basic electronics

by VAN VALKENBURGH,
NOOGER & NEVILLE, INC.

VOL. 6

JOHN F. RIDER PUBLISHER, INC., NEW YORK
a division of HAYDEN PUBLISHING COMPANY, INC.

THIRD PRINTING 1963

Library of Congress Catalog No. 55-6984

Printed in the United States of America

PREFACE

In its relation to Basic Electricity and Basic Electronics, the subject of transistors and FM fundamentals marks the most important addition to the Common-Core program since its inception. Beginning in 1953, when it was first adopted, and persisting until the present time at U.S. Navy specialty schools, the Common-Core program in Electricity and Electronics has been the means for instruction of over 100,000 Naval technicians. Results have been so outstanding as to attract the attention of the world.

Consequences are not limited to Naval technician training. Since release to civilians in 1954, thousands of industrial and commercial personnel, high school, college, and "after-hours" students have enjoyed like benefits and, as an implementing medium for the National Defense Education Act of 1958, these training tools will continue to improve this country's technical education and accelerate its growth.

Overflowing the boundaries of the United States and Canada, the Common-Core program has also taken solid root overseas. Many military, as well as industrial and commercial, organizations and civilian educational school systems have seen its advantages and adopted this program in various translated forms as a ready-made stepping stone for raising the level of technical education in foreign countries.

The Common-Core program has been described as pointing the way to the future and constitutes a distinct departure from the "classical" methods of technical education and textbook production. Its primary objective is to prepare a person to apply technology in a real job situation, and it concentrates on making technical education interesting and comprehensible, without the usual prerequisites of higher mathematics, physics, and the like.

The free world must foster technical education among its people, and this simple volume on transistors, solid state physics, and

FM techniques is an efficient medium for the extension of working knowledge on these subjects in those countries where the Common-Core program has been introduced.

A number of the processes, devices, and circuit arrangements described in this book are proprietary. The fact that they have been included does not imply that information is available, without permission, for use in design, manufacture, or sale. It is presented here solely for educational purposes. No one having part in the preparation or publication of this volume will be responsible for any liability resulting from the unlicensed use of this material.

For authorization to reproduce their designs, special appreciation is extended to the Radio Corporation of America and to the Heath Company, subsidiary of Daystrom Incorporated.

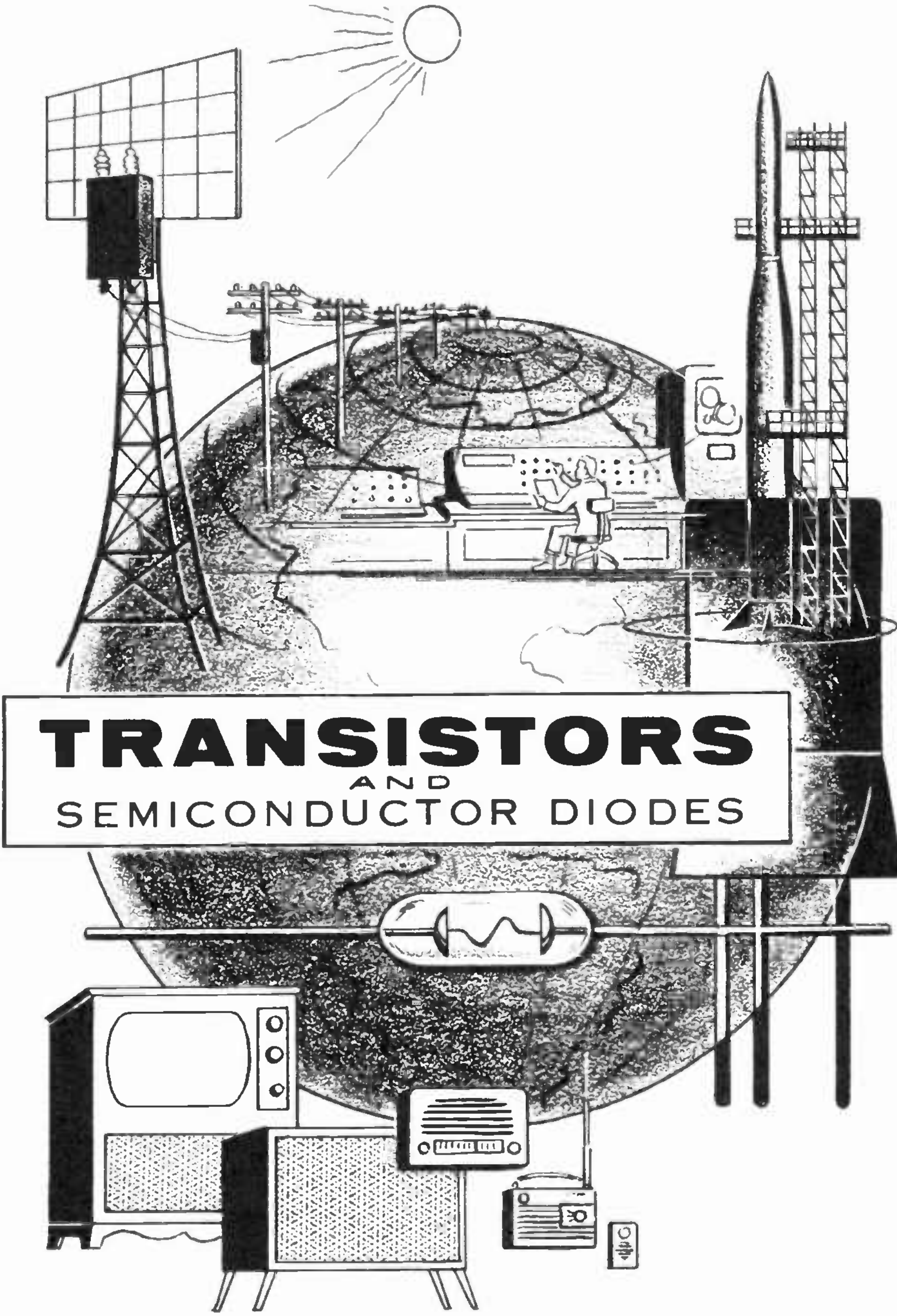
When Common-Core materials were released by the U.S. Navy in 1954, it was the intent that fundamental technical knowledge be placed within the grasp of men and women who would employ it for their own benefit and that of their country. In 1959, it is hoped that the addition of this volume on transistors and FM fundamentals will further the Navy's original objective.

Van Valkenburgh, Nooger & Neville, Inc.

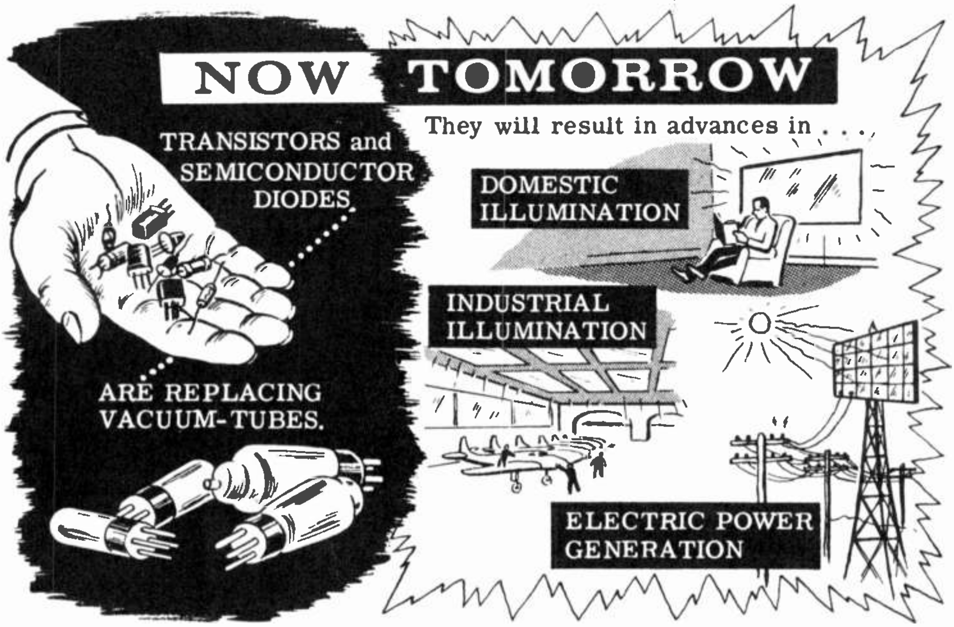
New York, N. Y.
September, 1959

TABLE OF CONTENTS

Vol. 6 — Basic Electronics


TRANSISTORS
AND
SEMICONDUCTOR DIODES

Brief History of Transistors and Semiconductor Diodes



The vast new field of solid-state electronics is represented by the transistor and the semiconductor diode—the first commercially available devices. These new devices are undoubtedly the most important developments in present-day electronics.

The semiconductor diode is a highly perfected development of the crystal detector used in the early radio receivers of 1906. This new device can detect, mix and rectify alternating current signals with excellent efficiency and has a wide variety of important new applications. The transistor was discovered in 1948 as the result of extensive studies of the operation of semiconductor diodes. Its name was derived from the words "transfer resistor" which describes the phenomenon which enables a completely solid device to amplify electrical signals.

Within the next few years these two devices will extensively replace vacuum tubes in many types of existing equipment, and they will be employed in a wide variety of domestic, transportation, industrial, scientific and military equipments which do not employ electronics at present. This is not the only significance of transistors and semiconductor diodes. As perfected representatives of the field of solid-state electronic devices they give a preview of new solid-state devices which are now in various stages of development. These devices will result in great advances in domestic and industrial illumination, electric power generation, conversion of electric power to mechanical motion, computer memory storage, ultra-high speed data transmission, detection and measurement of physical and chemical changes, electronic ignition and a vast variety of other aspects affecting our domestic and industrial life.

Brief History of Transistors and Semiconductor Diodes (continued)

Much investigation and experimentation was necessary before there could be any concept of the possibility of semiconductor diodes and transistors. The work of Volta, Ampere, Gauss, Faraday, and Hertz had to be accomplished and understood.

The discoveries that had been made in electricity raised many problems concerning the nature of matter. Astronomical investigations by Galileo, Tycho, Kepler and Newton also led to a great curiosity concerning the nature of matter. Investigations into matter itself by Romford, Davy, Carnot, Young, Fresnel, Maxwell, Hertz, Zeeman and Lorentz led to more questions than answers.

The first real breakthrough to the modern concept of matter came in 1897 when Sir J.J. Thompson discovered the electron while studying electric discharges through rarified gasses. Thompson's discovery was rapidly verified by other investigators. In 1913 Bohr evolved the basic theory of atomic structure, and that theory has been developed to our present-day concept of the nature of matter.

According to this theory all materials consist of various combinations of some one hundred different types of atoms. The atom is defined as the smallest unit into which an element may be divided before it loses its physical and chemical identity. Regardless of its specific element identification any atom consists of a positively charged nucleus around which one or more negatively charged electrons rotate. The electrons rotate in orbits which make up rings or shells located at various distances from the nucleus. The number of electrons in the various rings or shells are characteristic of the specific element to which the atom belongs. The electrons in the inner rings or shells have nothing to do with the ability of an atom to enter into chemical combinations with other atoms or to exhibit its various electrical characteristics; that is determined only by the electrons in the outer ring.

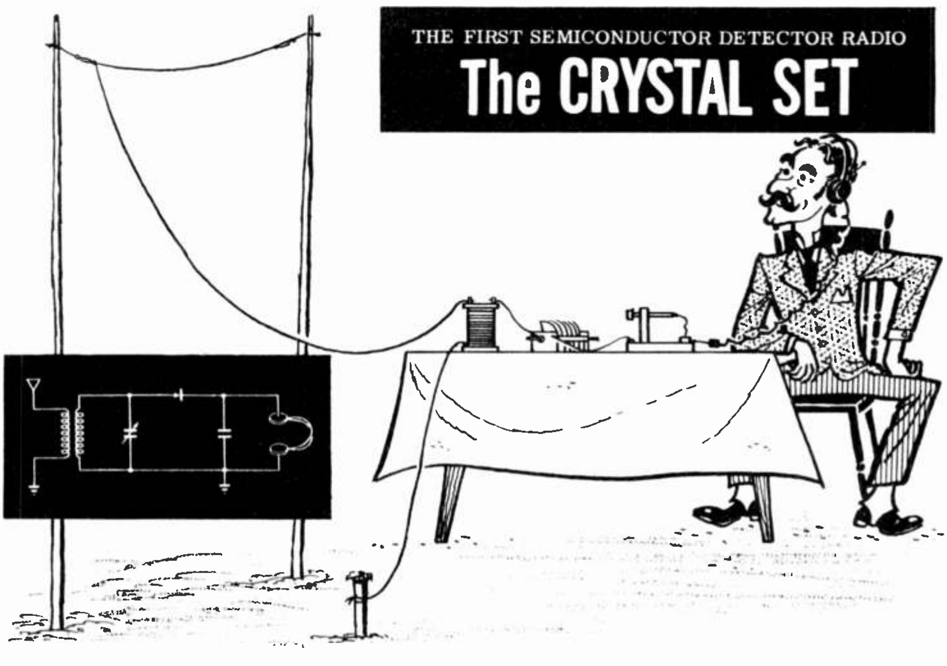
The electrical characteristics of an atom are determined by how tightly the nucleus holds onto its outer electrons. If the outer electrons are easily stripped off the atom by a low voltage electric field, the material will conduct a large amount of current; and the material is known as a "conductor." If a very large electric field is required to strip the electrons off the atom, the material is known as an "insulator."

Brief History of Transistors and Semiconductor Diodes (continued)

The material which is used in transistors and semiconductor diodes, and in most solid-state physics applications, is known as "semiconductor" material. This term is a general one and applies to all materials having an electrical resistance falling in the range between conductors and insulators. The materials which are of greatest use in present-day transistor applications are germanium and silicon, the characteristics of which will be considered shortly.

Interest in semiconductors began back in 1873 when it was discovered that rods and wires of selenium exhibited a decrease in electrical resistance when struck by sunlight. It was demonstrated that this was due to the presence of light and not due to heating, which normally result in an increase in electrical resistance. Later investigators found similar effects in other materials, but the change in resistance was so small that no practical applications could be found.

The next significant development concerning semiconductors was in 1906. At that time a variety of crystalline semiconductors were used as detectors of radio signals. Materials used for this purpose included galena (lead sulfide), silicon, iron pyrites and carborundum. The most common detector arrangement consisted of a piece of crystalline galena in contact with a short length of flexible wire, known as a "cat whisker." This device known as the "crystal detector" was used in the circuit arrangement shown in the diagram. This setup permitted the reception of radio signals, since the crystal acted as a rectifier which permitted easy current flow in only one direction.

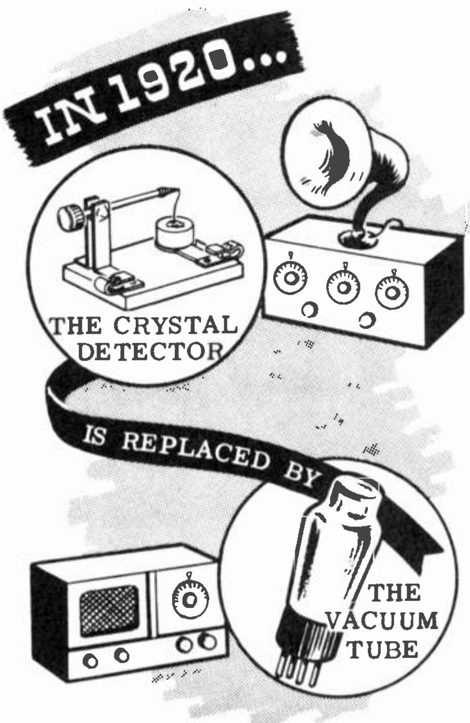


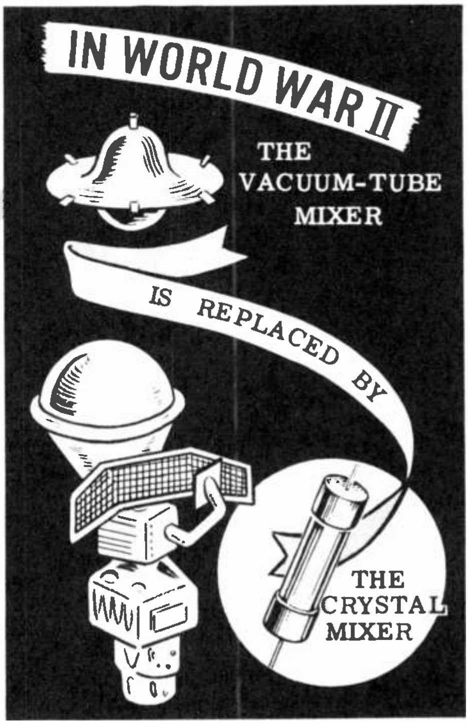
Brief History of Transistors and Semiconductor Diodes (continued)

The success of the crystal detector was short lived. The thermionic valve, more popularly known as the "vacuum tube" in the United States, was developed, and it served as a much more reliable detector than the crystal arrangement. In addition, vacuum tubes could amplify the detector output signal to an amplitude and power level sufficiently high to drive a loudspeaker.

During World War II radar was in a continuous state of development. One of the most important problems in radar was the detection of its extremely high frequency radio signals. Improvement in locating small targets required an increase in radar frequency, and each increase in frequency caused new problems in the vacuum-tube first detector (mixer) stage. New types of vacuum tubes were developed for the purpose, but eventually a frequency limit was reached beyond which vacuum-tube mixers would not operate. Crystal mixers were tried, and the silicon semiconductor type was found to be the most successful. Improved types of this mixer are widely used today in microwave radars.

While crystal mixers were being developed, a variety of semiconductor materials were investigated. Of these silicon and germanium were found to have very interesting properties, and these were investigated very extensively and systematically as soon as the war was over.



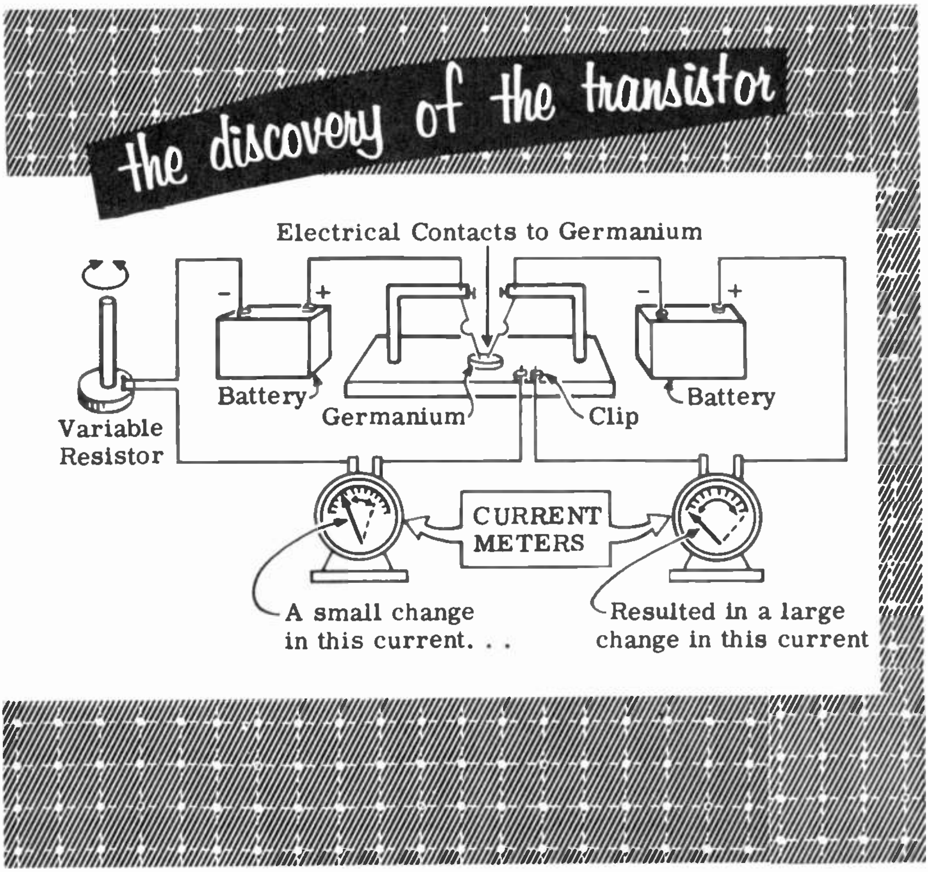


Brief History of Transistors and Semiconductor Diodes (continued)

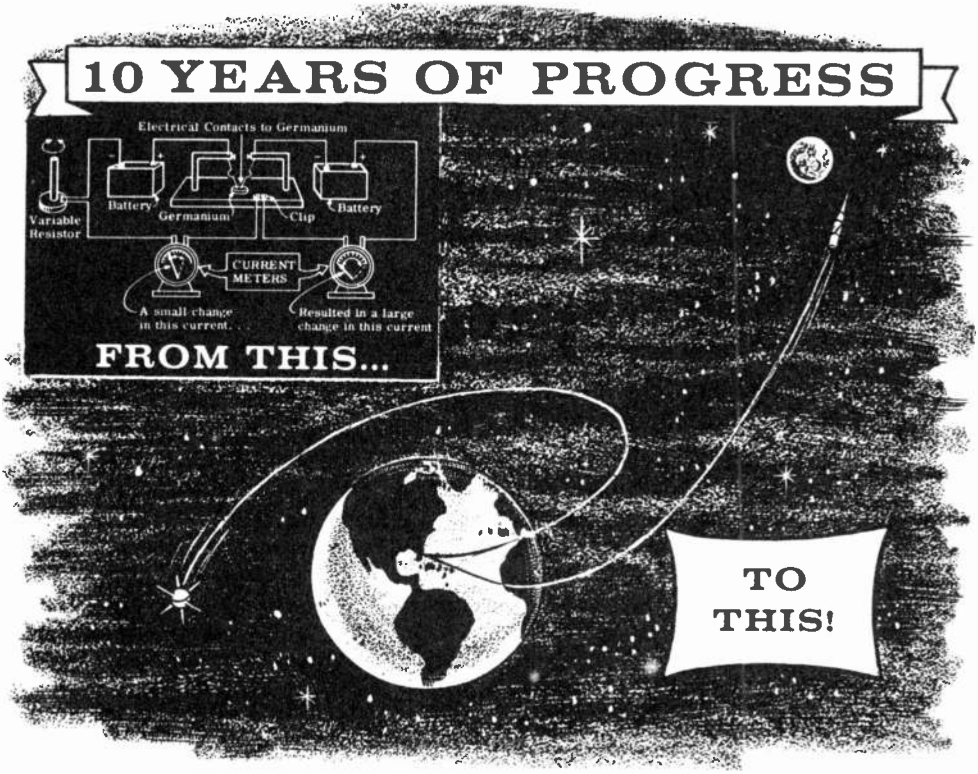
One of the first developments was a diode detector made of germanium. This detector was used in radio, television and miscellaneous electronics applications, but there was only a very limited need for a detector of this type.

During the development of germanium detectors a very important discovery was made. It was found that when two very close electrical contacts were made with a piece of germanium, the current flow through one of the contacts affected the amount of current flow through the other contact. This effect was found to be similar to the signal amplification occurring in a vacuum tube, except that no heated cathode and no vacuum was required.

An enormous amount of experimentation was conducted with this arrangement at the Bell Telephone Laboratories. The result was that in 1948 there was announced the development of the first solid-state amplifying device, the transistor. This discovery led to a new interest in semiconductor diodes, and their perfection led to the development of a wide variety of important new uses.



Introduction to Solid-State Electronics

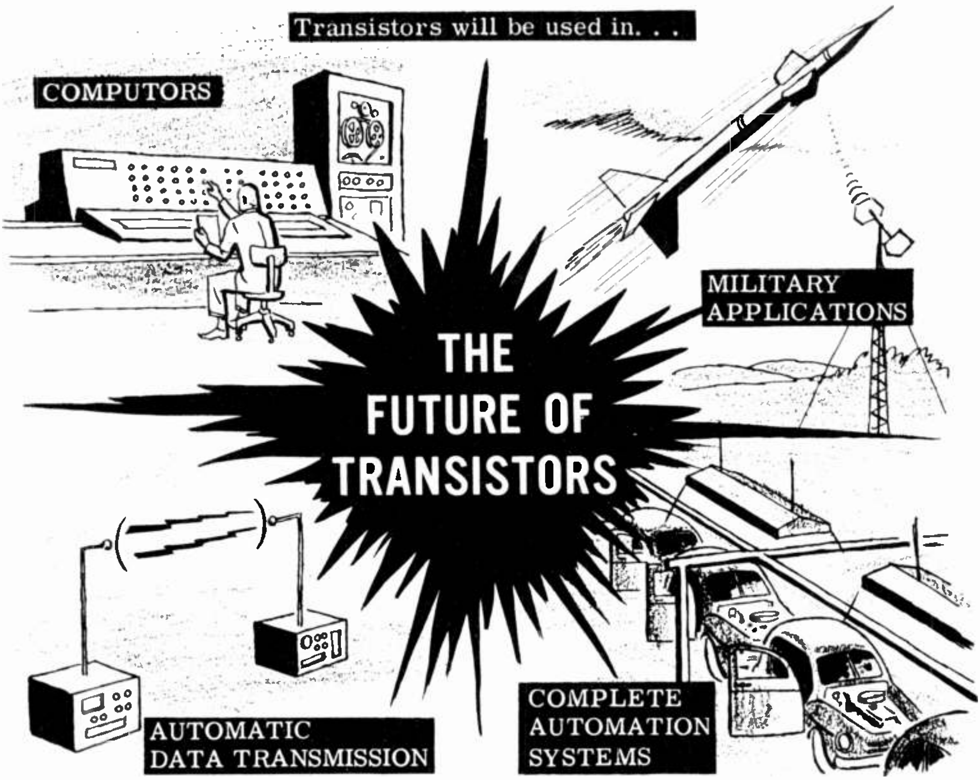


Learning about transistors and semiconductor diodes is the first step in preparing for future important work in the field of solid-state electronics. Extremely conservative predictions indicate that within the next ten years this field will encompass a broader and larger volume of applications than is contained today in the entire area of vacuum-tube electronics.

To give some idea of the progress that can be expected in the field of solid-state electronics, make a survey of the impact that transistors alone have had upon our industry and technology.

The first successful transistor was produced in 1948 at the Bell Telephone Laboratories. Ten years later transistors were successfully used to relay to earth information collected by the missle-launched satellites and space vehicles. Transistor sales amounted to about 70 million dollars, and in the next year the sales rose to just under 100 million dollars. Conservative estimates indicate that by 1968 the total sales of transistors will exceed 500 million dollars, with a total production quantity of over 500 million transistors. Note that these figures do not include the cost of the equipment into which these transistors will be installed. Semiconductor diodes which are closely related to transistors, are expected to have a sales level exceeding that of transistors.

Introduction to Solid-State Electronics (continued)



It should be completely obvious to those who are interested in transistors that their efforts in learning about these new devices will be greatly rewarded. Today there are enormous opportunities for interesting and important work in the fields of transistor and semiconductor diode applications. These opportunities will be multiplied vastly in the field of solid-state electronics that is just beginning to be developed.

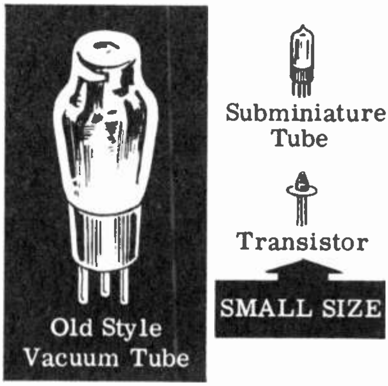
It may be asked why transistors are so important today, since most of the major developments are yet to come. At the present time the most important aspect of transistors is their replacement of vacuum tubes in a wide variety of applications. About 65 percent of present day transistors are used in the field of entertainment and domestic applications, about 25 percent of the transistors are used in the industrial and commercial fields, and about 10 percent of transistor production goes to military applications.

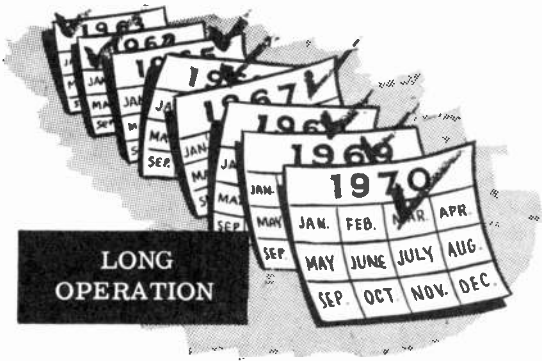
In a few years transistors will become well established in fields where vacuum tubes have only a tentative foothold today. These fields include industrial controls, complete automation systems, computers, and automatic data transmission systems. It is expected that there soon will be an enormous expansion of the use of transistors in military applications.

Advantages of Transistors

There may be some question as to why transistors are being accepted so widely as replacements for vacuum tubes. There are six basic advantages that transistors have over vacuum tubes. These advantages are outlined briefly in the paragraphs that follow.

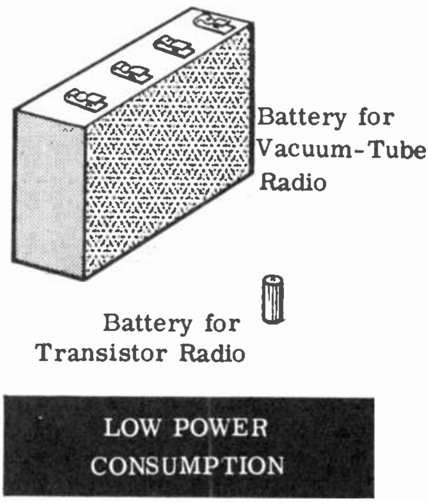
First, transistors are extremely small in size, ranging from the dimensions of subminiature vacuum tubes down to less than a quarter of that size. This small size permits the miniaturization of equipment, which is of great convenience in the entertainment and industrial fields and of vital importance in military applications.



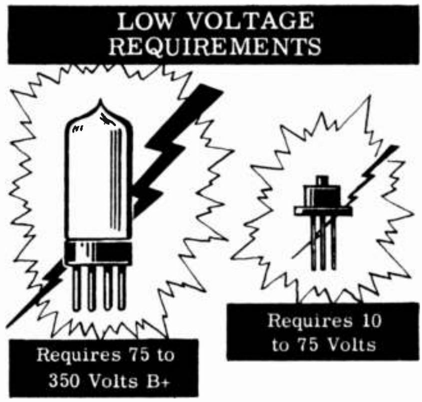


Second, transistors are inherently capable of performing their function for an indefinite period of time without deterioration of operating characteristics, as is the case with vacuum tubes.

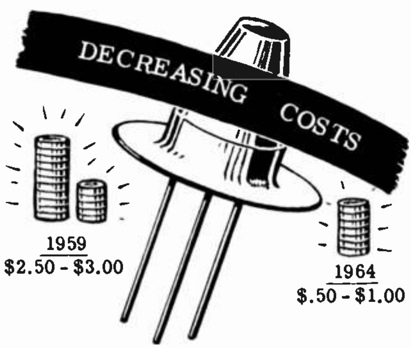
Third, transistors have a much lower power consumption than vacuum tubes. The reason for this is that transistors operate without the need for the heated cathode that is required by vacuum tubes. Heating the cathodes of vacuum tubes accounts for a large proportion of the power requirements of vacuum-tube equipment. These large power requirements make it difficult to produce portable battery-powered vacuum-tube equipment that has a reasonably long operating time. With transistors the same equipment can be made not only lighter and smaller, but the operating time can be in the order of five times longer.



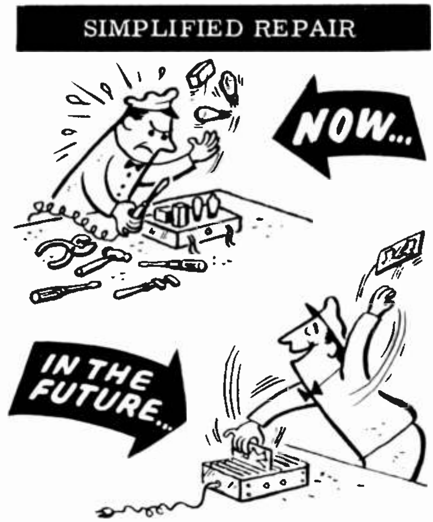
Advantages of Transistors (continued)



Fourth, transistors require lower B+ power supply voltages: from 10 to 75 volts as compared to 75 to 350 volts generally required by vacuum tubes. Use of these lower voltages reduces the filtering, hum shielding and voltage rating requirements of the power supply. The lower insulation requirements permit the use of R, L and C components significantly smaller in size than those used in conventional vacuum-tube equipment, permitting the miniaturization of transistorized equipment.

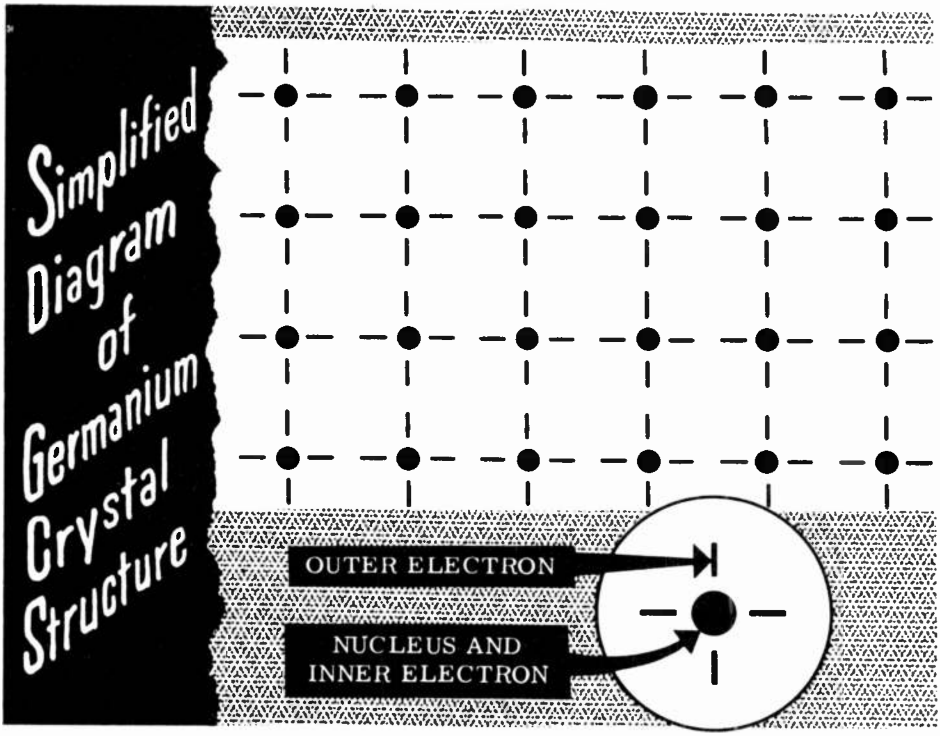


Fifth, when transistors reach the mass production level presently held by vacuum tubes, it is expected that their prices will be lower than the equivalent tube. At the present time the average transistor retail price is between $2.50 and $3.00. By 1964 it is expected that this average price will be from $.50 to $1.00. When this stage is reached, there will be no longer any considerations of higher transistor costs to retard their use in equipment.



Sixth, a transistor circuit is generally simpler and requires fewer components than an equivalent vacuum-tube circuit. When this feature is combined with the small size and low power requirements of transistor circuit components, it permits the construction of subminiature equipment and subminiature modules for larger equipment systems. Since mass production can be used to provide such modules at low cost, they can be made available to simplify the task of keeping equipment in operation. They provide means for the speedy and economical replacement of complete sections of the equipment, rather than going through the slow and henceforth costly process of trouble localization and component replacement.

Semiconductor Materials



To understand how transistors and semiconductor diodes operate it is first necessary to become familiar with the characteristics of the basic and modified materials that are in use.

The basic materials in commercial use are purified germanium and silicon which have been processed specially to the crystal state. These materials are excellent insulators because the crystalline structure very effectively bonds in place all of the outer electrons which would normally be free to enter into current flow.

The diagram shows a simplified picture of a pure crystalline semiconductor material such a germanium or silicon. Each atom has four outer electrons, which are shown as small minus signs. The inner electrons which are bound to the nucleus, and the nucleus itself, are shown as a single solid black dot. Because of the crystalline structure the nuclei are aligned in a symmetrical arrangement, and each outer electron shares an orbit with one outer electron from a neighboring atom. It is this orbit-sharing arrangement which effectively locks each electron in place, and not any unusually strong bond between the electron and its nucleus.

For an applied voltage to cause electron current flow, it would have to be sufficiently high to break the electron bonds before those electrons would be free to flow towards the positive voltage terminal. In breaking these bonds the voltage would also destroy the crystalline structure.

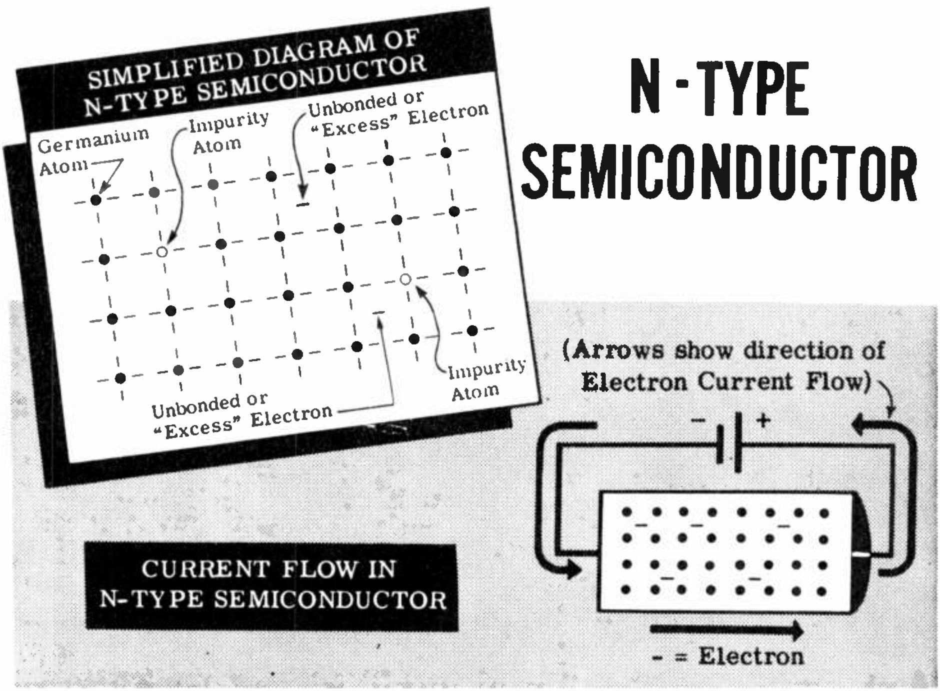
Semiconductor Materials (continued)

Since no electric current will flow through the pure crystalline material described, that material must be modified to obtain a controllable amount of electric current flow.

One method of obtaining current flow is to add a small number of atoms which have five outer electrons. Atoms which are suitable for this purpose include phosphorous, antimony, and most frequently arsenic. These atoms are distributed through the pure basic material as it is being processed into the crystal state, and the resulting structure is shown in the diagram. The proportion of impurity atoms added is in the order of one part per hundred million. A larger proportion produces a current flow that is not precisely controllable.

The impurity atom enters into the structure in the same manner as the atoms of the basic material. The one important difference is that the extra outer electron of each impurity atom remains unbonded to the crystal structure. If a DC voltage is connected across the ends of a piece of such material, those unbonded electrons are free to flow through the crystal structure towards the positive terminal. The total number of unbonded electrons in the crystal always remains the same – each electron that leaves the crystal at the positive terminal is replaced by one that enters at the negative terminal. As a result there is a continuous flow of current.

Since the current flow in this material consists of excess negative particles (electrons), the material is known as an "excess" or "N-type" semiconductor.



Semiconductor Materials (continued)

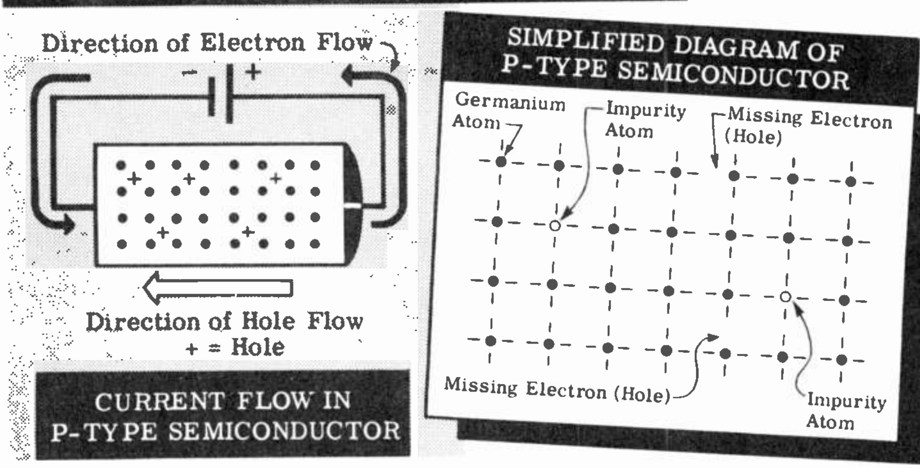
There is another method of modifying the pure basic crystalline material in order to obtain a controllable amount of electric current flow. During the processing of the basic material, impurity atoms such as aluminum, boron or indium are added in small amounts. These impurity atoms have only three outer electrons, and they enter into the crystalline structure as shown in the diagram.

Comparison of the diagram with that of the pure basic material shows that the modified structure has a missing electron for each impurity atom. The space in the structure caused by the missing electron is known as a "hole." Note that the hole is not necessarily located in the immediate vicinity of the impurity atom. During processing the impurity atom attracts a nearby outer electron to fill the gap in the surrounding crystal structure and the hole "moves" elsewhere. A succession of outer electrons may leave their nuclei to fill the gap, and the hole may move a considerable distance before it reaches a state of equilibrium.

If a DC voltage is connected across the ends of a piece of such material, the hole has the characteristics of a positive charge and flows towards the negative terminal of the voltage source. The total number of holes in the crystal always remains the same. Each hole that reaches the negative end of the crystal is neutralized by an electron which leaves the negative terminal and enters the crystal. This gives the crystal an excess negative charge. A neutral charge is regained by the crystal when it discharges an electron to the positive voltage terminal and creates another hole. The new hole flows towards the negative terminal, and the result is the continuous flow of holes through the crystal and a continuous flow of electrons through the connecting wires.

