining that an observer in space "sees" tiny segments of the moving electric and magnetic loops of force passing by him. They appear as straight lines positioned at right angles to each other. The plane of the electric lines of force relative to the horizontal surface of the earth indicates the polarization of the radiation. In the illustration, the radiation is horizontally polarized, in which case the magnetic lines of force are perpendicular to the earth. Horizontally polarized waves are radiated by horizontally positioned antennas. In vertical polarization, the vertically polarized wave has its electric lines of force perpendicular to the earth's surface, and its magnetic lines of force parallel to the earth's surface. It is radiated by antennas which are similarly positioned. The changes in the direction of action of the lines of force of the field correspond to the changes in polarity of the driving voltage and current in the antenna which radiated the energy. Of greatest significance in the presentation of the electromagnetic field moving in free space is the fact that the directions of the electric and magnetic field lines of force are mutually perpendicular (differ by 90°) in space, but are in phase in time (both pass through maximum and zero together). Both are at right angles to the direction of advance of the energy.

The Free-Space Wave (Velocity, Frequency, and Wavelength)

The RADIO WAVE of a PULSE leaves a POINT SOURCE as an EXPANDING SPHERE of ENERGY — with the energy uniformly distributed ALONG the SURFACE of the SPHERE



Note: Velocity of 3×10^8 meters per second = 984 feet per μsec

Electromagnetic energy has a constant velocity in free space – the velocity of light. It is approximated at 299,744,000 meters per second, roughly 300,000,000 (3×10^8) meters per second, or 186,000 miles per second. When explaining propagation we assume a theoretical point source of radiation from which the energy spreads out uniformily in all directions, just as if it were an expanding sphere with the energy lying along the surface. Inasmuch as the energy delivered to the antenna, and from the antenna to space is fixed in amount every second, only this much energy is available to be spread along the surface of the expanding sphere. Hence, as the sphere gets bigger as the energy is propagated, the amount of energy lying along any unit area of the sphere, decreases. Conforming with the geometric law that the surface area of a sphere increases as the square of its radius, the signal strength at any sampling point decreases inversely as the square of the distance to the transmitter. If the distance is doubled the signal strength decreases to one-fourth; if it is tripled, it decreases to one-ninth.

The frequency of the radiated wave is determined by the frequency of the voltages and currents applied to the radiating antenna. Concerning the electrical wavelength of the signal in free space, the term has exactly the same meaning as for the electromagnetic wave moving through a transmission line, with one difference. This is that the velocity of propagation of free space is 1, whereas for all other media it is some fraction of 1.

The Relationship Between VELOCITY (V), FREQUENCY (f), and WAVELENGTH (λ)

$$V \text{ (meter/second)} = f \text{ (cps)} \times \lambda \text{ (meters)}$$

$$f \text{ (cps)} = \frac{\text{velocity (meter/second)}}{\lambda \text{ (meters)}}$$

$$\lambda \text{ (meters)} = \frac{\text{velocity (meter/second)}}{f \text{ (cps)}}$$

Example: A 20 mc signal in free space

$$V \text{ (meter/second)} = 20{,}000{,}000 \times 15 = 300{,}000{,}000 \text{ meter/second}$$

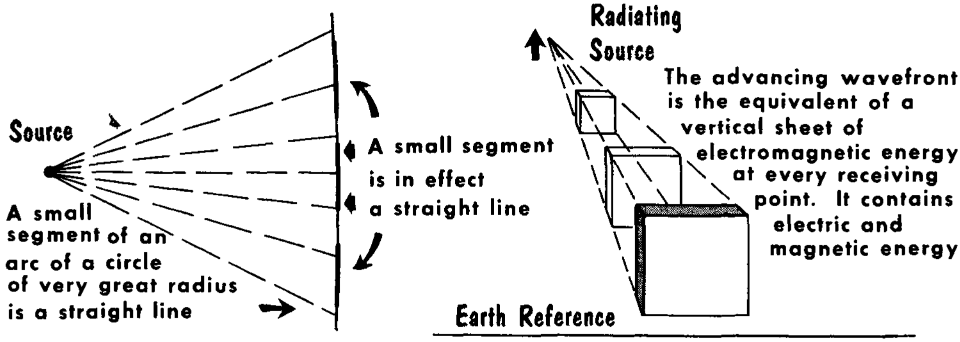
$$f \text{ (cycles)} = \frac{300{,}000{,}000}{15} = 20{,}000{,}000 \text{ cps}$$

$$\lambda \text{ (meters)} = \frac{300{,}000{,}000}{20{,}000{,}000} = 15 \text{ meters}$$

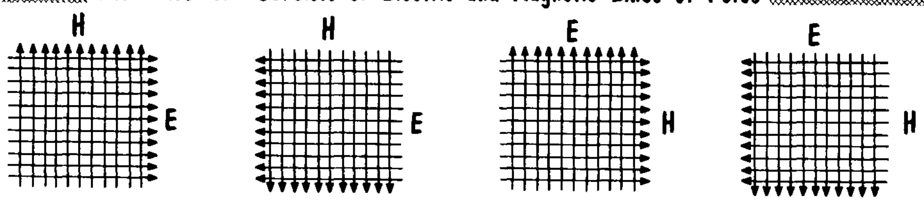
The Wavefront

Although we have spoken of spherical propagation of electromagnetic waves, an observer located in space (or at a receiving antenna) would "see" the radio wave approaching as a transverse plane surface or sheet of radio energy perpendicular to the direction of advance. We call the arriving sheet of energy the wavefront. Signal energy arriving at any receiving point arrives there as a wavefront. The wavefront is considered a plane rather than a

A RADIO WAVE APPROACHES as a SHEET of ENERGY



The Wavefront Consists of Electric and Magnetic Lines of Force



The electric lines of force are the equivalent of voltage in space

curved surface because any tiny segment of a circle or sphere of great radius is for all intents and purposes a straight line or plane.

If we could see the makeup of the wavefront, it would consist of electric and magnetic lines of force (E and H lines) at right angles to each other. If at one instant the E lines pointed to the right, the H lines would point upward; if the E lines pointed to the left, the H lines would point downward. Polarization of the radiated energy is indicated by the plane of the E lines in the wavefront; horizontal E lines indicate horizontal polarization, whereas vertical E lines indicate vertical polarization. In either case the E lines are the equivalent of voltage in space, which if given the opportunity to act on an object lying in their path will induce a difference of potential (a voltage) in the object. Metallic objects (antennas) are subject to the maximum flow of current because of the voltage induced in the antenna. Horizontally positioned antennas are subject to the greatest difference of potential by horizontally polarized E lines, vertical antennas by vertical E lines in the wavefront.

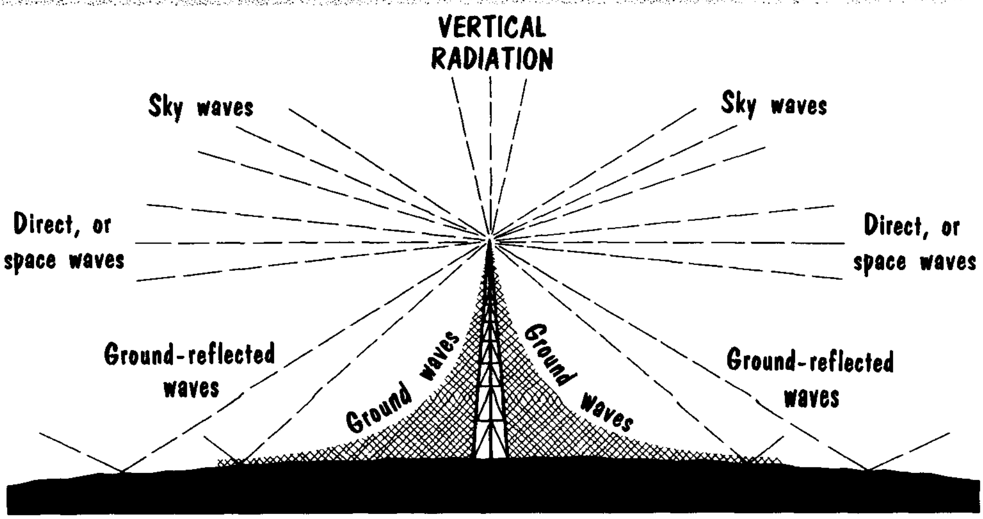
Practical Wave Propagation

Although radio propagation takes place in all directions from a theoretical antenna, the distribution of radiated energy from the practical antenna is not always uniform in all directions. Radio energy leaves the practical antenna in a number of significant directions which greatly influence the ability to communicate by this means. In showing the radiation we use the ray technique; actually the radiated energy fills in the spaces between the rays.

Some of the radiated energy is directed upward from the antenna. This is vertical radiation, and is generally of limited use. Some of the energy (sky waves) leaves the antenna at such angles relative to the surface of the earth that it advances towards the sky and contributes a great deal to long distance communication. Some of the energy (direct or space) leaves the antenna parallel to the earth's surface for perhaps 30 to 75 miles. Then, some of the energy (ground-reflected) leaves the antenna at such an angle as to strike the earth at many places within 20, to perhaps, 75 miles from the antenna, bounce off the earth, and advance upward. Finally, some of the energy leaves the antenna and travels along the earth's surface – even penetrating the surface somewhat. This radiation is called the ground wave or surface component.

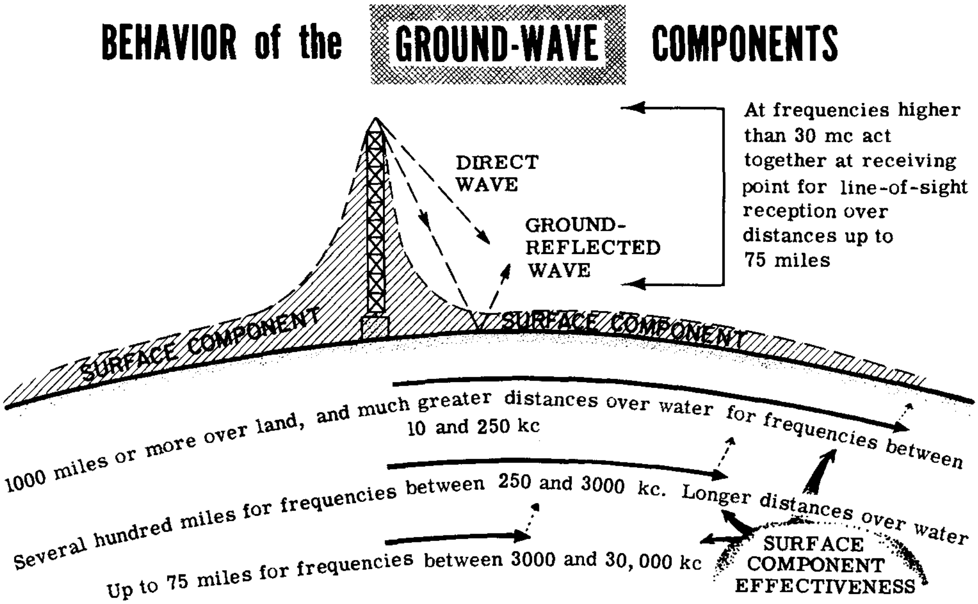
The usefulness of these radiations is determined by numerous factors. Frequency is very important. It determines, for example, the extent to which the ground wave is useful; it also determines which of the radiations directed towards the sky will return to earth, and which will be lost in space. (The time of the day and the season of the year are important.) It also determines the distance over which the direct and ground-reflected waves are useful.