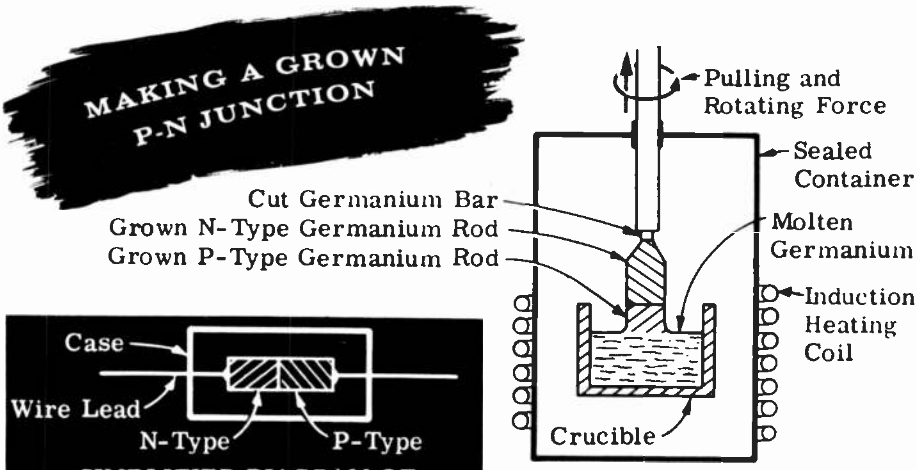
Since the current flow in this material is caused by defects (holes) in the crystal structure and these defects simulate positive charges, the material is known as a "defect" or "P-type" semiconductor.

P-TYPE SEMICONDUCTOR



CURRENT FLOW IN P-TYPE SEMICONDUCTOR

Basic Construction



SIMPLIFIED DIAGRAM OF GROWN JUNCTION DIODE

A semiconductor diode consists essentially of P- and N-type semiconductor materials in close contact with each other.

There are two basic types of semiconductor diodes in use today—the junction and point-contact types. There are several fundamental variations of each basic type which should also be considered at this time.

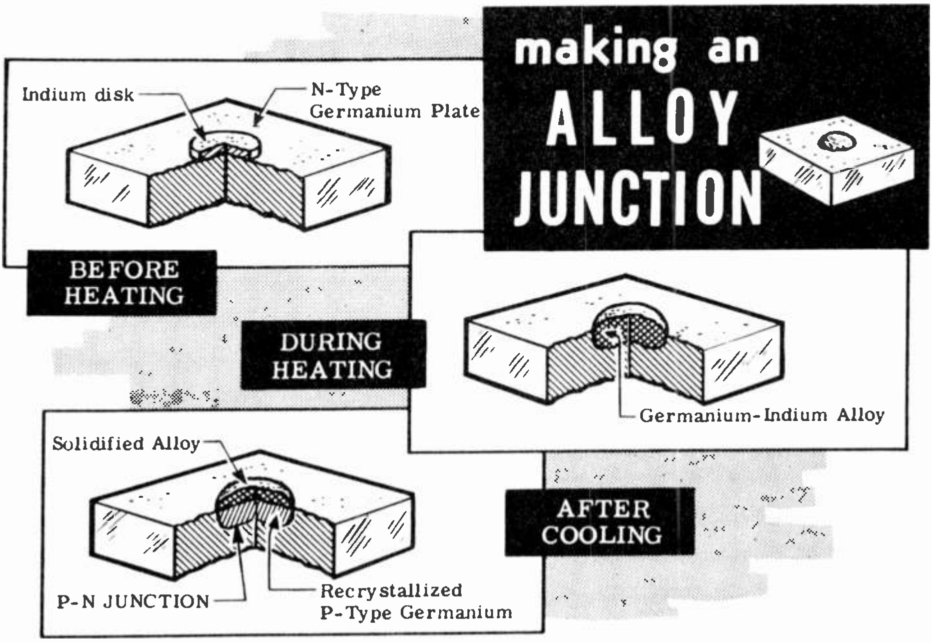
Two different types of junction constructions are in common use. In one type the junction is "grown" into the diode, and in the second type the junction is formed by diffusion.

A simplified diagram is shown of the arrangement for making a grown junction. A crucible containing pure germanium is suspended inside a sealed container which can be evacuated or filled with inert gas. An induction heating coil is used to heat the germanium to the melting point. To begin the formation of the diode an N-type impurity is added, and it diffuses throughout the melt. A small bar cut from single-crystal germanium is dipped down to touch the surface of the melted germanium and then it is slowly withdrawn and rotated. The melted germanium solidifies at the point of contact with the solid bar, and the withdrawing process causes the growth of a rod of N-type germanium. This rod is actually a single perfect crystal with a diameter in the order of one inch.

The junction is formed after the rod is grown to the length of about a half inch. Sufficient P-type impurity is added to neutralize the N-type impurity and to convert the germanium to the P-type. The withdrawing process is continued, and the remainder of the rod is of P-type germanium.

The entire rod is a single crystal of germanium, and the only difference is the type of impurity in the two halves. The P-N junction region is cut out of the rod and is diced up into as many as a hundred or more small junctions. Each piece has wire leads fused or soldered to it, and the assembly is mounted in a container which gives mechanical protection and isolation from contaminating atmospheres.

Basic Construction (continued)



There are several methods of making junction diodes by diffusion. The "alloy-junction" method of construction has been widely accepted because it lends itself to product uniformity and quantity production techniques.

In this method a small disk of P-type material (indium) is placed on a somewhat larger flat plate of N-type germanium. The materials are placed in a graphite holder and heated to a temperature of about 500^{o} Centigrade. The indium disk melts at about 155^{o} Centigrade, and upon reaching the higher temperature it dissolves away some of the germanium beneath it. A germanium-indium alloy is formed.

In the molten region the indium neutralizes the N-type impurities in the germanium, and leaves an excess of P-type impurities. After being subjected to heat for several minutes, an equilibrium condition is reached; and no more dissolving action takes place. The amount of P-type germanium that is formed is determined by the temperature and the size of the original indium disk, and the time is unimportant. This fact is important in achieving product uniformity.

Once the equilibrium condition has been established, the assembly is allowed to cool very slowly. The dissolved P-type germanium begins to recrystallize out of the alloy onto the N-type germanium base. The recrystallization follows the same atomic arrangement as that in the N-type germanium base, and a uniform P-N junction is formed.

After the assembly has cooled, electrical connections are bonded to the germanium base and to the indium disk. The assembly is mounted in a small container, and the alloy-junction semiconductor diode is complete.

Basic Construction (continued)

The point-contact type of construction resembles that of the crystal detector used in early radio receivers. It consists of a pointed wire pressed into contact with a small plate of semiconductor material. The assembly is sealed in a small container in much the same manner as the junction types.

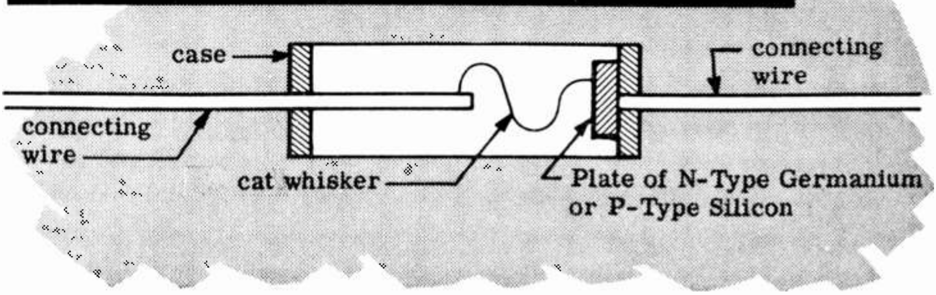
Platinum alloy, tungsten, phosphor-bronze, and other types of wire are used to make the contact. Several bends are made in the wire to give it a spring-like shape which presses the point against the semiconductor surface. It is the flexible nature of this wire that is the reason for it being called a "cat whisker." The pressure applied must be sufficient to hold the point in place without motion. The semiconductor plate usually consists of either P-type silicon or N-type germanium.

It was stated previously that a semiconductor diode consists basically of a junction between P- and N-type semiconductors. On first examination there appears to be no P-N junction in the point-contact construction. To be completely objective about the matter the operation of the point-contact diode is not very well understood. There are a number of fairly involved theories on the subject which are too complex to be reviewed here. In one way or another, regardless of the various special assumptions that are made, these theories boil down to the fact that there is something in the point-contact region that operates in a manner similar to a P-N junction.

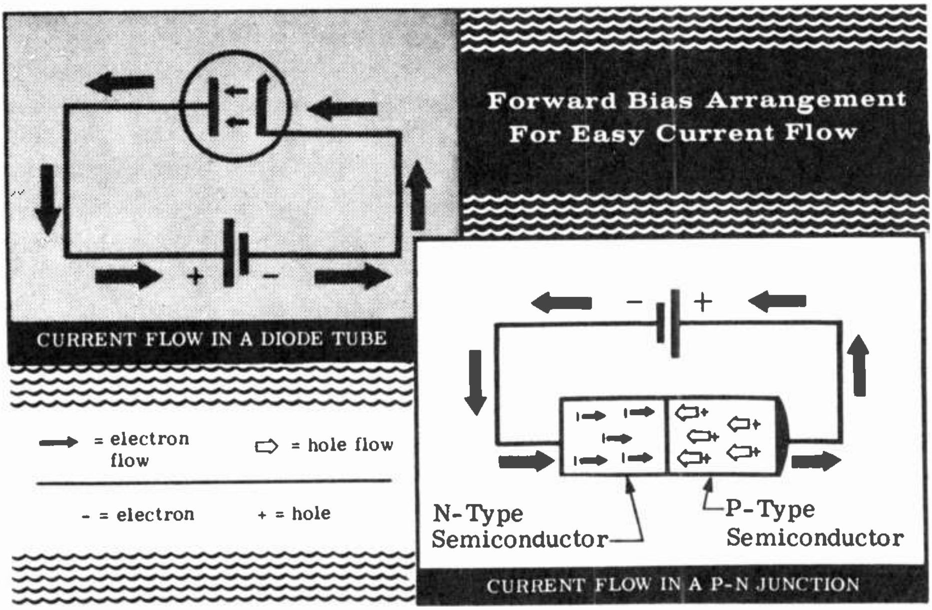
One verification of this theory is the fact that N-type germanium diodes using this construction generally operate better after "forming." Forming consists of passing a large pulse or current through the diode. After forming, the point of the cat whisker is found to be bonded to the semiconductor plate. The heavy current apparently melts the semiconductor material in the region of the point. This rapid melting and cooling apparently causes a localized conversion of N-type material to P-type material and a P-N junction is formed. The reasons for this conversion are difficult to explain, but exacting tests demonstrate that the conversion does take place.

Thus, for the purposes of this book it is sufficient to explain the operation of all semiconductor diodes on the basis of a P-N junction. The manner in which such a junction operates will be described in the pages that follow.

SIMPLIFIED CONSTRUCTION OF POINT-CONTACT DIODE



Operation of a P-N Junction



Semiconductor diodes consist basically of junctions between N and P semiconductors. The net effects which take place at a junction of this type are equivalent to the results produced by a diode vacuum tube. This equivalence can be demonstrated by comparing the results of connecting a DC voltage across a P-N junction and across a diode tube.

First consider the results when the positive and negative terminals of the voltage source are connected respectively to the plate and cathode of the tube and to the P and N semiconductors. In the tube, electrons flow from the negative voltage terminal to the cathode, through the vacuum to the plate, and on to the positive voltage terminal.

In the diagram of the junction, electrons are shown as minus signs, and holes are shown as + signs. Holes in the P-type material flow away from the positive voltage terminal towards the negative terminal, and electrons in the N-type material flow away from the negative voltage terminal towards the positive voltage terminal. At the junction there arrives a continuous flow of holes from one direction and a continuous flow of electrons from the other direction. When the electrons meet the holes at the junction, they neutralize the charge on each other. This neutralization of charge permits the formation of more holes at the positive end of the P-type material and the entry of more electrons at the negative end of the N-type material. All the requirements of a continuous current flow are met and such a continuous current flow does take place.

The direction of current flow in the connecting leads of the semiconductor diode is the same as for those of the vacuum tube. The polarity used for connecting the applied voltage is known as "forward bias," and is also known as the "direction of easy current flow."

Operation of a P-N Junction (continued)

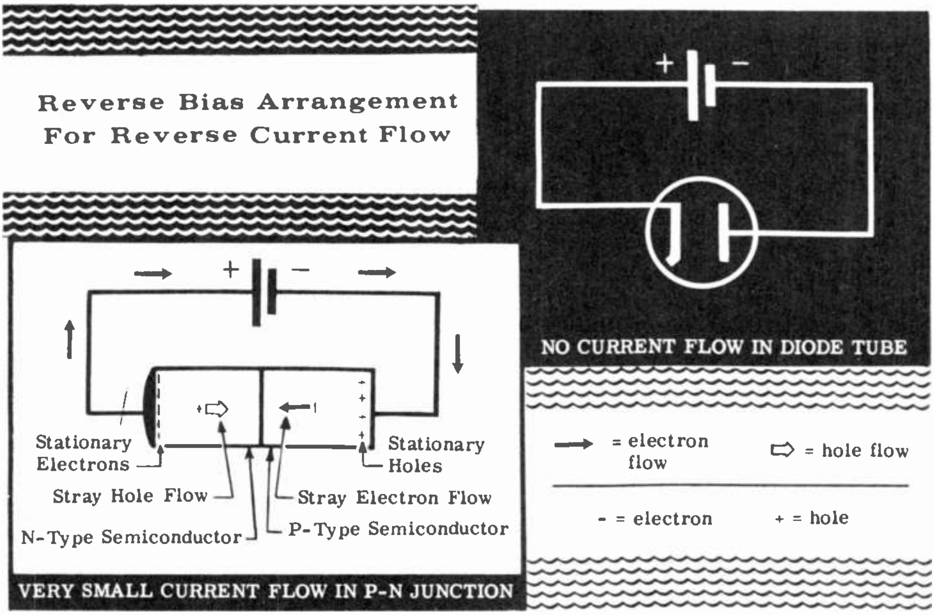
When the positive and negative terminals of the voltage source are reversed with respect to those of the previous arrangement, a completely different set of conditions takes place.

In the vacuum-tube circuit the plate is negative with respect to the cathode. Since the electrons emitted by the cathode are negatively charged, they are repelled by the negatively charged plate. No current flow takes place in the connecting wires.

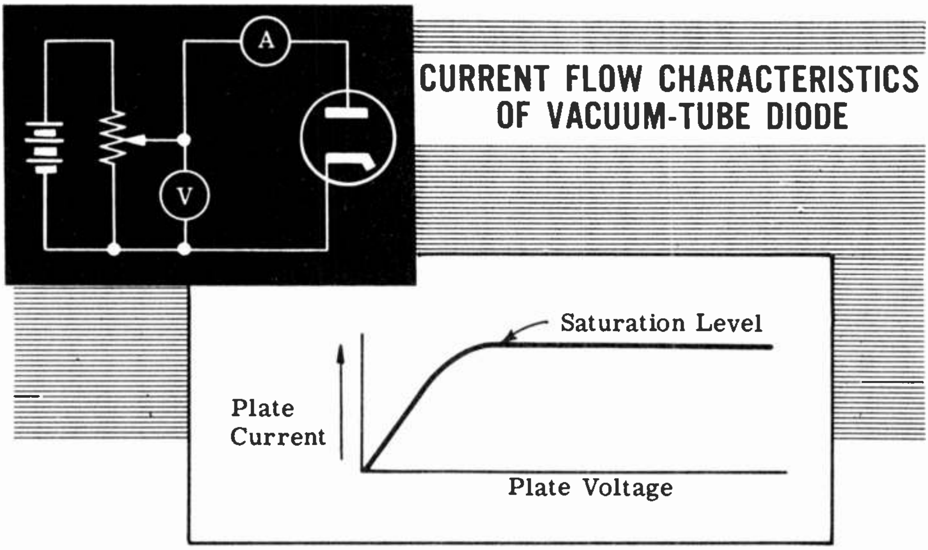
In the P-N construction holes in the P-type material are attracted towards the negative voltage terminal, and electrons in the N-type material are attracted towards the positive voltage terminal. This biasing arrangement has no provisions for the flow of current carrying holes or electrons to the junction, and no current flow can take place in the connecting wires.

Actually a very small amount of current does flow through the connecting wires. The reason for this is that N-type material does contain a small number of holes, and P-type material does contain a small number of electrons. These charges are able to flow in the direction required to maintain a steady current flow, as described for the forward bias condition. The reason for the existence of these stray charges is not due to a defect in the manufacuring process, but due to the breakdown of a few bonds in the crystal structure due to thermal agitation. As the temperature increases the number of these stray charges increases and the current increases.

Since the polarity used for connecting the applied voltage is the opposite of that used in the forward bias condition, this method of connection is known as "reverse bias," and it is also known as "reverse current" connection.



Characteristics of Semiconductor Diodes



It has been shown that the amount of current flow through a semiconductor diode varies in an outstanding manner when there is a reversal in the polarity of the biasing voltage. It is now necessary to find out the details of the relationship between current flow and biasing voltage. A comparison with the corresponding vacuum-tube characteristics will help to clarify the unusual current flow characteristics of a semiconductor diode.

In your study of diode vacuum-tubes there was a review of an arrangement used to learn about the plate voltage - plate current characteristics of a diode vacuum tube. In essence, the plate-to-cathode voltage is varied while a current meter is used to measure plate current. The corresponding plate currents and plate voltages are plotted on graph paper, and the result is a curve such as that shown in the diagram.

The heated cathode emits electrons which collect in a space charge around the cathode. When the plate is made negative with respect to the cathode, no current flows from the cathode to the plate because the negative plate repels the electrons. Current cannot flow from the plate to the cathode since the plate does not emit electrons. When the plate and cathode are at the same potential, the plate neither attracts nor repels electrons; the current remains at zero. When the plate is made slightly positive with respect to the cathode, a small portion of the electrons are attracted out of the space charge and flow to the plate and through the outside circuit. As the plate is made increasingly positive, the current flow becomes larger.

Eventually the current flow is so large that electrons are attracted to the plate as fast as the cathode can emit them. Further increase in plate voltage causes no further flow of plate current, and a state of saturation is reached.

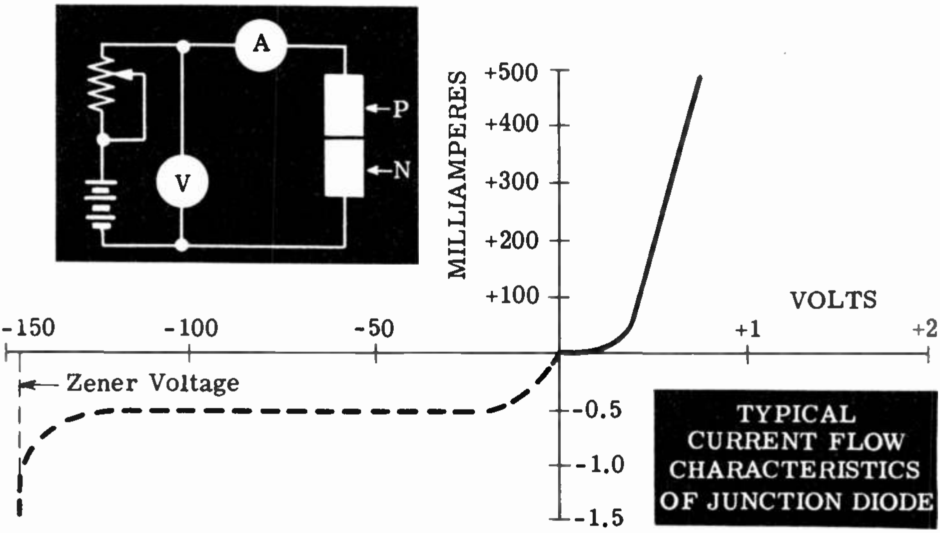
Characteristics of Semiconductor Diodes (continued)

When the same procedure is used to study the voltage and current characteristics of a semiconductor diode, somewhat different results are obtained. Shown here is a typical voltage-current curve for a junction diode. Examination shows that it is quite different from the curve of a typical diode vacuum tube, but the same general type of rectifying action is obtained.

First review the characteristic curve of the junction diode. When voltage is applied in the forward direction, the current varies as shown by the solid-line curve. Note that only a few tenths of a volt are required to cause a current flow in the order of 100 milliamperes. Further increase in forward voltage causes a current rise that is almost linear in relationship to the applied voltage, and the maximum rated current is reached before one volt is applied.

When voltage is applied in the reverse direction, the current varies as shown by the dashed-line curve. Large increases in voltage cause only very small rises in current. In fact, the current is so low that a different set of graph scales is required to show the change. The extremely small current flow is due to the fact that there are very few current carriers under reverse bias conditions; and once all of these current carriers are flowing, a state of saturation takes place.

This state of saturation does not continue indefinitely. As higher reverse voltages are applied, more current begins to flow. Eventually a condition is reached where the diode resistance drops very rapidly, and a very large reverse current increase takes place with no further increase in reverse voltage. Damage to the diode can be prevented by conducting away the heat generated and by decreasing the reverse voltage. The reverse voltage at which this effect takes place is called the "Zener" voltage, named after the man who predicted this effect. This effect is useful only in certain special applications.



TYPICAL CURRENT FLOW CHARACTERISTICS OF JUNCTION DIODE

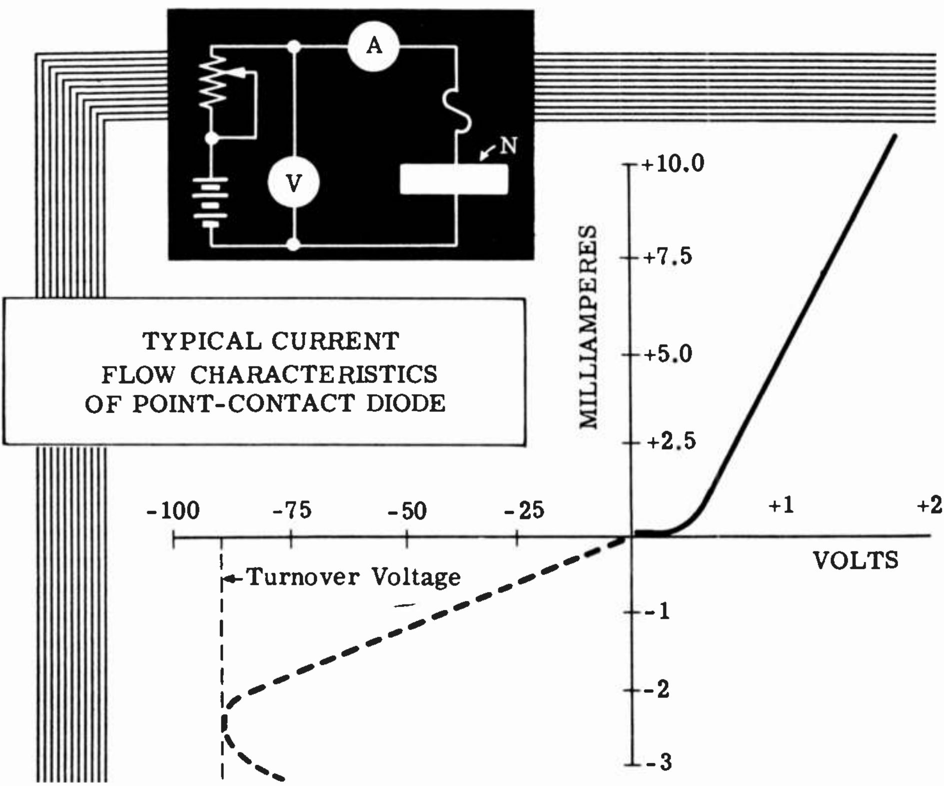
Characteristics of Semiconductor Diodes (continued)

Shown here is a voltage-current curve of a typical point-contact diode. The curve has many similarities to that of the junction diode considered previously, but there are several significant differences to be considered.

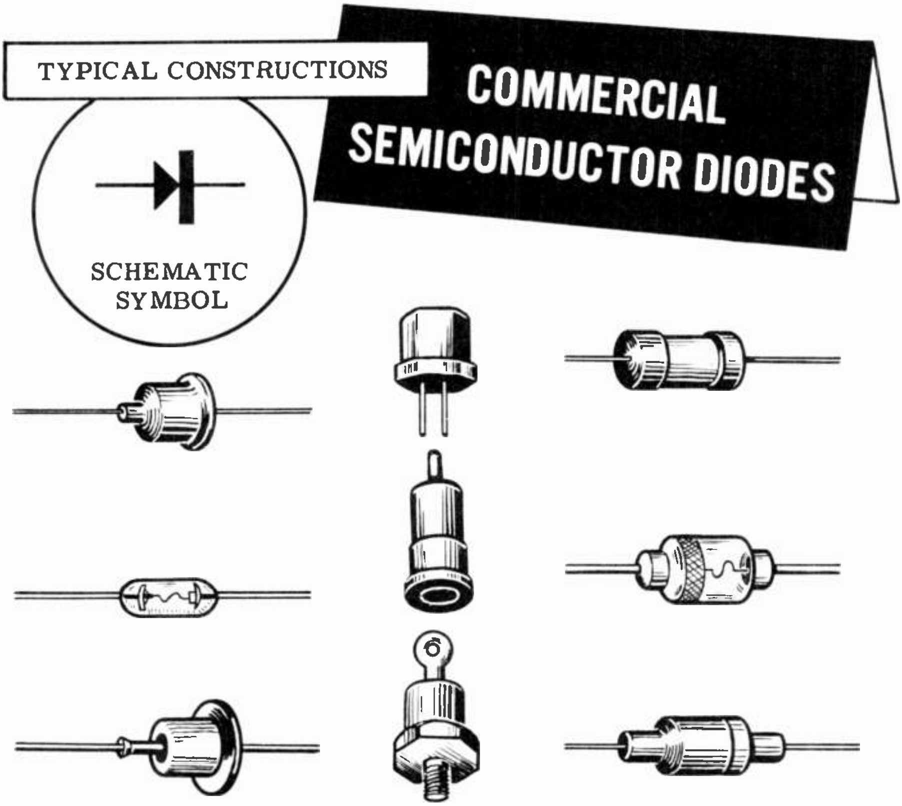
First, the rated current flow in the forward direction is only a small fraction of that obtainable from the junction diode. The reason for this is that the active junction area in the point contact construction is much smaller than that in a junction diode.

Second, the reverse current flow is several times larger than that of the junction diode. In addition, the reverse current does increase steadily with reverse voltage, and there is no sharp saturation effect as in the case of the junction diode.

Third, a different effect is obtained as the reverse voltage is increased. Instead of the Zener effect described for the junction diode there is a "turnover" effect. At the turnover voltage the internal resistance of the junction appears to become negative, rather than dropping to zero. Therefore, the current increases very rapidly, and continues to rise even though the reverse voltage is lowered. There is no satisfactory explanation for this effect; and it is not useful in practical applications, since the diode is destroyed when the effect occurs.



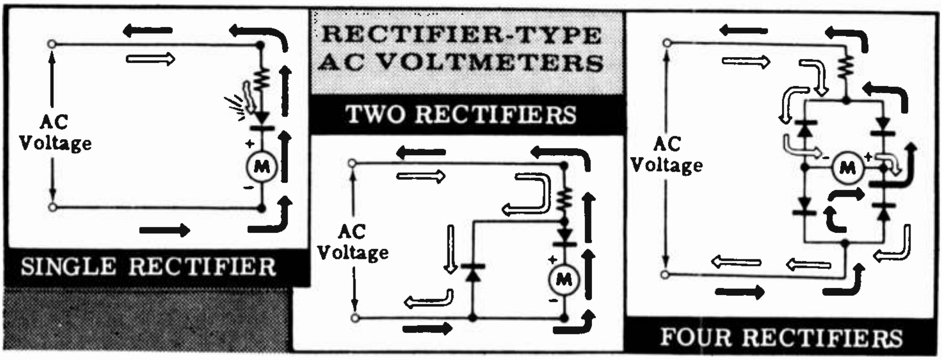
Commercial Semiconductor Diodes



Shown on this page are pictorial diagrams of the various types of semiconductor diodes that are available from commercial sources. It can be seen that there are a wide variety of physical constructions available. Included among these are ceramic jackets with metal ends, glass tubes with metal ends, all-glass jacket, plastic cases, plastic-coated metal cases, and metal jackets with a screw mounting. Some of these outward variations are due mainly to the preferences of the individual manufacturer. Other features have a specific function, such as the screw-mounting which may be employed to dissipate the heat generated by power rectifiers.

Although not always obvious without close examination, many of the semiconductor diode cases are marked with an arrow. The arrow shows the direction of easy current flow as indicated by a DC ammeter. The reason for this method of marking is that it permits technicians and repairmen a positively reliable method of checking the connections required. This eliminates the necessity for deducing this information from a schematic diagram, which may be confusing in some special cases. Many schematic diagrams indicate semiconductor diodes marked to correspond with this arrangement.

Applications of Semiconductor Diodes



Semiconductor diodes have great flexibility of application. They can be used in all those applications where dry-metal rectifiers or diode vacuum-tubes are presently employed, and they have some unusual applications of their own. The advantage of using a semiconductor diode as a replacement is that it generally is smaller, more efficient and operates at significantly higher frequencies than the tube or dry-metal rectifier; and no filament power is required as in the case of the tube.

The most elementary semiconductor diode circuit is one you learned about in your study of AC meters in Volume 3 of Basic Electricity. This circuit makes it possible for a basic DC voltmeter circuit to be used to measure AC voltage.

The simplest arrangement contains a resistor, rectifier and DC meter movement. Electron flow indicated by the black arrows passes through the meter movement and causes the pointer to move up-scale. This electron flow results from one half-cycle of the line voltage. The electron flow resulting from the alternate half-cycle of the line voltage is shown by the white arrows. Although only pulses of current flow through the movement, the pointer cannot move rapidly enough to follow the rise and fall; and the average value of the current pulses is indicated. The resistor is often made adjustable so that the pointer reading can be made to correspond with that of a precision meter. If a semiconductor diode is used as the rectifier, the meter can be calibrated at power line frequencies and will give accurate voltage readings, without a correction factor, at all frequencies up to thousands of megacycles.

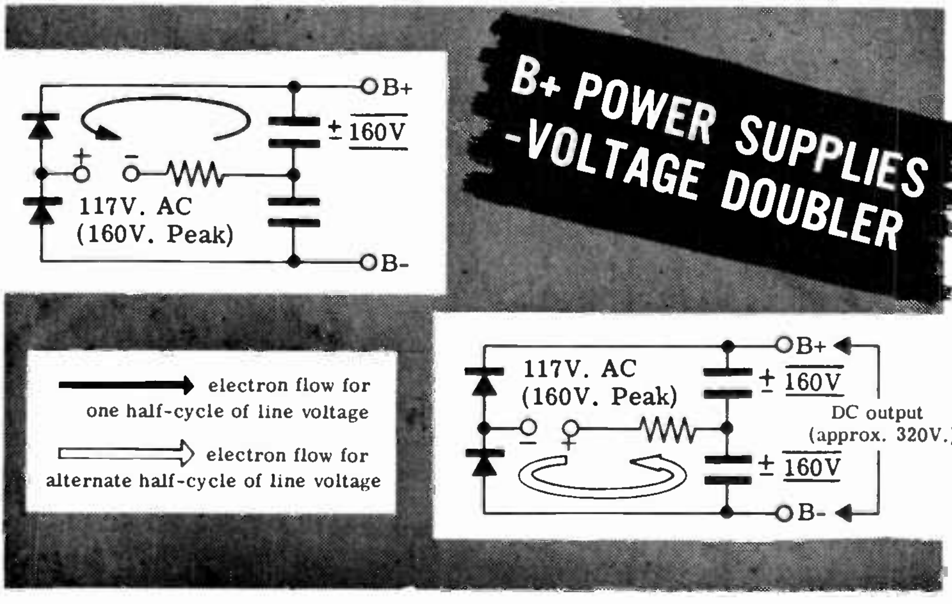
The AC voltmeter circuit considered above presents a low resistance to one half-cycle of the applied voltage and a high resistance to the alternate half-cycle. This is of no consequence in measuring voltages in power circuits. In AF and RF circuits, however, this lack of uniform loading may cause inaccurate readings and disturb the operation of the circuit. By adding a second rectifier to the circuit, the half-cycle that is not used is presented a low resistance path around the meter, and fairly uniform loading is achieved. A bridge circuit of four rectifiers can be used, as shown in the diagram, so that both half-cycles of the AC current flow through the meter in the same direction. This results in a balanced load to both half-cycles of current.

Applications of Semiconductor Diodes (continued)

Other applications of semiconductor diodes include their use in power supply circuits. In such applications semiconductor diodes have the advantage of being rugged, long lived, small in size and capable of large current output. Only recently semiconductor diodes were more expensive than equivalent dry-metal rectifiers, and there was a limited selection of types available for large power output. At present, semiconductor diodes are often used in preference to the dry-metal types because of the savings in size and efficiency, and because there is little difference in cost.

If semiconductor diodes are to be used in power supply circuits the half-wave rectifier and bridge rectifier circuits are the most common. These types are equivalent to the meter circuits described previously. The purpose of the resistors in series with the rectifiers is to prevent excessive current flow from damaging the rectifier, as would occur in the event of a short circuit or overload in the equipment to which the rectifier is connected. Either an RC or LC filter may be placed between the rectifier and the load.

Also available for use is the voltage doubler circuit shown in the diagram. This circuit was explained in detail in Volume I of Basic Electronics, and only a brief review will be included here. The voltage doubler consists of two half wave rectifier circuits. During one half-cycle of the line voltage the upper diode conducts and charges the upper capacitor to peak line voltage. During the alternate half-cycle the lower diode conducts and charges the lower capacitor to peak line voltage. Since the two capacitors are charged in series with the DC output terminals, the DC output voltage is equal to twice the peak of the line voltage. For a 115-volt line the DC output voltage is approximately 320 volts.



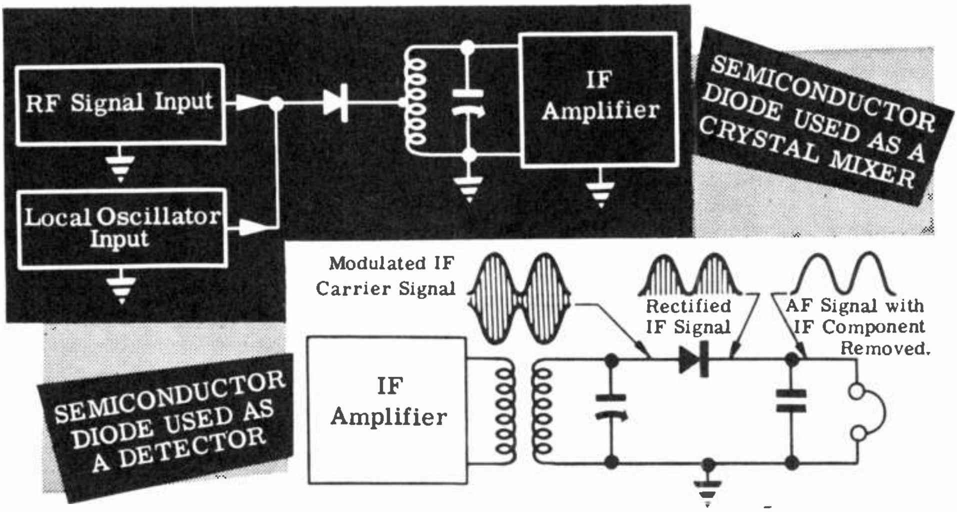
Applications of Semiconductor Diodes (continued)

In receiver circuits the semiconductor diode can be used as an efficient mixer or detector. Examples of both of these circuits will be considered now.

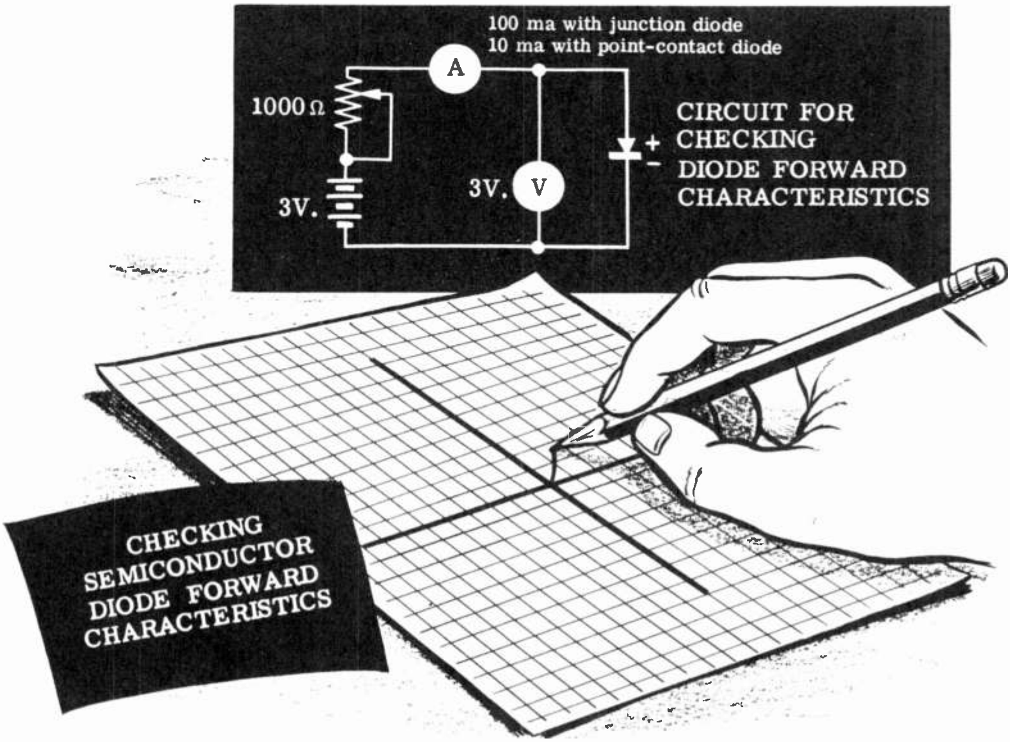
Shown in the diagram is a very simple type of semiconductor diode mixer. This mixer operates well through the broadcast, television, and microwave bands. It is not used frequently in the broadcast or television bands, since its gain is less than one; and significant gain can be obtained by means of vacuum-tube or transistor mixers. At microwave frequencies, however, the semiconductor diode mixer operates efficiently where other circuits fail.

When the circuit is in operation, the local oscillator applies a constant voltage to the rectifier. The result is a constant flow of current through the semiconductor mixer, a current flow consisting of unidirectional pulses at the frequency of the local oscillator. Also applied to the mixer is the incoming RF signal from the antenna. Heterodyning action takes place in a manner similar to that of a standard mixer circuit, and the output of the mixer consists of four different frequencies: the frequency of the RF signal from the antenna, the local oscillator frequency, the sum of these incoming signals, and the difference between these incoming signals. As in the case of a standard mixer, the IF transformer is tuned to only the difference signal; and amplification of the modulated signal takes place at the lower frequency.

In detector applications the semiconductor diode circuit is essentially the same as in the vacuum-tube circuit and in the meter circuit considered earlier. When the amplitude-modulated IF signal is rectified, the result is a pulsating unidirectional current which carries an IF signal component and an audio signal component. The IF signal component is bypassed to ground by a capacitor which is too small to bypass the audio signal component. The result is that the audio signal component is applied to the input of the audio amplifier, and detection has taken place.



Experiment—Semiconductor Diode Operation



The purpose of this experiment is to show you the outstanding characteristics and basic application of semiconductor diodes. A number of components and items of equipment are required, and these are clearly marked in the diagrams referred to on this page.

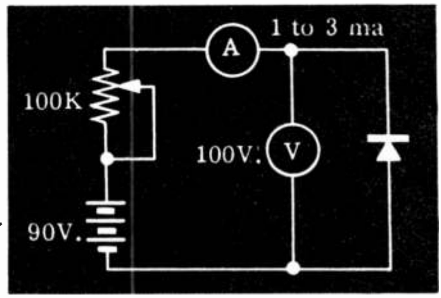
The first part of the experiment is designed to demonstrate the forward and reverse current characteristics of semiconductor diodes of the junction and point-contact type.

To begin, make the connections shown in the "Forward Characteristics" diagram shown here, but do not connect the diode into the circuit. Set the potentiometer to make the voltmenter reading equal to zero. Connect a junction diode into the circuit, as shown in the diagram. By turning the potentiometer slowly, increase the forward voltage in steps of 0.1 volt and plot the corresponding current readings on a sheet of graph paper. Do not exceed the maximum current rating indicated on the data sheet accompanying the diode. The completed graph will show that less than 1 volt is required to cause maximum rated current flow in the forward direction.

Reset the potentiometer for a zero voltage reading and repeat the experiment with a point-contact diode. As the results are plotted on a second sheet of graph paper, it will be seen that the results are similar to those obtained with the junction type, except that slightly higher voltages are required.

Experiment--Semiconductor Diode Operation (continued)

To complete the first part of the experiment make the connections shown in the "Reverse Characteristics" diagram at right. Set the potentiometer to make the voltage reading equal to zero and connect the junction diode into the circuit. By turning the potentiometer slowly, increase the reverse voltage in steps of 10 volts. Plot the corresponding current readings on the sheet of graph paper previously used for the junction diode. Do not exceed the maximum current rating indicated on the data sheet accompanying the diode.

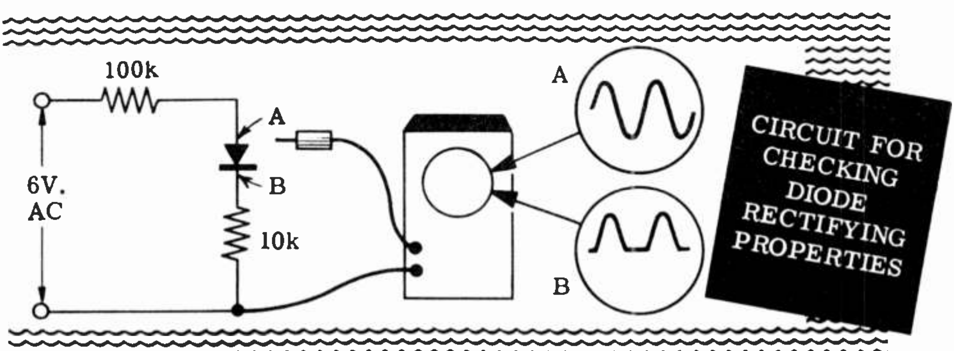


CIRCUIT FOR CHECKING DIODE REVERSE CHARACTERISTICS

If the Zener voltage described on page 6-19 is reached, immediately reset the voltage to zero. The completed graph will show that reverse current increases very slowly with the applied voltage. Depending upon the maximum current rating of the particular junction diode that has been selected, the Zener characteristics may or may not be visible.

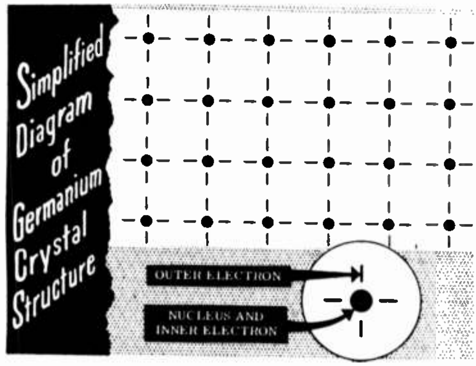
Reset the potentiometer for a zero voltage reading and repeat the experiment with a point-contact diode. Plot the results on the sheet of graph paper previously used for this diode. Be careful not to approach the turnover voltage too slowly because this will destroy the diode. The completed curve will show that reverse current increases very slowly with the applied voltage, although somewhat more rapidly than in the case of the junction diode.