Radio energy that leaves the antenna is identified in terms of its direction of advance



Ground, Ground-Reflected, and Direct Waves

It is generally felt that the ground wave embraces all radiations other than the sky waves and vertical radiation. Taking this approach, ground-wave radiation consists of three components: the surface component, which moves along the surface of the earth with the bottom edge of its wavefront in contact with the surface of the earth, in which it induces minute earth currents; poor



conductivity, dry, sandy soil tends to attenuate the earth currents thus limiting the range of communication. Moist soil or salt water are high-conductivity paths which aid propagation, and therefore extend greatly the distance that can be spanned by this radiation. The most useful frequencies are those up to about 3 mc, although the surface component is effective over shorter distances at frequencies up to 30 mc.

When the frequency of radiation is higher than 30 mc, and the distance to be spanned is within about 75 miles, the signal reaching the receiving antenna is a combination of the direct and the ground-reflected waves. This is called line-of-sight reception because the transmitting and receiving antennas can usually "see" each other. Two phenomena are significant in line-of-sight reception. The direct and ground-reflected waves leave the transmitting antenna with the same signal phase but travel different paths to the receiving antenna. These paths may be of different length. Because the ground-reflected signal suffers a 180° phase reversal at the point of reflection, it is conceivable that the two signals may aid or oppose each other in the receiving antenna. The resultant signal may be stronger or weaker than the direct-path signal alone. Varying the height of the receiving antenna can improve a weakened signal situation.

Line-of-Sight Propagation: Temperature Inversion

The effectiveness of line-of-sight propagation depends on the relative heights of the transmitting and receiving antennas, that is, the ability of the two devices to "see" each other. The amount of power radiated by the transmitting antenna will have a great influence on the distance covered. Transmitters rated at 1 or 2 watts or less will seldom span these distances. Sometimes, they will be effective only over a mile, or possibly less. Extremely sensitive noise-free receivers greatly extend the range. An approximation of the optical horizon or useful range of transmission can be made by applying the equation:

$$\begin{array}{c}\text{distance}\\\text{(in miles)}\end{array} = 1.41 \times \left(\sqrt{\begin{array}{l}\text{transmitter}\\\text{antenna height (in ft)}\end{array}} + \sqrt{\begin{array}{l}\text{receiver}\\\text{antenna height (in ft)}\end{array}} \right)$$

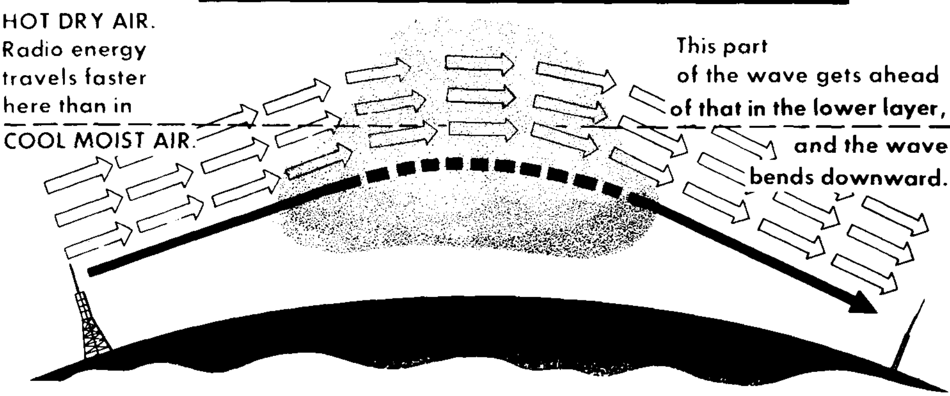
Given a transmitting antenna 400 feet above ground, and a receiving antenna 25 feet above ground, the maximum theoretical useful distance for communication is

$$\begin{aligned}\text{distance (in miles)} &= 1.41 \times (\sqrt{400} + \sqrt{25})\\ &= 1.41 \times (20 + 5) = 1.41 \times 25 = 35.25 \text{ miles}\end{aligned}$$

Temperature Inversion

On occasion, transmission of frequencies above 30 mc is accomplished over many times the normal range. (This distance may be 500 miles or more.)

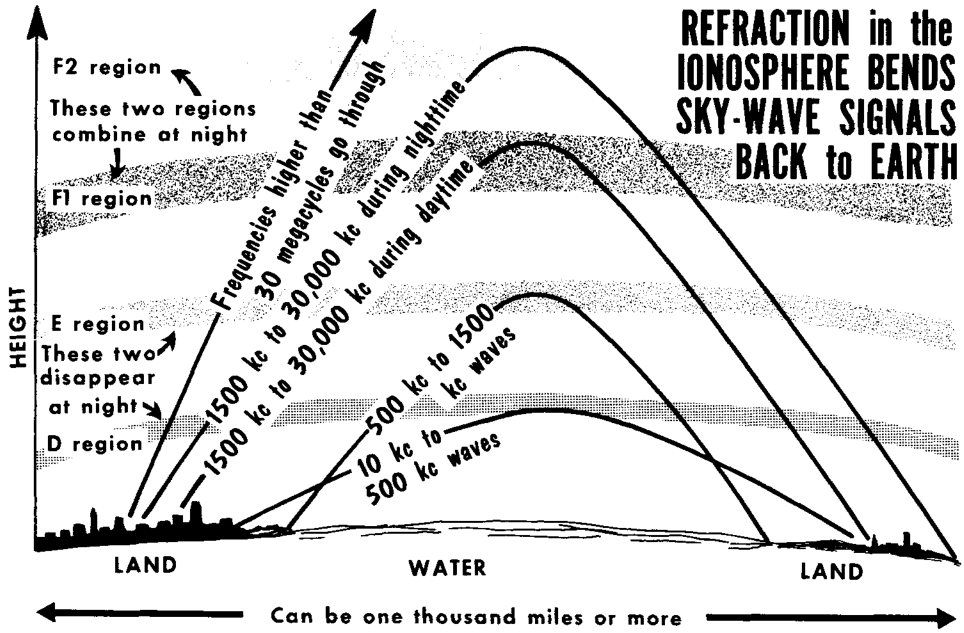
THE EFFECT OF TEMPERATURE INVERSION



The relatively long distance transmission is attributable to a phenomenon known as temperature inversion. At certain times during the year, usually Spring and Fall, sometimes in the summer, layers of hot, dry air may be at higher elevations above the earth than cool, moist air. Radiation that normally would pass over antennas near the ground is bent back to earth, striking receiving antennas far beyond the optical horizon as seen from the transmitting antenna.

Sky-Wave Propagation (The Ionosphere)

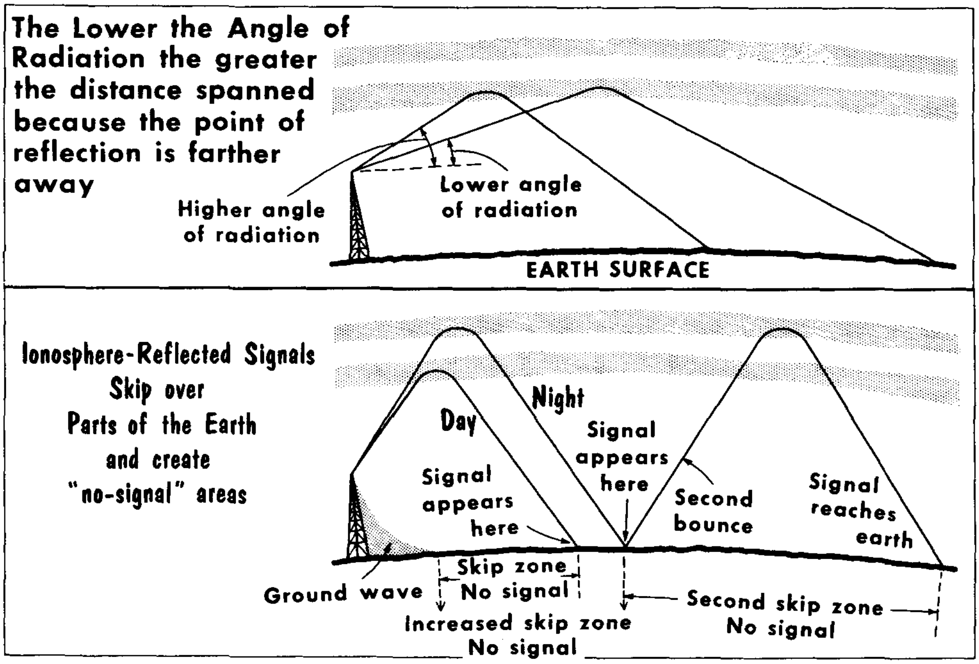
In the early days of radio communication the reasons underlying the reception of signals from long distances were not understood. Two scientists, an Englishman named Heaviside, and an American named Kennelly, suggested that far above the earth as part of our atmosphere, was an electrified layer of gaseous particles which enveloped the earth and reflected radio waves back to earth at a long distance from their point of origin. This became known as the Kennelly-Heaviside layer. Subsequent research established that there was not one layer, but several deep regions of ionized particles which bear the single name ionosphere. Today, we know it as the D, E, F_1, and F_2, regions. The regions exist at various heights above the earth, moving up and



down at different times of the day, week, month, and year. The F_1, and F_2 regions combine at night during winter and summer into a single F region, whereas the D and E regions apparently disappear at night.