The second part of the experiment demonstrates the rectifying capability of the semiconductor diode. It is this property which is fundamental to all the applications of this diode. Connect a junction diode in series with a 100,000-ohm resistor across source of 6-volts AC. With an oscilloscope examine the voltage waveform present at points A and B. Before the diode, at point A, the full sine wave of the AC voltage is visible. After the diode, at Point B, a half wave rectified waveform is visible. This demonstrates that the diode conducts current in one direction only, and is equivalent in function to a dry-metal or vacuum-tube rectifier. Substitute a point-contact semiconductor in the circuit and the same results will be obtained.

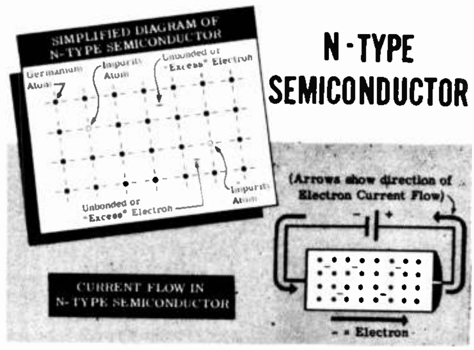


CIRCUIT FOR CHECKING DIODE RECTIFYING PROPERTIES

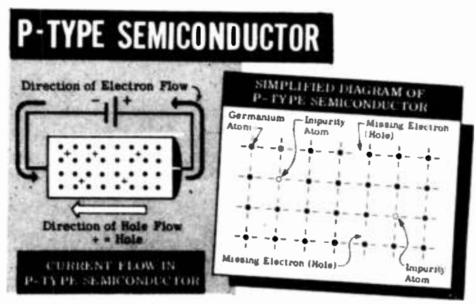
Review of Semiconductor Diodes



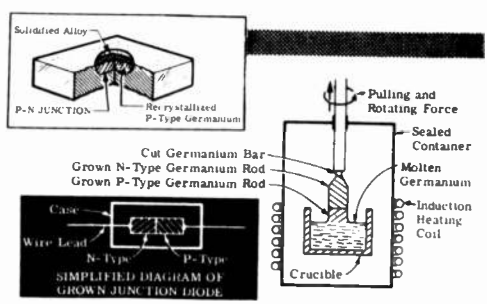
SEMICONDUCTOR MATERIALS— Purified crystalline germanium and silicon are the basic materials commonly used in semiconductor diodes and transistors. These materials are excellent insulators because the crystalline structure bonds all the outer electrons in place.



N-TYPE SEMICONDUCTOR— Semiconductor material can be made to conduct by adding impurity atoms which enter the crystalline structure but which have excess outer electrons which are not bonded to the structure. Current flow is conducted by the excess negatively charged electrons which flow through the crystal to the positively charged terminal.

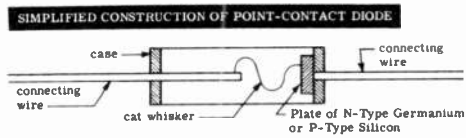


P-TYPE SEMICONDUCTOR— Conduction can also be obtained by adding impurity atoms which do not have sufficient outer electrons to fill all the crystal bonds. The unfilled spaces, or "defects" are known as "holes" and have the characteristics of positive charges. An applied voltage causes the holes to flow through the crystal to the negatively charged terminal.

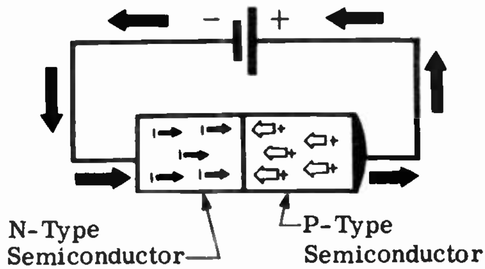


JUNCTION DIODE— A junction diode consists of P- and N-type semiconductor materials in close contact. The junction can be formed during the crystal growing process (grown junction) or formed by a dissolving and recrystallization process (alloy-junction).

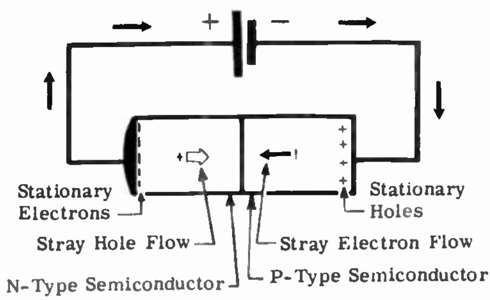
Review of Semiconductor Diodes (continued)



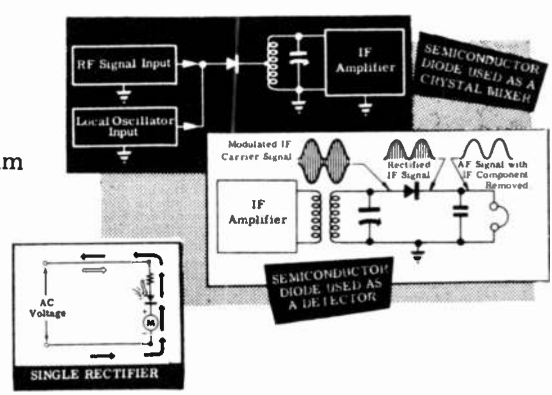
POINT-CONTACT DIODE—A point-contact diode consists of a plate of N- or P-type semiconductor material in contact with a pointed metal wire. The contact region can be regarded as a P-N junction.



FORWARD BIAS—The P-N junction biasing arrangement shown is known as "forward bias." Only several volts are required to cause all holes and excess electrons to flow to the junction and result in maximum rated current flow.

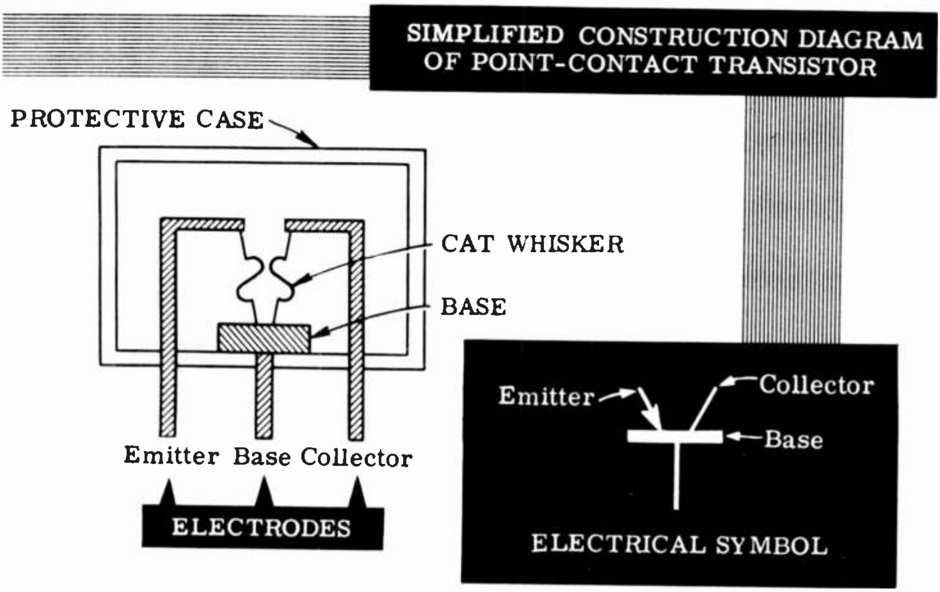


REVERSE BIAS—When the junction biasing connections are the reverse of forward bias, all holes and excess electrons flow away from the junction and do not enter into a continuous current flow. Only stray holes and electrons can enter into a continuous current flow. High voltages are required, and the maximum current is only a small fraction of that obtained with forward bias.



SEMICONDUCTOR DIODE APPLICATIONS—Semiconductor diodes can be used in all applications suitable for vacuum tubes or dry-metal rectifiers. The circuits with which you are familiar include meter rectifiers, power supply circuits, mixers and detectors.

Basic Construction — Point-Contact Type



There are two basic types of transistor construction in use today. These include the "point-contact" and "junction" types. Both of these have a number of variations. Only the basic construction of the most common variations are shown here, and the details are sufficiently fundamental to apply to the production differences introduced by the various manufacturers.

The point-contact construction is the earliest, and is no longer in widespread use. The arrangement is similar to that of a point-contact diode with a second "cat whisker" in contact with the germanium block. It is necessary that the two point contacts be separated in the order of a few thousandths of an inch apart, otherwise the desired operation cannot be obtained. As in the case of the point-contact diode, a simple explanation of the principles of operation requires the quite valid assumption that there is a P-N junction in the region of each cat whisker point.

The germanium block is known as the "base" since it is the foundation to which all the electrical contacts are made. All dimensions of the base are in the order of a few hundredths of an inch. The base material is almost always of the N-type; using P-type base material is theoretically possible but no great success has been achieved in making such an arrangement.

One of the contact wires is known as the "emitter" and the other is known as the "collector." These names are derived from the functions of the two wires at their points of contact with the base. When proper voltages are applied, the emitter causes the generation of current-carrying charges at its contact point. The collector accumulates current-carrying charges at its contact point and provides a terminal for conducting electric current through the outside circuit. Also shown in the diagram is the transistor electrical symbol.

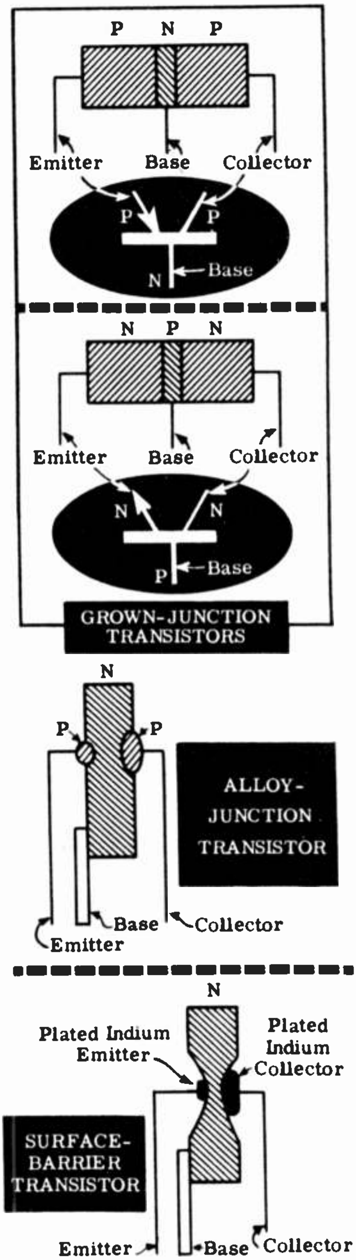
Basic Construction — Junction Types

The junction-type transistor also consists of a base, emitter and collector. Two basic types in general use are made in a manner similar to the "grown-junction" and "alloy-junction" methods of construction previously described in connection with junction semiconductor diodes. In both cases the result of the manufacturing process is essentially the same — two P-N junctions are formed and they are located several thousandths of an inch apart. Other features of construction are similar to those of the corresponding junction diode. The diagrams on this page show the major features of both types. Note that in the grown-junction type the semiconductor materials may be arranged in a P-N-P or N-P-N sequence.

Transistor manufacturers are continuously conducting research in order to achieve greater product uniformity plus speed and economy of production. Present efforts are concerned with methods of producing P-N junctions of easily controlled size and spacing. The alloy junction method shows great promise, and automatic machinery is being developed to control precisely and speed up all stages of the manufacturing process.

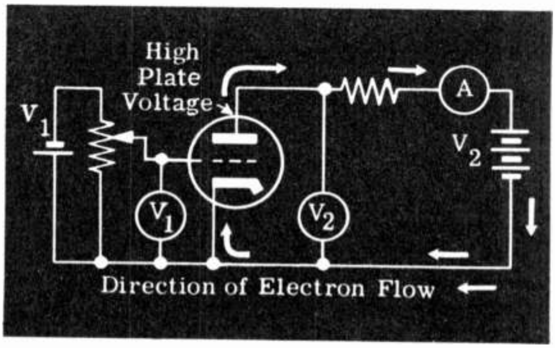
Another method in development is to electrolytically etch away two spots on opposite sides of an N-type germanium plate until the etched surfaces are several ten-thousandths of an inch apart. Then the electrolytic action is used to plate an indium spot on each etched surface. The result is a "surface-barrier" transistor in which an indium collector and an indium emitter are separated by several ten thousandths of an inch of N-type germanium. Although a complex "surface-barrier" theory is used to explain the operation of this type, the contact area between the indium and germanium has characteristics almost identical to that of a P-N junction; and the operation of this transistor validly can be explained on that basis.

SIMPLIFIED CONSTRUCTION DIAGRAMS OF JUNCTION TRANSISTORS



Operating Principles — N-P-N Transistor

In the previous review of basic types it was shown that there are two fundamental arrangements of semiconductor materials used in transistor construction. There can be a sequence in which P-material is located between surfaces of N-material, in an N-P-N arrangement; or there can be N-material located between surfaces of P-material, in a P-N-P arrangement. In either case the transistor is made up of two closely situated junctions of N- and P-type semiconductors.

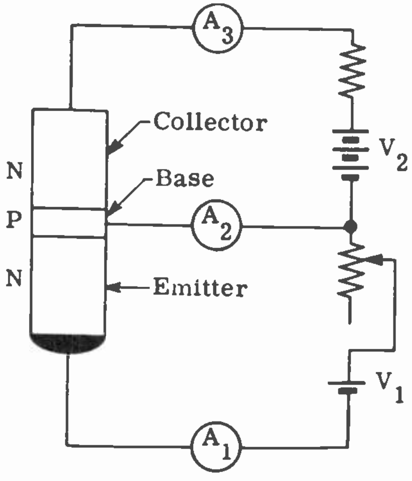


TRIODE TUBE CIRCUIT

To be considered now is how either of these arrangements can be used to amplify an electrical signal. The operation of an N-P-N transistor will be described first, since its operation most closely resembles that of the triode vacuum tube you learned about in Volume 2 of "Basic Electronics." In the explanations that follow it is important to note that the triode operation is under the control of the signal voltage applied to its input, and there is no grid current flow under ordinary conditions of operation. Transistor operation, however, depends upon signal current flow through its input circuit, and the transistor is essentially a current control device.

The circuit arrangements used to compare vacuum-tube and transistor operation are both shown on this page. Both arrangements are quite similar. Two voltage sources, V_1 and V_2, are connected across the elements of the tube and the transistor, and appropriate voltage and current meters are connected into the circuit to measure the results.

N-P-N TRANSISTOR CIRCUIT



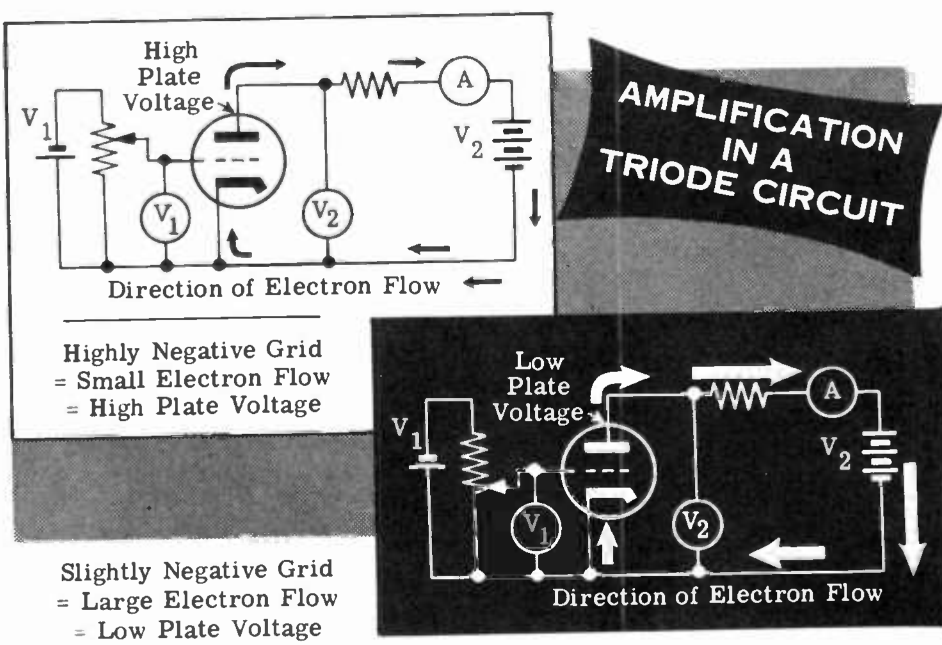
CIRCUIT FOR ANALYZING TRANSISTOR OPERATION

Operating Principles — N-P-N Transistor (continued)

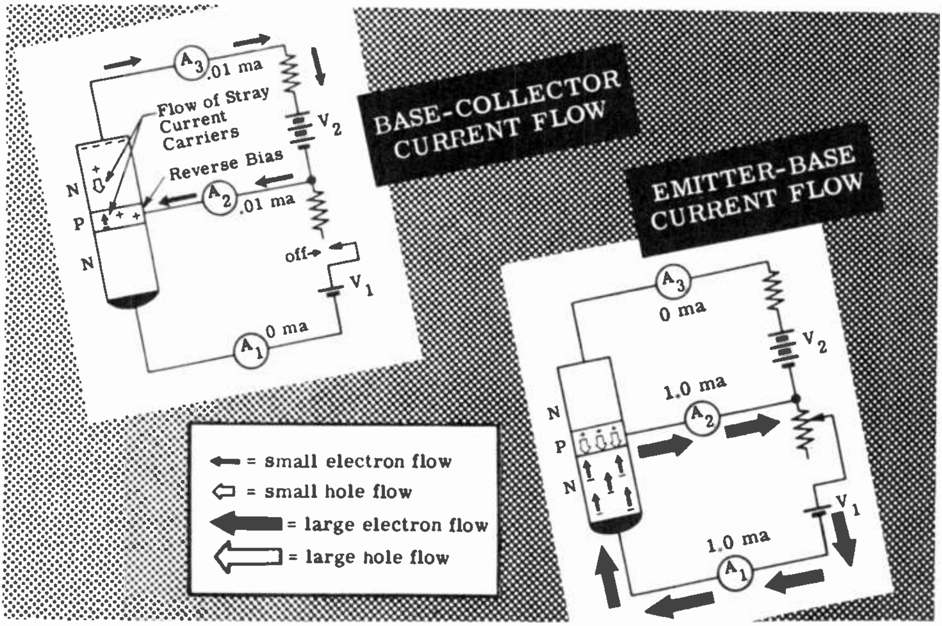
In the triode circuit electrons flow in the direction shown. They flow from the negative cathode, through the retarding negative electric field of the grid, to the positive plate, and through the outside circuit back to the cathode. The flow of electrons through the outside circuit can be increased by making the grid less negative; thus reducing the effectiveness of the grid in retarding the flow of electrons from cathode to plate. Similarly the flow of electrons through the outside circuit can be decreased by making the grid more negative; thus increasing the effectiveness of the grid in retarding the cathode-to-plate electron flow.

Amplification is obtained since a very small change in grid voltage causes a large change in plate current. Since the plate current can be passed through a large plate resistor, the change in plate current causes a large change in the voltage drop across the plate resistor. Thus, a small change in grid voltage produces a much larger change in plate voltage, and the result is signal voltage amplification. Amplification also can be obtained by passing the plate current through a step-up transformer. In this case the large change in plate current can be used to generate a large signal voltage at the output terminals of the transformer secondary winding.

In a vacuum tube power amplifier the grid bias is such that the input signal can drive the grid positive for part of the signal cycle. Since a positive grid attracts electrons, there is a flow of current in the grid circuit; and power is consumed from the source supplying the signal. The circuit is known as a power amplifier because a small amount of input power can be used to control a large amount of output power. This type of vacuum-tube operation more closely resembles that of a transistor.



Operating principles — N-P-N Transistor (continued)

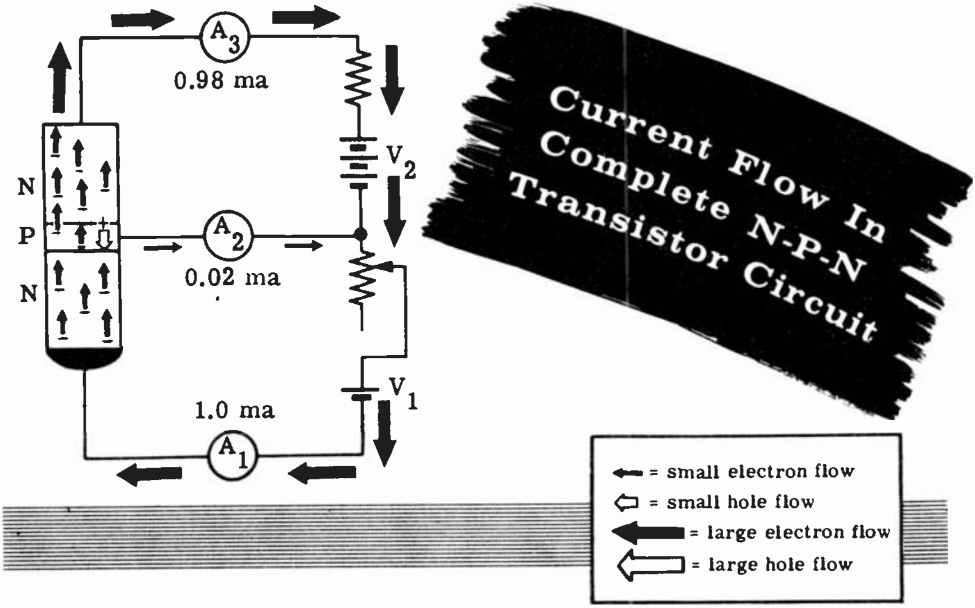


To begin an analysis of N-P-N transistor operation examine what happens when the emitter-to-base variable resistor is set to its off position. Now only the collector-to-base voltage is applied across the transistor. A look at the polarity of the voltage across the P-N junction between the base and collector shows a familiar set of conditions.

These conditions are identical to those of a semiconductor diode biased in the reverse direction. The positive terminal of the voltage source attracts the negatively charged electrons in the N-type collector, and the negative terminal of the voltage source attracts the positively charged holes in the P-type base. None of these current carriers can combine at the junction, as they do in the case of the forward biased semiconductor diode. The result is that the only current flow is that due to the stray holes in the collector and the stray electrons in the base, as in the case of the reverse biased semiconductor diode. Under these conditions the current indicated by current meters A_2 and A_3 will be very low, 0.01 milliampere for example.

To continue the analysis examine what happens when V2 is disconnected, and the variable resistor is set to the position of highest resistance. Under these conditions the emitter and base are connected as a semiconductor diode biased in the forward direction. The electrons in the emitter and the holes in the base are attracted towards the junction where they combine to maintain an appreciable current flow. For example 0.1 milliampere of current may be indicated on both A_1 and A_2 with the variable resistor set to maximum resistance. When the variable resistor is set towards minimum resistance the current flow through A_1 and A_2 may increase to 1.0 milliampere.

Operating Principles — N-P-N Transistor (continued)



The conditions change significantly when both voltage sources are connected simultaneously. If the variable resistor is set so that A_1 indicates 1.0 milliampere, A_3 will indicate approximately 0.98 milliampere, and A_2 will indicate approximately 0.02 milliampere.

There has been no increase in the voltage applied across the base and collector. With no current flow in the emitter-to-base circuit A_3 indicated 0.01 milliampere. However, with current flowing in the emitter-to-base circuit, the current through A_3 is almost identical to that flowing through A_1.

The reason for this new condition can be seen by examining the electrical conditions in the region of the base. Because of the forward bias conditions between the emitter and base there are a large number of free electrons in that region. Because the base is so thin, several thousandths of an inch thick, electrons penetrate through the base structure and come under the influence of the positively charged collector before they can combine with the holes in the base. Once the electrons are in the collector they rapidly flow through the positive terminal of V_2 and on through the outside circuit.

Since a small portion of the electrons from the emitter do combine with the holes in the base, there is a small base-to-emitter current. If the base were thicker, nearly all of the emitter electrons would combine with the base holes; and the result would be a large base-to-emitter current and a small base-to-collector current. The general rule is that the emitter electrons are divided into the two current flows shown, and the proportions of the division are determined essentially by the base thickness and the base-to-collector voltage.

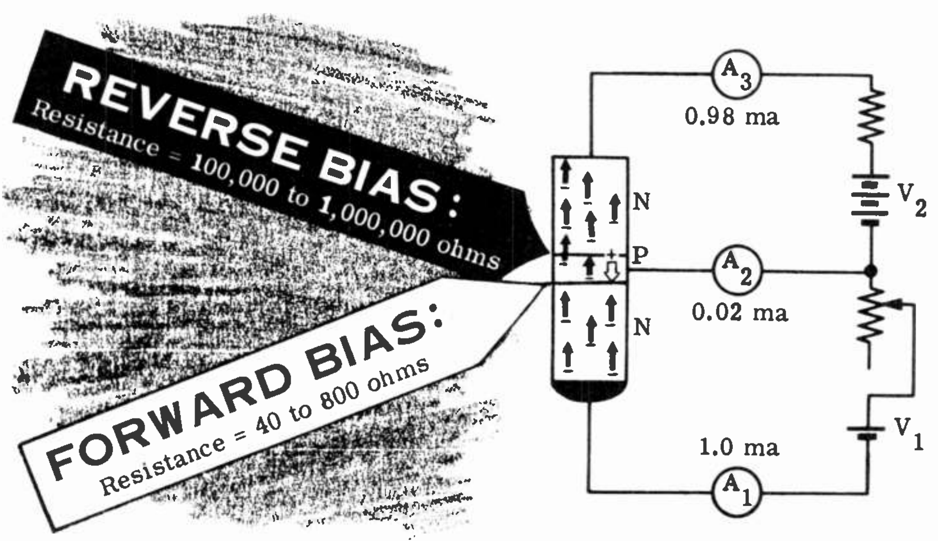
Operating Principles–N-P-N Transistor (continued)

Upon preliminary examination there may seem to be no advantage to the transistor operation as described, since 1.0 milliampere of current change in the input circuit (emitter to base) is required to produce 0.97 milliampere of current change (0.98 - 0.01) in the output circuit (collector to base). There is, in fact, a term "current gain" or "alpha" (α) applied to this condition. Current gain is defined as the output current change divided by the input current change. In this case the current gain is 0.97, and for junction transistors in general alpha falls in the range of 0.95 to 0.99; the current gain is always less than 1.

Although no useful current amplification is produced by this method of connecting a transistor, significant voltage and power gain is produced. To see why this is it is only necessary to compare the current and resistance conditions in the input and output circuits. It was seen that the current in the input and output circuits were almost identical. However, the resistance in the input and output circuits are enormously different. The bias across the emitter and base is in the forward direction, giving the junction between them a low resistance — such resistances generally range from 40 to 800 ohms.

On the other hand the bias across the base and collector is in the reverse direction, giving the junction between them a high resistance — such resistances generally range from 100,000 ohms to 1 megohm.

In Basic Electricity you learned that the voltage developed across a resistance is equal to the current multiplied by the resistance (E = RI), and you also learned that the power developed in that arrangement is equal to the square of the current multiplied by the resistance. Since almost identical currents flow in the input and output circuits and since the output circuit resistance is in the order of a thousand times higher than the input circuit resistance, it can be seen that voltage and power gains in the region of a thousand times have been produced.



N-P-N Transistor—Voltage and Power Gains

A very simple mathematical demonstration clearly indicates how voltage and power gain are produced, and it also reveals some very useful relationships and fundamental terms.

The voltage gain in any amplifying device is:

$$\text{VOLTAGE GAIN} = \frac{\text{Output Voltage}}{\text{Input Voltage}} = \frac{\text{I out} \times \text{R out}}{\text{I in} \times \text{R in}}$$

Since I out/I in (current gain) has already been defined as alpha (α)

$$\text{VOLTAGE GAIN} = \alpha \times \frac{\text{R out}}{\text{R in}}$$

For a typical transistor:

$$\text{VOLTAGE GAIN} = 0.98 \times \frac{500{,}000}{250} = 0.98 \times 2000 = 1960 \text{ Times}$$

Thus for a typical transistor connected in the manner shown previously a voltage gain of about 2000 times can be achieved. This gain is not due to any current amplification. Instead, it is entirely due to the high resistance in the output circuit as compared with the low resistance in the input circuit. Amplification is achieved because the semiconductor arrangement has transferred a current, with almost no loss, from a low resistance circuit to a high resistance circuit. This transfer through resistance is the reason that the unit is know as a transfer resistor or transistor.

N-P-N Transistor—Voltage and Power Gains (continued)

A similar mathematical demonstration shows how power gain is achieved by the transistor. The power gain in any amplifying device is:

$$\text{POWER GAIN} = \frac{\text{Output Power}}{\text{Input Power}} = \frac{I^2\ \text{out} \times R\ \text{out}}{I^2\ \text{in} \times R\ \text{in}} = \alpha^2 \times \frac{R\ \text{out}}{R\ \text{in}}$$

For a typical transistor:

$$\text{POWER GAIN} = (0.98)^2 \times \frac{500{,}000}{250} = .9604 \times 2000 = 1920.8\ \text{Times}$$

Since the ratio between output and input resistance (R out/R in) is so frequently used in voltage and power calculations, this ratio has become known as the "resistance gain" (Rg). For junction transistors resistance gain falls within the range of from 500 to 5000 times.

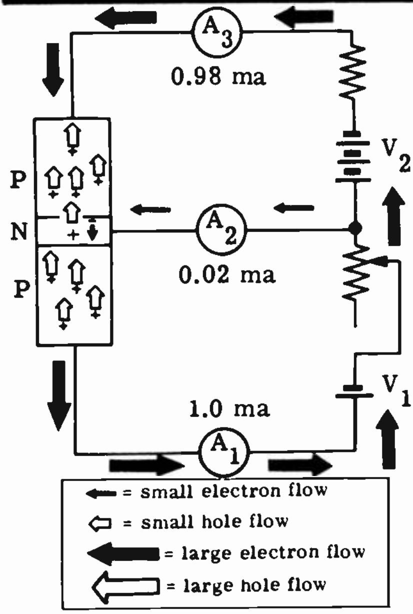
Additional useful relationships can be obtained through the direct use of the current and resistance gains:

$$\text{VOLTAGE GAIN} = \alpha\, Rg$$

$$\text{POWER GAIN} = \alpha^2\, Rg$$

Operating Principles of P-N-P Transistor

CURRENT FLOW IN COMPLETE P-N-P TRANSISTOR CIRCUIT



All the previous information presented for the N-P-N transistor also applies to the P-N-P transistor. The magnitudes of the emitter, base and collector currents are the same, and the same relationships exist for the current, voltage, resistance and power gains.

The major difference to be noted is that in the case of the N-P-N transistor the major part of the current flow through the unit is due to the movement of electrons. However, in the case of the P-N-P transistor the major part of the current flow is due to the movement of positively charged holes. Because of this feature of the P-N-P transistor it is necessary to reverse the connections of the voltage sources in order to achieve the desired current flow.

The diagram shows a P-N-P transistor connected in an arrangement equivalent to that considered previously. At the junction between the emitter and base the bias is in the direction of easy hole current flow, and holes flow into the base with very little resistance. At the junction between the base and collector the junction is biased in the reverse direction, and there is a very high resistance against the flow of the free current carriers that are normally present in the base and collector.

However, since the thickness of the base is several thousandths of an inch or less, the holes from the emitter can penetrate deeply through the base before they can combine with the free electrons in the base. Once these free holes come under the influence of the collector, they are attracted towards the negative terminal of the collector voltage source. Consequently, about 98 percent of the holes from the emitter flow into the collector, and only about 2 percent of the holes flow into the base. Except for the reversal of the voltage sources and current flows, the same relationships exist as for the N-P-N transistor considered previously.

Note that all commercially available point-contact transistors operate according to the same principles as those described for the P-N-P transistor. An unusual feature of these point-contact transistors is that current gains (α) of up to 3 times or higher are commonly achieved. No satisfactory explanation for this has been found, and the various theories are too complex to be included here. In any event, although the current gain is of some minor assistance, the resistance gain remains the major factor in determining the voltage and power gains.

In a typical point-contact transistor the resistance gain is between 65 and 70, and typical voltage and power gains are approximately 175 and 400 respectively.

Transistor Current Amplification

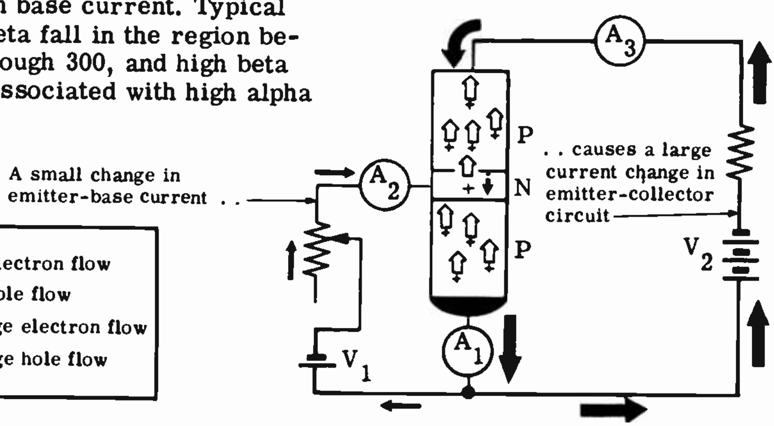
The circuit arrangements described previously indicated how voltage and power amplification could be achieved by means of a transistor. Current amplification also can be obtained and the general method used to accomplish this will be described here.

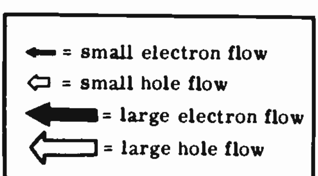
It is characteristic of transistor operation that the emitter, base and collector can be compared respectively to the cathode, grid and plate of a vacuum tube. The circuit considered previously was one in which the current change through the collector was induced by changing the current flow through the emitter. In a vacuum tube this corresponds to changing plate current by changing cathode current. This arrangement was considered first because it is the only circuit that is generally satisfactory for use with point-contact transistors and was the only arrangement used before junction transistors became available.

With a junction transistor it becomes possible to change emitter current by changing the current flow in the base circuit. This corresponds to a vacuum-tube arrangement in which plate current is varied by adjustments in grid current.

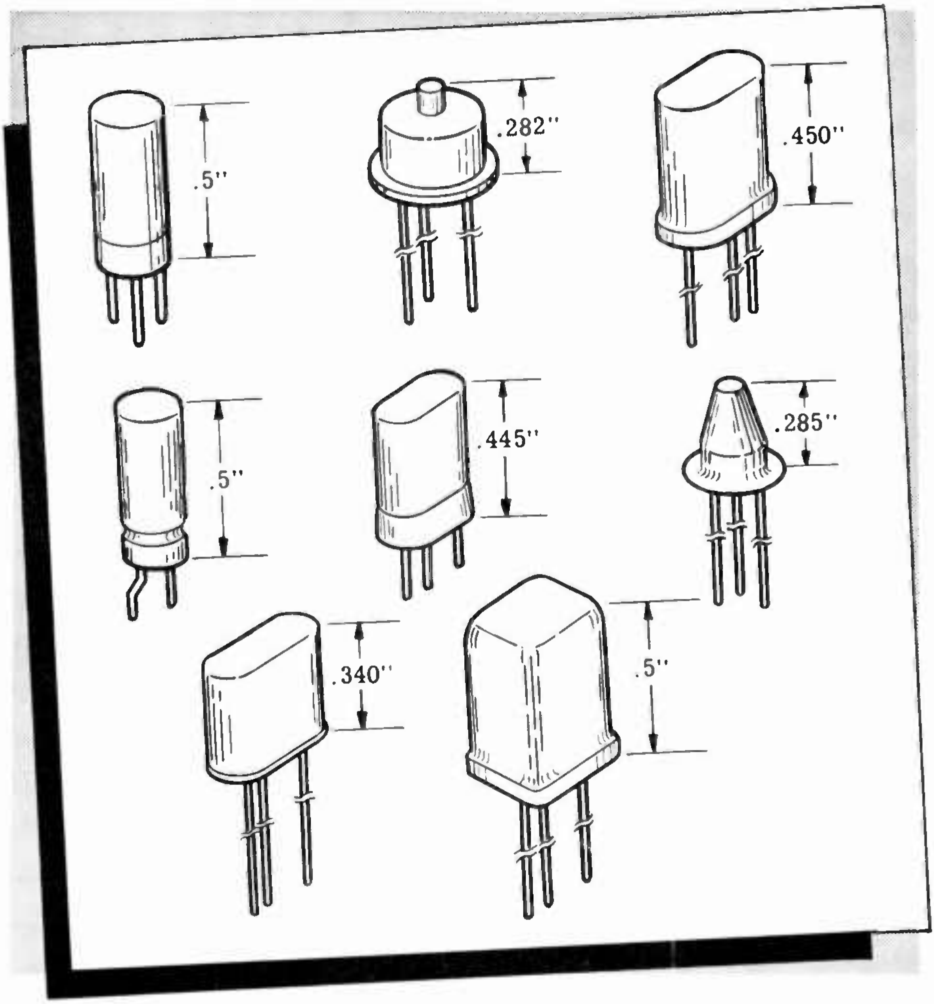
In the review of the previous circuit arrangement, it was seen that the base current generally amounts to less than 5 percent of the total current through the emitter and collector circuits. By increasing or decreasing the current through the base circuit alone, through A_2 as shown in the P-N-P transistor diagram, it becomes possible to obtain much larger current changes in the collector circuit. The reason for this is that the flow of holes through the base-emitter junction depends upon the existence of a forward bias in that region. If there is no current flow in the base circuit (through meter A_2) there can be no forward bias at the base-emitter junction, and there will be no current flow in the emitter circuit. When the current flow through A_2 increases, this indicates the presence of an increasingly strong forward bias at the base-emitter junction; and both the base and emitter currents will increase greatly. Except for the change in the direction of the bias, the same conditions apply to the N-P-N transistor.

The type of current gain produced in this manner is known as "beta" (β) current gain. β is equal to the change in collector current divided by the change in base current. Typical values for beta fall in the region between 25 through 300, and high beta values are associated with high alpha values.



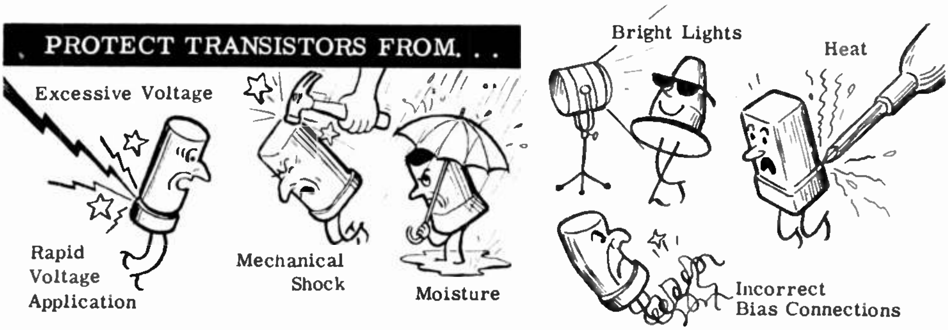


Commercial Transistors—Appearance



The illustrations on this page show the outstanding external characteristics of a wide variety of commercial transistors. The internal construction and basic operating characteristics have been described previously. Each transistor drawing has one dimension indicated in inches to give some indication of the relative sizes. Examination of these drawings show that there are a wide variety of shapes, sizes and arrangements of terminals or leads. Except for the fact that those intended for power amplification are larger and sometimes have flanged bases for conducting heat away, the variations in outside characteristics are largely due to the preferences of the different manufacturers. In all cases the manufacturer's data sheets should be the only guide concerning the recommended operating conditions and lead or terminal indentification.

Commercial Transistors—Care and Handling



Since demonstrations and experiments to be presented shortly require the use of commercial transistors, there will be included here the most important rules concerning their care and handling. Failure to follow these rules may make a transistor fail in operation or may change considerably its characteristics.

1. Unless a transistor is of the hermetically-sealed type it should not be used (or stored) in a damp place.
2. Do not drop transistors or subject them to unnecessary mechanical shocks. Although transistors will withstand considerable vibration and shock when mounted in equipment, rough handling can often apply even greater mechanical strain.
3. Some semiconductor constructions are sensitive to light. The fact that some transistors may be encased in a transparent structure does not imply that they will not be damaged if examined under a strong light. If such units must be used in brilliant light, they should be shielded by a covering of black tape or other suitable material.
4. Before installing a transistor in any circuit check the manufacturer's data sheet. Be sure to identify the emitter, base and collector terminals. Check the bias and other operating requirements and make sure that the maximum limits will not be exceeded in the circuit.
5. Always switch off the power before making or breaking transistor circuit connections. This precaution is not only for your personal safety but for that of the transistor. Applying voltage to one or two terminals before the others may damage the transistor.
6. Always check the voltage and polarity of the circuit bias supplies before connecting a transistor into the circuit. Previous changes in the circuit may apply excess or incorrectly polarized bias to the transistor.
7. Sudden application of voltage, in previously unchecked circuits, may damage a transistor. When the circuit contains controls for increasing or decreasing bias, set these controls to zero bias before connecting the transistor. After the connection is made, raise the biases slowly to the required operating point.
8. Remember that a transistor is sensitive to heat. Do not place a transistor in hot places or next to hot circuit components. When soldering the transistor into a circuit, use duck-bill or other pliers with a wide grasping area to conduct heat out of the lead before it reaches the transistor body.

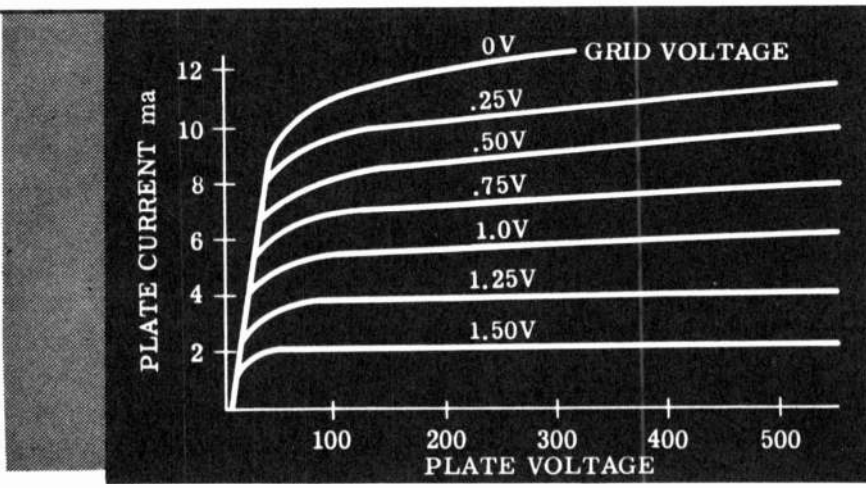
Experiment—Transistor Operating Characteristics

The purpose of this experiment is to make it possible for you to see for yourself the outstanding characteristics of transistor operation. Instead of restricting the experiment to the information presented on the previous pages, the equipment can be used to become familiar with the characteristic curves of transistors. The equipment required includes a junction transistor and point-contact transistor together with the manufacturer's data sheets. Also required are at least three sheets of graph paper, two 0-25 milliampere current meters, a 0-100 microampere current meter, and a vacuum-tube voltmeter with a polarity reversing switch. Two bias supplies, as described in a later paragraph are also required. In performing the experiment observe the precautions, previously described, for the care and handling of transistors. In addition, when plotting the characteristic curves do not exceed the current values indicated on the manufacturer's curves. To present the curves in their most useful form it is advisable to use the manufacturer's curves as a guide and to work to reproduce these curves during the course of the experiment.

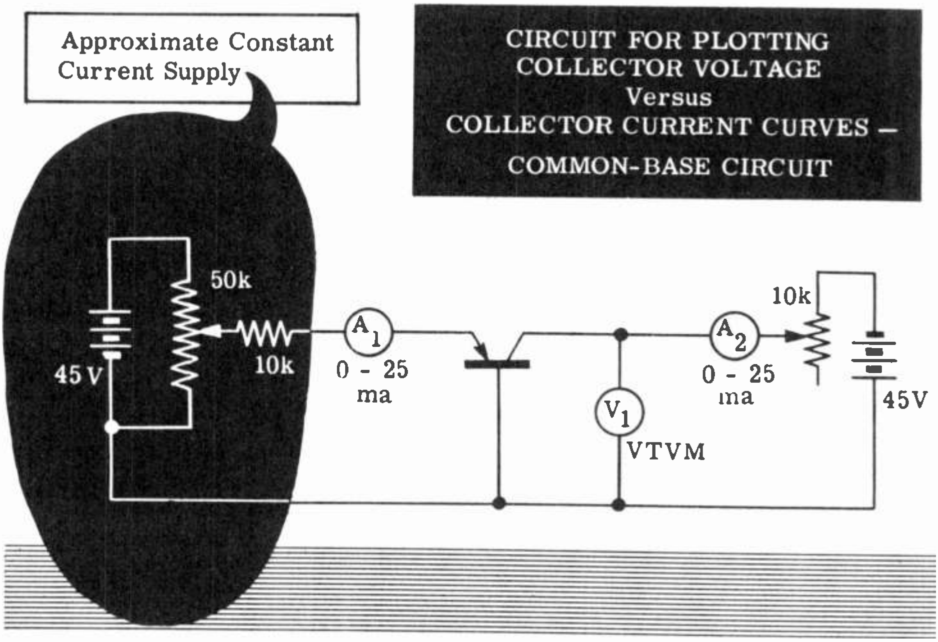
In your study of amplifier tubes in Volume 2 of Basic Electronics you saw how plate current versus grid voltage curves could be used to study the characteristics of the tube. In addition, load lines can be drawn upon those curves to learn about the characteristics of an amplifier stage based upon the use of that tube. Equivalent curves can be made for transistors, and they can be used for similar purposes. Since transistors can be connected in a variety of methods, as will be seen later, the characteristic curves are identified by the type of circuit arrangement that is used.

Transistor curves are also marked with + and - signs to indicate the polarity of the voltages developed and the direction of current flow. In marking the direction of current flow the accepted standard is that indicated by a DC current meter; it is the same as conventional current flow and opposite to electron flow.

PENTODE TUBE CHARACTERISTIC CURVES



Experiment—Transistor Operating Characteristics (continued)



An arrangement similar to that first used to learn about the relationship between emitter and collector current can be used to draw up a series of characteristic curves. One addition to this arrangement is the vacuum-tube voltmeter, which can be reversed in polarity, to measure collector voltage.

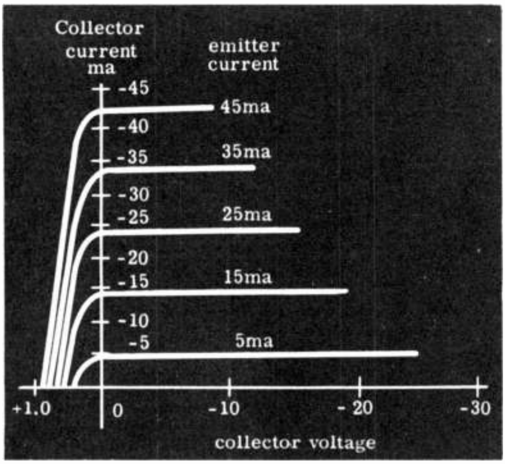
Another variation is that "constant-current" supplies are almost an absolute necessity for both bias supplies, particularly for the input circuit bias. The reason for this is that the input and output circuits of a transistor are not isolated from each other, and a current change in any one of the elements will affect the current in the other two. Such current changes are undesirable in making characteristic curve measurements, since one of the currents usually must be maintained constant while the other is varied. Any changes in the current that is expected to remain constant requires a readjustment of that current, which makes the measuring process tedious and inaccurate.

If no constant current supply is available, an approximation can be obtained by using a high voltage DC supply with the output connected in series with a large value of fixed resistance. The resistance should be at least 100 times higher than that of the transistor circuit being supplied. Thus the current in the circuit is determined by the series resistor rather than by the transistor resistance as shown in detail by the diagram. While a 45-volt battery with a 10,000 ohm series resistor is adequate for most input circuit applications, this arrangement is not practical for use in collector circuits.

Experiment—Transistor Operating Characteristics (continued)

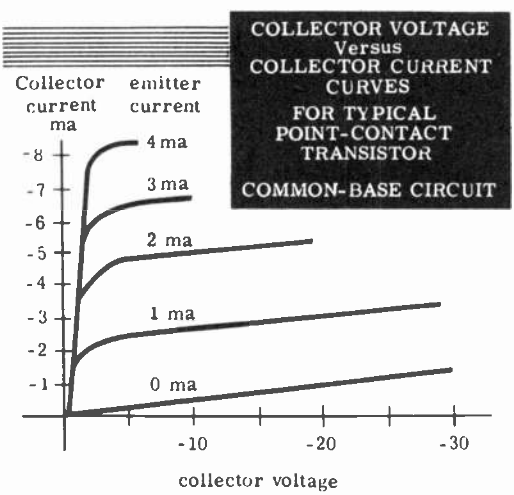
COLLECTOR VOLTAGE Versus COLLECTOR CURRENT CURVES FOR TYPICAL P-N-P JUNCTION TRANSISTOR

COMMON-BASE CIRCUIT



In making a plot of the collector current characteristic curves, the emitter current is maintained constant while the the collector current is varied. For each value of collector current the collector voltage is measured and plotted on a graph. Once a condition of collector current saturation begins for one particular value of emitter current, the process is repeated with another value of emitter current, until a complete set of curves is obtained. Note that the procedure is different from that used with vacuum tubes, where the current is measured while the voltage is varied.

When this procedure is followed with a P-N-P junction transistor, a set of curves such as those shown in the diagram is obtained. Note the resemblance between these curves and those of a pentode vacuum tube shown earlier. It is easy to see that only a small amount of collector bias is required to get all of the available current carriers to the collector and to bring about a condition of saturation. Examination of the curves shows that a change in emitter current produces a slightly smaller change in collector current—indicating a current gain (alpha) of less than one.



COLLECTOR VOLTAGE Versus COLLECTOR CURRENT CURVES FOR TYPICAL POINT-CONTACT TRANSISTOR

COMMON-BASE CIRCUIT

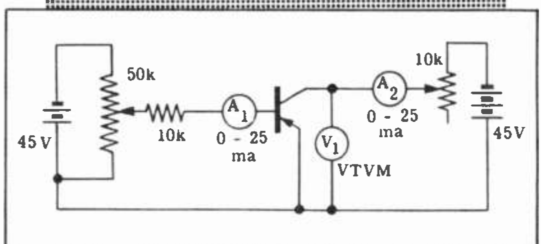
Further examination of the curves shows that the collector current does not have to be reduced to zero for the collector voltage to become zero. In fact when the collector current is reduced to zero, the collector voltage is slightly positive. This means that holes flow from the emitter through the base and into the collector even though there may be a small opposing bias across the base and collector, due to the high concentration of holes in the base and the thinness of the base.

Experiment — Transistor Operating Characteristics (continued)

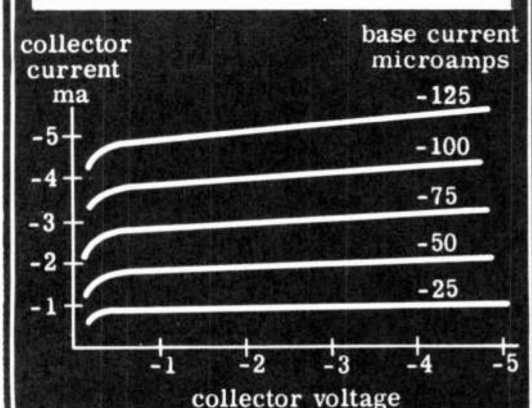
The diagram on 6-44 bottom indicates similar curves for a point-contact transistor. The major difference is that it can be seen that a change in emitter current produces a collector current change three times as large, indicating an alpha (current gain) of about three times. Also note the lack of reversal in collector voltage.

The next diagram shows the characteristic curves of collector current versus collector voltage for different values of base current. The circuit used to plot such curves is shown to the side and can be recognized as similar to that used to explain the beta (β) gain characteristics of a transistor.

CIRCUIT FOR PLOTTING COLLECTOR VOLTAGE Versus COLLECTOR CURRENT CURVES — COMMON-EMITTER CIRCUIT



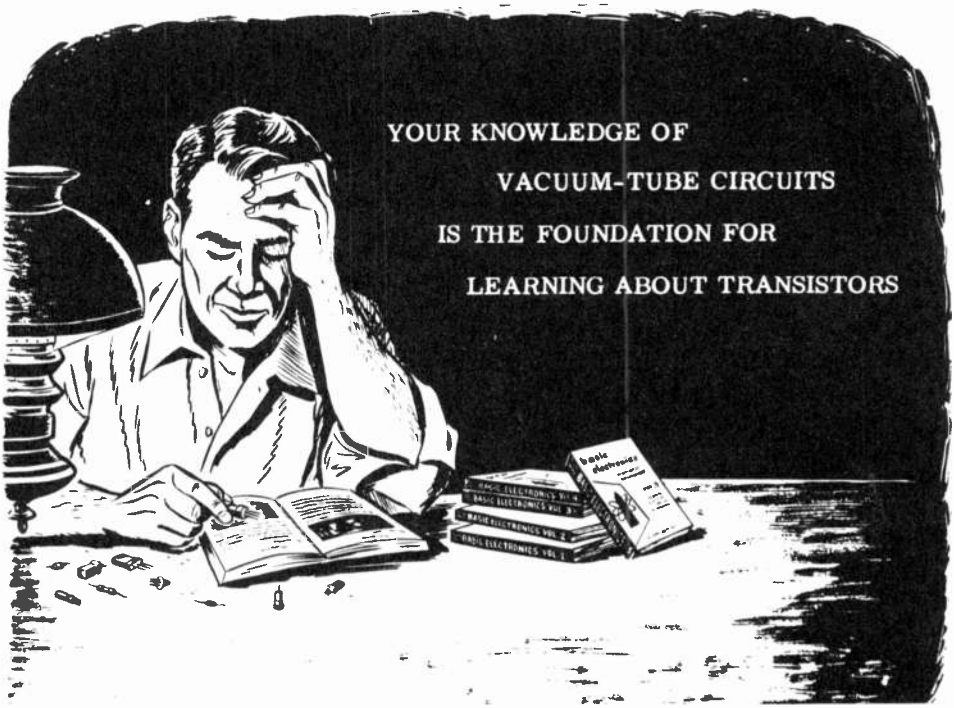
COLLECTOR VOLTAGE Versus COLLECTOR CURRENT CURVES FOR TYPICAL P-N-P JUNCTION TRANSISTOR COMMON-EMITTER CIRCUIT



To plot curves of this type maintain a fixed value of base current and vary the collector current over a predetermined range. For each value of collector current record the collector voltage. When the collector current begins to show signs of saturation, select another base current and repeat the process.

Examination of the complete family of curves shows that very small changes in base current produce large changes in collector current. It is not unusual for a 25 microampere change in base current to produce a 1 milliampere change in collector current, indicating a beta (β) gain of 40 times.

Introduction to Transistor Circuits



Now that you are familiar with the basic operating characteristics of transistors, you are ready to learn how transistors can be used in practical equipment. To accomplish this you will learn about transistor circuits in a sequence similar to that used in learning about vacuum tubes. The process will begin by examining how single transistor stages can be compared to the equivalent single-stage vacuum-tube audio amplifier. Once that is accomplished there will be a review of the methods of coupling transistor stages together to make up several types of complete audio amplifiers. Following this will be a review of transistor RF circuits, and this phase of learning about transistors will end with an analysis of a completely transistorized superheterodyne broadcast receiver.

In learning about transistor operating characteristics there were several instances where a comparison was made between transistor and vacuum tube circuits. In these comparisons it was indicated that the emitter, base and collector of a transistor could be thought of as corresponding to the cathode, grid and plate of a vacuum tube. Further comparisons of this type are helpful in bridging the gap between transistor and vacuum tube circuitry. It should be noted, however, that a point will be reached where this comparison ceases to be useful. Better progress can be made at that time by basing the explanation of more advanced transistor circuits upon the knowledge gained about the more basic transistor circuits.

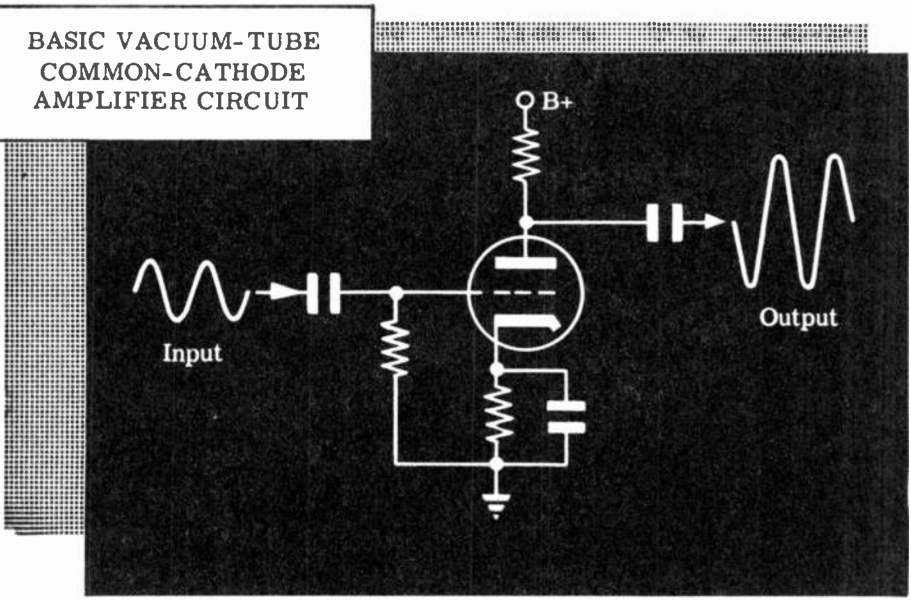
Basic Single-Stage Amplifiers

In learning about transistor operating characteristics two methods were found for obtaining the effects of amplification from a transistor. The circuits that were learned about can be converted easily into practical audio amplifier stages. Actually there are three methods of obtaining signal gain from a transistor, and these will be presented in a sequence which is based upon your knowledge of vacuum-tube audio amplifier stages.

The vacuum-tube amplifier circuit with which you are most familiar is the one shown in the diagram. In this arrangement the input signal is applied across grid and ground, and the output signal appears across the plate and ground. Actually, because of the cathode bypass capacitor, the input is across grid and cathode; and the output appears across plate and cathode. Because of the circuit arrangement this amplifier is popularly known as the "grounded-cathode" circuit. More correctly it should be called the "common-cathode" circuit, since both plate current and grid current (when present) flow through the cathode circuit, and since any one of the tube elements may be grounded without changing this characteristic.

A very brief review of this circuit will clarify the explanation of its transistor equivalent. When the input signal makes the grid more positive (or less negative), the plate current increases. The result is an increase in the voltage drop across the load resistor, and the plate voltage decreases. When the input signal makes the grid less positive (or more negative), the plate current decreases. The result is a decrease in the voltage drop across the load resistor and the plate voltage increases. Since the change in plate voltage is always in a direction opposite to that of the grid voltage, there is a 180-degree phase reversal between the input and output signal.

BASIC VACUUM-TUBE COMMON-CATHODE AMPLIFIER CIRCUIT



Basic Single-Stage Amplifiers (continued)

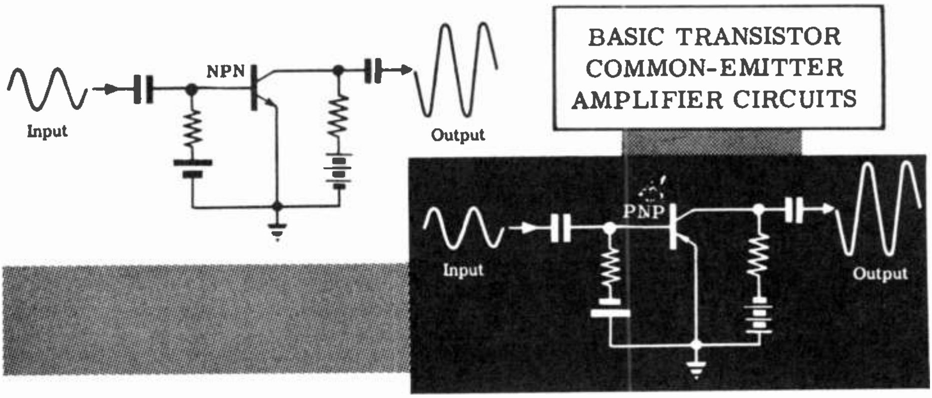
The equivalent transistor stage resembles the circuit used to learn about beta current amplification. It can be seen that the emitter, base and collector of the transistor correspond respectively to the cathode, grid and plate of the vacuum-tube circuit. Because of the circuit arrangement this circuit is popularly known as the "grounded-emitter" circuit, although it is more correct to call it the "common-emitter" circuit. The operation is the same whether a P-N-P or N-P-N transistor is used, although the polarities of the bias voltages would be reversed as shown in the diagram.

First consider the operation with an N-P-N transistor. When a positive-going input signal drives the base more positive, the forward bias across the base-emitter junction is increased. More electrons flow from the emitter into the base, and, as described earlier, there is an increase in collector current. This increase in current results in an increased voltage drop across the load resistor and the collector voltage decreases (becomes less positive). When the input signal becomes negative, some of the positive bias on the base is cancelled; and the forward bias across the base-emitter junction is decreased. Fewer electrons flow from the emitter into the base, and there is a decrease in collector current. This decrease in current results in a decreased voltage drop across the load resistor, and the collector voltage increases (becomes more positive).

Since the change in collector voltage is always in a direction opposite to that of the base voltage, there is a 180-degree phase reversal between the input and output signal.

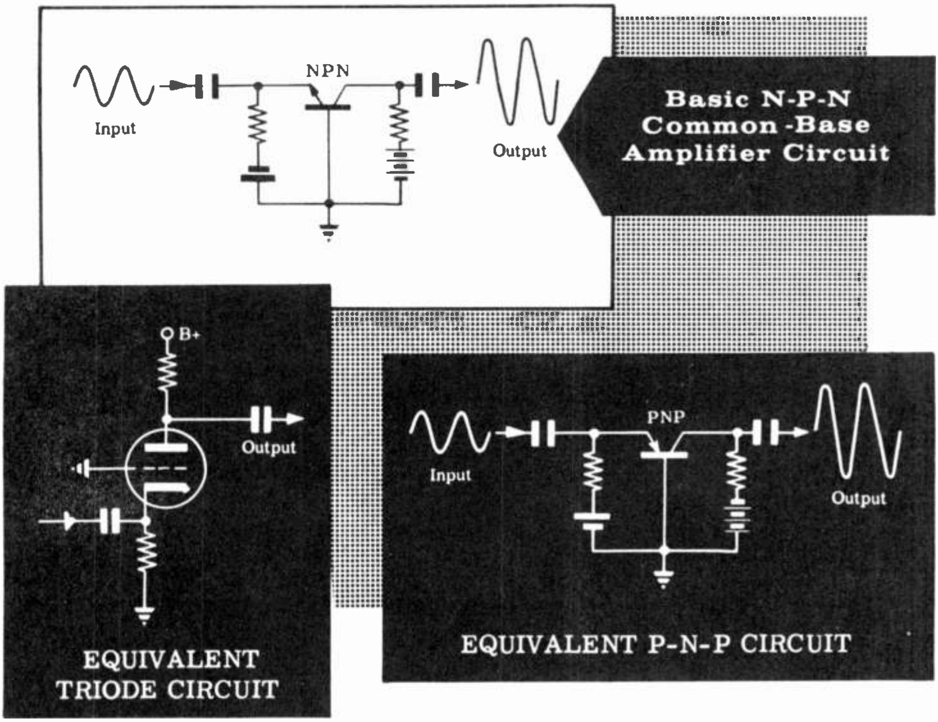
The same overall results take place when a P-N-P transistor is used, although all of the bias changes are reversed. When a positive-going signal makes the base less negative, fewer holes flow from the emitter into the base, the collector current decreases, and the collector voltage increases (becomes more negative). A negative-going signal makes the base more negative, more holes flow from the emitter into the base, the collector current increases, and the collector voltage decreases (becomes less negative).

The outstanding advantage of this circuit is that it produces higher power amplification than the other types to be considered. Disadvantages are that this circuit has the greatest tendency to oscillate, and has impedance-matching problems that will be considered later.



BASIC TRANSISTOR COMMON-EMITTER AMPLIFIER CIRCUITS

Basic Single-Stage Amplifiers (continued)



Basic N-P-N Common-Base Amplifier Circuit

EQUIVALENT TRIODE CIRCUIT

EQUIVALENT P-N-P CIRCUIT

The second basic amplifier arrangement is known as the "grounded-grid" or "common-grid" circuit in its vacuum-tube form. In its transistor equivalent the arrangement is known as the "grounded-base" or "common-base" circuit. This arrangement resembles the circuit used to learn about alpha current amplification.

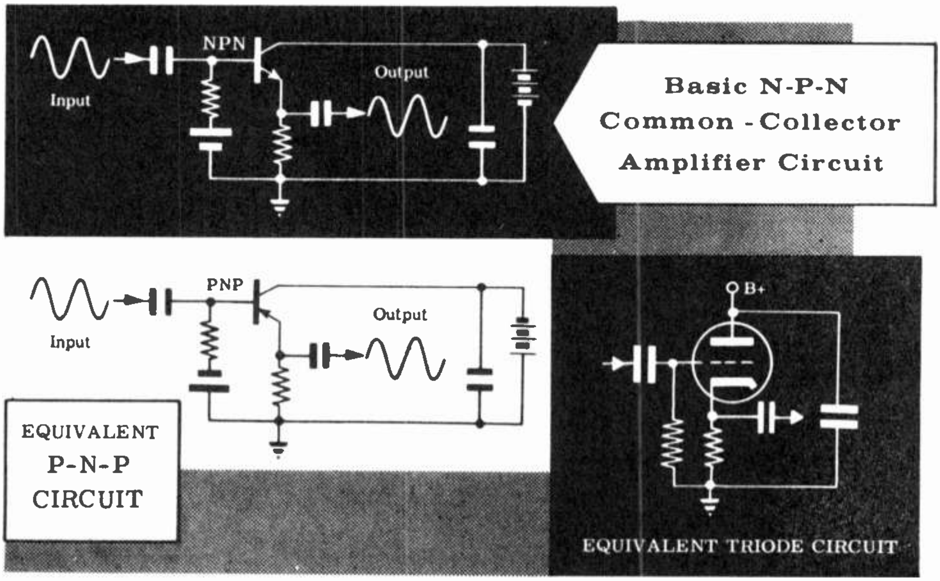
Because of previous coverage of this material a description of the N-P-N transistor form of this circuit is sufficient, and the operation of the vacuum-tube and P-N-P transistor variations are quite easily understood.

When the input signal becomes positive, part of the forward bias across the emitter-base junction is cancelled. The result is a decrease in emitter current and a decrease in collector current. The voltage drop across the load resistor decreases, and the collector voltage becomes more positive.

When the input signal becomes negative, there is an increase in the forward bias across the emitter-base junction. The result is an increase in both emitter and collector current. The voltage drop across the load resistor increases, and the collector voltage becomes less positive.

Since the collector voltage rises and falls with the input signal, there is no phase reversal between the input and output signals. The outstanding advantages of this circuit are that it produces the least noise and has the least tendency to oscillate. The amplification produced is not as high as for the common-emitter circuit.

Basic Single-Stage Amplifiers (continued)



The third basic amplifier arrangement is known as the "grounded-plate" or "common-plate" circuit in its vacuum-tube form. In its transistor equivalent the arrangement is known as the "grounded-collector" or "common-collector" circuit. This arrangement is essentially identical to the single-tube phase inverter learned about in your study of audio amplifiers, except that only the cathode signal circuit is used.

Again, because of your familiarity with this circuit and because of previous comparisons, a description of the N-P-N transistor form is sufficient explanation for all forms of this circuit.

When the input signal becomes positive, the forward bias across the emitter-base junction is increased. The result is an increase in emitter current and an increase in collector current. The voltage drop across the load resistor (emitter to ground) increases, and the emitter voltage becomes more positive.

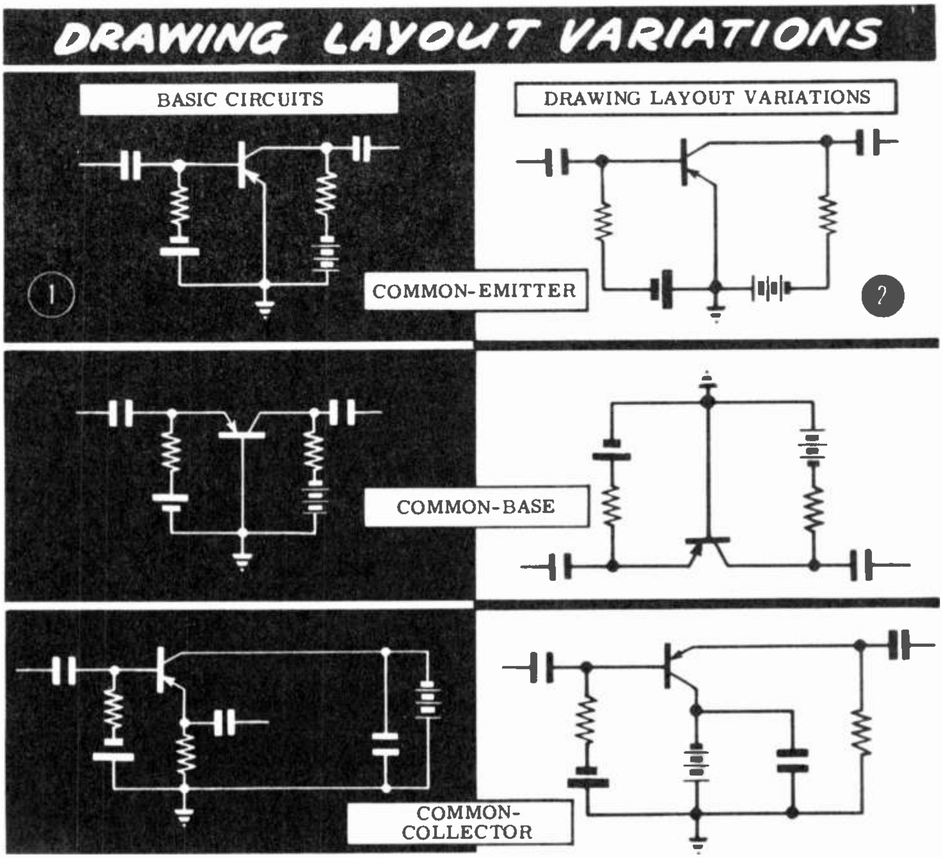
When the input signal becomes negative, part of the forward bias across the emitter-base junction is cancelled. The result is a decrease in emitter current and a decrease in collector current. The voltage drop across the load resistor (emitter to ground) decreases and the emitter voltage becomes less positive (more negative).

Since the emitter voltage rises and falls with the input signal, there is no phase reversal between the input and output signals. The outstanding characteristic of this circuit is that, as in the case of the phase inverter, the voltage gain is always less than one, although power gains of over ten times may be produced. The circuit has input impedances in the order of 250,000 ohms and the output impedance is often below 1,000 ohms. This makes the circuit useful for matching high impedance sources to low impedance outputs.

Amplifier Circuit Variations

As you learned from your previous study of vacuum-tube circuitry, there are many common variations of any simple circuit. Not only are these variations due to different methods of drawing the circuit, they are also due to actual differences in the number of components used, in the types of components used, and in the wiring of the various components. The fact that there are variations should not lead to confusion, because the basic circuit operation is fundamentally the same, and the basic circuit can be recognized in spite of the changes. Learn to recognize the basic circuit in spite of the variations.

To see how this affects the transistor circuits you have learned so far, examine the diagrams on this and the next page. Detail 1 on this page shows three types of circuit arrangements (common emitter, common base and common collector). Detail 2 shows one variation for each type, and the difference is only in the manner the various circuit components are arranged in the circuit. To prove that the difference is only one of drawing layout you can check the connections of each component with those in Detail 1.



Amplifier Circuit Variations (continued)

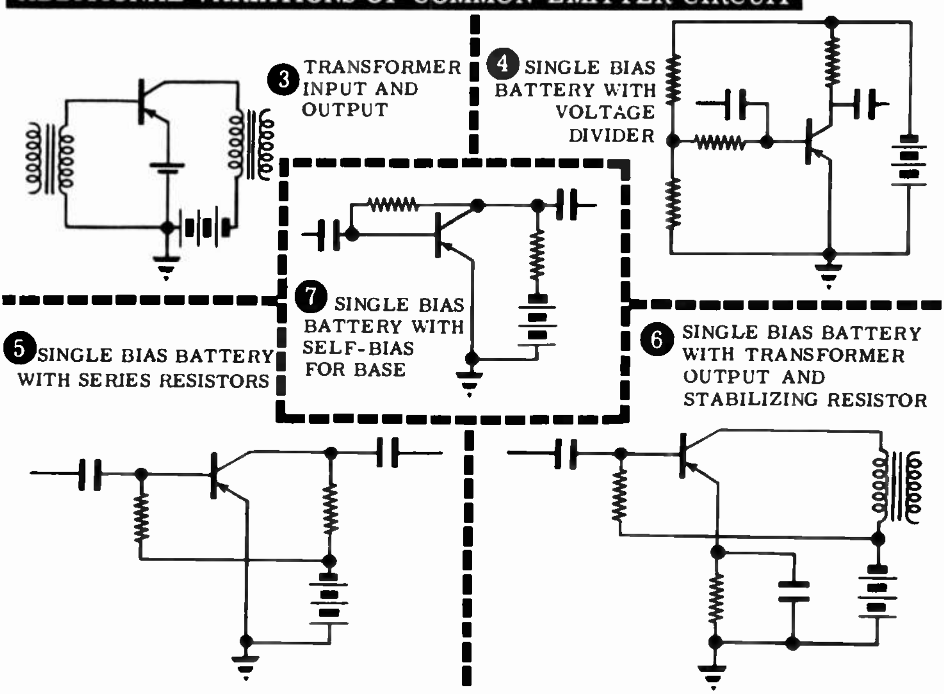
The drawings on this page show five more variations for the common-emitter circuit shown in Detail 1 on the previous page. Similar variations exist for the common-base and common-collector circuits.

Detail 3 shows a second variation of the common-emitter circuit. Although there are drawing layout differences, the main difference is that transformer coupling, rather than RC coupling, is used at the input and output.

Detail 4 includes drawing layout differences, but the major change is that one bias battery is used instead of two. In transistor applications one battery is often used instead of two. The two different biases that are required can be obtained from a single battery, either by means of a voltage divider arrangement (as in Detail 4) or by means of individual series resistors from the battery to the collector and base (as in Detail 5).

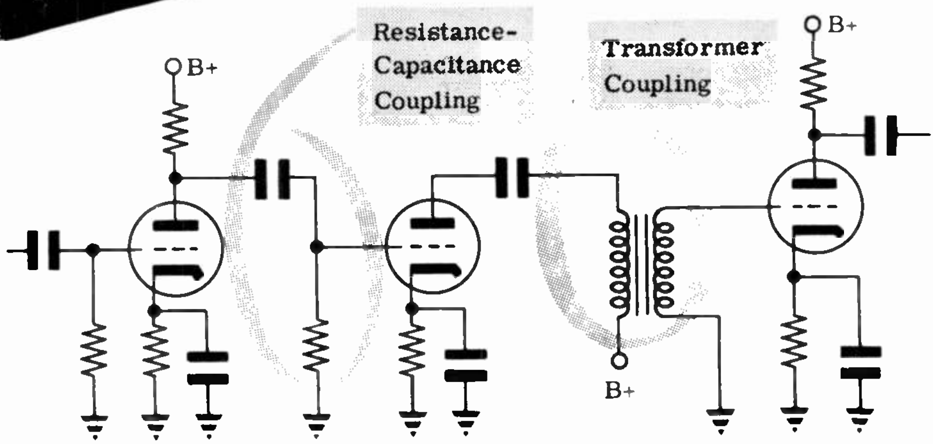
A variation (Detail 6) shows how the primary winding of an output transformer can be substituted for the battery-to-collector resistor. Also shown in that detail is a bypassed resistor in the emitter-to-ground circuit. This resistor is known as a "stabilizing" resistor; it generates a self-bias voltage, similar to cathode bias, which maintains the desired operating characteristics in spite of current variations in the other circuit elements. Detail 7 shows how self-bias can be generated by a resistor between base and collector. The base bias resistor increases and decreases the base current as required to maintain stable operation.

ADDITIONAL VARIATIONS OF COMMON-EMITTER CIRCUIT



Multiple-Stage Transistor Amplifiers

BASIC VACUUM - TUBE COUPLING METHODS

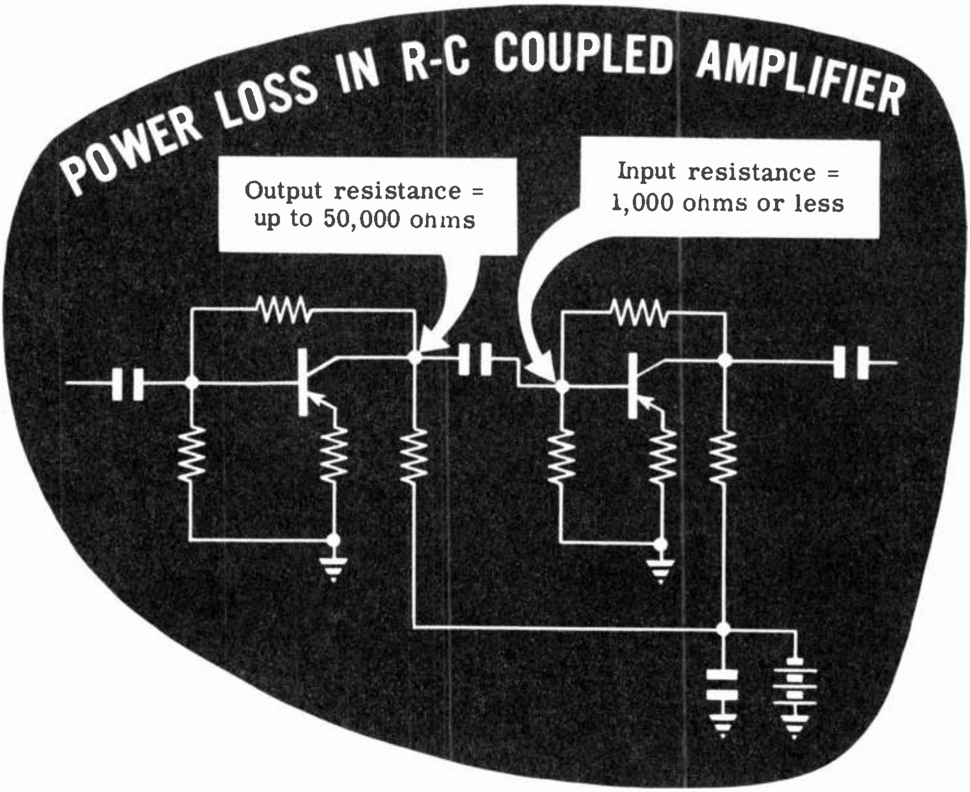


Although large amounts of power, voltage and current amplification can be produced by a single transistor stage, the requirements of practical equipment generally demand even higher amounts of amplification between the input and output. In your study of vacuum-tube circuits you gained experience with the amplification requirements of audio amplifiers, video amplifiers and radio receivers. During that study you learned about the various methods of achieving the needed amplification by coupling together or "cascading" a number of amplifier stages. The same general methods are used in multiple-stage transistor amplifiers.

As in vacuum-tube amplifiers, the input stage must be impedance-matched to the signal input source—the microphone, antenna or signal generator; and the output stage must be impedance-matched to the output load —the loudspeaker, earphones, transmission line or meter. The methods that can be used for coupling the amplifier stages to the input source and output load, or for coupling amplifier stages to each other, include transformer coupling and resistance-capacitance coupling. To refresh your memory the diagrams shown here illustrate these two methods of coupling vacuum tubes.

Another coupling method, direct coupling, makes use of a direct connection between the plate of one stage and the grid of the next stage. Since the plate and grid have vastly different bias requirements, complex supplementary biasing must be used to cancel the excess plate voltage from the grid. Because of these complications, direct coupling is rarely used with vacuum tubes; but in transistor circuits it becomes simple enough for practical use, as will be seen shortly.

Multiple-Stage Transistor Amplifiers (continued)



It is true that multiple-stage transistor amplifiers are based upon the same general principles as vacuum-tube amplifiers and employ the same general interstage coupling techniques. It is not true that such a transistor amplifier can be understood clearly from the facts you have learned about vacuum tube amplifiers.

In multiple-stage transistor amplifiers the problem of assembling the various stages together is complicated by the peculiar characteristics of the different types of individual amplifier stages. Both the grounded-base and grounded-emitter types have input resistances of 1000 ohms or lower, and the grounded-base type input resistance may be as low as 25 ohms. On the other hand, the output resistances of these stages may be as high as 50,000 ohms for the grounded-emitter type and 500,000 ohms for the grounded-base type. If resistance-capacitance coupling is used in cascading amplifier stages of these types, there is a very large difference between the output resistance of one stage and the input resistance of the next stage. Your knowledge of impedance-matching shows you that there is a very large power loss in the coupling network between the stages; and such losses result in significantly reduced gain through the overall amplifier. To make up for these losses it is sometimes necessary to add one or more stages to the amplifier.

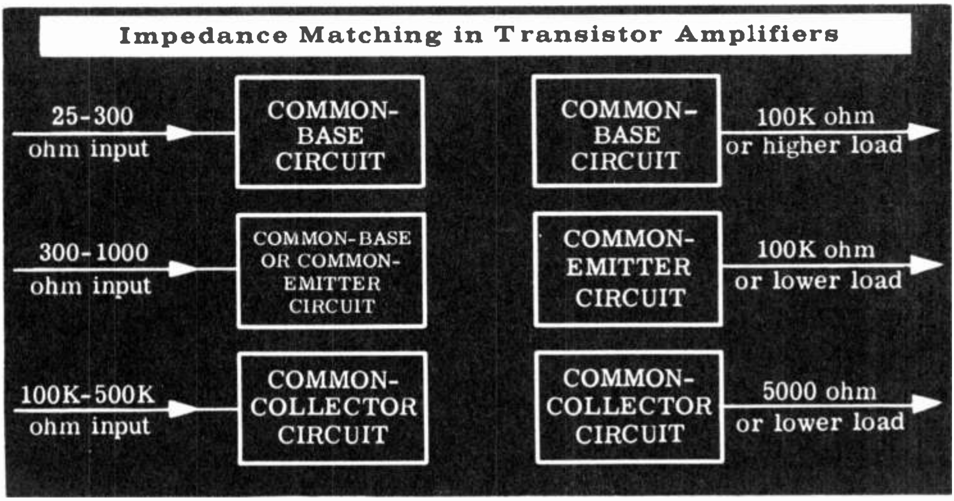
Multiple-Stage Transistor Amplifiers (continued)

Although the use of resistance-capacitance coupling results in a large loss in amplification, it is simpler and less expensive than the other methods available. Because of these desirable characteristics several methods have been devised for retaining this type of coupling while partially overcoming the important problem of mismatch.

Matching the first amplifier stage to the signal source can be accomplished by selecting a type of amplifier circuit that has an input resistance essentially equal to the source. When the input signal source has a resistance between 25 and 300 ohms (as is the case with many transmission lines, low impedance microphones or phonograph pickups) a common-base type of input stage is used. If the characteristic resistance of the signal source is in the range between 300 and 1000 ohms, as is the case with some transmission lines and special transducers, either a common-base or common-emitter type of input stage may be used. In the event that the signal source has a resistance in the order of several hundred thousand ohms, a common-collector type of input stage is used.

A similar technique is used to match the final amplifier stage to the output load. For load resistances below 5000 ohms or below 100,000 ohms the common-collector and common-emitter types of output amplifier stages are used respectively. For load resistances above 100,000 ohms the common-base type of output amplifier stage is used. Unfortunately, there is no very efficient direct method of transmitting high levels of signal power to a load of less than 500 ohms; a transformer is generally required in such cases.

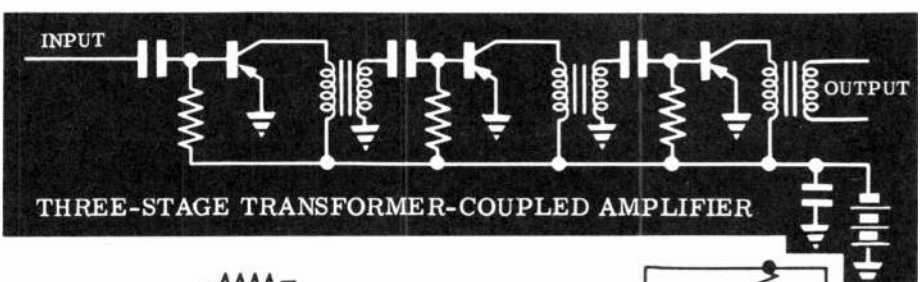
In power amplifiers a common-collector stage with its high input resistance (100,000 to 350,000 ohms) and low output impedance (700 to 25,000 ohms) can be used to present a fairly good impedance match between stages of the other types. Although the voltage gain is less than one, there is good power gain (about 10 times), and the mismatch losses are eliminated effectively.