The ionized state is produced by bombardment of the upper regions of our atmosphere by gamma rays, ultraviolet rays, electrons, and other particles emitted by the sun, as well as cosmic rays from outer space. By virtue of the variation in density of the atmosphere at different heights above the earth, different degrees of ionization exist during the day and night, at different times of the year. The importance of the changing ionization and the changing height of the ionized regions above the earth is in the way in which they permit sky-wave radiations of certain frequencies which enter these regions to pass through them, while others are bent back (refracted) towards the earth. Sometimes we say that the energy is reflected back to earth.

Angle of Radiation and Skip Distance



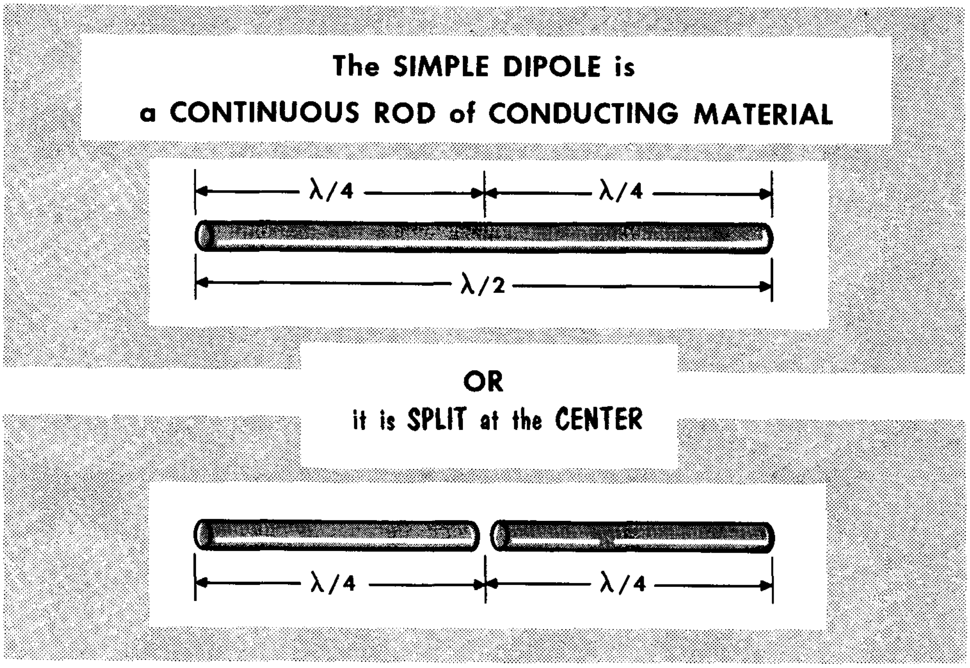
The distance spanned by one "hop" of a radio signal into the ionosphere and back to earth varies according to the angle of radiation of the transmitting antenna, and the height of the reflecting region in the ionosphere. The angle of radiation is the angle made between the center of the wavefront radiated toward the sky by the antenna and a line drawn parallel to the earth's surface at the antenna. The lower the angle of radiation the greater the distance embraced between the point of origin and point of return to earth of the reflected wavefront. The angle of radiation of an antenna of any given frequency is a function of the height of the transmitting antenna above the surface of the earth, the angle becoming lower as the antenna height is increased.

It is characteristic of ionosphere reflected waves that the point of nearest return to earth is at a greater distance from the transmitting antenna than the distance covered by ground-wave radiation. Thus, there is a silent or skip zone between where the surface ground-wave radiation becomes ineffective, and the reflected sky-wave signal is first detected. This skip distance may be many hundreds of miles or more. As the signal frequency is increased, skip distance is increased. Long-distance coverage is also accomplished by multiple hops of the radiated energy. After being reflected back to earth from the ionosphere the first time, the signal bounces off land or sea surface toward the sky, to return to the earth again at a much greater distance. Thus, a signal leaving New York City could reach Los Angeles on the first bounce; Hawaii on the second bounce and farther distances by additional reflections. All radiations reflected to earth again bounce off, but the energy may not always be sufficient to be detected at the second and third points of return.

The Fundamental Antenna

A transmitting antenna is a structure made of metal which liberates (radiates) radio-frequency current and voltage supplied to it as electromagnetic energy into space. A receiving antenna also is a structure made of metal which intercepts electromagnetic energy moving through space, and converts it into voltage and current of corresponding frequency.

The fundamental radio antenna is a metal rod or tubing which has a physical length approximately equal to one-half wavelength in free space at the fre-



quency of operation. Such a structure is known as a half-wavelength dipole, which name is often abridged to half-wave dipole. It is also known as the Hertz antenna. Sometimes the half-wave dipole is called a half-wave doublet. Dipole and doublet mean the same thing as long as the electrical dimensions of the two are the same. A dipole antenna is defined as "a symmetrical antenna in which the two ends are at equal potential relative to the midpoint".

A half-wave dipole (or doublet) is usually positioned horizontally relative to the earth's surface, but, it can, if properly supported, be mounted vertically or obliquely. Another identity given the half-wave dipole is "zero-db gain antenna". This identification is useful only when some other structure used as an antenna affords certain advantages in the concentration of radiation, and is compared with the half-wave dipole. Such concentration of radiation is, as you shall see, the equivalent of having obtained more energy from the transmitter, without actually doing so.

The Half-Wavelength Dimension

The wavelength symbolized by the Greek letter (λ) called lambda, of any signal in free space is given by

$$\lambda \text{ (meters)} = \frac{\text{velocity of propagation (meters per second)}}{\text{frequency of signal (cycles)}}$$

$$= \frac{300,000,000}{\text{(cycles)}}$$

Since the unit meter is not the most convenient to work with, especially when dipole antenna lengths are frequently given in feet and inches, the following equivalents for conversion will allow direct use of the determined dimensions:

$$1 \text{ meter} = 39.37 \text{ inches} \qquad 1 \text{ mc} = 1,000,000 \text{ cycles}$$

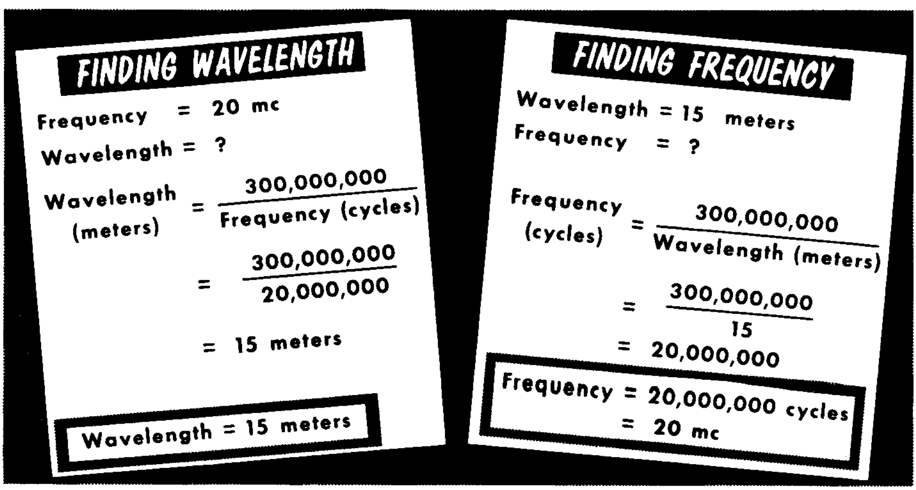
$$\lambda \text{ (inches)} = \frac{300,000,000 \times 39.37}{1,000,000 \times f \text{ (cycles)}} \qquad \frac{300 \times 39.37}{f \text{ (mc)}} = \frac{11,811}{f \text{ (mc)}}$$

$1/2\ \lambda$ is one-half of this value, or, $1/2\ \lambda$ in inches = 5906/f (in mc). If it is desired to find $1/2\ \lambda$ in feet, divide by 12, which leads to $1/2\ \lambda$ (feet) = 492/f (in mc).

Practical conductor materials up to about 1 inch in diameter used for transmitting antennas, allow the movement of electromagnetic energy along the metal at about only 95% of its velocity in free space. Thus, the equation for the $1/2\ \lambda$ dimension in free space is modified by the VP factor 0.95. Then,

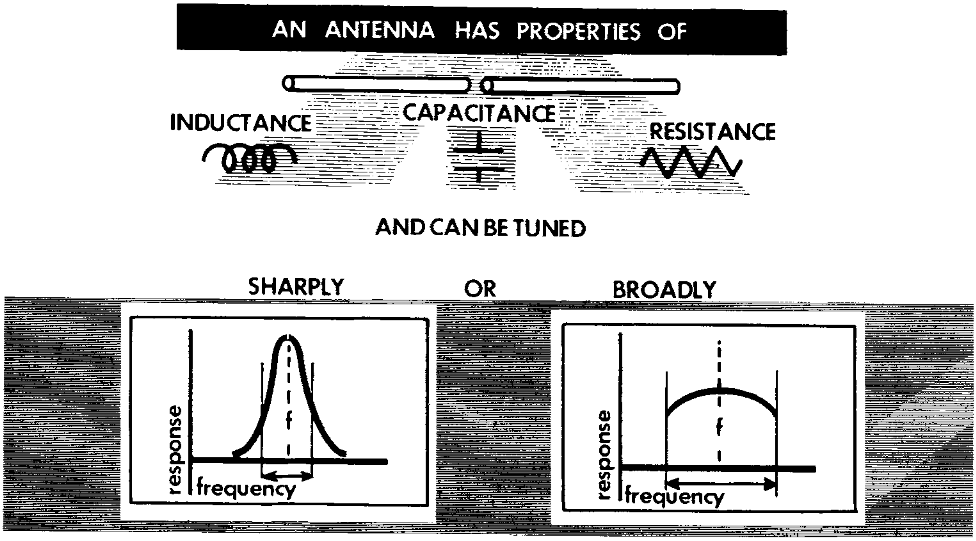
$$1/2\ \lambda \text{ for thin conductors (feet)} = \frac{492 \times 0.95}{f \text{ (mc)}} \qquad \frac{468}{f \text{ (mc)}}$$

$$\text{(inches)} = \frac{5906 \times 0.95}{f \text{ (mc)}} = \frac{5609}{f \text{ (mc)}}$$