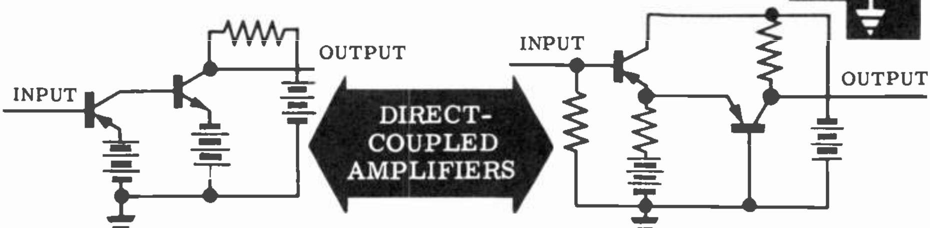


Multiple-Stage Transistor Amplifiers (continued)

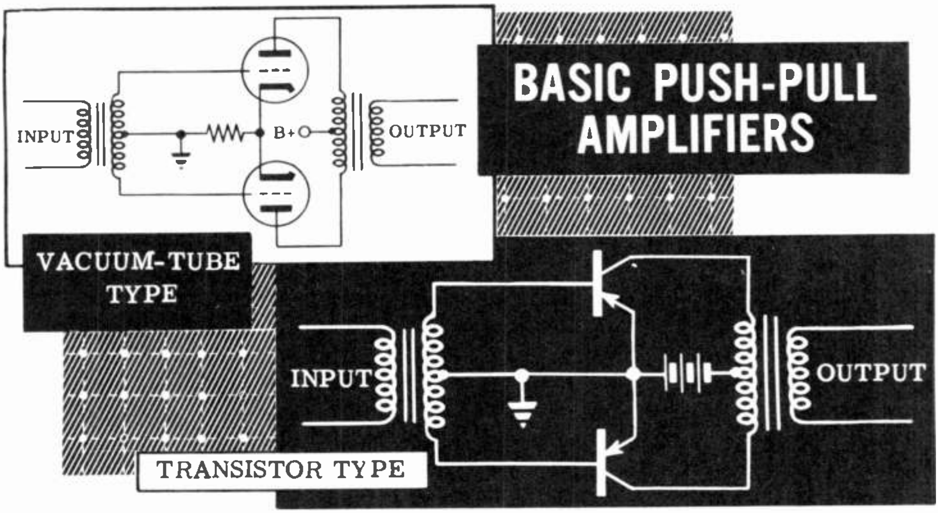
Transformers offer the most efficient method of matching transistor stages to their inputs and outputs. Not only is this method efficient but it is fairly simple, as can be seen by the diagram. Notice that the diagram shows a capacitor placed between the base and the transformer secondary. The purpose of this capacitor is to prevent the bias on the base from being shorted to ground. The reason for the efficiency of transformers is that they can be manufactured to have almost any desired combination of input and output impedances, and almost perfect matches can be achieved. The advantage of the use of transformers is evidenced by the fact that one or two less transformer-coupled amplifier stages are required than in a resistance-capacitance amplifier of equivalent gain. Using transformers has the disadvantage of added cost, space and weight. Even though mass production techniques are used in manufacturing such transformers, and miniaturized models are widely available, it is a rare transistor amplifier that employs transformer-coupling exclusively.

Direct-coupling techniques, as shown in the diagrams, are used in special amplifiers where the phase shift introduced by other coupling methods cannot be tolerated. Direct coupling is also used in the amplification of signals at frequencies too low to be coupled efficiently by the previously described methods. Output devices such as meters, relays, earphones and other units can be direct-coupled by connecting them in series with the output element of the final amplifier stage. If the device has a high impedance, it can be used in conjunction with the common-base and common-emitter types; and if it has an impedance below 10,000 ohms it can be connected in series with the output of a common-collector stage. Note that in a direct-coupled amplifier P-N-P and N-P-N transistors may be used in conjunction with each other, resulting in a simplified biasing arrangement without impedance-matching problems. This method is known as the "complementary symmetry" circuit arrangement, and further examples will be seen in the review of push-pull amplifier arrangements.





Push-Pull Audio Amplifiers



In your study of vacuum-tube amplifiers you learned that the power output requirements are sometimes more than can be produced by a single output stage. Under such conditions the power output can be increased by using two tubes in the output stage, and the push-pull arrangement produces the highest power output with the least distortion. Shown in the diagram is a push-pull power amplifier with which you are very familiar.

In this circuit the output signal of the previous stage is applied to the primary of a transformer. The secondary of the transformer is center-tapped and applies equal and opposite voltages to the grids of the push-pull tubes. Because opposite ends of the transformer secondary are applied to the two grids, the signal voltage on one grid is becoming more positive while the voltage on the other grid is becoming more negative. Thus the plate current in one tube increases while that in the other tube is decreasing. Because the DC plate currents in the output transformer primary flow in opposite directions, the transformer core is not easily saturated; and high operating efficiency can be obtained from the transformer.

In audio amplifiers the power output that can be obtained from a single output stage employing ordinary transistors is usually limited to well below 150 milliwatts. An output at such a low level is quite adequate for driving earphones but is normally considered insufficient for use with a loudspeaker. By using a push-pull power transistor arrangement in the final amplifier stage, outputs of from 250 milliwatts up to 10 watts can be obtained readily with Class B operation. This is quite adequate for driving the loudspeaker of a portable radio.

The transistorized equivalent of the vacuum-tube circuit is also shown in the diagram. It can be seen that the two circuits are almost perfect duplicates, and the explanation given previously applies equally well to both circuits. It should be noted that the operating characteristics of the two transistors must be carefully matched to obtain efficient operation and low distortion.

Push-Pull Audio Amplifiers (continued)

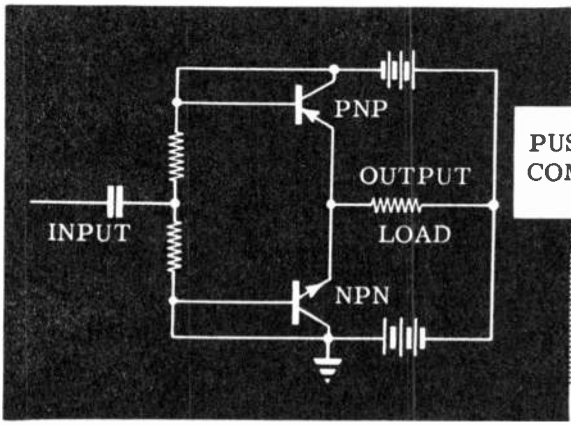
A unique and useful method of obtaining push-pull operation is by means of complementary symmetry. Circuits making use of this arrangement do not contain many of the coupling components usually required. These circuits operate on the basis that P-N-P and N-P-N transistors can be made with equivalent operating characteristics while demanding oppositely polarized bias sources. The arrangement shown here operates in a satisfactory manner without either input or output transformers.

Both transistor bases are driven without phase reversal by the output signal of the previous stage. The push-pull input effect is obtained because the emitters of both stages are connected to the load. Because of the balanced voltage divider arrangement shown at the left of the circuit, the two bases are equally and oppositely biased with respect to each other. The center resistors are low in value, so that there is little loss in signal amplitude; and equal signals appear on both bases.

The transistor circuits can be recognized to be of the common-collector type. Since one circuit is of the P-N-P type while the other circuit is of the N-P-N type, push-pull operation is obtained even though the bases are driven in phase with each other. When there is no signal applied, the currents through the load resistor are equal and opposite; and they cancel each other. The biasing selected usually cuts off the collector current when zero-signal is applied, giving Class B operation for maximum power output.

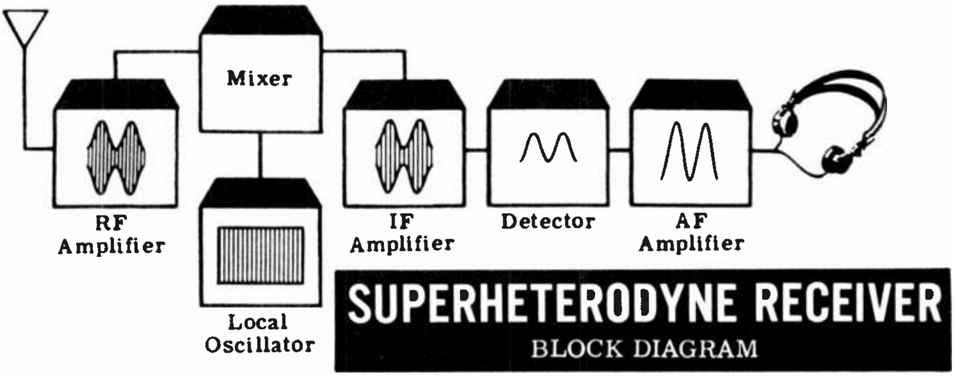
When the input signal becomes positive, the upper transistor does not conduct. However, the lower transistor conducts and current flows from the emitter, through the load resistor and into the battery center tap. When the input-signal becomes negative, only the upper transistor conducts; and current flows from the battery center tap, through the load resistor to the emitter. Thus push-pull operation is achieved.

Since a common-collector circuit is employed, a low impedance output is achieved without an output transformer. The negative feedback obtained through the common load resistor balances out small differences in transistor operating characteristics, and precise matching of the two transistors is not required.



PUSH-PULL AMPLIFIER USING COMPLEMENTARY SYMMETRY

Introduction



Now that you are acquainted with some of the outstanding features of transistor audio amplifiers, you are ready to learn about transistor superheterodyne receiver circuits. Because you are already very familiar with the vacuum-tube version of these circuits, only a brief review of their purposes will be necessary.

The superheterodyne receiver antenna picks up a small portion of the transmitted signal and delivers it to the input transformer of the RF Amplifier. The RF Amplifier steps up the amplitude of the signal to which it is tuned and does not amplify signals at other frequencies. Thus the purpose of the RF Amplifier is to add selectivity and sensitivity to the receiver. The RF Amplifier is often eliminated in table-top and portable receivers and the antenna signal is led directly to the converter.

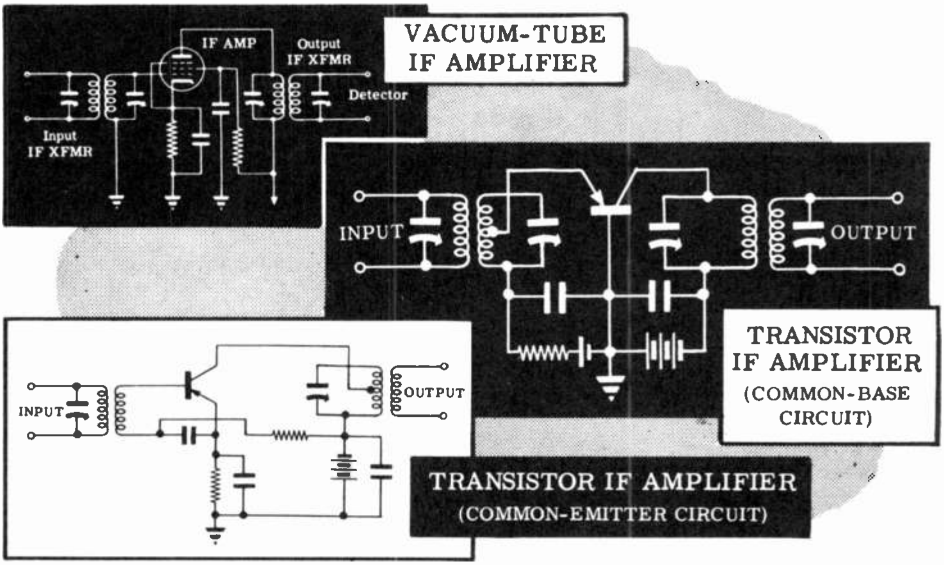
In the converter, often known as the Mixer or first detector, the incoming signal in the CW output of a Local Oscillator are mixed together. The plate current of the converter is varied according to both of these signals which are at different frequencies. A beat, or difference signal appears at the output; and this signal is fed to the IF Amplifier, which is tuned to this difference frequency.

The IF signal has the same modulation as the RF carrier, and the only change is a lowering of the frequency of the carrier signal. When the RF Amplifier and converter, or the converter alone, are tuned to the incoming signal, the oscillator is also tuned. The tuning is arranged so that the difference signal is always at the same frequency, and the IF Amplifier can be fixed-tuned.

The IF Amplifier is permanently tuned to the fixed-frequency signal coming from the Mixer stage. Since no variable tuning is required, the IF Amplifier is designed for maximum amplification and high selectivity. From one to three IF Amplifiers are normally included in a superheterodyne receiver, and the intermediate frequency normally used in broadcast receivers is 456 kilocycles.

The Detector stage, sometimes known as the second detector, receives the output of the IF Amplifier. Acting as a rectifier, the detector removes the IF carrier signal and leaves only the audio signal. This audio signal is stepped up in amplitude by the Audio Amplifier, the output of which drives the earphones or loudspeaker.

IF and RF Amplifier Stages



As is the case with their vacuum-tube counterparts, transistor IF and RF Amplifiers have basically the same design. Also as in the case of their vacuum-tube counterparts, the major design problem in these stages is achieving adequate gain at high frequencies. In both cases adequate gain is normally accomplished by the use of tuned circuits.

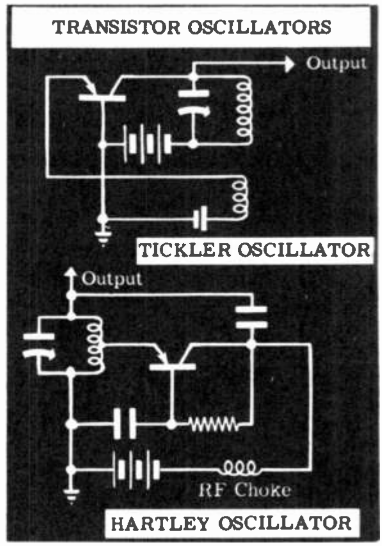
The first diagram shows a vacuum-tube IF Amplifier with its transistor equivalent. The transistor circuit is based upon the common-base arrangement. Only two significant differences can be seen in the transistor amplifier. First are the differences due to the change in bias requirements. Second is the fact that the emitter circuit is connected to a tap on the secondary winding, rather than to the end of that winding. The purpose of this is to provide a good impedance match between the two stages; the high output impedance of the collector circuit is effectively matched to the low input impedance of the emitter which follows, while maintaining the normal tuning requirements of an IF Transformer.

Also shown is a variation of the transistor IF Amplifier. In this two-stage circuit a common-emitter arrangement is used. The major difference here is that the impedance match is accomplished in the transformer primary. Instead of being connected to the end of the primary winding, the collector is connected to a tap on the primary. In addition the secondary winding is untuned, as is sometimes done in vacuum tube circuits.

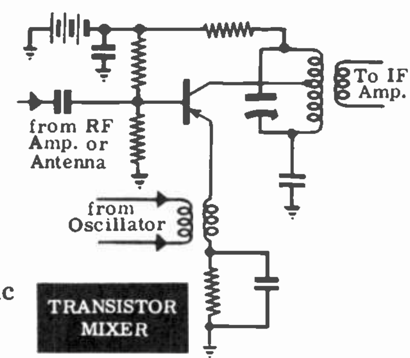
Except for the fact that their frequency range is higher and wider, transistor RF Amplifiers resemble the transistor IF Amplifiers shown in the diagram. To make a variable-tuned RF Amplifier the variable capacitors in the diagram should be mechanically ganged and connected to a tuning knob and dial. Because of the complexities of ganging the tuning capacitors of variable-tuned amplifiers, there is usually only one tuned circuit in such an arrangement.

Oscillator and Mixer Stages

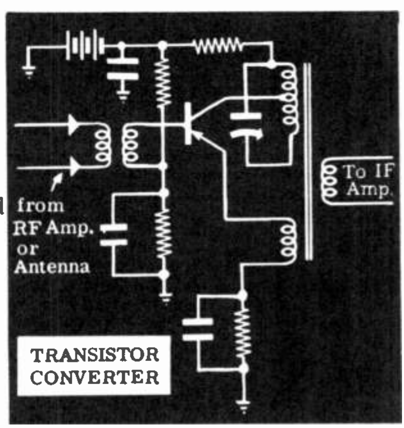
As in the case of vacuum-tube oscillators, transistor oscillators operate by means of a positive feedback network. This arrangement takes a portion of the output signal and feeds it back to the input circuit. The phase relationship is such that the input signal is reinforced, and oscillation takes place at a frequency determined by the tuned circuit. The basic principles involved are identical to those described in Volume 3 of Basic Electronics.



Just as there is a wide variety of vacuum-tube oscillators, there are many kinds of transistor oscillators. There are transistorized versions of the tickler and Hartley oscillators, which are shown in the diagrams. The operating principles are the same in both cases, and the only significant differences are those required for maintaining the proper bias on the various elements. In addition, there are a number of oscillator circuits peculiar to transistors. These can always be recognized in a schematic diagram by the fact that each contains a series- or parallel-tuned circuit with some form of coupling between the input and output circuit.

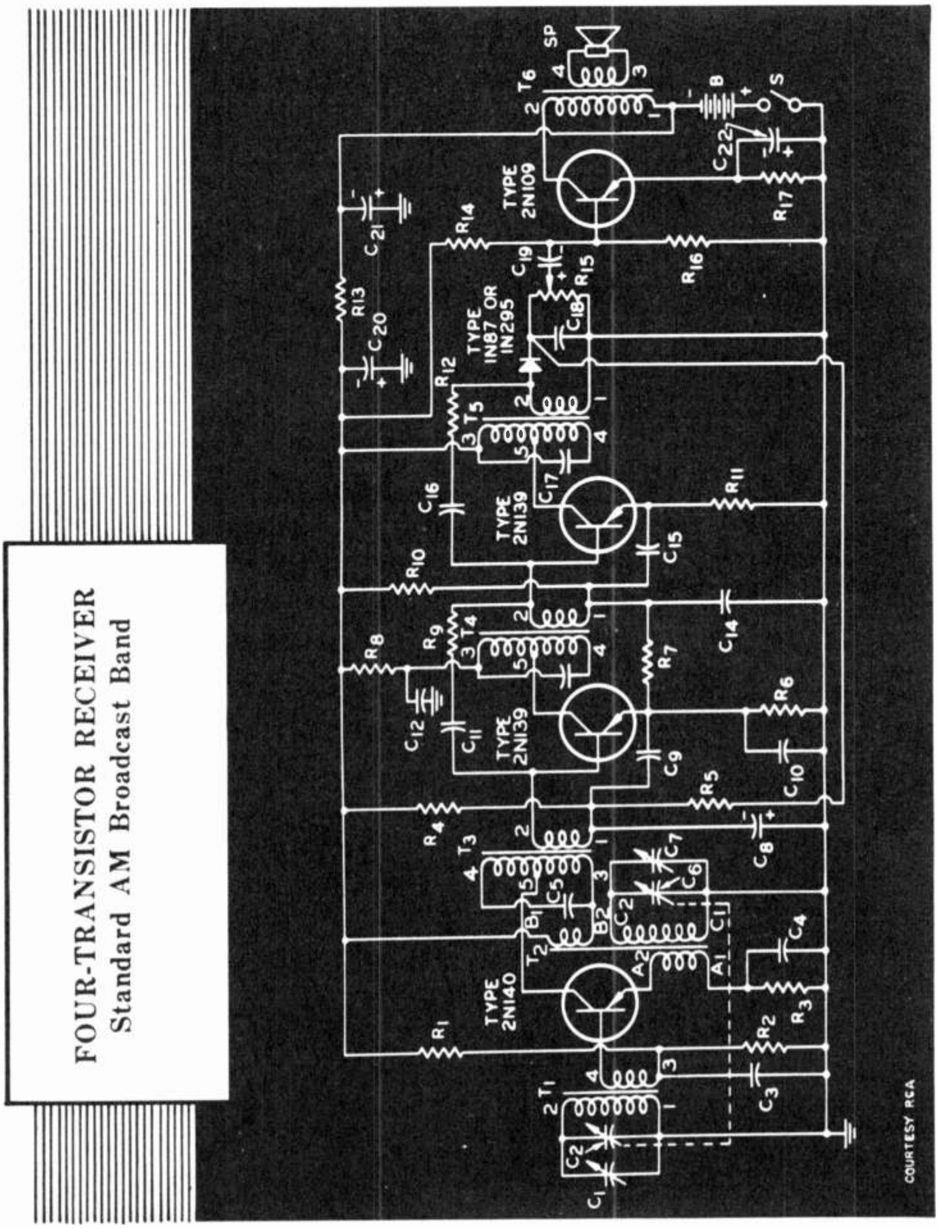


Also shown on this page is the schematic diagram of a typical mixer circuit. The emitter is frequently biased by means of a bypassed resistor to ground, and the local oscillator signal is injected into it by means of transformer coupling. The modulated RF signal from the antenna or RF Amplifier is coupled to the base. Operation is explained in the same manner as for the vacuum-tube mixer in Volume 5 of Basic Electronics. It was also described in that volume how an oscillator and mixer circuit could be combined to make a converter stage which combines the functions of both. The transistor version of this stage is shown in the last diagram.



Experiment—Transistor Receiver Operation

The purpose of this experiment is to examine a typical battery-powered superheterodyne receiver and to observe its operation. The equipment required is the receiver shown in the schematic diagram on this page, an RF signal generator equipped with amplitude modulation, an audio signal generator, plus an oscilloscope or audio signal output meter.



B = 9 volts, VS300 or VS301
C1 = Trimmer capacitor, 0-20 μμf
C2 = Variable capacitor, 12-230 μμf
C3 = 0.05 uf, paper, 150 v.
C4 = 0.04 μf, mica, 150 v.
C5 = 220 μμf, mica, 150 v.
C6 = Variable capacitor, 10-105 μμf
C7 = Trimmer capacitor, 0-20 μμf
C8 = 10 μf, electrolytic, 15 v.
C9 = 0.05 μf, paper, 150 v.
C10 = 0.05 μf, paper, 150 v.
C11 = 75 μμf, mica, 150 v.
C12 = 0.05 μf, paper, 150 v.
C13 = 220 μμf, mica, 150 v.
C14 = 0.05 μf, paper, 150 v.
C15 = 0.05 μf, paper, 150 v.
C16 = 33 μμf, mica, 150 v.
C17 = 220 μμf, mica, 150 v.
C18 = 0.05 μf, paper, 150 v.
C19 = 2 μf, electrolytic, 15 v.
C20 = 50 μf, electrolytic, 15 v.
C21 = 50 μf, electrolytic, 15 v.
C22 = 100 μf, electrolytic, 15 v.
R1 = 33,000 ohms, 1/8 watt
R2 = 24,000 ohms, 1/8 watt
R3 = 820 ohms, 1/4 watt
R4 = 120,000 ohms, 1/8 watt
R5 = 18,000 ohms, 1/8 watt
R6 = 1200 ohms, 1/4 watt
R7 = 75,000 ohms, 1/8 watt
R8 = 560 ohms, 1/4 watt
R9 = 560 ohms, 1/4 watt
R10 = 150,000 ohms, 1/8 watt
R11 = 1200 ohms, 1/4 watt
R12 = 2000 ohms, 1/8 watt
R13 = 100 ohms, 1/4 watt
R14 = 10,000 ohms, 1/8 watt
R15 = Volume control, 2500 ohms, 1/4 watt
R16 = 5600 ohms, 1/8 watt
R17 = 220 ohms, 1/4 watt
SP = Speaker

T1 = Antenna transformer wound on the largest feasible ferrite core to provide the following characteristics:

Primary Inductance (With secondary open)	353 μh
Primary Q at 1 Mc, mounted on chassis with secondary open	200
Equivalent output resistance across secondary terminals, at 1 Mc with primary tuned	600 ohms

The primary should be wound on one end of the ferrite rod with spacing between turns equal to the thickness of the wire.
The secondary should be wound on the opposite end of the ferrite rod with no spacing between turns (close wound). The end of the primary winding nearest the secondary is the ground end. Use #7/41 Litz wire. A ferrite rod about 8″ long and 3/8″ in diameter will provide excellent results.

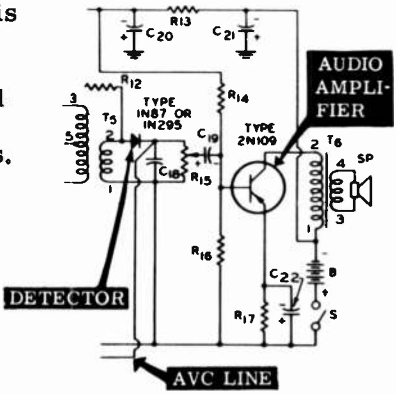
T2 = Oscillator transformer. Wind as follows:
Using a threaded resinite coil form about 1″ long with internal threads to match a 1/4″-diameter ferrite core about 3/8″ in length, wind 4 turns near the center of the coil form. This winding is located between the terminals marked A1 and A2 on the circuit diagram. The winding between terminals B1 and B2 consists of a 10-turn winding and is wound on top of the 4-turn winding. The winding between terminals C1 and C2 is wound on top of the other two windings and is a multilayer 115-turn winding. This winding should extend over a length of about 3/8″. All the windings are universally wound with #5/44 Litz wire. A resinite collar with 6-terminals may be fitted over one end of the coil form to provide secure anchorage for the ends of the windings and convenient terminals for making connnections to the windings. When the transformer is completed, connect a 120-μμf capacitor across the 115-turn winding and tune the ferrite core to obtain resonance at 1.455 Mc with the other two windings open circuited. The Q of this winding under the same conditions should be 100 or greater.

T3, T4, T5 = Intermediate-frequency transformers. Using materials like those described for the oscillator transformer T2, wind T3, T4, and T5 to meet the following requirements:

	T3	T4	T5
Tuned resistance at primary tap	118,000	15,300	17,200 ohms
Primary (terminals 3 and 5): Reflected resistance with secondary terminated	206,000	29,000	10,900 ohms
Secondary (terminals 1 and 2): Reflected resistance with primary terminated	1,000	500	1,000 ohms
Turns Ratios:			
Terminals 4 and 3 to terminals 5 and 3	1.17	2.48	3.16
Terminals 5 and 3 to terminals 2 and 1	14.35	7.62	3.3
Core (Ferrite):			
Unloaded Q (mounted in chassis)	110	61	110
Loaded Q (mounted in chassis)	35	35	35

T6 = Class A output transformer, primary impedance = 500 ohms, secondary impedance = 12 ohms.

Experiment—Transistor Receiver Operation (continued)



Two basic procedures will be used in this experiment. First, each stage will be identified to determine its purpose, and then the operation of that circuit will be checked by using the signal injection techniques described in Volume 5 of Basic Electronics.

To begin the experiment, examine the diagram to identify the audio output amplifier. The output amplifier can be identified at the right of the diagram by the fact that it is connected to the loudspeaker. This stage can be identified as a common-emitter circuit by the fact that the input signal is connected to the base, and the output transformer is connected to the collector. Voltage is supplied to the collector by connecting it through the output transformer primary winding to the negative terminal of the battery. Bias is supplied to the emitter circuit by passing its current through a capacitor bypassed resistor coming from ground. A voltage divider formed by R14 and R16 provides negative bias to the base. This bias is obtained from the negative terminal of the battery through the RC filter formed by C20, R13 and C21.

Check the operation of the output stage by signal injection techniques. Connect the oscilloscope or output meter across the output transformer secondary winding terminals 3 and 4. Connect the audio signal generator across R16. Turn on the receiver by closing power switch S. Turn on the audio signal generator and vary its frequency; the audio signal will be heard coming from the loudspeaker and will cause an indication on the oscilloscope or output meter.

The detector stage can be seen in the diagram to the left of the audio output amplifier. The detector stage consists of a single semiconductor diode arranged in a half-wave rectifier output circuit identical to that described previously. The input signal is provided by the IF output transformer to the left, and the audio output signal appears across the volume control, R15.

Note the lead that connects the junction between the diode and C18 to the second stage through R5. From your knowledge of the vacuum-tube superhet receiver, this circuit can be recognized as one which supplies automatic volume control (AVC) voltage to the IF Amplifier.

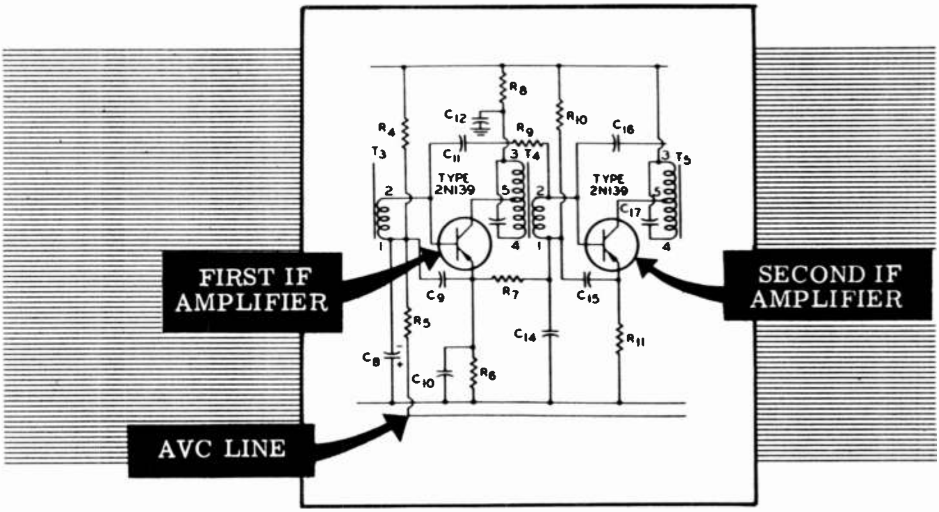
To check the operation of the detector stage connect the RF signal generator across the secondary winding of IF Transformer T5, across terminals 1 and 2. Turn on the signal generator and set it to 455 kilocycles. When the signal generator amplitude modulation is turned on, an audio signal will be heard coming from the loudspeaker and will indicate on the oscilloscope or output meter.

Experiment—Transistor Receiver Operation (continued)

Further examination of the schematic diagram shows that there are two almost identical stages to the left of the detector stage. Closer examination shows that these stages contain fixed-tuned transformers, and that is the basis for identifying them as IF Transformers. Both of these stages can now be recognized as being equivalent to the IF stages described previously. The two stages are of the common-emitter type, and obtain stabilizing emitter bias by means of their emitter-to-ground resistors R6 and R11. The first stage achieves maximum gain by employing bypass capacitor C10 to conduct the signal from ground to the emitter around R6. The second stage does not contain such a capacitor across R11, thereby resulting in negative feedback which achieves stability in that stage. Both collectors obtain their bias from the filtered B- line, and a decoupling network formed by R8 and C12 prevents interstage feedback which might result in oscillation.

Note that both IF Amplifiers include neutralizing circuits (C11 and R9, also C16 and R12). These circuits provide negative feedback between the output and input, in order to prevent oscillation.

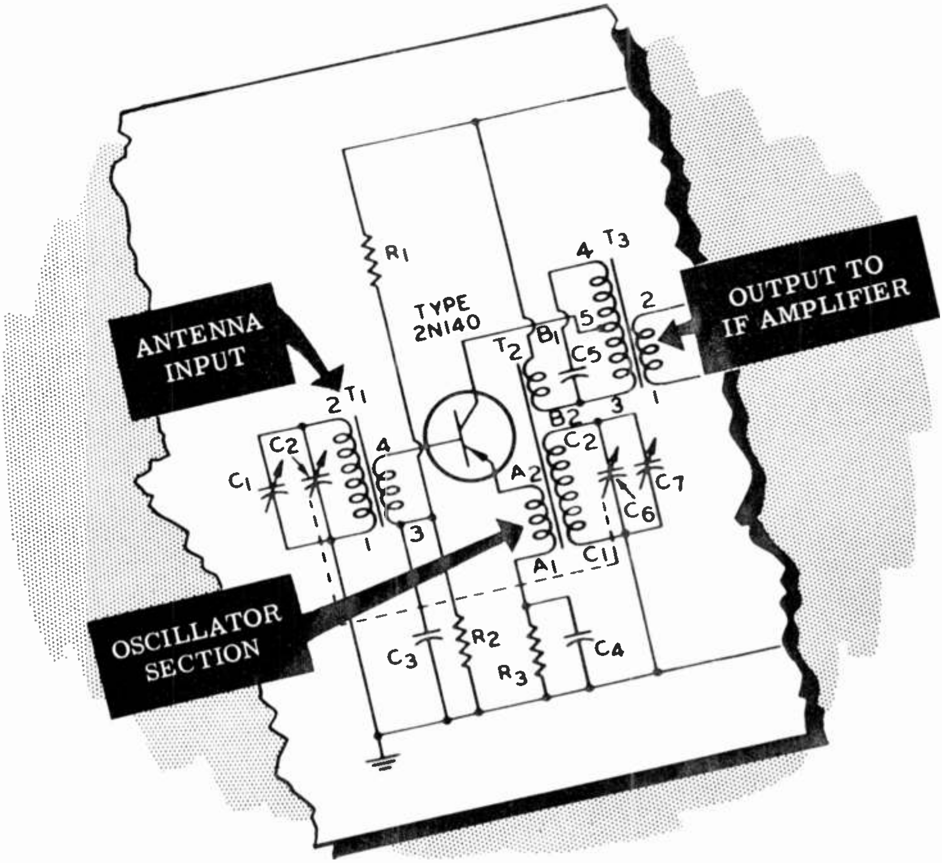
To test the operation of the If Amplifiers turn on the RF signal generator and allow its setting to remain at 455 kilocycles with the amplitude modulation on. Connect the generator across base and ground in each IF Amplifier, and the audio signal will be heard coming from the loudspeaker. Vary the signal generator frequency about its initial setting, and the loudness of the sound coming from the loudspeaker will be heard to vary. In addition, the amplitude of the oscilloscope or output meter indication will be seen to vary. It will be noticed that there is a maximum audio output when the signal generator is set to 455 kilocycles, the frequency to which the IF Transformer are tuned. The audio output decreases rapidly as the signal generator frequency is tuned away from 455 kilocycles, showing that the IF Amplifiers have a narrow bandpass.



Experiment—Transistor Receiver Operation (continued)

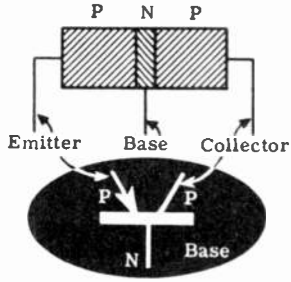
The last stage to be considered is the one shown to the extreme left of the diagram. This stage can be immediately recognized as the converter stage considered previously on page 6-61. The oscillator function can be recognized by the fact that transformer T2 couples the output and input circuits. The mixer function can be recognized by the fact that the oscillator signal is injected into the emitter circuit, while the incoming RF signal is injected into the base circuit. Collector voltage is supplied from the filtered B- line through windings of T2 and T3. Emitter bias is supplied by means of the capacitor-bypassed resistor R3.

Operation of the converter stage, and of the complete receiver can be observed by using the RF signal generator to inject a broadcast-band signal into the primary winding of T1, across terminals 1 and 2. Set the RF signal generator to a frequency of 1 megacycle, with amplitude modulation on, and tune the receiver until the audio signal is heard from the loudspeaker and seen on the oscilloscope or output meter. Since the primary winding of T1 acts as an efficient loop antenna, radio signals will be heard if the receiver is tuned with the signal generator disconnected.

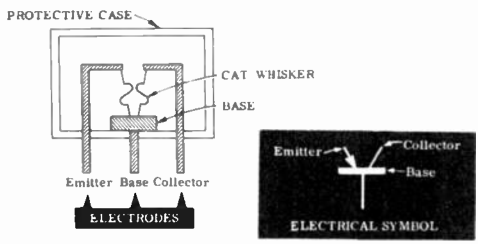


Review of Transistors

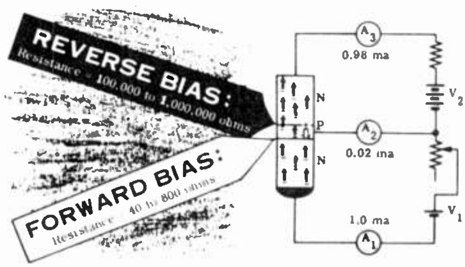
JUNCTION TRANSISTOR—A junction transistor consists of semiconductor materials in an N-P-N or P-N-P sequence. The junctions may be formed during the crystal growing process (grown junctions), by a dissolving and recrystallization process (alloy junctions), or by an etching and plating process.



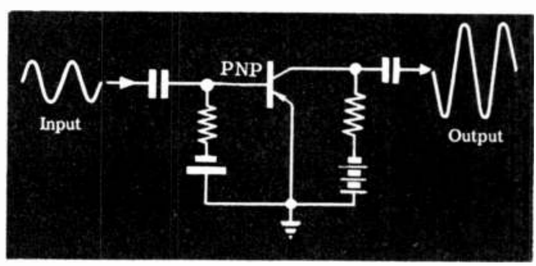
POINT-CONTACT TRANSISTOR—A point-contact transistor consists of a block of N-type semiconductor in contact with two closely spaced pointed metal wires. The construction can be considered the equivalent of a P-N-P junction arrangement.



TRANSISTOR AMPLIFICATION-Amplification is achieved because the transistor transfers current from a low resistance circuit to a high resistance circuit. This function is the reason that the unit is known as a transfer resistor or transistor.

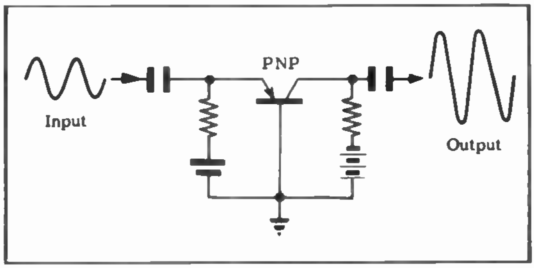


COMMON-EMITTER AMPLIFIER—This type of circuit operates in the same general manner as the common-cathode vacuum-tube arrangement. This circuit has the highest power amplification and the greatest tendency to oscillate. Input resistance is 1000 ohms or lower; output resistance is as high as 50,000 ohms.

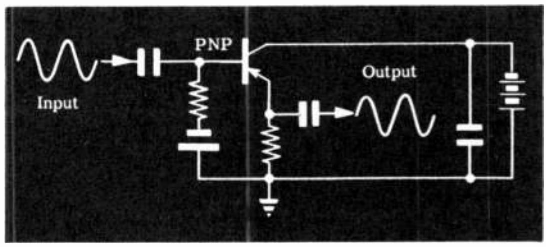


Review of Transistors (continued)

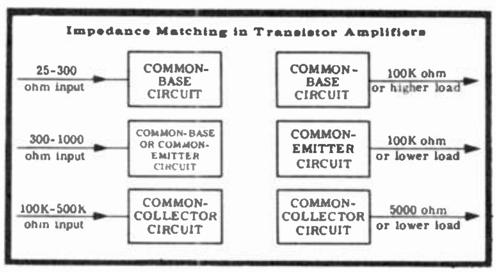
COMMON-BASE AMPLIFIER—This circuit operates in the same general manner as the grounded-grid vacuum-tube amplifier. It produces lower power amplification than the common-emitter arrangement and has the least tendency to oscillate. Input resistance may be as low as 25 ohms; output resistance is as high as 500,000 ohms.



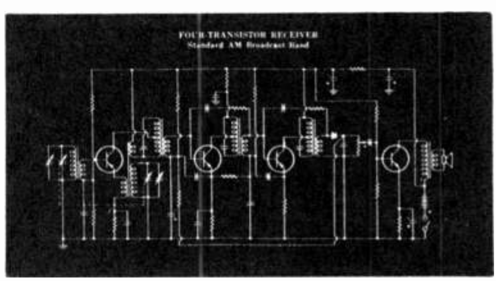
COMMON-COLLECTOR AMPLIFIER—This circuit operates in the same general manner as the cathode-follower vacuum-tube arrangement. The voltage gain is less than one; power gains of over ten times can be produced. Input resistance is in the order of 25,000 ohms; output resistance is as low as 1000 ohms.



MULTIPLE-STAGE AMPLIFIERS—Coupling transistor amplifier stages together involves important impedance-matching problems which are solved by transformer-coupling or by careful selection of the types of stages used in sequence. The same coupling methods are used as in vacuum-tube amplifiers.



TRANSISTOR APPLICATIONS—The basic transistor circuits can be modified to duplicate the functions of all the vacuum-tube circuits learned previously. Your knowledge of vacuum-tube circuits is of great assistance in identifying the function of the various transistor stages used in complex equipment.

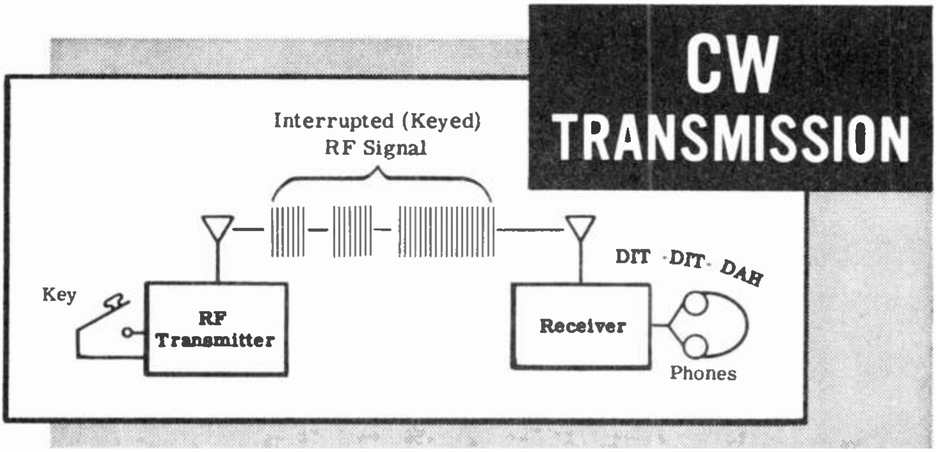



Fundamentals of FM
RADIO TAXI
POLICE

PART 2

FUNDAMENTALS OF FREQUENCY MODULATION

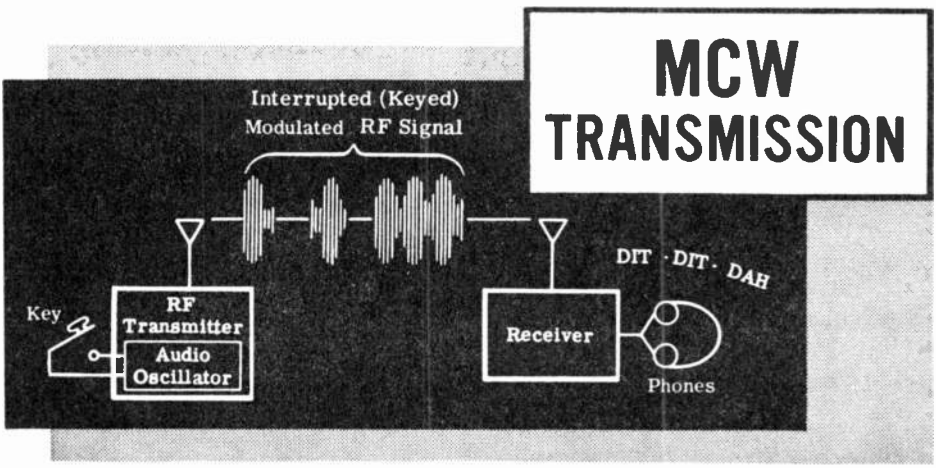
What Frequency Modulation Is



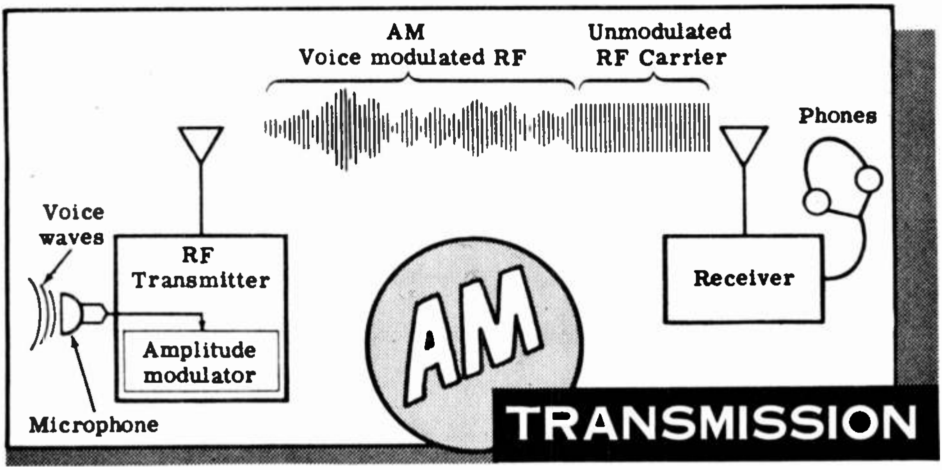
There are four basic methods by which a message can be superimposed upon the RF carrier signal of a radio transmitter. You will review these methods in the paragraphs which follow.

In continuous wave (CW) transmission, the carrier signal is interrupted, or turned on and off, with a hand key. The result is the familiar dot-dash, or "dit-dah", signal. This method is used primarily for long distance communication. The signals can be picked up by communications receivers or by home-type receivers to which a special oscillator stage has been added.

In modulated continuous wave (MCW) transmission an audio frequency signal of constant amplitude is superimposed upon the carrier. A key is used to turn the carrier on and off, as in CW transmission. MCW is used mainly for emergency transmission, and the signals can be picked up by home-type receivers.



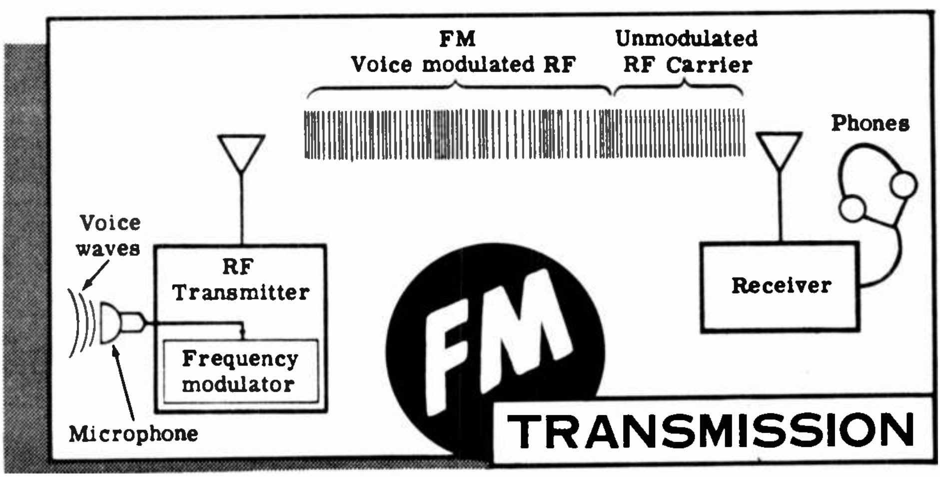
What Frequency Modulation Is (continued)



In amplitude modulated (AM) transmission the amplitude of the carrier is varied up and down to correspond with the audio signal generated by means of a microphone. This is the type of transmission picked up by a standard home-type receiver.

Frequency modulation (FM) is another method of transmission used to send voice and other sound signals. The frequency of the RF carrier signal is shifted or deviated to a higher and lower number of cycles per second, and the rate of deviation is equal to the frequency of the sound signal.

The techniques of CW, MCW and AM transmission and reception were reviewed in Volumes 4 and 5. The remainder of this volume is devoted to the methods of FM transmission and reception.

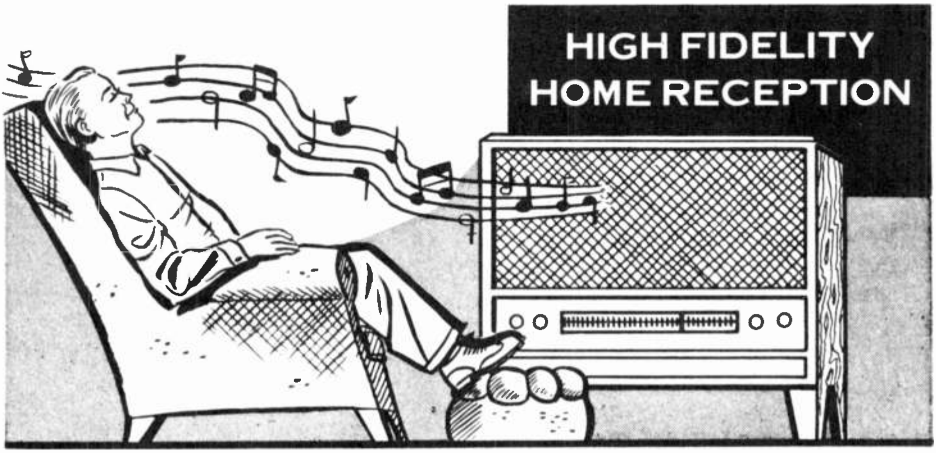


Applications of Frequency Modulation

Most applications of FM are based upon one or both of its two outstanding advantages. FM permits the transmission of an extremely wide range of audio frequencies. In addition, it provides almost complete freedom from the annoying sounds produced by natural and man-made interference that are frequently heard coming from AM receivers. These characteristics provide the basic requirements for important military, commercial and home applications.

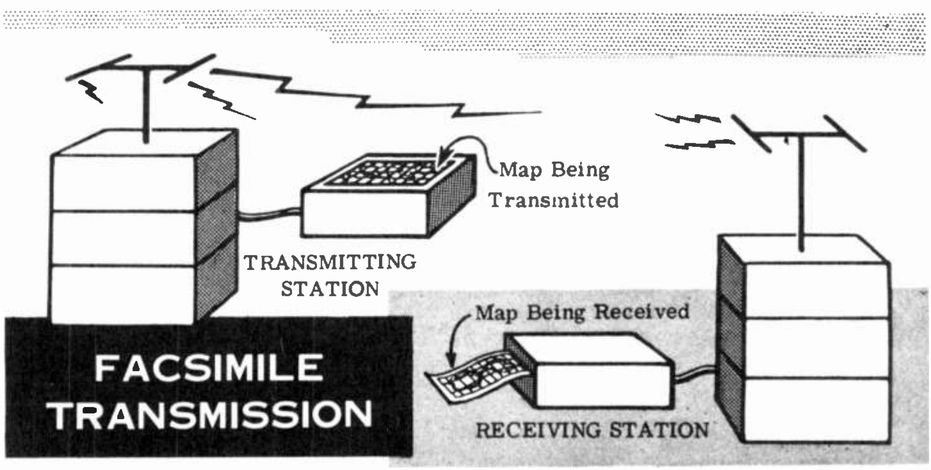
The exchange of messages under tactical military conditions requires clear message reception without danger of the content being obscured by noise. This is important in communication between moving vehicles, particularly tanks, where there are very high interference levels produced by a variety of electrical equipments.

Applications of Frequency Modulation (continued)

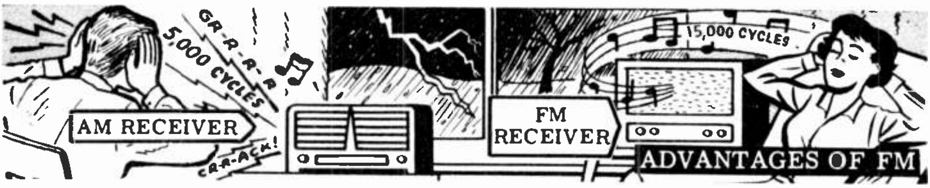


In commercial applications there are a number of instances in which instrument outputs and other electrical signals must be transmitted with perfect reception and freedom from interfering noise. One instance is in high speed, high quality facsimile transmission of detailed maps, photographs and printed information. In such applications FM provides faithful reproduction of the required wide band of video frequencies and prevents noise from blotting out the fine details.

Perhaps the best known application of FM is its use in "high-fidelity" radio receivers which provide excellent, noise-free reception of voice and music. The high quality sound reception that is obtained with television receivers is due to the fact that the sound portion of television programs is transmitted and received by FM techniques.



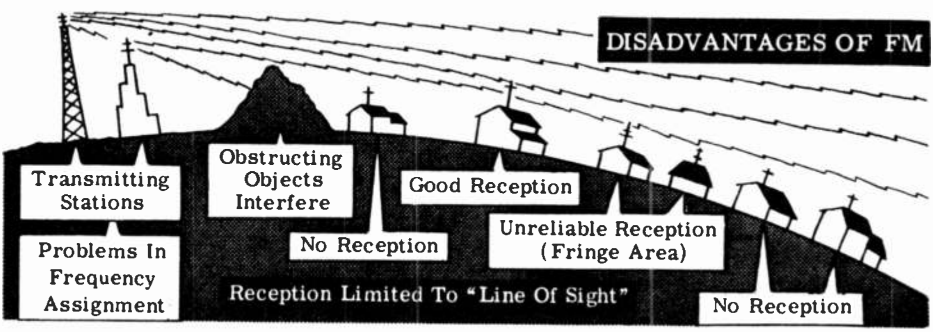
Advantages and Disadvantages of FM



It was mentioned earlier that the two outstanding advantages of FM are first that it permits the transmission and reproduction of an extremely wide range of audio frequencies and second that it permits reception that is free from man-made and natural interference. It is entirely a practical matter for an FM receiver to reproduce audio signals at 15,000 cycles or higher without distortion and without noise during a violent electrical storm. Under present broadcasting conditions an AM receiver cannot reproduce audio signals higher than 5,000 cycles; and during a violent electrical storm, the "static" completely drowns out the signal. The audio frequency limitation of AM reception is not inherent in AM transmission, but due to broadcasting regulations intended to restrict the transmitter bandwidth and thereby prevent interference between AM stations.

FM has its disadvantages. For FM to perform its function the transmitter carrier frequency must deviate over a wide band of frequencies. Although good quality FM broadcasting does not require that the transmitter deviate the maximum allowable limit of 75,000 cycles above and below the assigned center frequency, this limit is closely approached by those stations featuring high-fidelity FM broadcasting. Frequency ranges of such width are not available in the commercial radio broadcasting band, and commercial FM transmission has been assigned to the region between 88 and 108 megacycles. In this frequency band FM faces the same limitations that are well known to all television viewers: FM reception is essentially limited to locations that are in "line-of-sight" of the transmitting antenna, and "fringe area" reception quality varies widely at different times of the day.

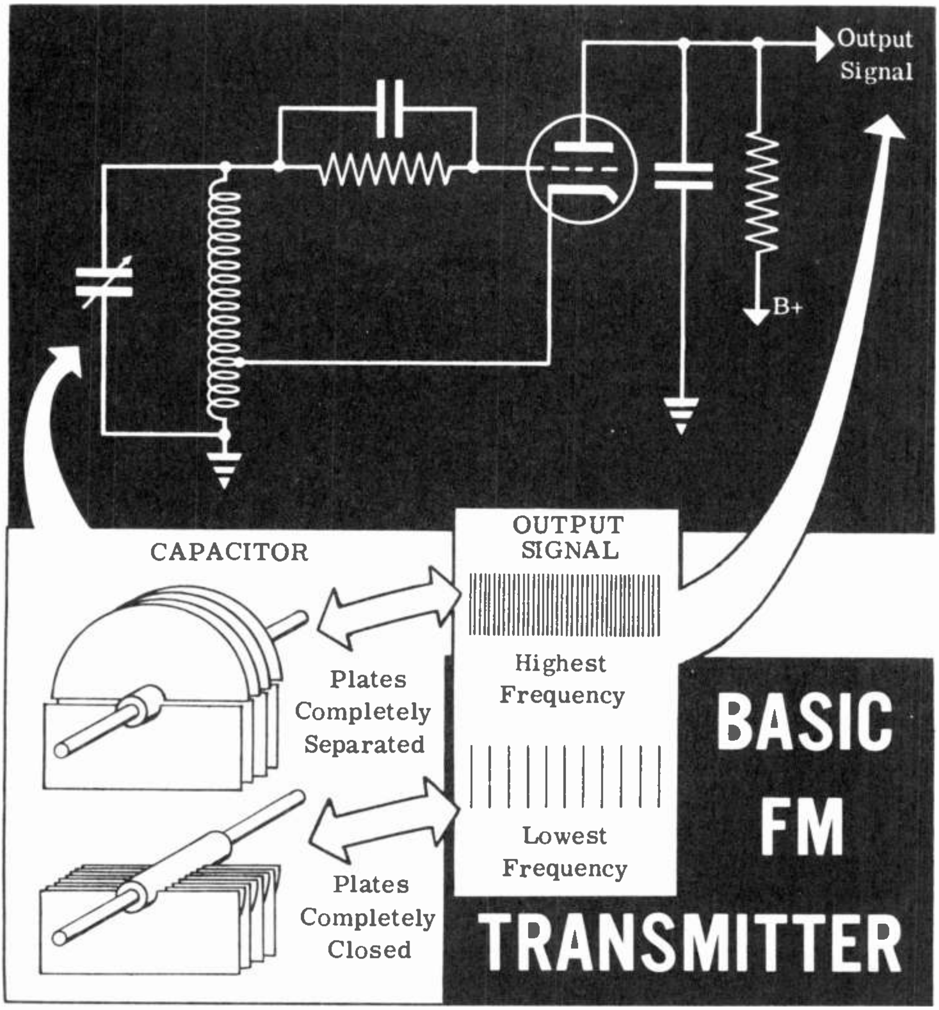
Another disadvantage of FM is that each transmitter requires a wide frequency band, and in the case of closely overlapping bands only the strongest transmitter will be heard. This requires careful assignment of transmitter operating frequencies over the entire country in order to prevent any possible overlap. In the case of fixed stations the problem need be solved only once. In the case of mobile transmitters, especially in military operations involving large numbers of mobile transmitters, very careful frequency assignment is necessary; and such assignments must be under constant review to prevent interference.



Simplest Form of an FM Transmitter

A rapid understanding of the nature of frequency modulation may be achieved through a review of the operation of the simplest form of an FM transmitter. Consider an RF oscillator, such as a Hartley oscillator, coupled to an antenna. The buffer amplifiers, frequency doublers and power amplifiers usually placed between the oscillator and antenna are omitted for purposes of simplicity.

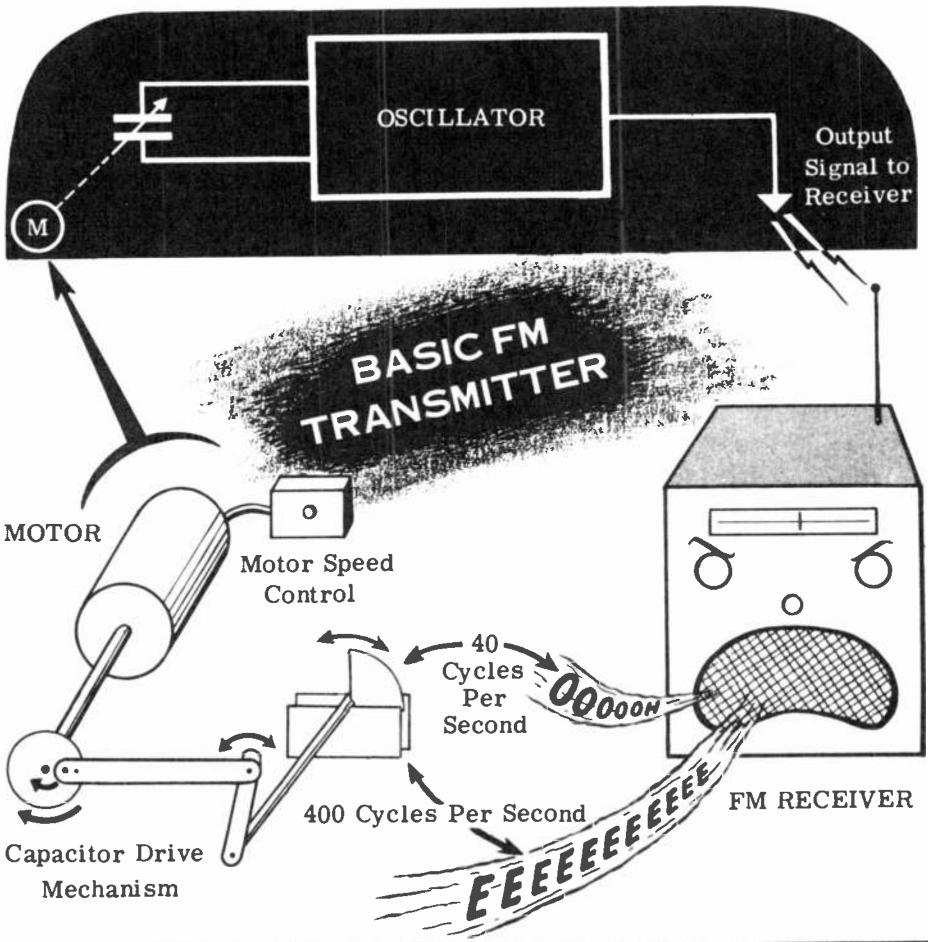
It should be remembered from your previous study of transmitters, that the signal sent out by this arrangement is an RF sine wave carrier signal of constant amplitude. The frequency of this transmitted signal is determined by the setting of the variable capacitor. The frequency increases as the capacitor is set towards its minimum capacitance, and the frequency decreases as the capacitor is set towards its maximum capacitance.



Simplest Form of an FM Transmitter (continued)

Assume, for example, that when the variable capacitor is set to the center of its range, the frequency generated is 1,000,000 cycles (1 megacycle). To obtain the effect of frequency modulation the hand, or a suitable vibrator mechanism, may be used to turn the variable capacitor shaft rapidly back and forth about its center position.

If the shaft is mechanically oscillated about its center position at a rate of 40 cycles per second, an FM receiver tuned to the center frequency will put out a 40-cycle tone from its loudspeaker. If the shaft is turned at a rate of 400 or 4,000 cycles per second, the receiver loudspeaker will put out tones of 400 or 4,000 cycles respectively. The tone from the FM receiver loudspeaker is always a duplicate of the rate of carrier frequency shift.



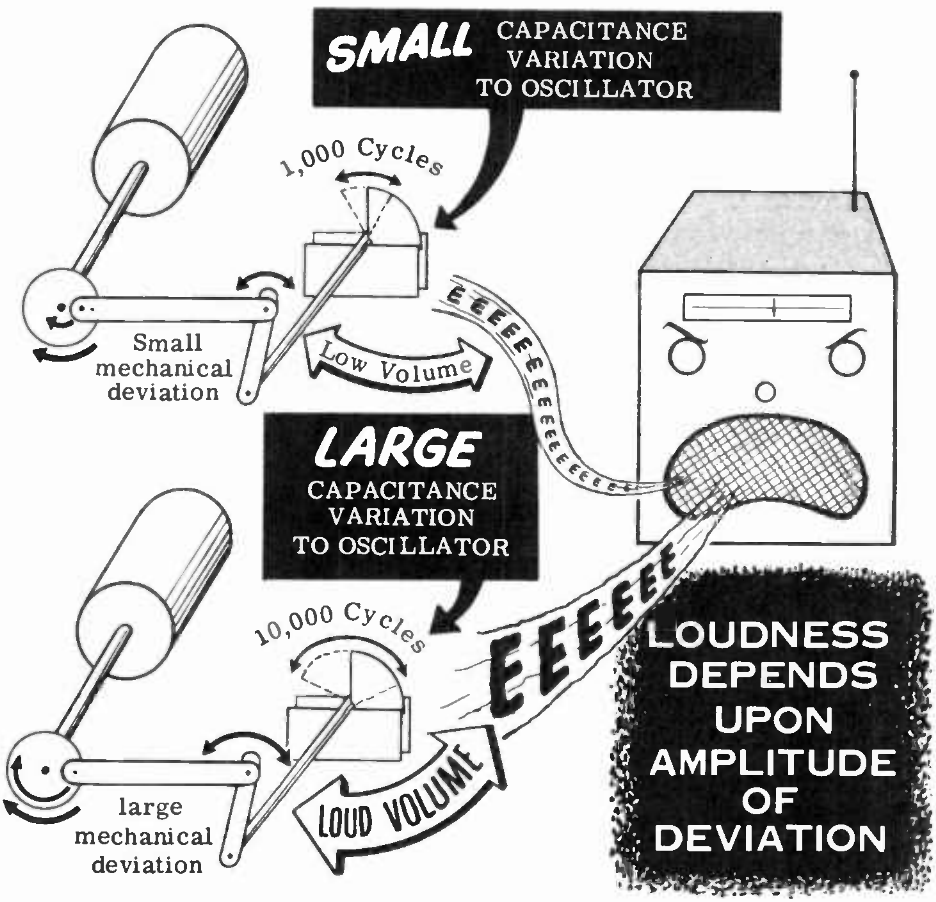
TONE DEPENDS UPON RATE OF DEVIATION

Simplest Form of an FM Transmitter (continued)

As shown on the previous page, the tone coming from the FM receiver loudspeaker depends upon the rate of the carrier frequency deviation. This is true regardless of the magnitude of the deviation. The transmitter capacitor shaft may be turned by only a small amount to each side of its center position — varying the frequency 1,000 cycles above and below the 1,000,000-cycle center frequency. On the other hand, the capacitor shaft may be turned at the same rate but by a larger amount — varying the frequency 10,000 cycles above and below the 1,000,000-cycle center frequency.

In both cases the same tone comes out of the FM receiver loudspeaker. However, the loudness of the loudspeaker tone increases with the amplitude of the frequency deviation. The tone produced by the ± 1,000-cycle shift will be quite low in volume, while the same tone produced by the ±10,000-cycle shift will be much louder.

Thus the frequency and loudness of the sound coming from the FM loudspeaker are determined respectively by the rate and amplitude of the transmitter frequency shift.

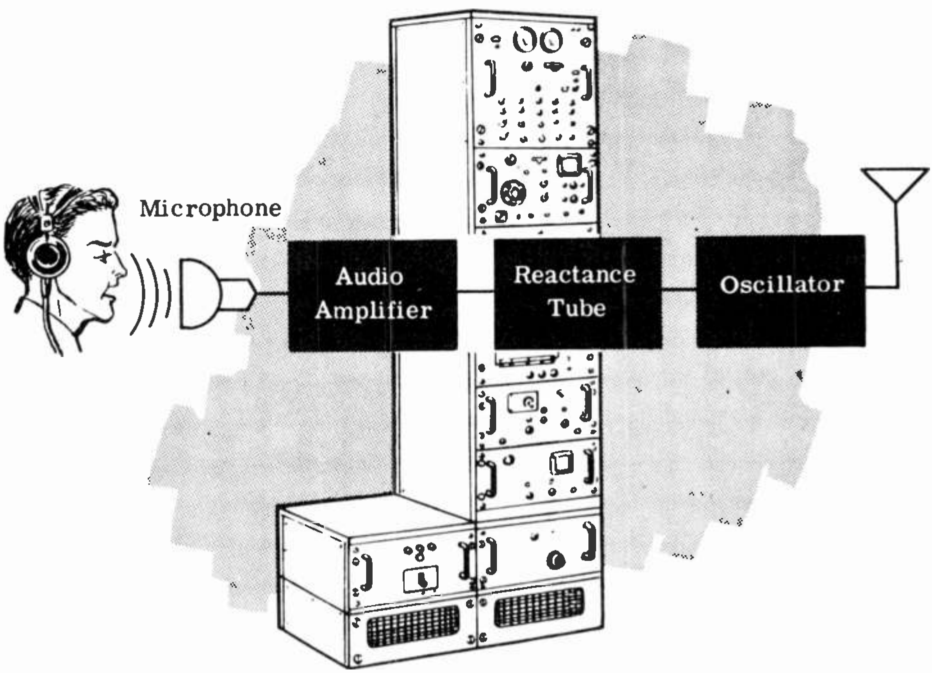


Simple Reactance Tube FM Transmitter

The FM transmitter just described can be used to transmit tones of any desired frequency and loudness. Since mechanical methods are used to produce frequency shifting, the arrangement is known as a "mechanical modulator". While such a system can be used for test purposes, what is required is an arrangement that can be used to transmit voice and other sound signals. To accomplish this a microphone used in conjunction with an audio amplifier must be able to cause deviations in the transmitter carrier frequency. The rate of the frequency shift must be equal to the frequency of the sound going into the microphone, and the amplitude of the frequency shift must be in proportion to the loudness of the sound.

One fundamental and very widely used method of accomplishing these desired results is based upon the use of a reactance tube. The basic arrangement of a reactance tube FM transmitter is shown in the block diagram. Note that the frequency multipliers and power amplifier normally following the oscillator are omitted, for the present, in order to achieve simplicity. In the circuit arrangement shown a microphone drives an audio amplifier which, by means of a reactance tube, shifts the frequency of the oscillator.

The operation of this system is very much the same as that of the mechanical modulator. In one case the oscillator frequency is shifted at a rate and amplitude determined by a mechanical arrangement. In the other case the oscillator frequency is shifted at a rate and amplitude determined by a reactance tube under the control of an audio signal.

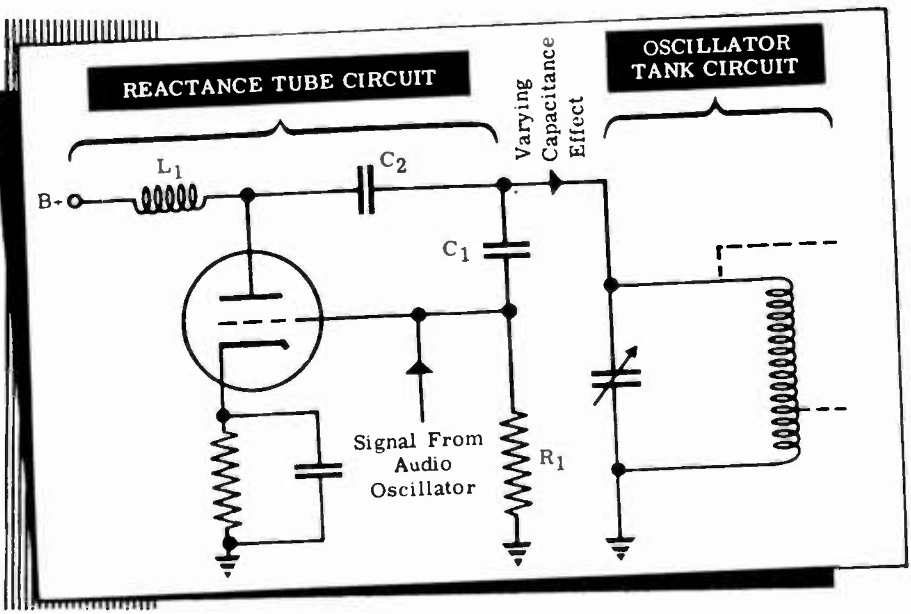


Basic Reactance Tube Circuit

The diagram on this page shows a basic reactance tube circuit. This circuit injects the effect of a varying capacitance into the oscillator circuit. The rate and amplitude of the capacitance change are determined by the frequency and amplitude of the amplified audio signal from the microphone.

The circuit operates by changing the capacitive reactance in the oscillator circuit and thereby achieves the effect of varying the capacitance of the tuning capacitor. From your study of capacitive reactance in Basic Electricity, you know that the outstanding characteristic of this quantity is that the voltage and current are 90 degrees out of phase with each other. The circuit arrangement of the reactance tube causes a condition wherein the AC plate voltage and plate current are 90 degrees out of phase and the magnitude of this effect is determined by the audio signal voltage on the reactance tube grid.

The AC plate voltage is supplied by the oscillator tank circuit. The DC plate voltage is obtained from the B+ line through RF choke L_1, and the B+ voltage is isolated from the reactance circuit by capacitor C_2. In order to make the AC plate current 90 degrees out of phase with the AC plate voltage, it is necessary to make the AC grid voltage have the same phase relationship. This is accomplished by connecting a capacitor (C_1) and resistor (R_1) in series from the plate to ground and by connecting the grid to the junction between the two. In such an arrangement, when low values of resistance and capacitance are used, the AC voltage across the resistor is 90 degrees out of phase with the AC voltage across the network. Since plate current is in phase with grid voltage, the AC plate current leads the AC plate voltage by 90 degrees. The effective result is that capacitance has been connected across the oscillator tank circuit.



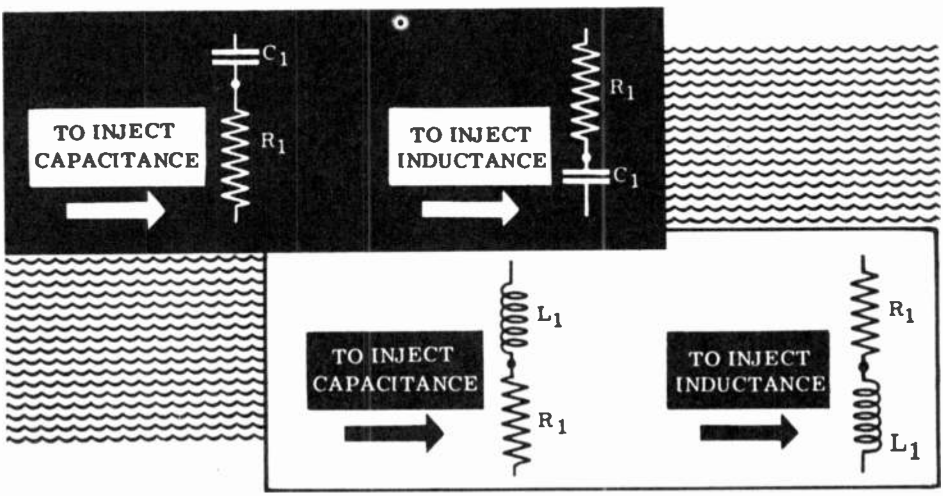
Basic Reactance Tube Circuit (continued)

Connecting capacitance across the oscillator tank circuit lowers the oscillator frequency in proportion to the amount of capacitance that is introduced. The quantity of capacitance that is injected into the oscillator tank circuit is equal to the transconductance (gm) of the reactance tube multiplied by the product of the special network resistance and capacitance (C injected = $gm \times R_1 \times C_1$).

It will be remembered from your introduction to vacuum tubes that transconductance is equal to the change in plate current divided by the change in the grid voltage which produced the plate current change. With the plate voltage remaining constant, changes in grid voltage produce changes in plate current. Thus the transconductance of the tube varies by significant amounts in accordance with the frequency and amplitude of the audio signal. As the transconductance changes, the injected capacitance varies according to the above equation; and the frequency of the oscillator is deviated as described previously.

Reactance tube circuits generally employ pentode tubes such as the 6AC7 or 6SJ7. The value of C is generally between 20mmf and 30mmf, and the value of R is usually between 500 ohms and 1000 ohms. With such an arrangement an audio signal of from 1 to 2 volts peak-to-peak on the grid will cause the capacitance injected into the oscillator tank circuit to shift from approximately 100mmf to 300mmf. Thus the frequency of the oscillator will shift in accordance with the frequency and amplitude of the audio signal.

The reactance tube arrangement that has been considered is only one of four basic types that can be used. If R_1 and C_1 are interchanged in the circuit, a variable inductance will be injected into the oscillator tank circuit; and this causes the desired frequency changes as effectively as the variable capacitance. If a coil, L_1, is substituted for C_1 in the original arrangement, the circuit will inject a variable inductance into the oscillator tank; and interchanging L_1 and R_1 will inject variable capacitance. Any of these arrangements can be used in an FM transmitter.

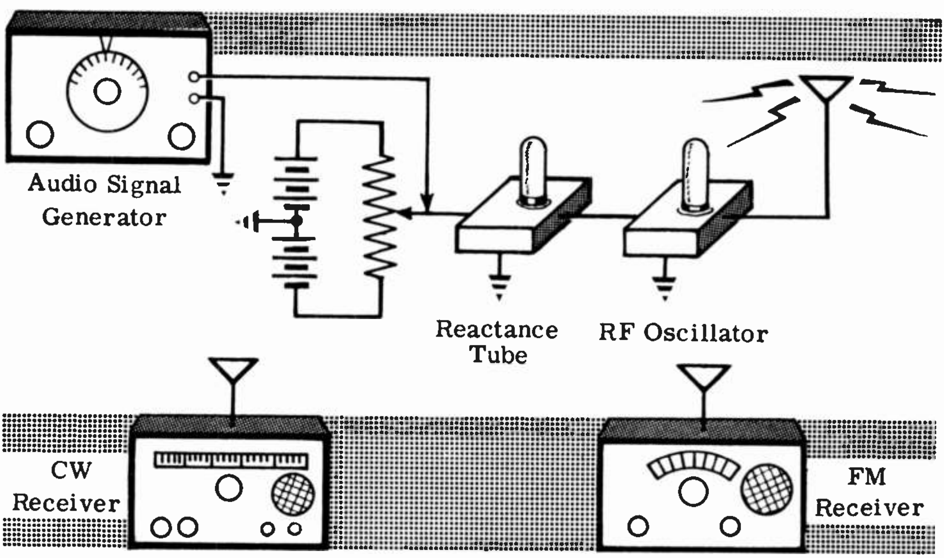


Demonstration—Basic Reactance Tube Circuit

The purpose of this demonstration is to indicate the outstanding characteristics of the basic reactance tube circuit. In accomplishing this the outstanding features of FM also will be indicated. The equipment required is a reactance tube connected to an oscillator, an audio signal generator, a pair of 6-volt batteries with a potentiometer connected across them, plus an FM receiver and a CW receiver which can be tuned to the frequency of the oscillator.

The first phase of the demonstration is begun by connecting the batteries and potentiometer to the reactance tube input. By adjusting the potentiometer the reactance tube input voltage can be varied over the range of from +6 to -6 volts. With the input voltage set to zero, the CW receiver can be tuned to the oscillator frequency. When the signal is heard, the receiver dial indicates the oscillator center frequency. When the input voltage is varied the receiver is retuned until the oscillator signal is heard. A record is kept of the CW receiver dial settings corresponding to the individual reactance tube input voltages. An examination of this record clearly indicates that the amount and direction of oscillator frequency deviation is closely proportional to the amplitude and polarity of the reactance tube input voltage.

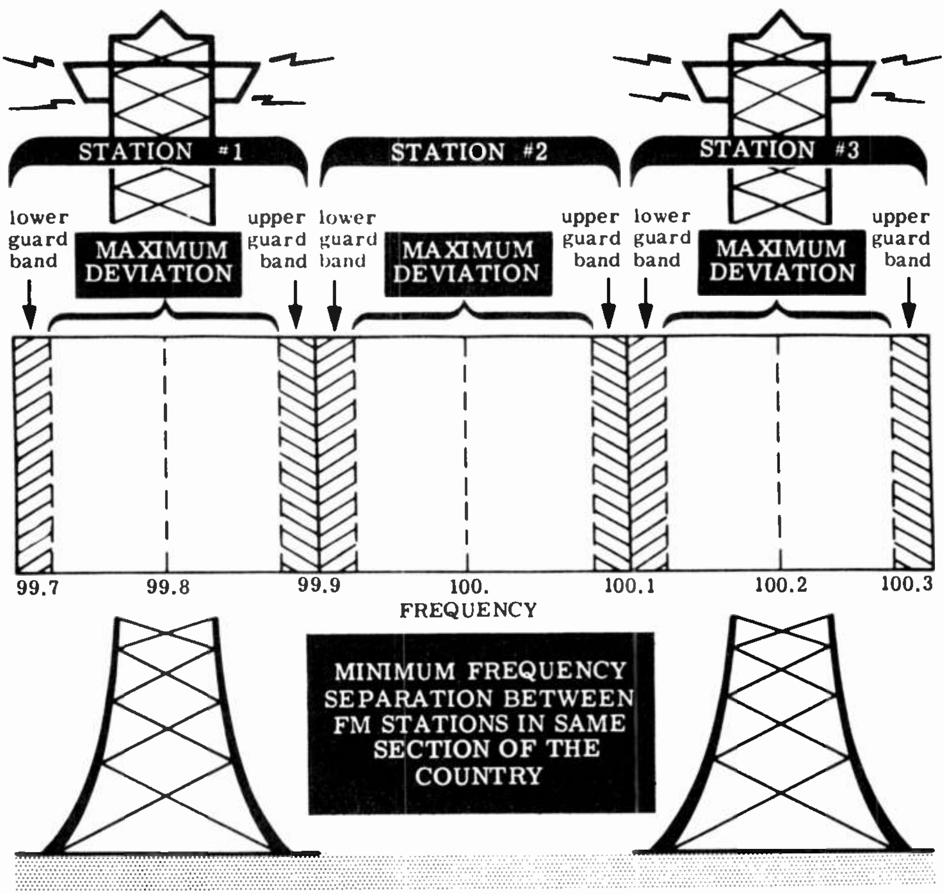
In the next phase of the demonstration the audio signal generator is connected to the reactance tube input. With the audio input voltage equal to about 0.5 volts at 1000 cycles, the FM receiver is tuned for the loudest signal. After this initial adjustment the amplitude and frequency of the audio signal input is varied. It is clearly observed that the signal heard from the receiver corresponds to the frequency of the audio signal generator and that the loudness of this signal corresponds to the amplitude of the audio signal.



Frequency Characteristics of FM Transmitters

Now that there has been a review of the basic features of a simple FM transmitter, it is time to consider the operating characteristics of commercial FM transmitters. After this the circuit arrangement of FM transmitters will be analyzed in greater detail.

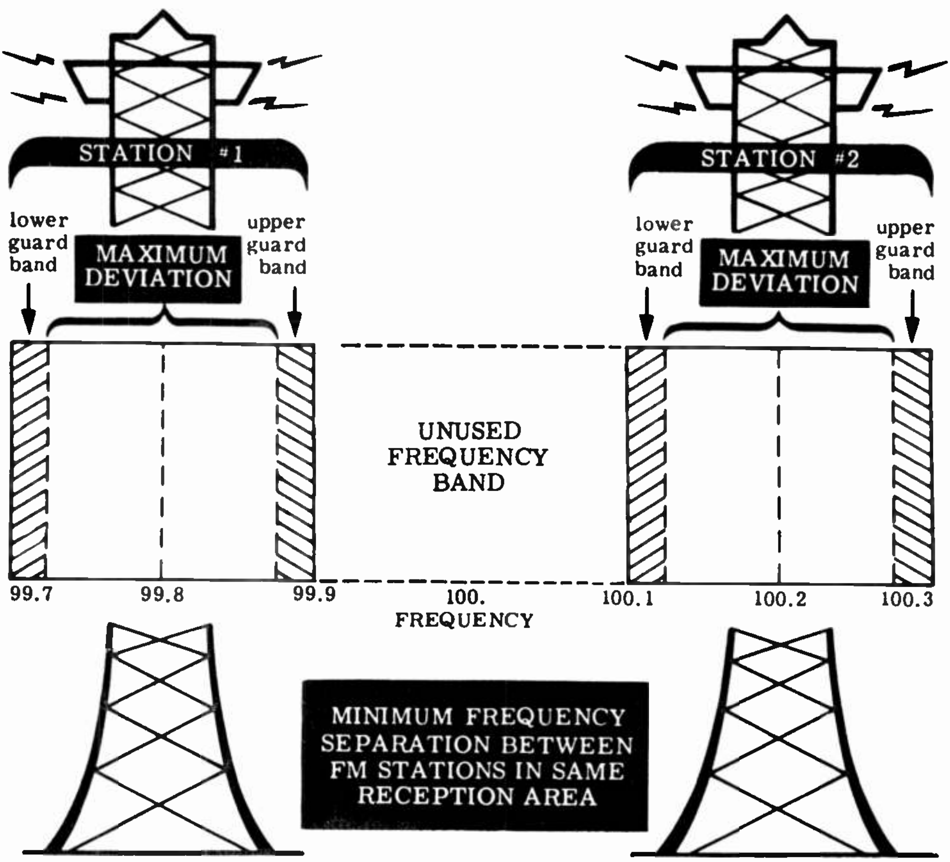
Commercial broadcast stations which transmit program material by means of frequency modulation operate within the frequency band of from 88 to 108 megacycles. In order to prevent interference between stations in the same section of the country, the Government has specified that the carrier frequency separation between stations in the same region can be no less than 200,000 cycles. This means that if a station is assigned to a center frequency of 100 megacycles, other regional stations cannot be assigned center frequencies closer than 99.8 megacycles on the low side or 100.2 megacycles on the high side.



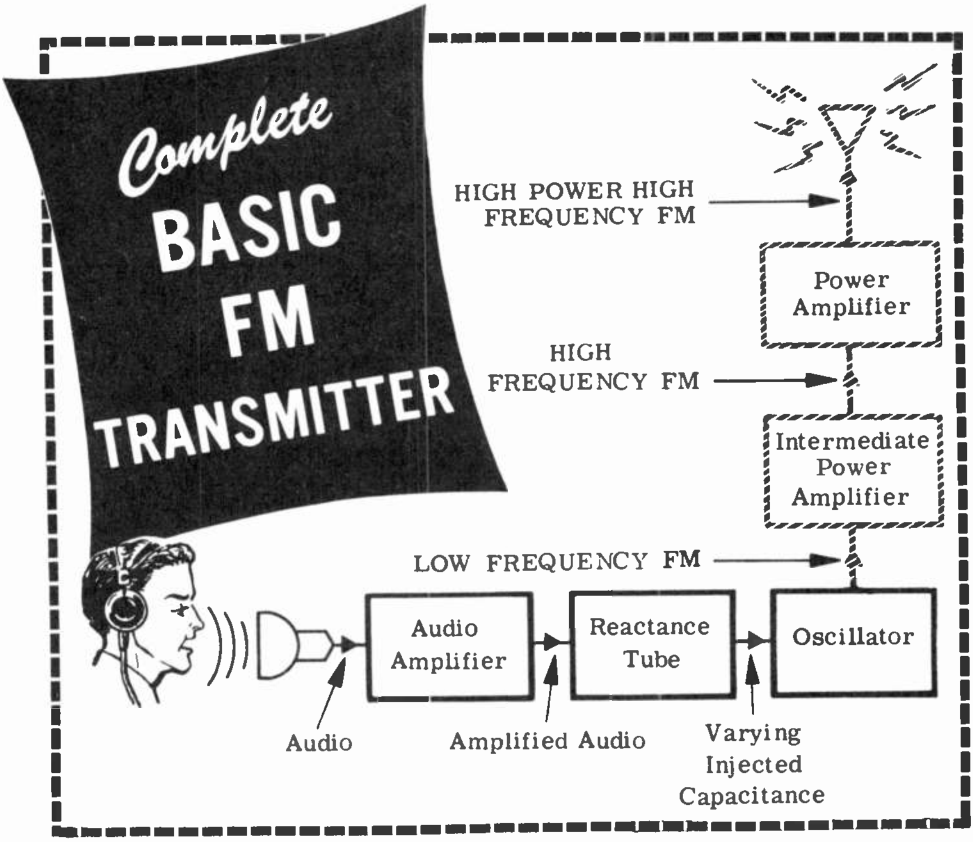
Frequency Characteristics of FM Transmitters (continued)

It is important to prevent any possible overlap between the frequency shifts of stations with adjacent center frequencies. If this should happen, a receiver tuned to one station might also pick up interfering signals from the adjacent stations on each side. To guard against such a possibility the maximum frequency shift that is permitted is 75,000 cycles on each side of the center frequency. No signal must be transmitted in the frequency band between 75,000 and 100,000 cycles on each side of the center frequency. Thus each station has a 25,000 cycle "guard" band on each side of its center frequency to prevent interference between stations, even though the center frequency of one or both of them may drift slightly.

Under actual conditions the 200,000 cycle separation between center frequencies is reserved for stations completely out of range with each other. Since there may be eight or more stations operating in densely populated areas, such stations have their center frequencies separated by 400,000 cycles. This places an unused frequency band of 100,000 cycles between adjacent guard bands. Including the guard bands there is an unused frequency band of 150,000 cycles between the actively used channels. Thus there is good assurance against interference, even though adjacent stations may stray considerably from their assigned operating ranges.