Resonance in the Half-Wave Dipole

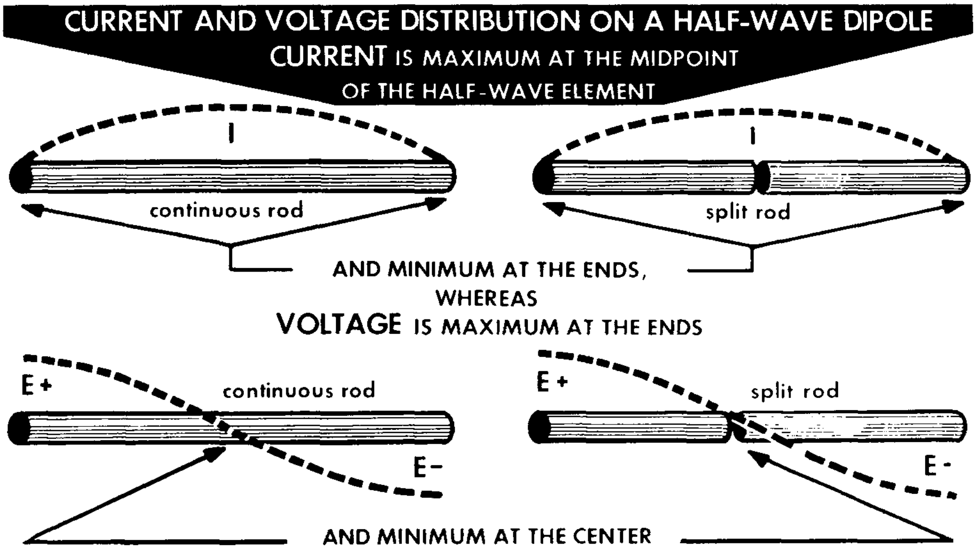
Every half-wave dipole (continuous rod or split) is the equivalent of a resonant circuit. It has distributed L, C, and R. The L is present in the metallic elements when current flows. The C exists between portions of the metallic elements of the antenna, as well as between the elements and the adjacent earth (ground). The R takes the form of electrical losses associated with the r-f currents rather than the d-c resistance of the metal, although this too exists.



OVER A RANGE OF FREQUENCIES ON EITHER SIDE OF THE FREQUENCY FOR WHICH IT WAS CUT.

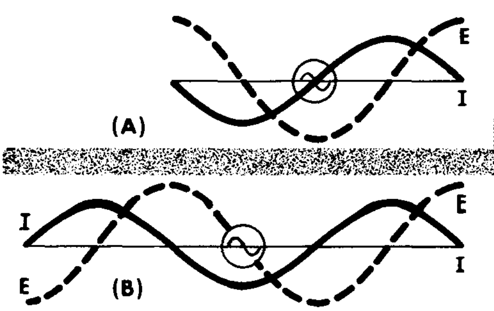
Like the conventional L-C circuit, a half-wave dipole antenna can be made resonant only to one frequency, the frequency for which it has been "cut". This frequency is determined by the L-C constants, principally by L, that is, the length of the antenna elements. To change the resonant frequency, the length of the elements is altered as needed. Although the half-wave dipole is dimensioned for a certain frequency it will, like any resonant L-C circuit, function over a narrow range of frequencies both sides of resonance. While it is impossible to set definite limits of the bandwidth of an antenna, and have it apply to every antenna, we can say that the usual bandwidths embrace several percent of the resonant frequency above and below resonance. Whether the antenna will "tune" broadly or sharply, is a function of its Q, which in turn is a function of the outside diameter of the antenna elements. The larger this diameter, the lower the Q of the antenna, hence, the broader its acceptance bandwidth. Equally important, an outside diameter of 2.5 inches or more can reduce the physical length equal to a 1/2 λ at any given frequency, by as much as 10% of that for a 0.5- or 1-inch diameter element.

Voltage and Current in the Half-Wave Dipole



With very few exceptions all types of transmitting antennas function with standing waves of voltage and current along the elements. Voltage is fed to the antenna and current flows to the open end. It cannot go further, hence becomes zero. The related magnetic field collapses back into the antenna and makes the voltage maximum at the ends. A fixed pattern of standing wave develops in the half-wave dipole as long as energy is supplied to the antenna. In the 1/2 λ metal rod (continuous or split) the standing wave pattern of voltage and current has the voltage maximum at the ends and minimum at the center. The current is maximum at the center and minimum at the ends.

The standing wave of voltage and current is not influenced by the orientation of the antenna; it can be physically positioned horizontally, vertically, or obliquely. Neither is the standing wave pattern influenced by the amount of energy supplied to the antenna; the pattern is the same for small and large amounts of power, and it is the same for small diameter or large diameter elements.



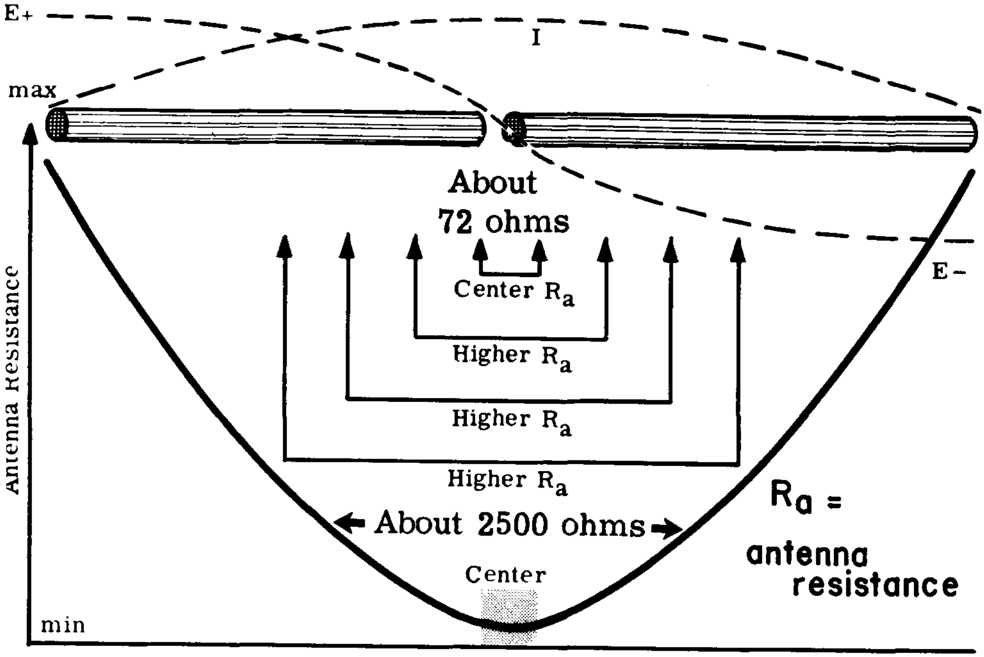
IN a WIRE ONE WAVELENGTH LONG (A) current I is minimum at the center, while voltage E is maximum

IN A WIRE THREE HALF-WAVELENGTHS LONG (B) the current I is maximum at the center while voltage E is minimum

Antenna Impedance

A characteristic of the standing waves of voltage and current on the half-wave dipole (or, on any antenna on which standing waves exist) is that the ratio between E and I, or E/I is not constant along the length of the antenna. The reason is that I and E do not change together. If we use the ratio E/I to express the impedance of the antenna, the impedance is not constant along the length of the element. Assuming a resonant half-wave dipole, the impedance

THE ANTENNA RESISTANCE OF A HALF-WAVE DIPOLE AT RESONANCE IS LOWEST AT THE CENTER AND INCREASES SYMMETRICALLY ON EACH SIDE OF THE CENTER.



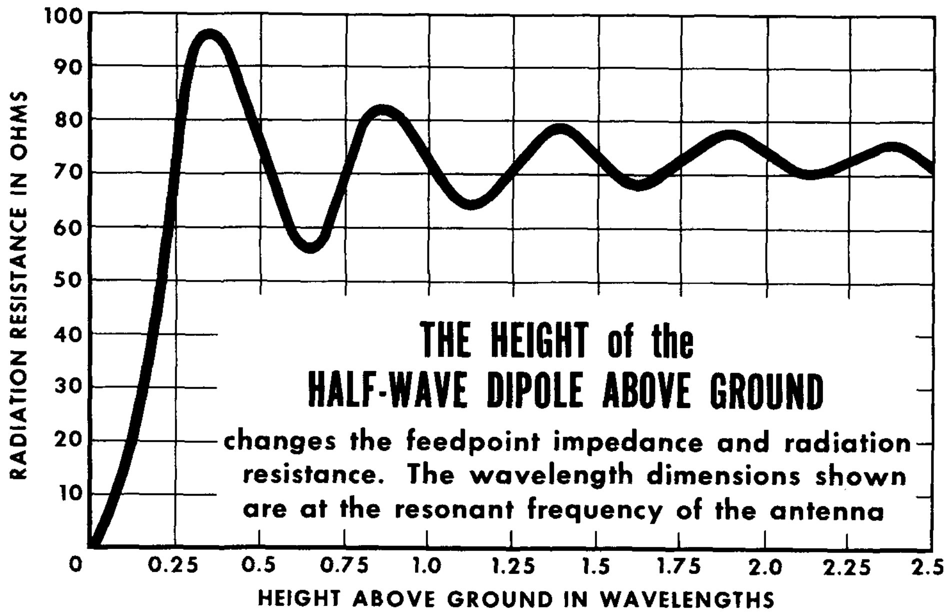
is lowest at the center, this being where the current is maximum. On the other hand, the impedance is maximum at the ends where the voltage is maximum yet the current is minimum. The center of the antenna is considered to be a narrow region, perhaps 1/2 to 3/4 inch each side of the exact middle of the rod.

Most half-wave dipoles are fed at the center because the maximum current point is the minimum voltage point, and it is easier to produce transmission lines for low voltage than for high voltage. Also, when the half-wave dipole is resonant, the capacitive reactance and inductive reactance cancel each other, leaving resistance only as the net impedance. Under such conditions the antenna impedance is resistive between any two points equidistant from the center along the antenna length.

Antenna Input Impedance and Radiation Resistance

The desired electrical relationship between a transmitter and an antenna is when as much as possible of the power generated in the transmitter is delivered to the antenna. Similarly, it is desirable to deliver to the receiver the maximum possible amount of energy intercepted by the antenna. For this to happen, the "input" to the antenna must match the characteristic impedance of the transmission line, which in turn is assumed to match the output impedance of the transmitter. Thus, the important aspect of antenna impedance is the impedance at the point where the transmission line is connected. This point is called by several names, antenna input impedance, feedpoint impedance, and radiation resistance. Of these three names the first two are self-evident. Radiation resistance is defined as a fictitious resistance which when substituted for the antenna would consume as much power as the antenna radiates.