Complete Basic FM Transmitter



Once the operation of the reactance tube is understood, there is little difference between a basic FM transmitter and the AM transmitters that were reviewed in Volume 4. The block diagram on this page shows a basic FM transmitter; and if the reactance tube is eliminated, it is the same arrangement used in an AM transmitter with modulation introduced into the oscillator.

To review the operation of the system, consider the function of each stage, beginning with the microphone. The microphone converts sound waves into a low voltage electrical signal having the same frequency and amplitude characteristics as the sound wave. This low voltage audio signal is fed to the audio amplifier, which steps up the signal voltage to a level suitable for driving the reactance tube. Without the reactance tube the oscillator generates an RF signal of constant frequency and amplitude. Due to the operation of the reactance tube the oscillator frequency is shifted at a rate equal to the frequency of the audio signal; and the magnitude of the frequency shift is in proportion to the amplitude of the audio signal.

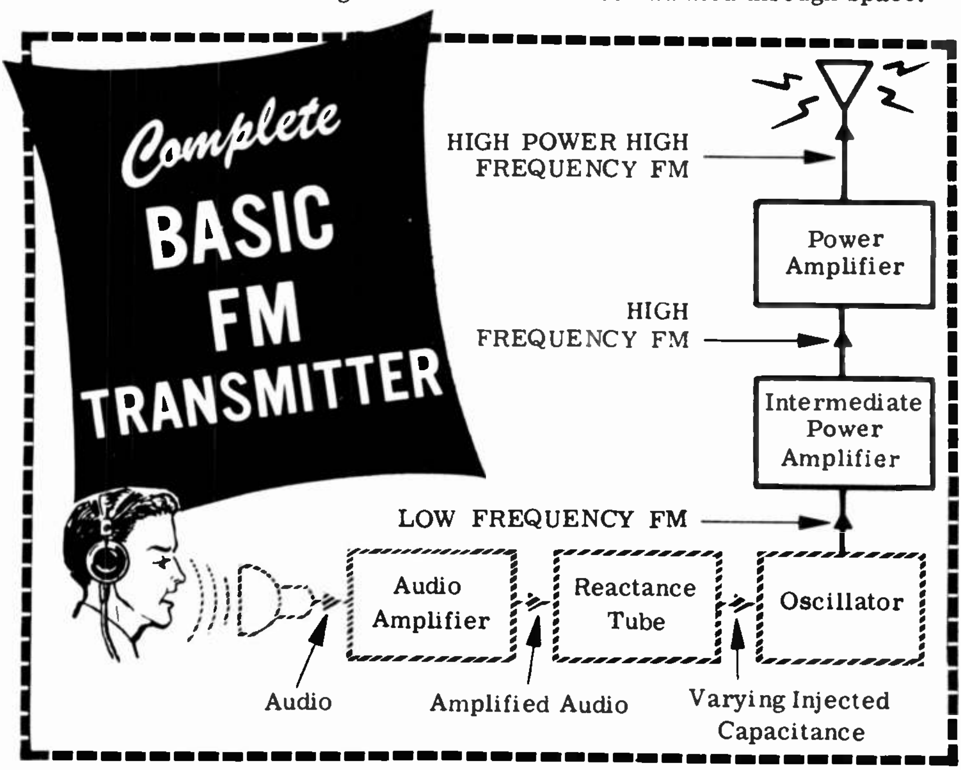
Complete Basic FM Transmitter (continued)

The purpose of the intermediate power amplifier (IPA) is to isolate the oscillator for improved frequency stability and to amplify the RF signal in order to drive the power amplifier efficiently. The IPA also frequently serves as a frequency multiplier. This permits the transmitter to send out the desired frequency while the oscillator operates at a much lower frequency, where it can maintain stable operation more easily.

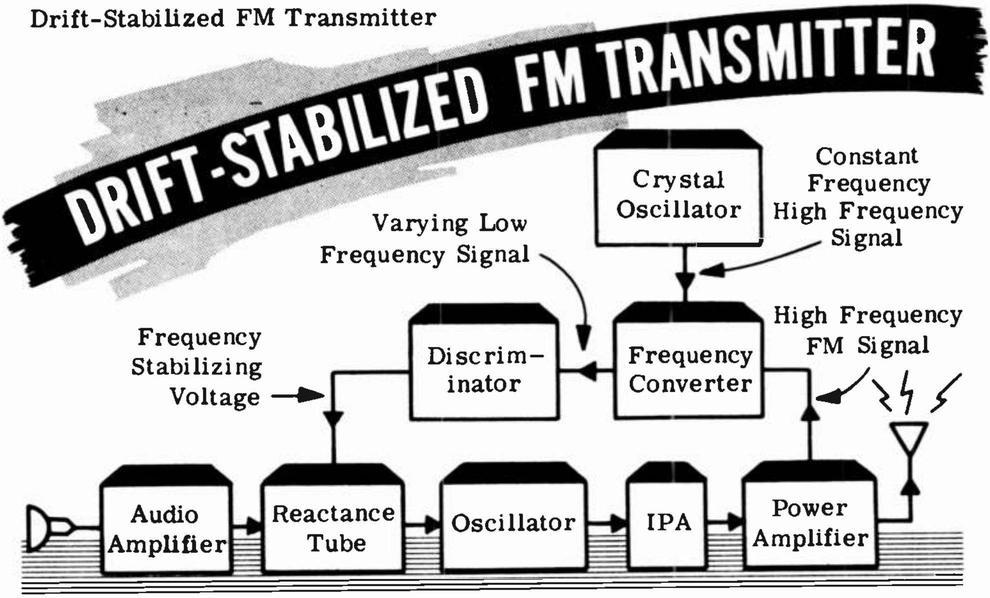
It should be recognized that in addition to multiplying the oscillator signal frequency, the IPA also multiplies the oscillator signal frequency deviation. Thus, if the oscillator resting frequency is 30 megacycles with a frequency deviation of $\pm$ 25,000 cycles, and the IPA is operated as a frequency tripler, the IPA output resting frequency will be 90 megacycles with a frequency deviation of $\pm$75,000 cycles. Consequently, the use of frequency multipliers permits the required frequency deviation to be obtained while the oscillator operates at a lower, more stable frequency and with a smaller, more easily obtainable frequency deviation.

The purpose of the power amplifier is to increase the power level of the IPA output signal. This achieves adequate power radiation by the antenna to assure good reception within the area being serviced. The power amplifier usually has its input and output signals at the same frequency in order to obtain maximum operating efficiency.

The antenna converts the signal delivered by the power amplifier to the form of an electromagnetic wave that can be radiated through space.



Drift-Stabilized FM Transmitter



The FM transmitter just reviewed is subject to oscillator frequency drift. The most important reasons for this drift are the dimensional changes that take place in the oscillator coil and capacitor due to temperature variations within the transmitter itself. In AM transmitters the most reliable method of preventing this drift is to use a crystal oscillator. However, this method cannot be employed in an FM transmitter, since the frequency of a crystal oscillator cannot be shifted.

To obtain the required frequency deviation while maintaining the necessary frequency stability it is necessary to make use of a crystal oscillator in an indirect manner. The method is shown in the block diagram. The FM transmitter previously described is used as the major part of the system. To this arrangement are added a crystal oscillator, a frequency converter (mixer), and a discriminator.

The frequency converter receives signals from both the power amplifier and the crystal oscillator. The frequency of the mixer output is equal to the difference between the two input signals. Thus, the mixer output frequency increases and decreases as the power amplifier frequency deviates higher or lower. The discriminator output is a "frequency stabilizing voltage" which rises and falls in accordance with the drift of the power amplifier signal. The operation of the discriminator will be analyzed in detail in the section on FM receivers. The frequency stabilizing voltage is injected into the reactance tube, and causes the reactance tube to bring the oscillator back to the predetermined center frequency. Consequently, the overall arrangement detects any drift in the oscillator frequency and corrects that drift.

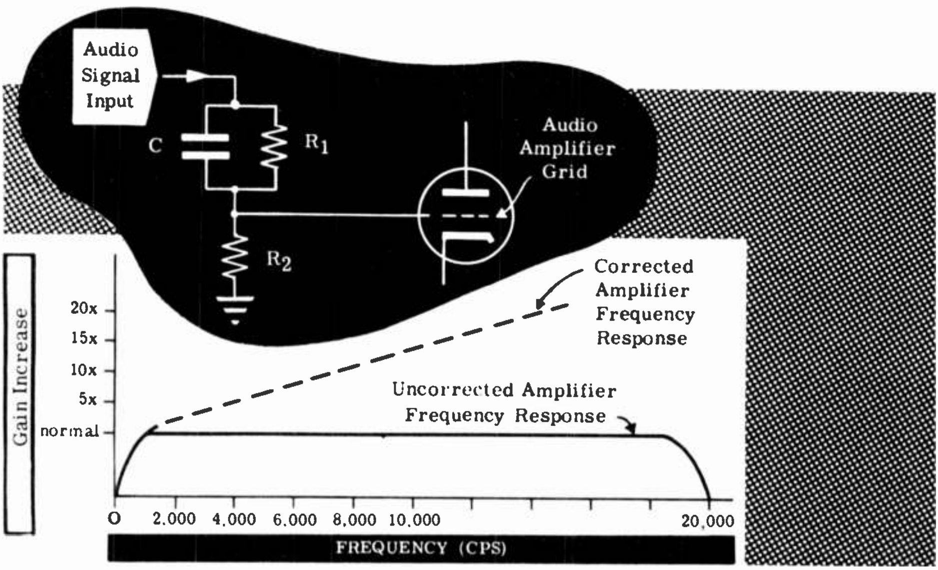
Pre-Emphasis

It has been mentioned that an important advantage of FM is that it can be used to transmit audio signals at frequencies up to 15,000 cycles or higher. The frequencies above 5000 cycles contain mainly the upper harmonics of the fundamental frequencies in voice or music. These upper harmonics are low in signal amplitude; but if they can be reproduced at the receiver, they add the unusually fine quality that is characteristic of FM reception.

Since the upper harmonics are low in amplitude, their signal-to-noise ratio is much lower than that of the signals below 5000 cycles. In addition, much of the noise is at the higher frequencies, which would tend to make the high frequency signals unbearably noisy.

Pre-emphasis is a method of improving the signal-to-noise ratio of the high frequency signals, so that they can be received without noise. The method of increasing the signal-to-noise ratio is simply to increase the high frequency gain of the transmitter audio amplifier as shown by the curve. Since the audio signal amplitude decreases rapidly with frequency, the amplifier gain must be increased as the frequency rises.

The simplest method of obtaining the desired gain characteristic in the audio amplifier is to use a high-pass filter in one of the amplifier stages. The most basic circuit which can accomplish this is the RC network shown in the diagram. Circuit operation is based upon the fact that the voltage divider has different characteristics at various frequencies. At low frequencies C has such a high impedance that the voltage division is determined only by R_1 and R_2. As the frequency increases the impedance of C decreases rapidly. Eventually the impedance of C is much lower than R_1, and the voltage division is determined only by C and R_2. The result of the filter voltage dividing action is that the high frequency signals are attenuated much less than the low frequency signals, and the desired frequency response characteristics have been introduced into the amplifier.

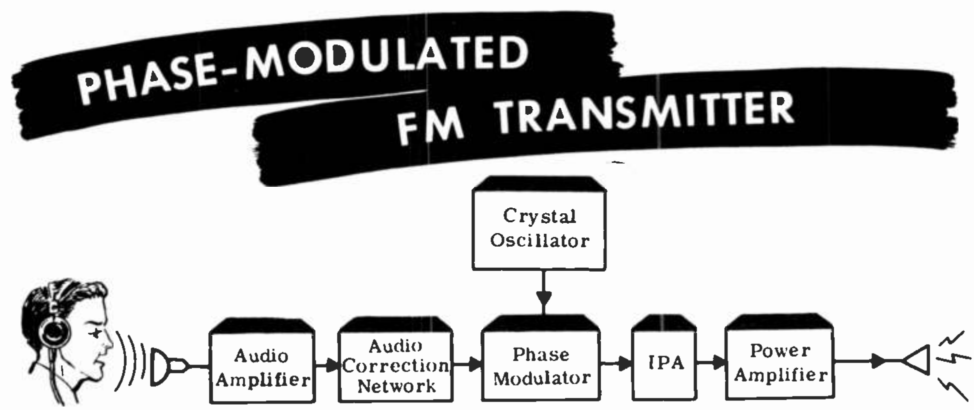


Phase-Modulated FM Transmitter

All the FM transmitter arrangements that have been considered up to this point are known as the "direct" type. In these transmitters the RF carrier signal is frequency modulated in the oscillator — the original source of the carrier signal. The oscillator frequency is deviated in accordance with the frequency and amplitude of the audio signal; and equivalent, but multiplied, frequency swinging takes place in the stages which follow.

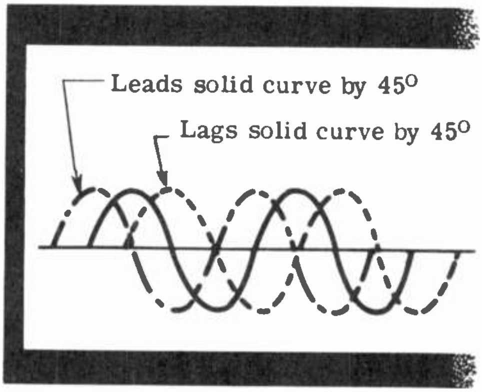
Another basic type of FM transmitter is in widespread use. It is known as the "indirect" type because the frequency deviation is not introduced at the source of the RF carrier signal (the oscillator). Instead, the oscillator frequency remains constant, and the frequency swinging is introduced in one of the stages following the oscillator. The basic advantage of this method is that the oscillator can be the crystal-controlled type, which will maintain a stable center frequency without the need for separate stabilizing circuits. Frequency swinging is not introduced in the same manner as in "direct" FM transmitters; a method known as "phase modulation" is used instead. It should be clearly understood that the final output signal of an "indirect" FM transmitter is the same as that produced by a "direct" FM transmitter. The differences involved are confined to the methods used to obtain the final output signal, and not in the characteristics of the output signal itself.

The phase-modulated transmitter consists of a crystal oscillator, frequency multipliers (IPA), power amplifiers (PA) and antenna; plus a microphone, audio amplifier, audio correction network and phase modulator. The RF carrier signal from the crystal oscillator is stepped up in frequency by the frequency multipliers, and the power amplifiers increase the signal power to a level suitable for adequate reception by all receivers within the area being served. The audio signal from the microphone goes into the audio amplifier, and the signal is amplified to a level suitable to operate the audio correction network and phase modulator. All but these last two circuits are understood on the basis of information reviewed earlier in this and previous volumes.



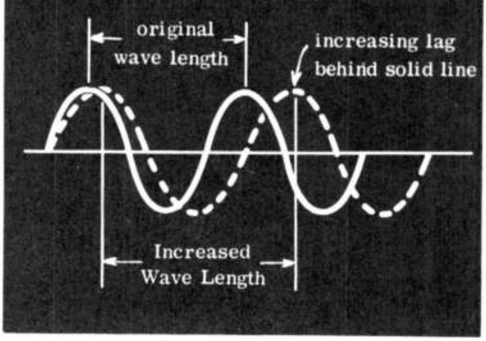
Phase-Modulated FM Transmitter (continued)

Advanced textbooks on the subject of frequency modulation go to great lengths to prove that phase modulation is the equivalent of frequency modulation. Mathematical analysis indicates that when the frequency of an oscillator is deviated, phase modulation also takes place; and, in addition, when the phase of the oscillator signal is shifted, frequency modulation takes place. It is beyond the scope of this book to repeat such a proof, and only a very simple graphic presentation of the similarity will be included here.



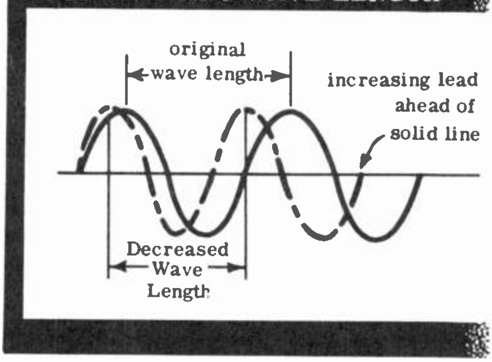
The first diagram shows a sine wave drawn in a solid line. The dotted curve shows a sine wave which lags 45 degrees behind the solid curve, and the dash-dot curve shows a sine wave which leads the solid curve by 45 degrees. These curves show that the peaks of a sine wave can be advanced or retarded in time by means of phase shifting.

PHASE LAG SIMULATES INCREASING WAVE LENGTH

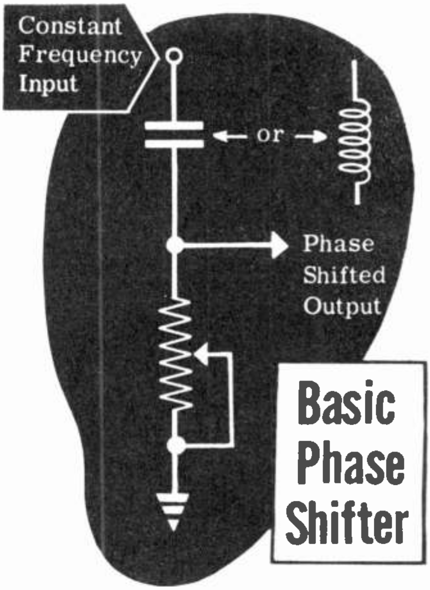


The second diagram shows the result of using a variable phase shifting network to smoothly shift the phase of the sine wave while it is being generated. When the phase is smoothly shifted so that it lags the original wave, the peaks occur later; this is equivalent to increasing the wave length or lowering the frequency. If the phase is smoothly shifted so that it leads the original wave, (see third diagram) the peaks occur sooner; this is equivalent to shortening the wavelength or increasing the frequency. Thus smoothly shifting the phase of a signal is equivalent to smoothly shifting its frequency.

PHASE LEAD SIMULATES DECREASING WAVE LENGTH

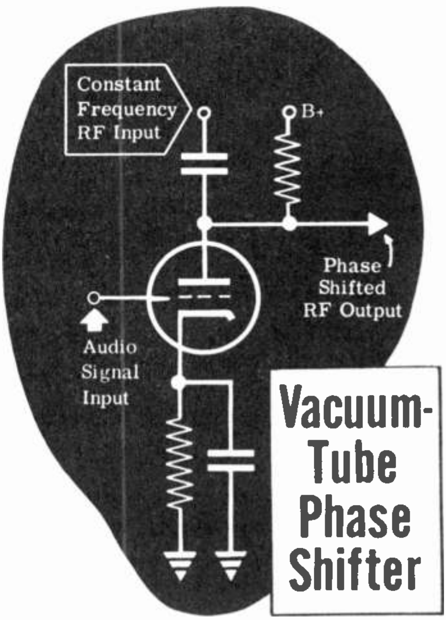


Basic Phase Modulator



A simple form of phase shifting network was analyzed in detail in Volume 4 of Basic Electricity. This circuit consists of a capacitive or inductive reactance connected in series with a resistor. When a signal of constant frequency is connected across the series combination, the signal output across the resistor is shifted in phase with respect to the input signal. If the resistance is varied, the phase shift also varies. Under conditions where the resistance is more than 10 times larger than the reactance, the phase shift is essentially equal to 0 degrees. As the resistance approaches one tenth the value of the reactance, the phase shift approaches 90 degrees. The constant-frequency RF signal from a crystal oscillator can be applied to the network, and its phase can be shifted by varying the resistance.

By replacing the resistor with a vacuum tube, the resistance in the phase shift network can be varied in accordance with the amplitude changes in an audio signal of constant frequency. As the audio signal voltage increases and decreases in amplitude, the tube plate current rises and falls. Rises in the plate current are equivalent to lowering the plate resistance of the tube, and this causes a maximum shift in the phase of the input RF signal. Decreases in the plate current are equivalent to raising the plate resistance of the tube, and this causes a minimum shift in the phase of the input RF signal.



Thus an RF signal of constant frequency, such as that coming from a crystal modulator, can be phase shifted in accordance with the amplitude of an audio signal of constant frequency. There are many variations of this arrangement, but they all involve changing either the reactance or resistance in a circuit of this type. In actual practice a single circuit of this type cannot produce a sufficient amount of phase shift, and four or more similar arrangements are connected in cascade to produce the required equivalent frequency deviation.

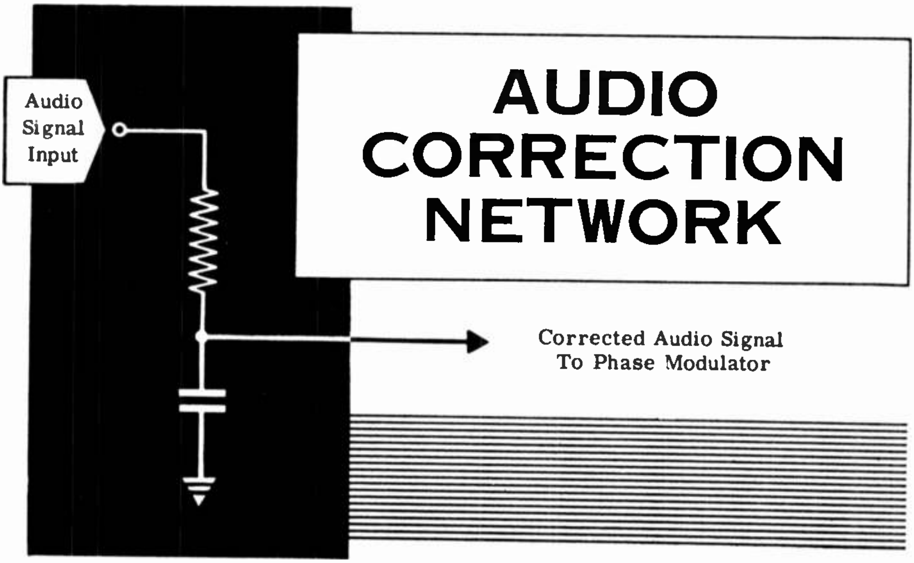
Audio Correction Network

It was shown that with the phase modulator the amount of carrier signal phase shift is in proportion to the amplitude of the audio signal. This is equivalent to saying that the extent of the equivalent frequency deviation is in proportion to the amplitude of the audio signal.

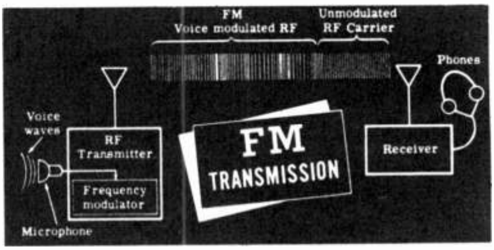
Because of the sensitivity of the circuit to phase changes, a change in the frequency of the audio signal causes additional phase shift in the RF carrier signal. In the circuit shown the rate of phase change, and the equivalent frequency deviation, rises as the audio frequency rises. Assume for instance that a 100-cycle audio signal of 1 volt amplitude causes an equivalent 1000 cycle frequency deviation in the RF carrier. If the audio frequency rises to 1,000 cycles and the amplitude remains at 1 volt, there will be an equivalent frequency deviation of 10,000 cycles in the RF carrier.

Such an effect is completely contrary to the working principles of FM transmission. According to the requirements of FM, the RF carrier frequency deviation must be proportional only to the amplitude of the audio signal, and only the rate of carrier frequency swing is in proportion to the frequency of the audio signal.

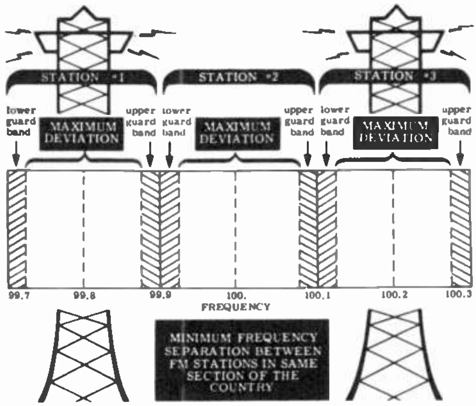
The basic method of eliminating the effect of increasing equivalent frequency deviation with increasing audio frequency is to decrease the amplitude of the audio signal in proportion to the frequency rise. The fundamental circuit for achieving this effect is the RC divider network shown in the diagram. The reactance of a capacitor decreases as the applied frequency increases, but the effect of the resistor remains constant at all frequencies. Consequently, the amplitude of the audio output signal decreases with increasing frequency. The values of R and C are selected to make the audio signal amplitude decrease exactly counteract the undesired frequency deviation increase.



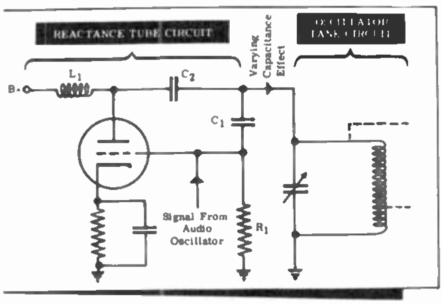
Review of FM Transmitters



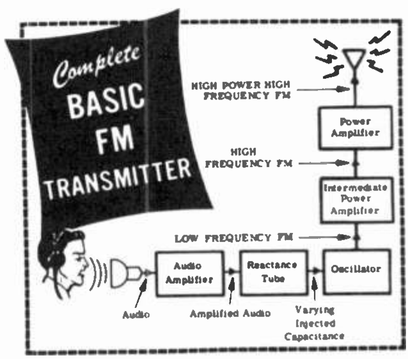
FREQUENCY MODULATION—In frequency modulation the frequency of the RF carrier signal is deviated at a rate and amplitude determined by the frequency and loudness of the audio signal being transmitted.



TRANSMITTER FREQUENCY CHARACTERISTICS—Commercial FM transmitters operate within the frequency band of from 88 to 108 megacycles. To prevent interference, the carrier frequency separation between stations in the same region can be no less than 200,000 cycles. The maximum permitted frequency shift is ±75,000 cycles.



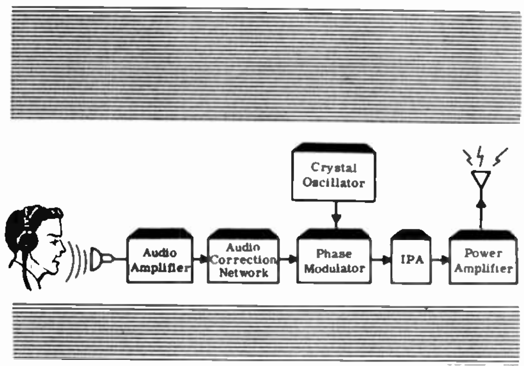
REACTANCE TUBE—The reactance tube achieves frequency modulation by injecting the effects of varying capacitance or inductance into the transmitter oscillator circuit. This is known as "direct" frequency modulation. The reactance injected into the oscillator varies at a rate and amplitude determined by the frequency and loudness of the amplified audio signal from the microphone.



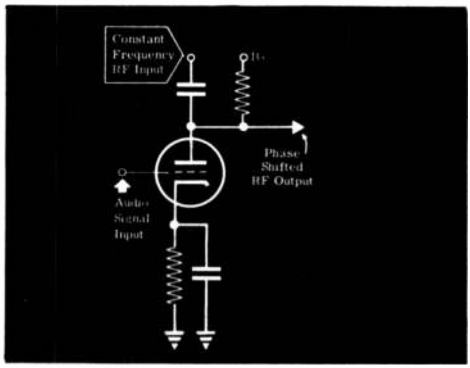
BASIC FM TRANSMITTER—Except for the addition of a reactance tube circuit, a basic FM transmitter is essentially the same as an AM transmitter with modulation introduced into the oscillator. Increasing the high frequency gain of the audio amplifier by a high-pass filter achieves pre-emphasis which increases the signal-to-noise ratio of the high frequency audio signals.

Review of FM Transmitters (continued)

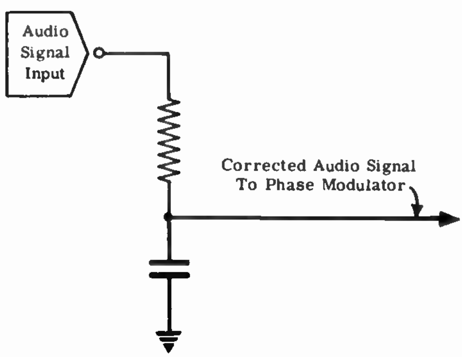
INDIRECT FREQUENCY MODULATION—FM transmission can be achieved "indirectly". The oscillator frequency remains constant, and the required frequency deviation is introduced in one of the stages following the oscillator. The advantage of this method is that a crystal-controlled oscillator can be used, and a stable center frequency can be achieved without the use of stabilizing circuits.



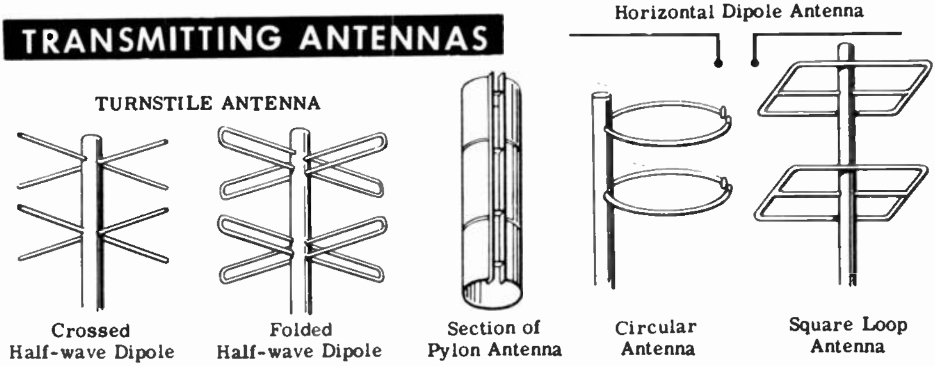
PHASE MODULATION—Indirect frequency modulation can be achieved by the use of a phase modulator which shifts the phase of the RF signal in accordance with the variations of the audio signal. The amplified audio signal from the microphone is used to drive a phase shifting network in one of the stages following the oscillator. The transmitted signal is exactly equivalent to an FM transmitted signal.



AUDIO CORRECTION NETWORK—When a phase modulator is used, the equivalent carrier frequency deviation is in proportion to the amplitude of the audio signal. Increases in the audio signal frequency cause additional undesired phase shift of the RF carrier signal. The undesired increase in phase shift is eliminated by using an RC network to decrease the amplitude of the audio signal in proportion to the frequency rise.



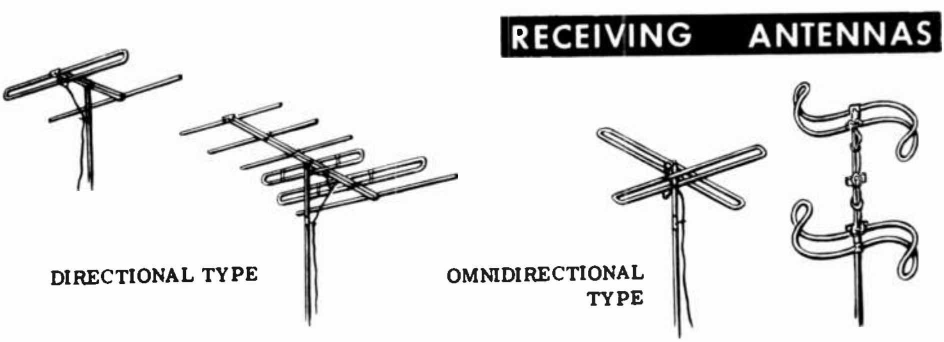
FM Antennas



The subject of antennas and wave propagation was considered in detail in Volume 4 of Basic Electronics. The information reviewed there also applies to FM radio signals, with the understanding the radio waves used in commercial FM transmission fall in the frequency range between 88 and 108 megacycles.

The types of transmitting antennas in common use are all variations of the horizontal dipole antenna considered in Volume 4. The turnstile type consists of multiple bays of crossed half-wave dipoles or folded half-wave dipoles. Another type, known as the circular antenna, consists of multiple bays of dipoles which are folded to the shape of an almost complete circle. The square loop antenna consists of multiple bays of square elements which effectively operate as dipoles and folded dipoles. Still another type, the pylon antenna operates like a very large number of circular elements stacked one above the other and in contact with each other.

Receiving antennas may be directional in those cases where there are only a few transmitting stations of interest, and they are all in the same general direction. In such an event the wide variety of television-type dipole antennas are quite suitable for use. When there are a number of stations in a region and they are located in different directions from the receiver, an omnidirectional antenna is used. This type is equally sensitive in all directions and consists of one or more bays of crossed dipoles or bent dipoles.

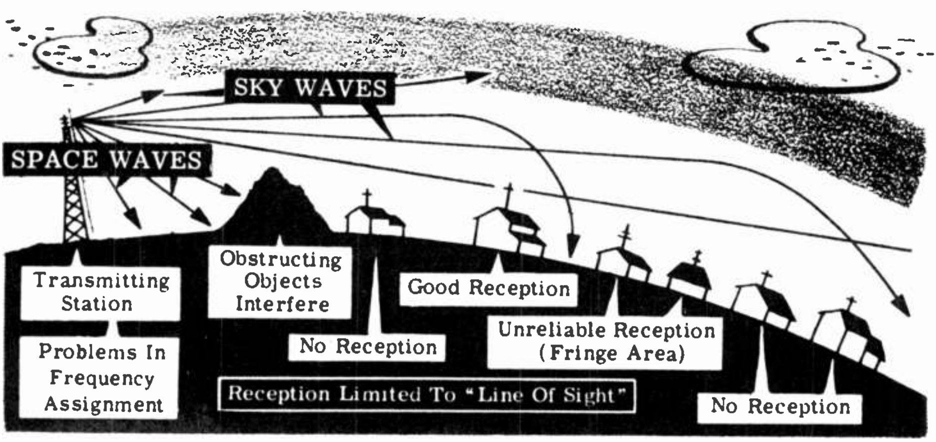


Wave Propagation

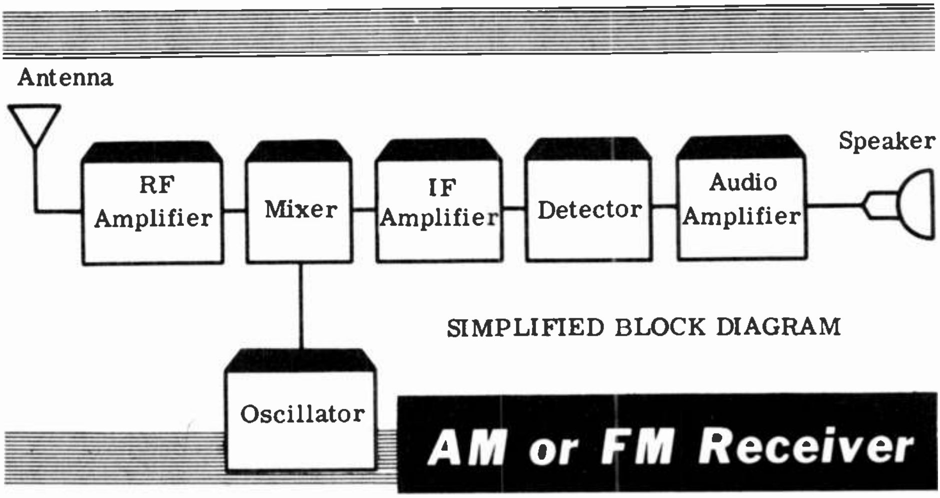
The function of the transmitter antenna is to radiate the transmitter signal into space. The radiated signal is electromagnetic energy, and it travels through space in a manner that is determined by its frequency. When a radiated wave leaves the antenna, part of the energy travels through the surface of the earth and is known as the "ground wave". The remainder of the energy is radiated into space in a pattern that is determined by the antenna design. Those waves which strike the ground between the transmitter and the horizon are known as the "space waves". Those waves which travel at an angle sufficiently high to pass over the horizon are known as "sky waves". At certain frequencies the upper layers of the atmosphere reflect and refract a portion of the sky wave back towards earth so that extremely long range reception is possible. At different frequencies some of these waves are more effective than others in transmitting signals.

At the frequencies used for FM transmission and television the ground waves are rapidly attenuated. In addition, the sky waves are not reflected or refracted back towards the ground so that they can be picked up by the receiver antenna. Because of this, reception is limited to the region within which the space waves can travel in a direct line of sight from the transmitting antenna to the receiving antenna. Thus reception is limited by the curvature of the earth. This range can be extended slightly when the transmitter is sufficiently powerful to cause some appreciable amount of sky wave to be bent back towards earth.

If both the transmitting and receiving antennas are at altitudes of 100 feet above sea level, FM transmission can take place at ranges in the order of 30 miles. If both antennas are raised to 1000 feet above sea level, FM reception can take place at ranges in the order of 80 miles. In mountain areas antenna elevations may be sufficiently high to permit reception at ranges over 150 miles. In cases where hills and buildings obstruct the line of sight this situation may cause severe decrease in the quality and range of reception.



FM Receivers



The most direct method for you to learn about FM receivers is to examine the block diagram of the most basic type and to compare it with that of the AM superheterodyne that was analyzed in detail in Volume 5 of Basic Electronics.

A preliminary look at these diagrams shows that both receivers have a number of common stages. Both have antennas to pick up the RF signal from the transmitter, and this signal is delivered to the main body of the receiver by means of a suitable lead-in line. Both receivers have an RF amplifier stage to step up the voltage of the signal from the antenna, although this stage is sometimes omitted in low-cost AM equipment.

The receivers both have an oscillator and mixer (converter) – sometimes as separate stages, sometimes in a single, dual-purpose stage. The purpose of these stages is to mix the amplified RF signal with the constant-frequency, constant-amplitude signal available from an oscillator. This results in a lower (intermediate) frequency signal which carries the same intelligence as that modulated upon the RF signal picked up by the antenna. The purpose of lowering the signal frequency is to permit high-gain amplification to be obtained more easily.

In both receivers the purpose of the stage following the IF amplifier is to remove the audio modulation from the intermediate frequency signal and to pass the resulting signal on to the audio amplifier. The general name for the stage following the IF amplifier is the "detector".

Following the detector is an audio amplifier, usually consisting of two or more stages. The purpose of this amplifier is to step up the audio signal to a level sufficient to drive the loudspeaker at satisfactory sound volume.

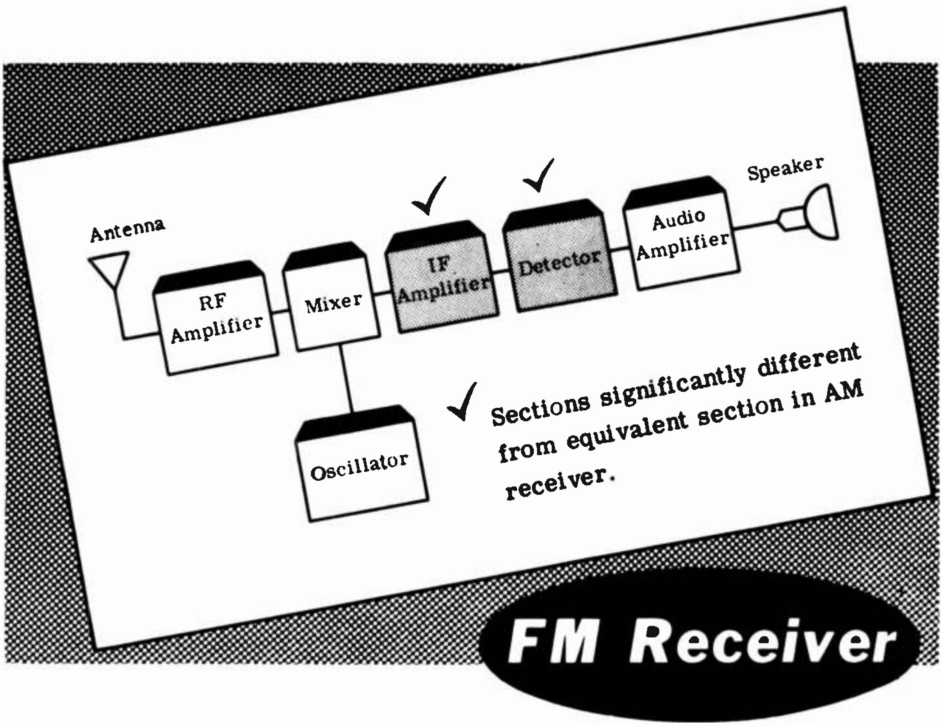
From this review it would seem that there are few differences between AM and FM receivers. Actually there are differences in all stages, with the most important differences being in the IF amplifier and detector.

FM Receivers (continued)

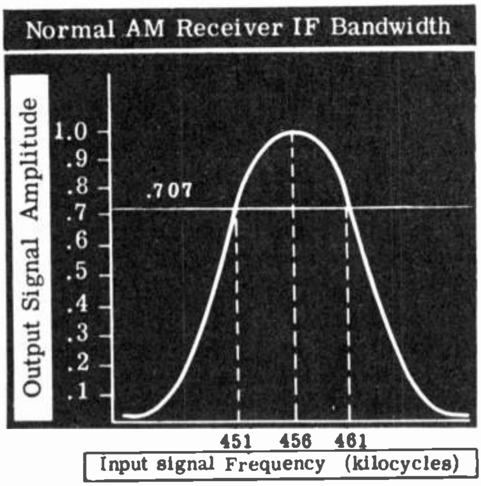
The antenna, RF amplifier and oscillator in an FM receiver must operate with incoming signals in the frequency range of 88 to 108 megacycles. This requires special consideration in the design and construction of these stages in order to achieve stable and reliable operation.

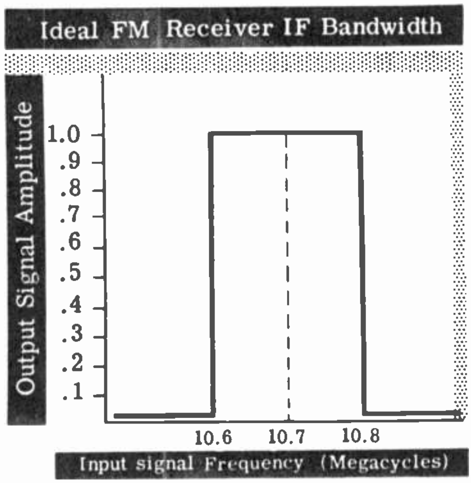
At the other end of the system, the audio amplifier and speaker in an FM receiver must be capable of a much wider frequency range than in an AM receiver. Because of the restrictions in the transmission of AM signals, 5000 cycles is the highest frequency that can be transmitted and received. To reproduce a 5000 cycle signal places no great requirements upon the audio amplifier and loudspeaker, and these components are capable of producing good results even though they are designed for economy rather than quality. In FM transmission, on the other hand, there is practically no limitation upon the range of audio signals that can be transmitted. Under the usual commercial FM broadcasting conditions, frequencies of up to 15,000 cycles are transmitted. Special consideration must be given to the design of the audio amplifier and loudspeaker in order to reproduce these signals in a satisfactory manner.

The various special considerations that have been mentioned up to this point have solutions that have been reviewed in the previous volumes. It is in the IF amplifier and detector sections that there are major differences between FM and AM receivers. These differences are so significant that they will be reviewed in detail in the pages that follow.