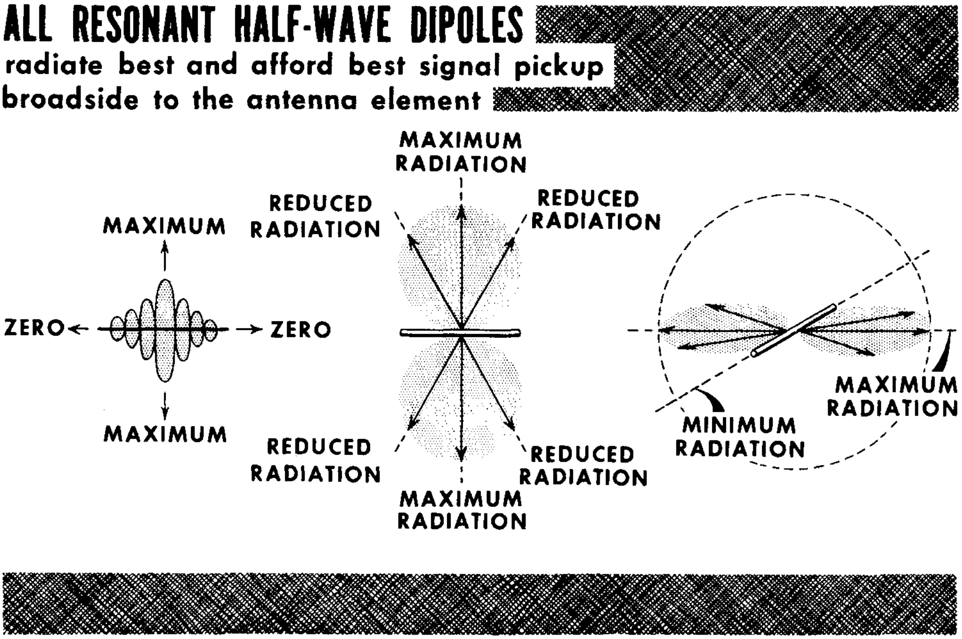
In a 1/2 λ resonant dipole, the antenna feedpoint impedance has a theoretical value of about 72 ohms. This is for an infinitely thin antenna in free space.



If it is a conventional element up to about 1-inch diameter, and located fairly close to ground, its value will fluctuate between about 60 and 90 ohms depending on the antenna height above ground. If the diameter of the antenna element is 2 or 3 inches, the feedpoint impedance can be as low as 35 ohms. While we say that the feedpoint impedance or the radiation resistance is purely resistive, this is true only when the wavelength of the dipole corresponds to the frequency of operation. Fortunately, however, there is enough latitude to permit the antenna to accept energy from the transmitter over a narrow band of frequencies each side of the resonant frequency.

Directivity of Resonant Half-Wave Dipole

While a point-source radiator located in space radiates equally in all directions, this is never true with practical antennas. All antennas radiate more energy in some directions than in others. The directions in which an antenna radiates best is the favorable direction for response as a receiving anten-



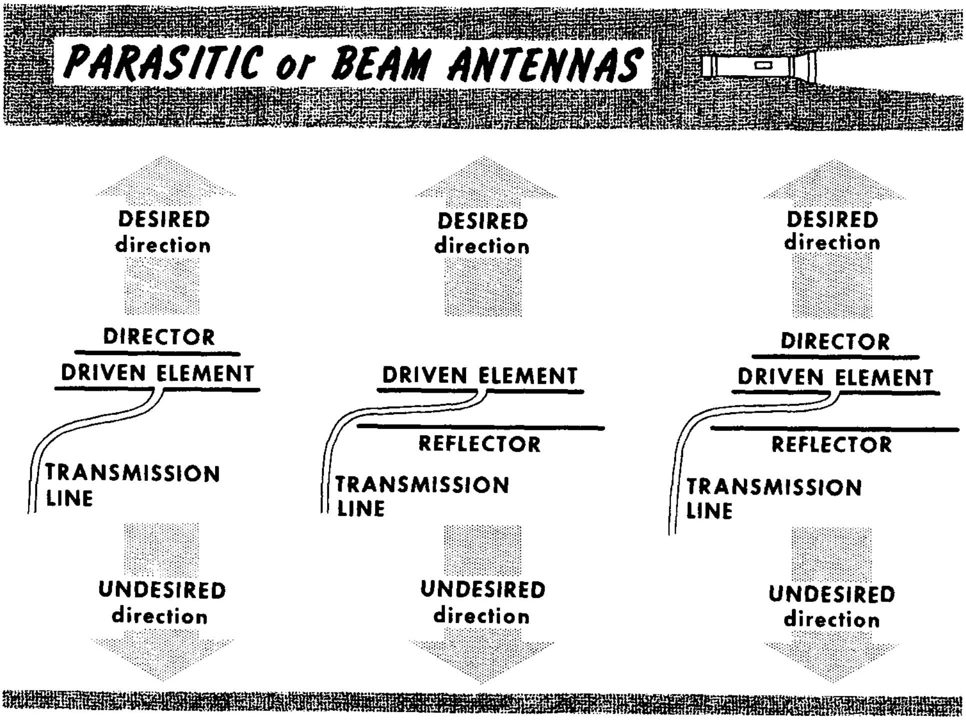
na. A pictorial representation of the pattern of radiation is called the directivity pattern, and applies to reception and transmission. The cause for directional radiation by a resonant half-wave antenna is that the radiation intensity is proportional to the square of the current in the antenna, but since the current is not the same everywhere in the antenna, more energy is radiated from certain parts than from others.

In the resonant half-wave dipole, radiation and signal pickup takes place perpendicular to the long axis of the antenna, with very little radiation (theoretically zero) in line with the long axis. An antenna mounted horizontally in space would have a radiation pattern that resembles a doughnut. If we cut a slice in the horizontal plane through the doughnut, the slice would resemble a figure 8, with the antenna positioned at the crossover point. This is the characteristic radiation and signal pickup pattern of a horizontally positioned resonant half-wave antenna. If the long axis pointed east-west, maximum radiation and signal pickup would be north and south, and vice-versa. It is standard practice to orient horizontally-positioned antennas so that an imaginary line drawn normal from the midpoint of the antenna element would point in the desired direction. The proximity of ground to the antenna affects the radiation pattern by changing the angle of radiation in the vertical plane.

Parasitic Arrays (Yagi Antennas)

The parasitic array is a way of using resonant half-wave dipoles to concentrate radiation in a desired direction and minimize radiation in undesired directions. This affords gain in the antenna. The reference is the half-wave resonant dipole previously identified as the zero db gain antenna. If an antenna design is said to afford a gain of 3 db, the result relative to radiated energy in a chosen direction is the same as if the zero db gain resonant half-wave dipole had been used with twice as much power obtained from the transmitter. (3 db = 2 time increase in power; 6 db = 4 time increase in power; 10 db = 10 time increase in power.) An antenna with gain tends to concentrate the radiated energy into a narrower beam, therefore the antenna must be pointed more accurately.

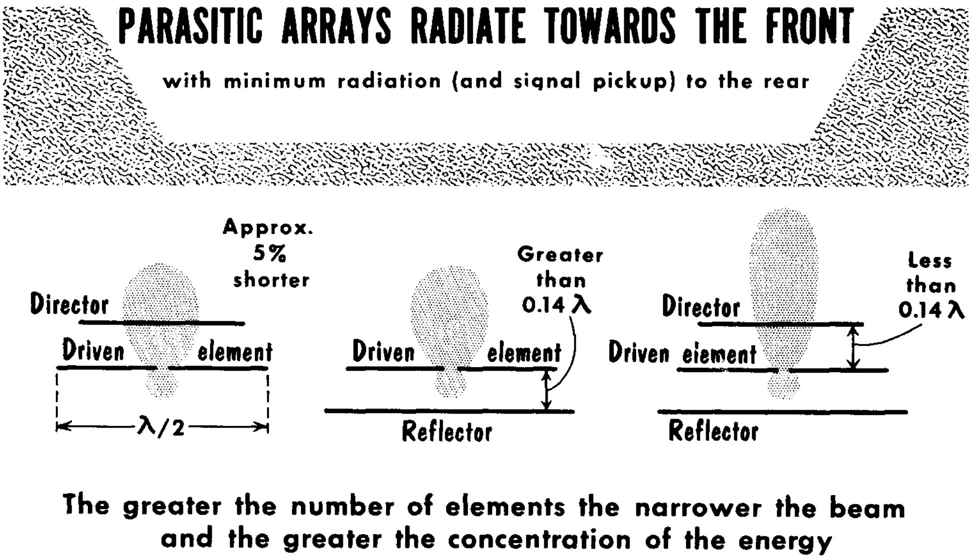
A parasitic array or beam antenna contains several elements: a resonant half-wave dipole that receives power from the transmitter, or which delivers



the received power to the receiver, and is called the driven element; also, one or more continuous metal rods (the parasitic elements) which are parallel to the driven element at the same "line-of-sight" level. The rods are electrically coupled to, but not connected to, the driven element. The rod in front of the driven element is the director. Sometimes, two or more directors are used. The rod in back of the driven element is the reflector, of which only one is used. Frequently both director and reflector are used with the driven element in the same antenna.

Operation and Characteristics of the Parasitic Array

When treated in terms of transmission, energy is delivered to the driven element. It radiates energy towards the front and the rear. Some of this energy induces current in the parasitic element(s), which in turn reradiate virtually all the energy. By suitable dimensioning of the parasitic relative to the driven element, as well as the electrical distance between them, the electrical energy reradiated by the parasitic is "timed" to reinforce the cur-

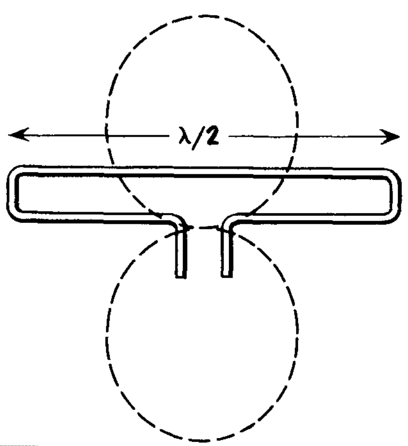


rent in the driven element. It also reinforces the radiation in front of the antenna, while tending to cancel the radiation towards the rear. The result is concentration of the radiation towards the front of the antenna i.e., towards the desired direction. When receiving, the parasitic element(s) and the driven element are acted upon by the approaching wavefront, but not at the exact moment because of the spacing between them. By suitable electrical timing (the spacing and the dimensions of the elements) energy received from the front of the antenna is reinforced in the driven element; energy arriving from the rear is effectively cancelled in the driven element.

Because the current in the driven element is a combination of that received from the transmitter as well as from the parasitic element(s), the feedpoint impedance of the driven element is very much less than the 72 ohms of the resonant half-wave dipole. It can be as low as 25 ohms. Therefore it requires special methods of matching to the transmission line. Also, the behavior of the elements of the antenna as coupled circuits tends to narrow the band of frequencies around the resonant frequency which can be accepted by the antenna. A beam antenna tunes more sharply than a single resonant half-wave dipole. Usually the length of the director is 5% shorter than the driven element, while the reflector is 5% longer.

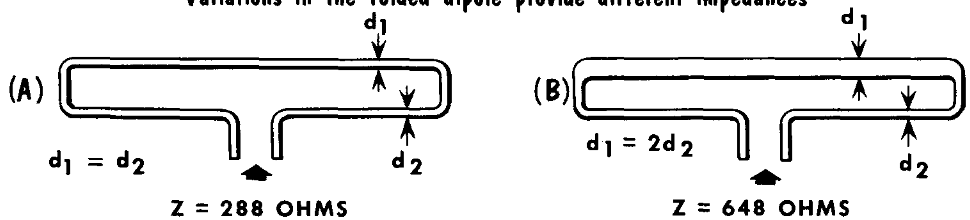
The Folded Dipole

A variation of the conventional half-wave dipole is the folded dipole. In effect it is two 1/2 λ dipoles – one a continuous rod, and the other split in the center – connected in parallel. The transmission line is connected to the split dipole. Its half-wave dimensioning is done exactly as for the ordinary half-wave dipole. Its behavior differs from the conventional half-wave dipole in several respects. The directivity of the folded dipole is bidirectional, but because of the distribution of the currents in the parts of the folded dipole its input impedance is very much higher. If all parts of the antenna are made of like diameter rod or tubing, the feedpoint impedance is 288 ohms. In this



THE FOLDED DIPOLE is bidirectional and a zero-db gain antenna. The amount of power it will accept from a transmitter is the same as for the ordinary half-wave dipole. Both will intercept the same amount of energy from an approaching wavefront.

Variations in the folded dipole provide different impedances



In general: $Z = 72\left(1 + \frac{d_1}{d_2}\right)^2$

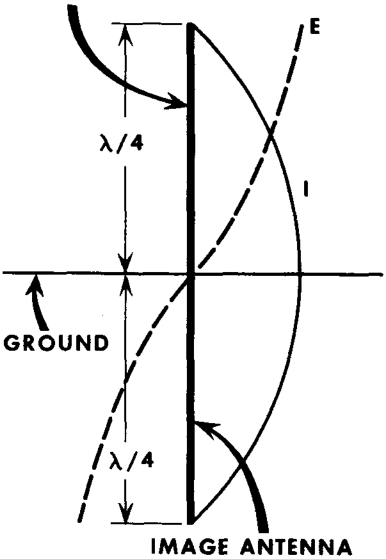
way it is an excellent match for 300-ohm twin-lead (ribbon) transmission line. Another characteristic of the folded dipole is that it is somewhat broader in the band of frequencies around resonance. Sometimes it is used as the driven element in parasitic antennas because the high impedance allows a reduction when used in the beam, while still being suitable for a match with a transmission line rated at from 50 to perhaps 75-ohms impedance.

The Vertical Antenna

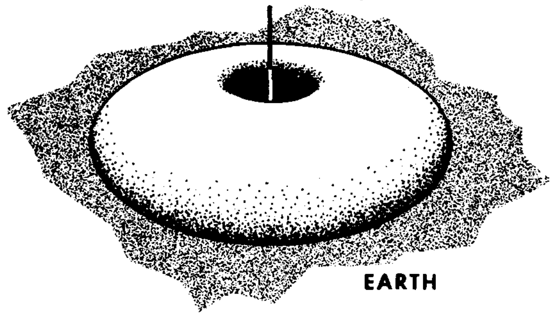
THE QUARTER WAVELENGTH ($\frac{1}{4}\lambda$) VERTICAL ANTENNA

is an omnidirectional antenna. It radiates equally on all sides except upwards in line with the long axis

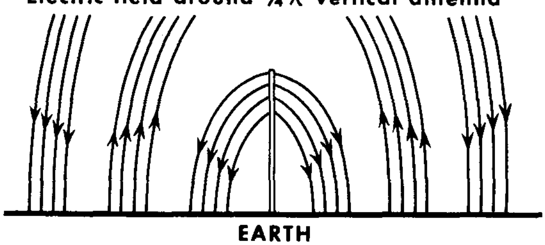
GROUNDED MARCONI ANTENNA



Solid pattern of radiation is equal all around

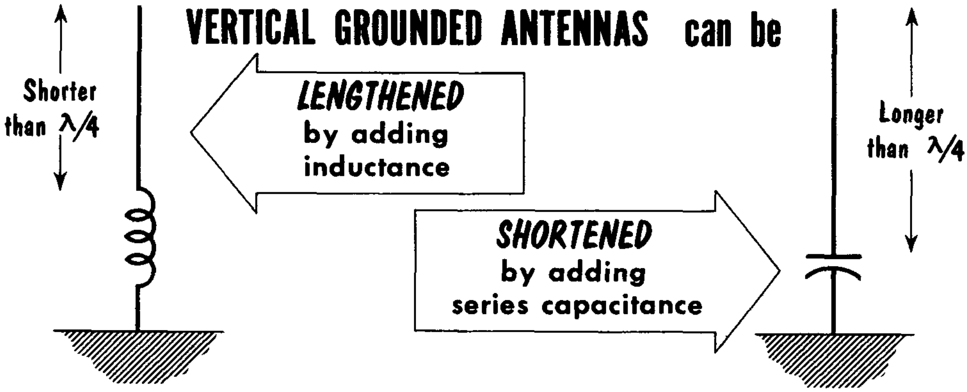


Electric field around $\frac{1}{4}\lambda$ vertical antenna



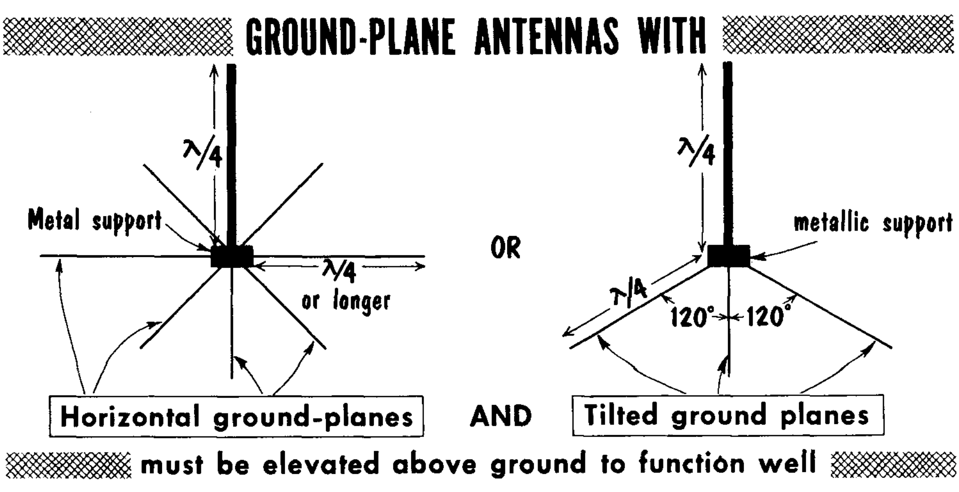
A standard form of antenna is the $1/4\,\lambda$ vertical, also known as the Marconi antenna. It is a vertical rod or tube $1/4\,\lambda$ long at the resonant frequency, and is "grounded" at the bottom. It is used for all frequencies beginning at about 500 kc and extending up into about 150 mc. When used as a "whip" antenna in a vehicle, the vehicle acts as the ground. The antenna radiates a vertically polarized wave at a low angle wherein the electric lines of force are positioned perpendicular to the earth. By connecting the lower end of the antenna to ground directly, the ground presents to the antenna a mirror image of itself. This mirror image is the equivalent of the missing $1/4\,\lambda$, so that the antenna then has a current and voltage distribution like that of the conventional resonant half-wave dipole. Current is minimum at the top of the antenna and maximum at ground, while voltage is maximum at the top and minimum at ground. The feedpoint for the antenna is to a point slightly up from the bottom of the antenna and to ground, the solid conductor of the antenna not acting as a short-circuit because of the resistance present everywhere in the antenna. The feedpoint resistance or radiation resistance of the resonant vertical $1/4\,\lambda$ antenna is about 36 ohms measured at the point of connection to the coupling device to the transmitter or receiver. Grounded vertical antennas which are shorter than $1/4\,\lambda$ suffer from lowering of the feedpoint impedance and radiation resistance. If the antenna is too long, the radiation resistance increases substantially above 36 ohms.

The Vertical Antenna (Contd.)



Quarter-wave vertical antennas are usable over a narrow band of frequencies around resonance. When it is necessary to modify the electrical length the electrically short antenna may be lengthened by adding series inductance; the electrically long antenna may be shortened by adding series capacitance. We said that the feedpoint impedance or radiation resistance of the $1/4\,\lambda$ vertical grounded antenna is 36 ohms; higher impedance is available by advancing upwards from ground along the length of the antenna. Thus, one point might be a feedpoint for a 36-ohm line, while a point somewhat higher up along the antenna would serve as a feedpoint for a 52-ohm line. Of course, ground is one of the connections for the transmission line in both instances.

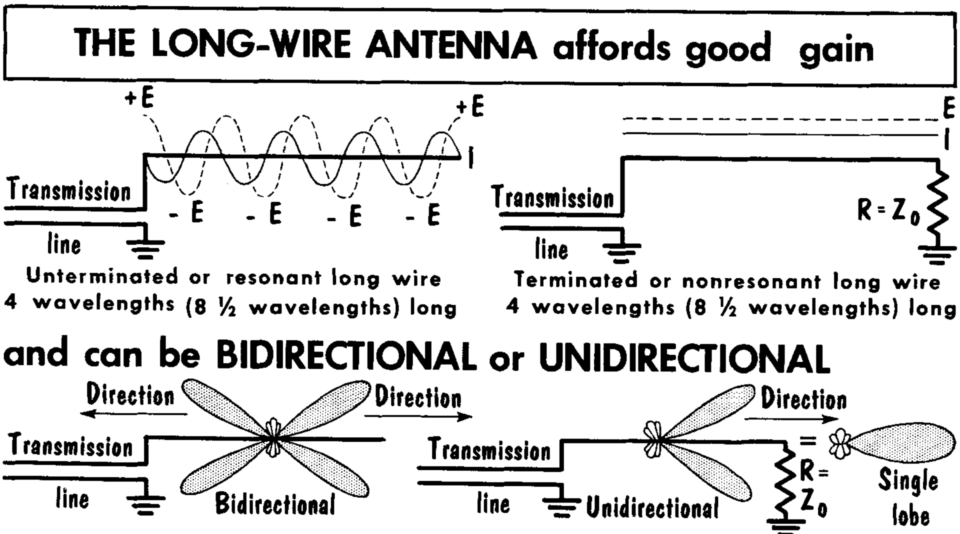
The ground-plane antenna is a vertical radiator for which several straight, or tilted-downward rods extending outwards from the bottom of the radiator, act as the ground. These rods, approximating a $1/4\,\lambda$ are the ground planes. In effect, they are an artificial ground sometimes referred to as a counterpoise, and are used as the ground point for the transmission line cables. The feedpoint impedance or radiation resistance of the ground plane antenna is about 30 ohms.



The Long-Wire Antenna

The long-wire antenna is a single long wire, usually two or more wavelengths (four or more $1/2\,\lambda$) long at the operating frequency. It is also known as an harmonic antenna. The greater the number of $1/2\,\lambda$ the antenna length will accommodate for the operating frequency, the more effective the antenna as to gain over a zero-db gain half-wave dipole, and the better its directivity. It radiates a horizontally polarized wave at relatively low angles, from about 17° to perhaps 25° relative to the earth's surface. The gain varies from about 1.5 db when the antenna is four $1/2\,\lambda$ long to about 4.5 db when the wire is sixteen $1/2\,\lambda$ long. The long-wire antenna is end-fed, with a feedpoint impedance of 500-600 ohms.

Long-wire antennas are of two types – unterminated (resonant), and terminated in the characteristic impedance of the antenna (nonresonant). The resonant antenna has standing waves along its length. The voltage reverses



The equation for determining the physical length (L) of the antenna wire for a number of $\frac{1}{2}\lambda$ is

$$\text{length (feet)} = \frac{492\ (N - 0.05)}{f\ (mc)}$$

where N is number of $\frac{1}{2}\lambda$

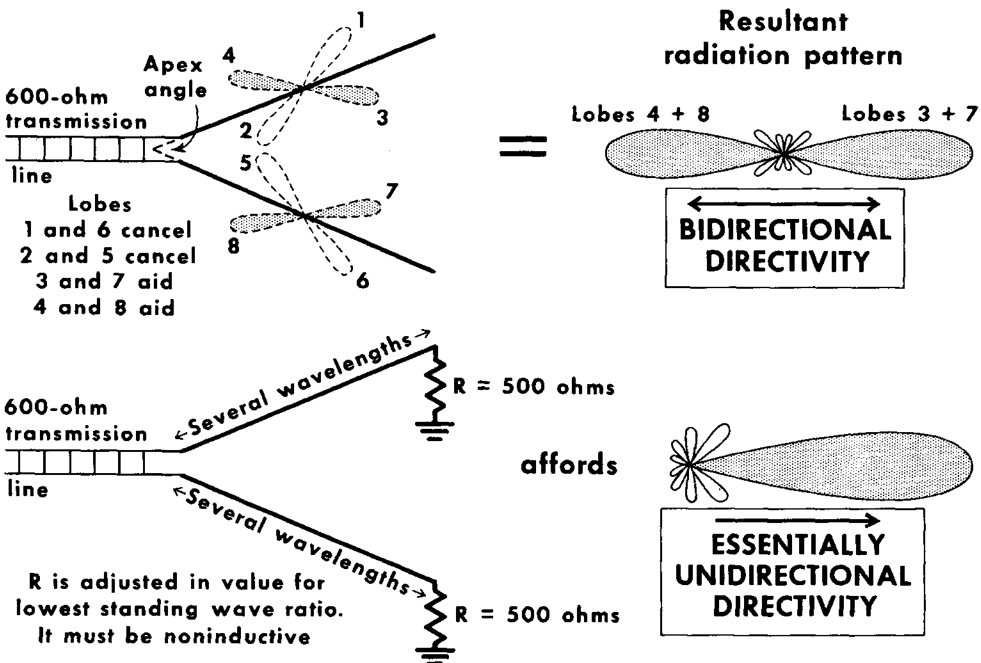
its polarity every $1/2\,\lambda$, and the current reverses its direction. The terminated antenna has substantially uniform current flow in the antenna to the termination. Parallels to these electrical conditions are the resonant and nonresonant transmission lines.

The essential difference in performance between the above two kinds of long-wire antennas is in the directivity. The resonant line is bidirectional at small angles to the long axis of the wire, while the terminated antenna is unidirectional in the direction towards the terminated end of the antenna. When the antenna length is five or more $1/2\,\lambda$ at the operating frequency, some radiation takes place at right angles to the long axis of the antenna.

The V Antenna

The V antenna is a version of the long-wire antenna. It is the equivalent of two long wires (legs) arranged in the form of a flat V, each wire being fed with a voltage 180° out of phase with the other. The advantages of this antenna are gain and directivity. It is achieved by cancellation between oppositely directed corresponding radiation lobes in each leg, and by aiding

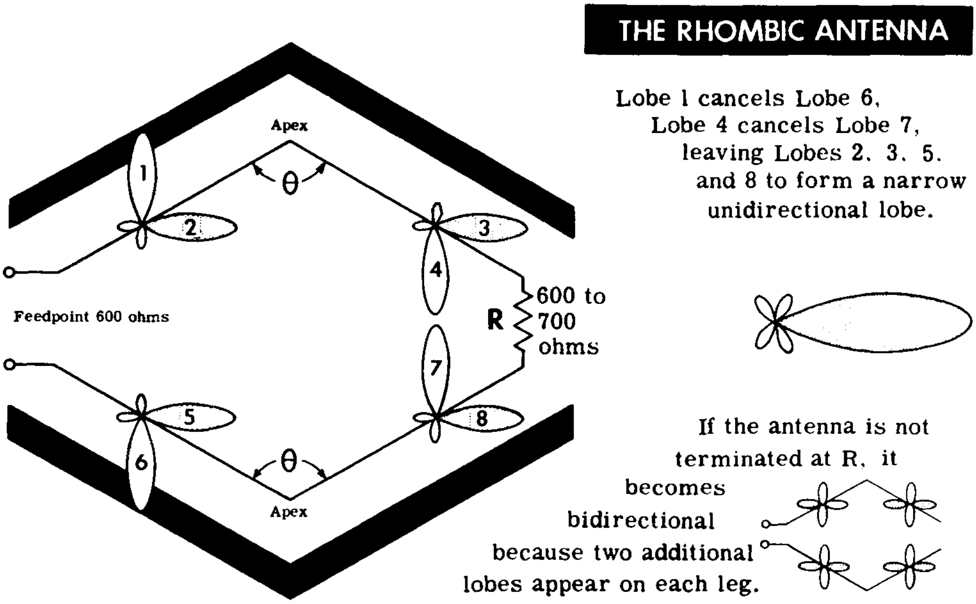
THE V ANTENNA has HIGH GAIN AND A CHOICE of BIDIRECTIONAL or UNIDIRECTIONAL DIRECTIVITY



action between similarly directed corresponding lobes in each leg. The final result is a bidirectional radiation pattern in which the lobes are much narrower than for the single long wire. Gain achieved with the V antenna is about twice that for the single long wire, which has a length equal to that of the legs of the V. Almost 12-db gain over a half-wave dipole is realized when each leg of the V is 8 λ long. The angle made at the apex of the figure is important. This angle varies from about 35° for the 8-λ structure to about 70° for the 2-λ wires. When the antenna is to be used over a wide range of frequencies, the apex angle is made the average between the optimum for the lowest and highest frequencies in terms of the number of 1/2 λ in each leg. At the lowest frequency the legs should be several wavelengths long. If each leg is terminated in a resistance of about 500 ohms the radiation pattern becomes unidirectional.

The Rhombic Antenna

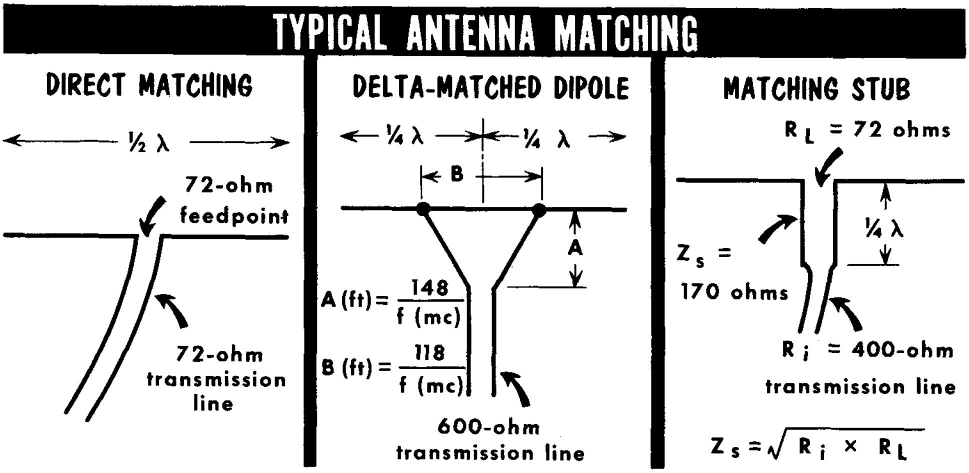
The rhombic is a very efficient antenna of broad frequency capability, and is prominent in all radio communication facilities where the space necessary for the large structure is available. It can be said to be a double V, in which case it becomes a diamond-shaped structure lying in the horizontal plane. It can be terminated or unterminated, the difference between the two being directivity. When used, the terminating resistance is between 600 and 700 ohms, noninductive, and both types are fed by open lines of 600 ohms characteristic impedance. Its advantages become evident when each leg of the rhombic is several wavelengths long, and if the angle θ at the apex is between 50° and 70°. In this event the antenna shows a power gain of from 8 to 12 db over a conventional half-wave dipole. When terminated, the radiation pattern is unidirectional in the direction of the terminated end, and has a very narrow



principal lobe. Because of this behavior the antenna must be laid out so that it is pointing in the desired direction of communication. It can be made bidirectional simply by removing the terminating resistance. It is an excellent antenna for operation over a wide band of frequencies, especially if the leg lengths are between 5 and 6 λ long at the lowest frequency of operation. Little is gained by making each leg longer than about 8 λ.

The highly directional characteristic of the rhombic arises from the fact that two principal lobes develop in each leg of the terminated arrangement, and from the condition that one lobe in each leg cancels the corresponding oppositely directed lobe in the other leg. This leaves four lobes pointing in the same direction, one in each leg. These are additive, and produce a single narrow lobe in the forward direction.

Feeding the Antenna (SWR)



It is extremely important that there be a proper impedance match between the transmission line and the antenna. If the characteristic impedance (Z_o) of the line differs from the feedpoint impedance (R_a) of the antenna, reflections occur at the feedpoint, and standing waves appear on the transmission line. The ratio between the maximum and minimum values of current or voltage on the line is called the standing wave ratio (SWR). The desired SWR is 1, where there is no change in the value of E or I along the line. Acceptable operation is usually possible with a SWR of as much as 1.75 with low-power transmitters. The SWR figure can also be obtained from the ratio of R_a/Z_o. If R_a is 70 ohms and Z_o is 50 ohms, the SWR is 70/50, or 1.4. Unfortunately, the feedpoint impedance is known only under certain conditions, such as at resonance, and for certain types of antennas, or must be measured. When the operating frequency is not at resonance, the feedpoint impedance not only varies greatly, but also becomes reactive. At resonance, the feedpoint impedance is resistive because the reactive effects balance out.

Transmitting antennas can be fed at the center, off center, or at the end. When the antenna contains an odd number of 1/2 λ at resonance, the current maximum (loop) is at the center, and the feedpoint impedance is equal to the radiation resistance, which in turn is a function of the number of 1/2 λ in the conductor. This is so at any point along the antenna where a current loop exists. When the antenna contains an even number of 1/2 λ at the resonant frequency, current loops appear on both sides of center. A feedpoint can be located at any current loop. However, it is only in the case of the half-wave dipole operated at resonance that the impedance at the center is 72 ohms. Antennas containing two or more 1/2 λ can be fed at the end – a high-impedance point. The high transmission-line impedance required for end-feeding can be obtained through a resonant line, or a 1/4 λ open line cut for the operating frequency, to act as an impedance transformer between the transmission line and antenna. With the parasitic array, the feedpoint must be determined by experiment. It is difficult to state a dimension each side of the center of the driven element that will serve as the connecting points for any given Z_o of the transmission line used.

A transmission line is a device for guiding electrical energy from one point to another. Such a line has electrical constants of inductance, capacitance, and resistance distributed along its length.

A line terminated in a resistance equal to its characteristic impedance is said to be terminated correctly.

The ratio of maximum to minimum voltage or current along a transmission line is called the standing wave ratio – this is a measure of the energy reflected.

A properly matched line is nonresonant. It produces no reflection of energy and there are no standing waves – there is maximum transmission of energy.

A delta-matching section used with a two-wire line is made by fanning out the end of the transmission line as it approaches the antenna.

Electromagnetic waves are propagated into space at nearly the speed of light, 300,000,000 meters per second, or 186,000 miles per second.

A radio wave may be described as a moving electromagnetic field, having velocity in the direction of travel.

The wavelength of a radio wave in free space is equal to its velocity divided by its frequency ($\lambda = V/f$).

The energy received at a distant point is the sum of the direct wave and the ground-reflected wave.

The velocity of wave travel on an antenna or transmission line is lower than in free space.

If an antenna is cut to a length of exact resonance, the reactance is zero, and the antenna impedance is purely resistive. If the antenna is made shorter, capacitive reactance is present; if made longer, inductive reactance is present.

A quarter-wave antenna operating in conjunction with ground operates as a resonant antenna.

An array is a combination of half-wave elements operating together as a single antenna. Arrays provide more gain and greater directivity than single-element antennas.

The polarization of a radio wave is determined by the direction of the electric flux lines with respect to the surface of the earth.

REVIEW QUESTIONS

1. What is meant by electromagnetic energy?
2. What is meant by ground, ground-reflected, and direct waves?
3. What is meant by skip distance?
4. What is the function of a transmission line?
5. Name three different types of transmission line.
6. What is meant by a wavelength?
7. What is an antenna?
8. What is meant by radiation resistance?
9. What is meant by the characteristic impedance of a transmission line?
10. What is the function of a parasitic element?
11. Explain the term polarization as applied to an electromagnetic wave.
12. What is the current and voltage distribution on a grounded quarter-wave antenna?

GLOSSARY

Amplitude Modulation: A system of transmission and reception of intelligence in which the amplitude of the carrier is made to vary in accordance with the intelligence.

Antenna: A device used to radiate or absorb r-f energy.

Array: An arrangement of antenna elements, usually dipoles, which results in desirable directional characteristics.

Attenuation: The reduction in the strength of a signal.

Breakdown Voltage: The voltage at which an insulator dielectric ruptures, or at which ionization and conduction take place in a gas or vapor.

Buffer Amplifier: An amplifier used to isolate the output of an oscillator from the effects produced by changes in voltage or loading in following circuits.

Carrier: The r-f component of a transmitted wave upon which an audio signal or other form of intelligence can be impressed.

Characteristic Impedance: The ratio of the voltage to the current at every point along a transmission line on which there are no standing waves.

Coaxial Cable: A transmission line consisting of two conductors concentric with and insulated from each other.

Continuous Waves (CW): Radio waves that maintain a constant amplitude and a constant frequency.

Coupling: The association of two circuits in such a way that energy may be transferred from one circuit to another.

Crystal: A natural substance, such as quartz or tourmaline, which is capable of producing a voltage stress when under pressure, or producing pressure when under an applied voltage. Under stress, it has the property of responding only to a given frequency when cut to a given thickness.

Dipole Antenna: Two metallic elements, each approximately one quarter wavelength long, which radiate r-f energy fed to them by the transmission line.

Director: A parasitic antenna placed in front of a radiating element so that r-f radiation is aided in the forward direction.

Electromagnetic Field: A space field in which electric and magnetic vectors at right angles to each other travel in a direction at right angles to both.

Electrostatic Field: A field of influence between two charged bodies.

Field Intensity: Electrical strength of a field.

Feedback: A transfer of energy from the output circuit of a device back to its input.

Filter: A combination of circuit elements designed to pass a definite range of frequencies, attenuating all others.

Frequency: The number of complete cycles per second existing in any form of wave motion; such as the number of cycles per second of alternating current.

Frequency Distortion: Distortion that occurs as a result of failure to amplify or attenuate equally all frequencies present in a complex wave.

Frequency Modulation: A system of transmission and reception of intelligence in which the carrier-wave frequency is made to vary in accordance with variations in the modulating signal.

Frequency Stability: The ability of an oscillator to maintain its operation at a constant frequency.

Grid Current: Current which flows between cathode and the grid whenever the grid becomes positive with respect to the cathode.

Ground: A metallic connection with the earth to establish ground potential. Also, a common return to a point of zero r-f potential, such as the chassis of a receiver or a transmitter.

Harmonic: An integral multiple of a fundamental frequency. (The second harmonic is twice the frequency of the fundamental or first harmonic.)

Harmonic Distortion: Amplitude distortion.

Interelectrode Capacitance: The capacitance existing between the electrodes in an electron tube.

Ionization Potential: The lowest potential at which ionization takes place within a gas-filled tube.

Ionosphere:A region composed of highly ionized layers of atmosphere from about 70 to 250 miles above the surface of the earth.

Matched Impedance: The condition that exists when two coupled circuits are so adjusted that their impedances are equal.

Modulation: The process of varying the amplitude (AM), the frequency (FM), or the phase (PM) of a carrier wave in accordance with other signals to convey intelligence.

Modulator: The circuit which provides the signal that varies the amplitude, frequency, or phase of the oscillations generated in the r-f portion of the transmitter.

Neutralization: The process of nullifying the voltage fed back through the interelectrode capacitance of an amplifier tube, by providing an equal voltage of opposite phase; generally necessary only with triode tubes.

Node: A zero point; specifically, a current node is a point of zero current, and a voltage node is a point of zero voltage.

Oscillator: A circuit capable of converting dc into ac of a frequency determined by the constants of the circuit.

Oscilloscope: An instrument for showing, visually, graphical representations of the waveforms encountered in electrical circuits.

Parallel Feed: Application of a d-c voltage to the plate or grid of a tube in parallel with an a-c circuit so that the d-c and a-c components flow in separate paths. Also called *shunt feed*.

Parasitic Suppressor: A resistor in an electron tube circuit to prevent unwanted oscillations.

Piezoelectric Effect: The effect of producing a voltage by placing a stress, either by compression, expansion, or twisting, on a crystal, and, conversely, the effect of producing a stress in a crystal by applying a voltage to it.

Plate Dissipation: The power in watts consumed at the plate in the form of heat.

Plate Efficiency: The ratio of the a-c power output from a tube to the average d-c power supplied to the plate circuit.

Plate Modulation: Amplitude modulation of a class-C r-f amplifier by varying the plate voltage in accordance with the signal.

Radiation Resistance: A fictitious resistance which may be considered to dissipate the energy radiated from the antenna.

Reflector: A metallic object placed behind a radiating antenna to prevent r-f radiation in an undesired direction and to reinforce radiation in a desired direction.

Screen Dissipation: The power dissipated in the form of heat on the screen grid as the result of bombardment by the electron stream.

Series Feed: Application of the d-c voltage to the plate or grid of a tube through the same impedance in which the ac flows.

Standing Wave: A distribution of current and voltage on a transmission line formed by two sets of waves traveling in opposite directions, and characterized by the presence of a number of points of successive maxima and minima in the distribution curves.

Thermocouple Ammeter: An ammeter that operates by means of a voltage produced by the heating effect of a current passed through the junction of two dissimilar metals. It is used for r-f measurements.

Unbalanced Line: A transmission line in which the voltages on the two conductors are not equal with respect to ground: for example, a coaxial line.

Wave: Basically, an electromagnetic impulse, periodically changing in intensity and traveling through space. Also, the graphical representation of the intensity of that impulse over a period of time.

Wavelength: The distance, usually expressed in meters, traveled by a wave during the time interval of one complete cycle. It is equal to the velocity divided by the frequency.

Wave Propagation: The transmission of r-f energy through space.

CUMULATIVE INDEX

NOTE: The first number of each entry identifies the Volume in which the information is to be found; the second number identifies the page.