FM IF Amplifiers





In an AM receiver the IF amplifiers are tuned to a fairly sharp peak. This permits high gain and good selectivity to be obtained with only two, and rarely more than three, simply-designed stages. Since the audio signal modulated upon the RF signal does not exceed 5000 cycles, mixing the RF and local oscillator signals results in an IF signal (usually 456,000 cycles) with a bandwidth of 10,000 cycles. The bandwidth is 10,000 cycles because amplitude modulating an RF signal with an audio signal of 5000 cycles results in sidebands both 5000 cycles above and 5000 cycles below the RF carrier frequency. Thus an IF amplifier can have a bandwidth as narrow as 10,000 cycles without significantly attenuating modulating signals with frequencies up to 5000 cycles.

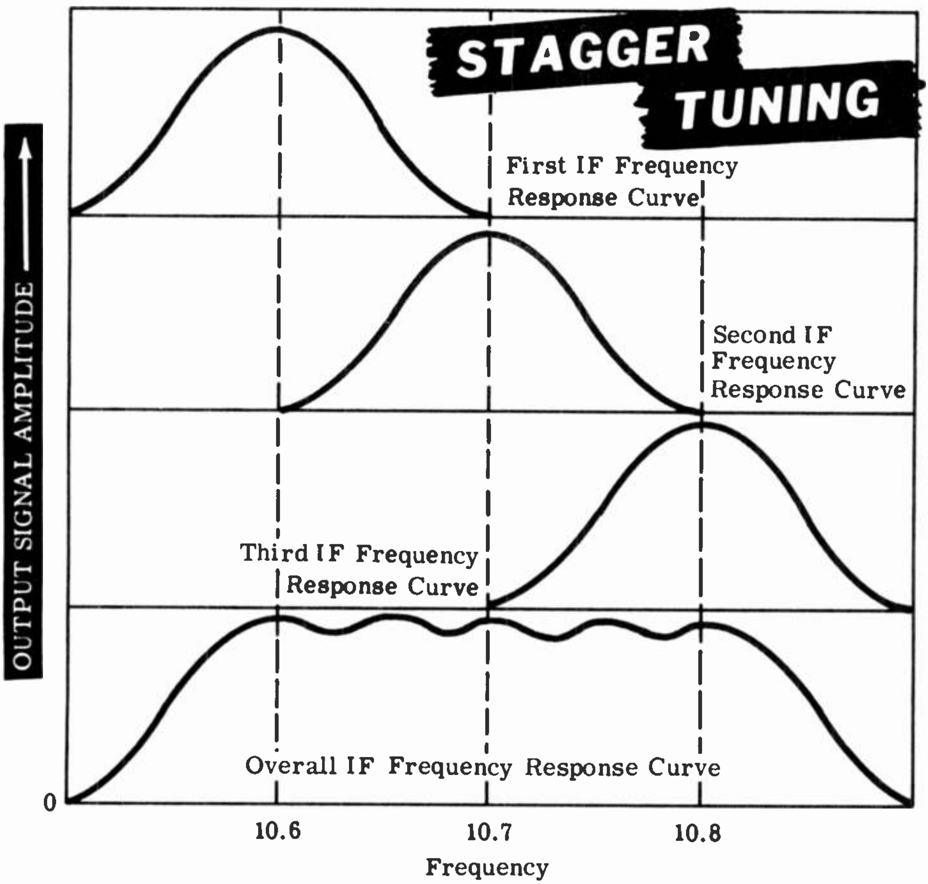
The situation in FM receivers is quite different. In the discussion of FM transmitting characteristics earlier in this volume it was indicated that a commercial FM transmitter is permitted a frequency deviation of 75,000 cycles above and below its rest frequency. When the received FM signal is mixed with the fixed-frequency oscillator signal, the resulting IF signal deviates 75,000 cycles above and below its center frequency. This means that 150,000 cycles is the minimum bandwidth that can be used in the IF amplifier. Furthermore, to pass the sidebands necessary to give a 150,000 cycle deviation it is desirable to design an FM receiver IF amplifier with a bandpass that is essentially flat for 100,000 cycles above and below the center frequency. The most common methods of obtaining this unusually wide frequency response will be considered next.

The usual FM receiver IF amplifier has a center frequency of 10.7 megacycles. It would be ideal for the amplifier to have a frequency response which is perfectly flat between 10.6 to 10.8 megacycles with a perfectly vertical dropoff outside of this frequency range. The reason for such a highly idealized response is that any deviation from it will produce different amounts of amplification for different signal frequencies. This introduces the effect of amplitude modulation in the signal and causes problems in the detector stage. This ideal frequency response is never achieved in practice, but three methods of obtaining a reasonable approximation are in general use.

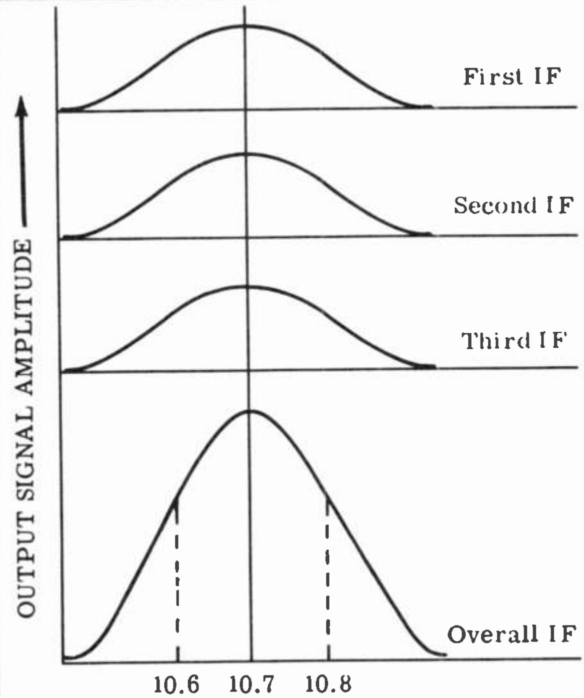
FM IF Amplifiers (continued)

The first method, known as "stagger tuning", gives an excellent frequency response curve. Three IF amplifier stages are used and they are tuned to 10.6, 10.7 and 10.8 megacycles respectively. The overall frequency response of this combination adds to produce the close to ideal response curve shown in the diagram. Note that at the intersection of the individual response curves of two adjacent amplifiers, the gain of the two amplifiers adds to produce a total that is about equal to the maximum gain of a single stage.

Although this method produces a very good frequency response curve, the overall gain produced by the three stages is no higher than that of a single stage. One way of counteracting this effect is to take special care to design the individual stages to have very high gain. Another way is to add a second group of three similarly tuned IF amplifiers — giving six stages in all. Neither of these solutions is economical to produce. In addition, special alignment techniques are required in either case. As a result, this method is not used as frequently as the two which will be described next.



FM IF Amplifiers (continued)



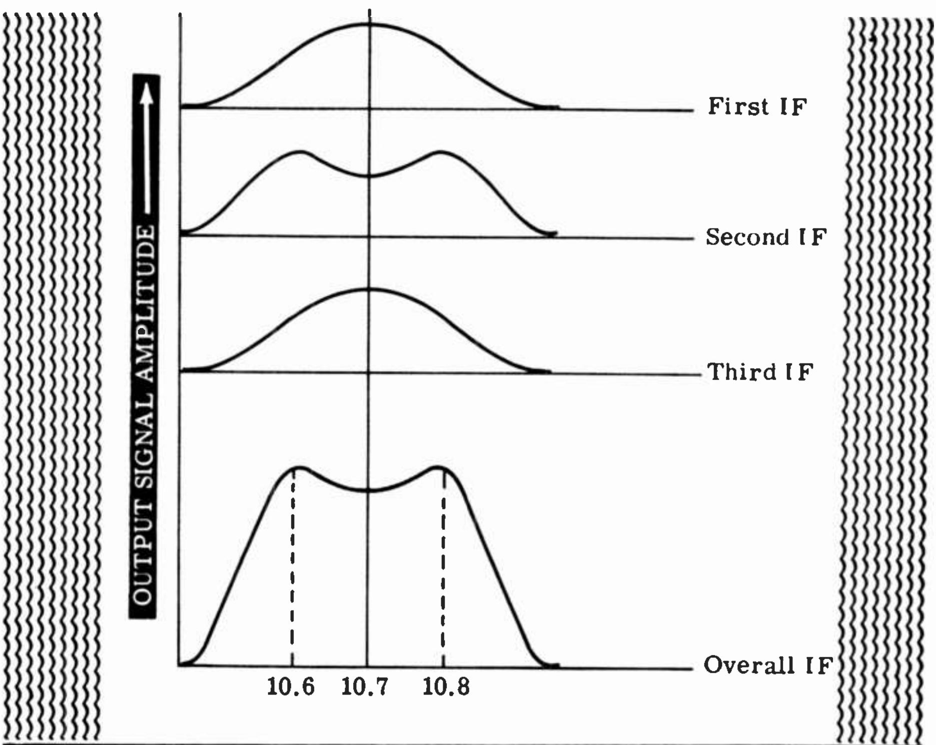
SUCCESSIVE SINGLE-PEAKED STAGES

The second method of approximating the desired broad frequency response is to use three IF amplifiers all tuned to the same center frequency. Ordinarily this would give a very sharply tuned frequency response curve. However, a broad response is obtained in each stage by using less than critical coupling in the transformers and by using transformer coils with low Q. Because of the broad frequency response of each amplifier the gain of each stage is low, but the overall gain of the three stages is sufficient to give the required signal amplification. Although this method produces only an approximation of the desired response curve and the overall gain is no more than adequate, the design and alignment involve very few problems. This method is frequently used in low-cost FM receivers. In the past few years vacuum tubes have been manufactured which are capable of producing unusually high gain at the IF signal frequencies in question. When these tubes are used in conjunction with careful circuit and component design, good gain and frequency response characteristics can be obtained with only two broadly tuned IF stages of the type described.

FM IF Amplifiers (continued)

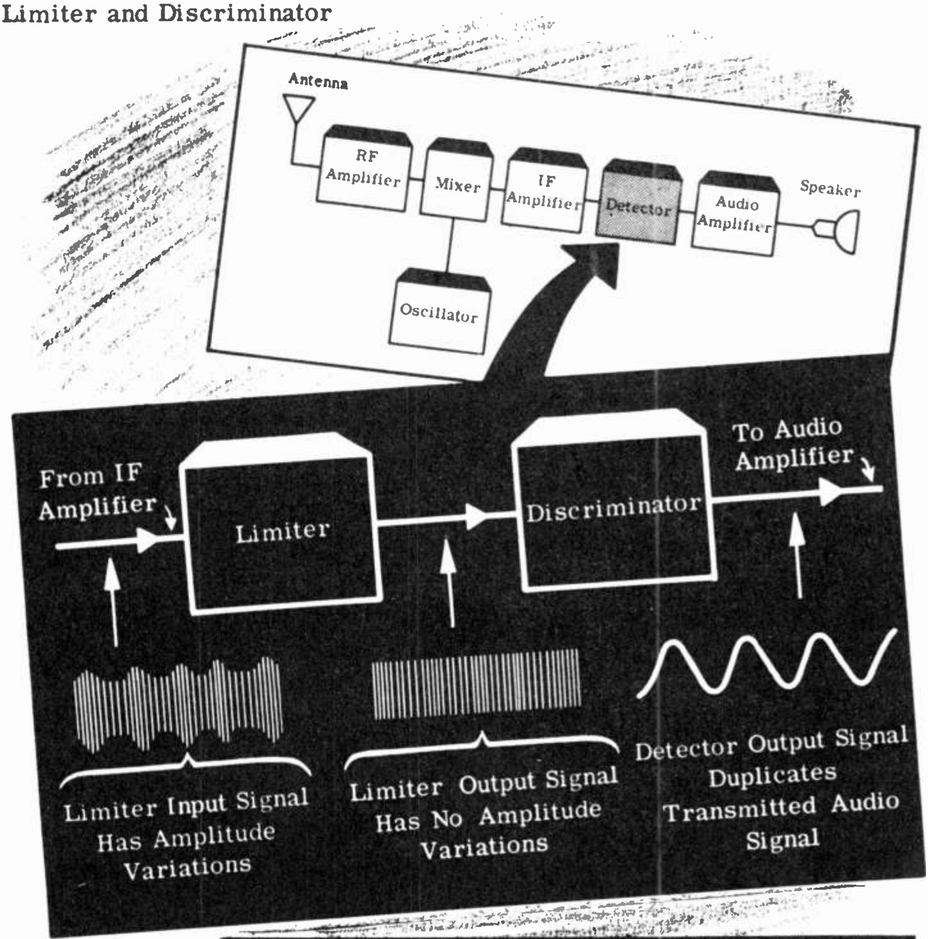
The third method also makes use of three IF amplifier stages tuned to the same center frequency. The first and third stages are designed and tuned as in the second method. However, the transformer of the second stage is overcoupled, resulting in the familiar double-peaked frequency response curve. When the individual response curves of the three stages are combined, the double peaks of the second stage have a significant effect in broadening the frequency response and producing a good approximation of the desired response curve. Although the overall gain is only adequate and the alignment is slightly more difficult than in the second method, the good response characteristics make this method a popular one. By using the new high-gain vacuum tubes, good gain and response characteristics can be achieved.

There are other methods of obtaining the approximate required response with adequate overall gain. These methods are usually combinations of the three methods that have been mentioned. When quality, rather than cost, is the prime consideration, one or more IF stages may be added to any of the combinations considered; and the result is a better response characteristic with higher gain.



ALTERNATE SINGLE-AND-DOUBLE-PEAKED STAGES

Limiter and Discriminator

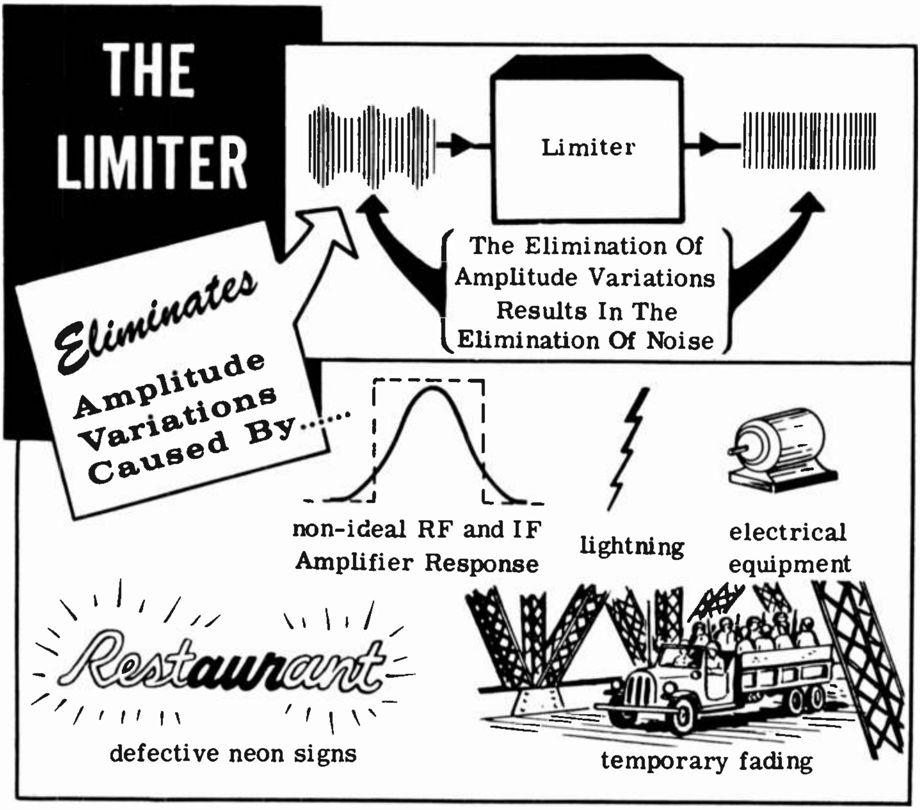


THE LIMITER- DISCRIMINATOR TYPE OF DETECTOR

The review of FM receiver IF amplifiers indicated to you that a variation was necessary from the arrangement used in AM receivers. The review of FM receiver detectors will indicate that the switchover from AM to FM is not a matter requiring variation, but a matter requiring completely new circuits. Two types of detector arrangements are in common use in FM receivers, and these will be analyzed in detail.

The first type of detector arrangement makes use of two stages with different functions: a limiter and a discriminator. The limiter serves the function of eliminating all amplitude variations in the received signal, so that the detector receives a signal that is a close duplicate of that sent out by the transmitter. The discriminator converts the FM signal output of the limiter to an audio signal, the frequency of which is equal to the frequency of the carrier signal swing and the amplitude of which is in proportion to the amplitude of the frequency variation.

Limiter and Discriminator (continued)



The purpose of the limiter is to eliminate all amplitude variations in FM signal. This is necessary since the discriminator is sensitive to amplitude variations and will reproduce them as signal distortion and noise in its audio output signal. Amplitude variations do exist in the FM signal for two basic reasons.

First, the RF and IF stages do not have a frequency response which is perfectly flat across the top and with perfectly vertical drop-offs. Any variations from this response causes different amounts of amplification for different signal frequencies and this adversely affects the operation of the discriminator, as will be seen later.

Another reason for amplitude variations are the interference signals caused by electrical equipment, lightning flashes, atmospheric disturbances, neon signs and a wide variety of other causes. In vehicles there is signal fading as hills, transmission lines, steel bridges and other large objects temporarily obstruct the signal path between the transmitter and receiver. If these variations reach the discriminator, they will produce noise or fading in the audio signal output. It is the action of the limiter to eliminate these variations that give FM receivers their well known freedom from noise.

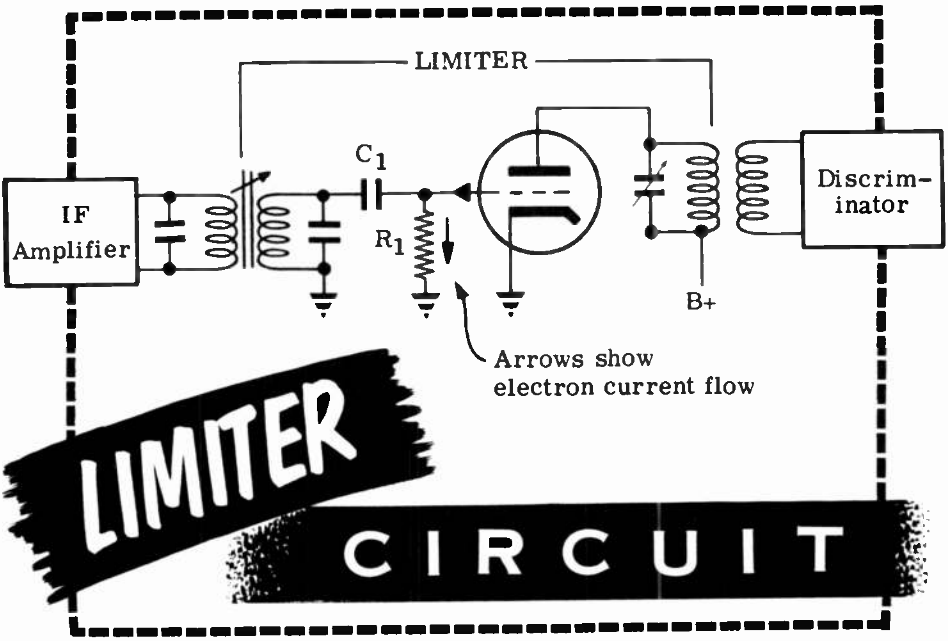
Limiter and Discriminator (continued)

The schematic diagram of a limiter circuit shows a close resemblance to both an IF amplifier and the grid-leak detector described in Volume 5. The circuit operates so as to limit the peak-to-peak voltage of the output signal to a fixed and pre-determined value. This operation takes place for all normal input signal levels, low or high, out of the IF amplifier.

Examination of a typical limiter circuit shows that the tube develops its grid bias by means of a grid resistor and a fixed capacitor in the grid circuit. The tube is of the sharp cutoff type and is operated with low plate voltage. The circuit diagram shows a triode tube for purposes of simplicity, although a pentode tube is generally used.

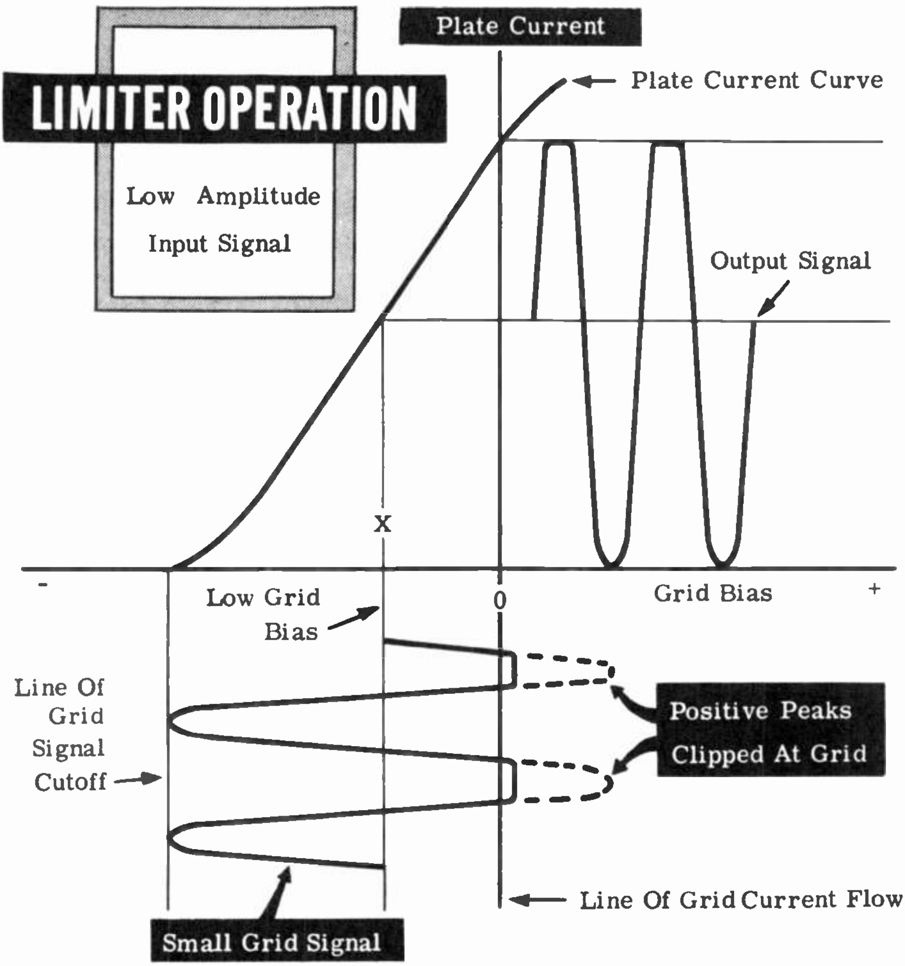
Note that the grid circuit arrangement is basically the circuit of a diode detector. The control grid operates as the diode plate, and the grid-leak resistor, R_1, replaces the diode load. The grid capacitor, C_1, serves both to couple signal to the grid and to take part in developing the grid bias.

When a positive signal peak from the IF amplifier reaches the grid, the grid becomes positive and attracts electrons from the cathode. The flow of electrons through the grid-leak resistor to ground produces a voltage drop across that resistor, and the flow is in such a direction as to make the grid negative with respect to the cathode. During the first few positive signal peaks from the IF amplifier, electrons also accumulate upon the capacitor plate next to the grid. Sufficient electrons accumulate upon that plate to maintain a stable current flow through the resistor during all parts of the cycle. Thus the magnitude of the grid bias is determined by the magnitude of the positive signal peak from the IF amplifier. The greater the amplitude of the input signal, the greater is the negative grid bias.



Limiter and Discriminator (continued)

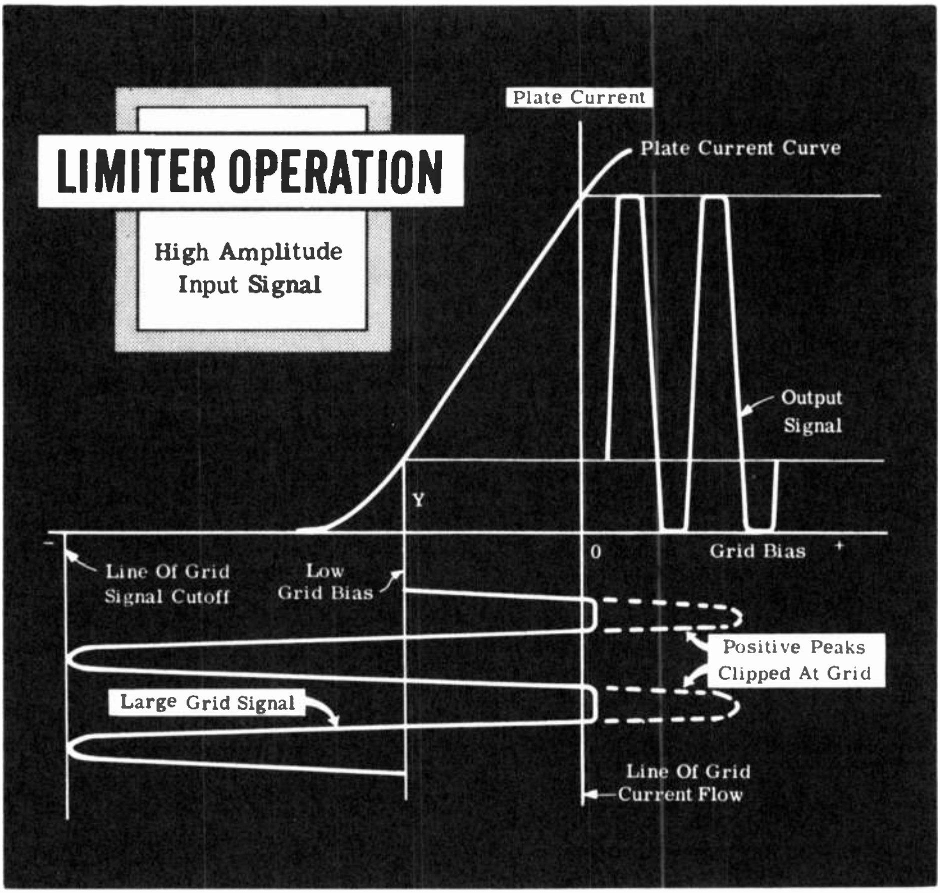
To see the effect of this grid bias examine the grid voltage versus plate current curves shown in the illustration. An input signal of low amplitude develops the low negative bias illustrated by line X. This bias voltage is less than the voltage of the positive signal peak. Consequently, the uppermost portion of the positive signal peak will drive the grid positive, and the grid will draw current. This current flow overloads the IF amplifier output transformer, and the current flow through the transformer internal impedance causes a drop in the output signal voltage. Thus the positive peak of the signal is clipped at the limiter grid, so that the peak can rise only a slight amount above zero volts. The negative peak of the grid signal drives the limiter plate current almost to the point of cutoff, just up to the point where clipping of the negative peaks begins to take place.



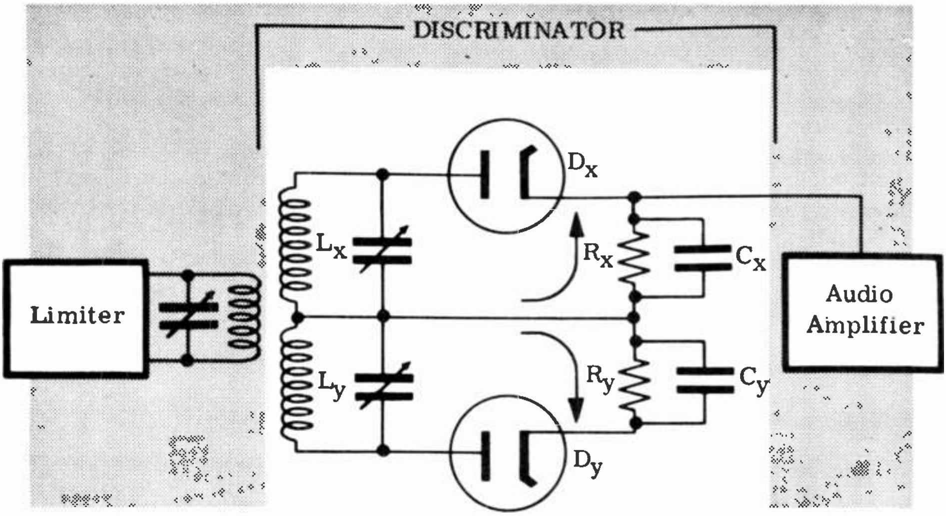
Limiter and Discriminator (continued)

When the input signal is at a higher amplitude, a large negative bias is developed, as illustrated by line Y. As large as this bias is, it is less than the voltage of the positive signal peak. Again the upper section of the signal peak drives the grid positive, causes grid current flow, and results in the clipping of the positive peak. Now the negative peak of the grid signal is of sufficient amplitude to drive the plate to cutoff, and the negative peak is clipped as shown in the diagram.

Thus any signal input which has an amplitude greater than the weak signal shown will have both its positive and negative peaks clipped. The limiter output will be the same for all such signals. Weaker signals will not have their negative peaks clipped, and there will be a small amount of amplitude variation. While the clipping action does distort the shape of the output signal, the discriminator is not sensitive to such distortion, and the desired frequency variations are preserved.



Limiter and Discriminator (continued)



Discriminator Circuit

Although there are a number of discriminator circuits in use today, they are mostly variations of the basic circuit to be considered here.

In this circuit the final IF transformer has a center-tapped secondary winding, and each half of this secondary winding has its own tuning capacitor. In order for you to understand the operation of the circuit it is important to consider exactly how the various windings are tuned.

The transformer primary is tuned to the resting frequency of the IF signal. Secondary winding L_x is tuned to a frequency 75,000 cycles higher than the resting frequency, and secondary winding L_y is tuned to a frequency 75,000 cycles lower than the resting frequency. Thus, these two windings are tuned to the extreme ends of the maximum frequency deviation of the IF signal. Low-voltage signals are developed across the ends of each transformer secondary winding when the incoming signal is at the resting frequency. As the incoming signal swings towards the resonant frequency of either tuned secondary winding, an increasingly large signal is developed across that winding.

The signal appearing across each transformer is rectified by a separate diode rectifier. Thus, DC voltages are generated across resistors Rx and Ry, and each voltage is proportional to the amplitude of the signal appearing across the associated transformer secondary winding. Electron flow through each resistor is in the direction shown by the arrows, and the voltages developed across the two resistors are in opposition. The resultant of these two voltages is the signal applied to the audio amplifier.

Limiter and Discriminator (continued)

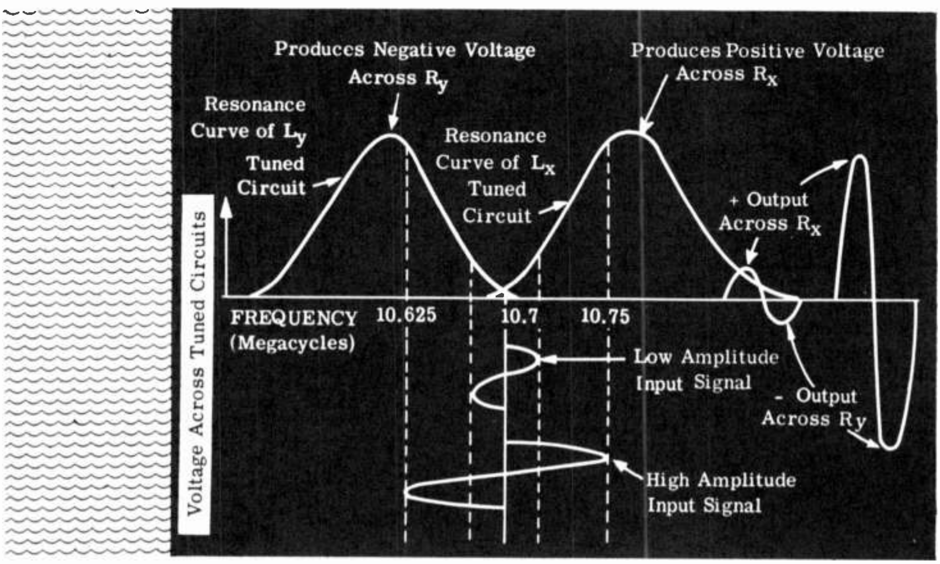
The resonance curves shown here indicate how an audio signal results from the FM signal coming out of the IF amplifier. It must be remembered that the frequency of the IF amplifier swings at a rate equal to the transmitted audio signal, and the deviation of the frequency is in proportion to the loudness of the audio signal.

Assume that the transmitted signal is a 1000-cycle note at low volume. In this event the IF signal will deviate approximately 15,000 cycles to each side of its resting frequency, and the frequency swing will be at a rate of 1000 times per second.

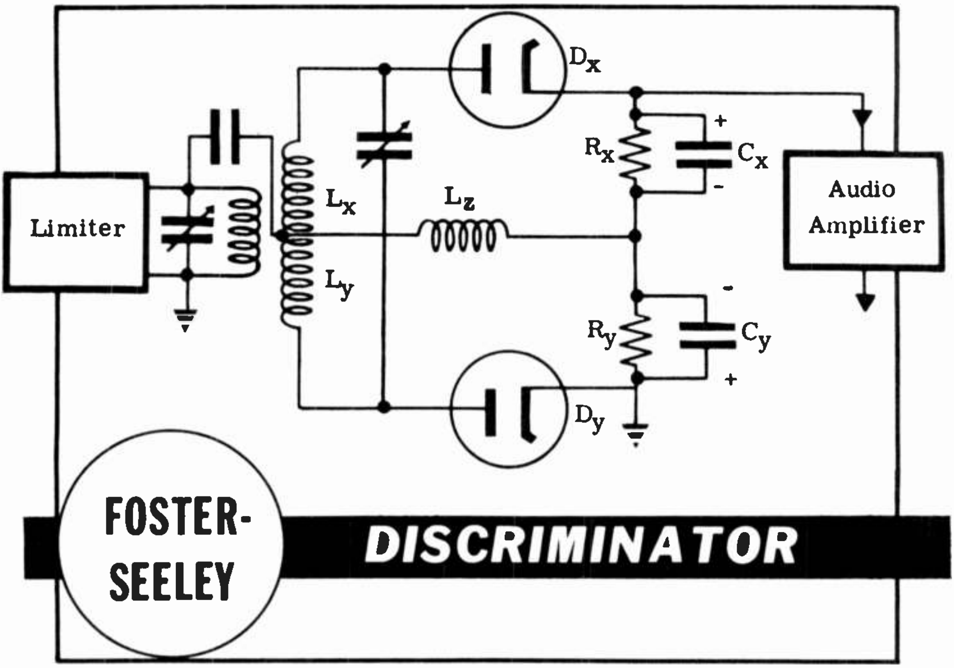
When the IF signal swings to 15,000 cycles higher than its center frequency, a moderate DC voltage is developed across Rx and a very low signal voltage is developed across Ry. These two voltages are opposed to each other; but since the voltage across Rx is the larger, a positive signal is applied to the audio amplifier input.

As the IF signal swings back towards its resting frequency, the voltage across Rx becomes lower; and the voltage across Ry becomes larger. At the resting frequency the two voltages are equal and opposite, and the voltage applied to the audio amplifier is equal to zero. As the IF signal swings towards 15,000 cycles lower than its resting frequency, the voltage across Ry becomes larger, while that across Rx becomes lower. The result is that the voltage applied to the audio amplifier input becomes increasingly negative. Since the IF frequency swings back and forth at a rate of 1000 cycles per second, a 1000-cycle signal of low amplitude is connected to the audio amplifier input.

If the frequency deviation increases, as it would with a louder transmitted audio signal, larger voltages are developed across Rx and Ry at the extremes of the frequency swing. The result is that an audio signal of high amplitude is connected to the audio amplifier input.



Foster-Seeley Discriminator



Although the discriminator described on the previous pages is the basis of a number of variations, it itself is seldom used. One reason for this is that the circuit is difficult to align, due to the two separated tuning frequencies of the transformer secondary winding. In addition, the transformer is more costly than the average IF transformer because of the difficulties of designing and manufacturing the unusual secondary arrangements.

The Foster-Seeley discriminator to be described here is a basic variation which is very widely used in modern FM receivers which make use of a limiter-discriminator. The major physical difference between it and the previous circuit is that the transformer secondary is center-tapped, rather than split; and a single tuning capacitor is used in the transformer secondary circuit. The diodes operate in the same manner as in the basic discriminator circuit, and the output signal is identical to that produced by the basic circuit. The main difference is that associated with the voltages across the secondary winding.

Both the primary and secondary windings are tuned to resonate at the resting frequency of the IF signal. Regardless of the frequency swing of the IF signal, the signal voltage across the upper half of the transformer secondary, Lx, is always equal to the signal voltage across the lower half of the secondary Ly. The signal voltage developed across the transformer primary is applied across the RF choke Lz. The voltage across Lz adds to both the voltages across Lx and Ly, as will be shown.

Foster-Seeley Discriminator (continued)

This arrangement results in an unusual phase relationship in the signal voltages applied across Lx, Ly and Lz. This phase relationship varies as the IF frequency deviates, and an audio signal voltage is developed across the discriminator output.

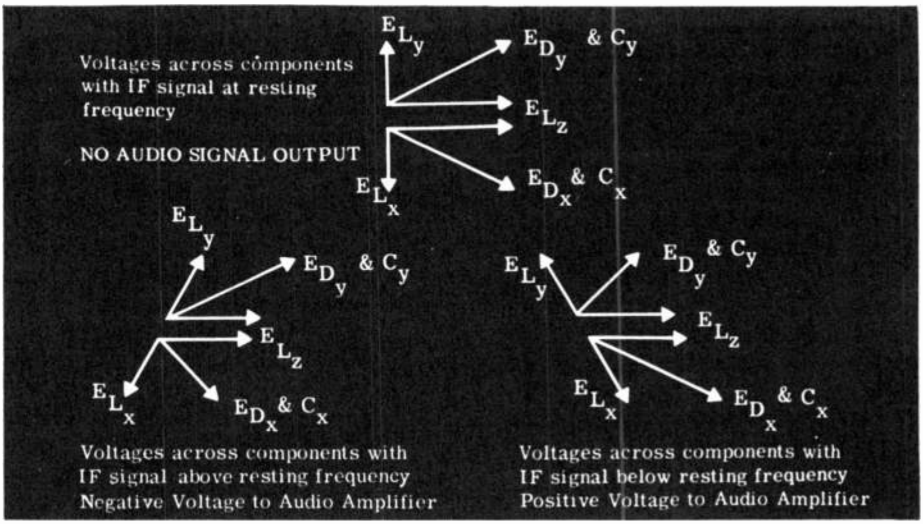
The various phase relationships in tuned circuits were analyzed in detail in Volume 4 of Basic Electricity. The statements which follow below describe only the results of the IF frequency deviation; and the reasons for these effects can be found by reference to that earlier volume. At present it is important to note that the vectorial addition of the voltages across Ly and Lz produce the voltage across diode Y and Ry, and the vectorial addition of the voltages across Lx and Lz produce the voltage across diode X and Rx.

When the IF signal is at its resting frequency, the signal voltages across Lx and Ly are equal and 180° out of phase with each other. In addition, they are 90° out of phase with the signal voltage across Lz. As a result the voltages across diode X and diode Y are equal, and the output to the audio amplifier is equal to zero.

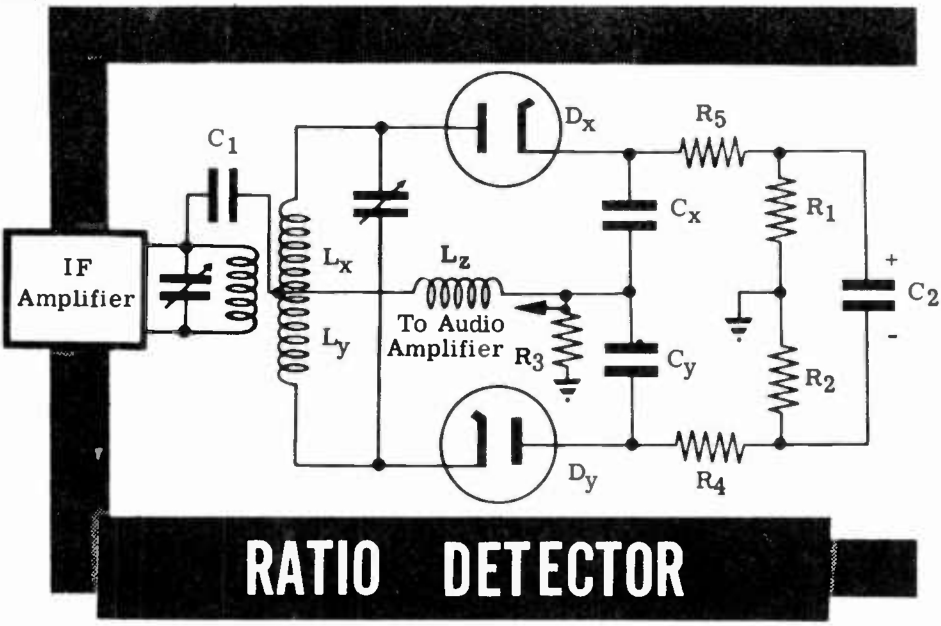
When the IF signal rises, the reactance in the secondary winding becomes increasingly inductive. Although the signal voltages across Lx and Ly remain equal and 180° out of phase with each other, they are no longer 90° out of phase with the voltage across Lz. The resultant voltage across diode Y is higher than that across diode X. Because of this, a negative voltage is applied to the audio amplifier.

When the IF signal falls below that of the resting frequency, the reactance in the secondary winding becomes increasingly capacitive. The voltages add as shown in the diagram, and that across diode X is higher than that across diode Y. Thus a negative voltage is delivered to the audio amplifier.

As the IF frequency swings back and forth, an audio signal is developed to correspond to the frequency and amplitude of the IF frequency deviation. The end result is identical to that of the discriminator considered previously.



Ratio Detector



Since the discriminator circuits considered previously are sensitive to amplitude variations in the IF signal, it is necessary that they be used in combination with a limiter which eliminates such amplitude changes. The ratio detector has been developed to provide FM detection without the need for a limiter stage. Because the ratio detector permits the elimination of one stage in the FM receiver, it is in widespread use today.

It can be seen from the diagram that the ratio detector has a close resemblance to the discriminator circuits considered previously. The major difference is that in this circuit the two diodes are connected together in a plate-to-cathode arrangement. In addition, the ratio detector has two resistors (R_1 and R_2) across which the audio output signal of each diode is developed. An arrangement which has not been seen in the discriminators reviewed previously is the resistor and capacitor network (R_1, R_2, C_2) shown in the right of the diagram. It is this arrangement which provides the limiting action which required an additional stage in the previous circuits.

Because of the way in which the diodes are connected, the incoming IF signal is rectified in such a manner that the upper part of the RC network is charged positively, while the lower part of that network is negatively charged. The voltage developed across the RC network is determined by the average IF signal voltage. The time constant of the network is generally selected to be in the order of 0.1 seconds. With such a time constant normal audio reproduction can be obtained; and short-term amplitude variations, such as produced by noise, will have no effect in changing the voltage across this network. Long-term increases and decreases in IF signal level cause corresponding increases and decreases across the RC network.

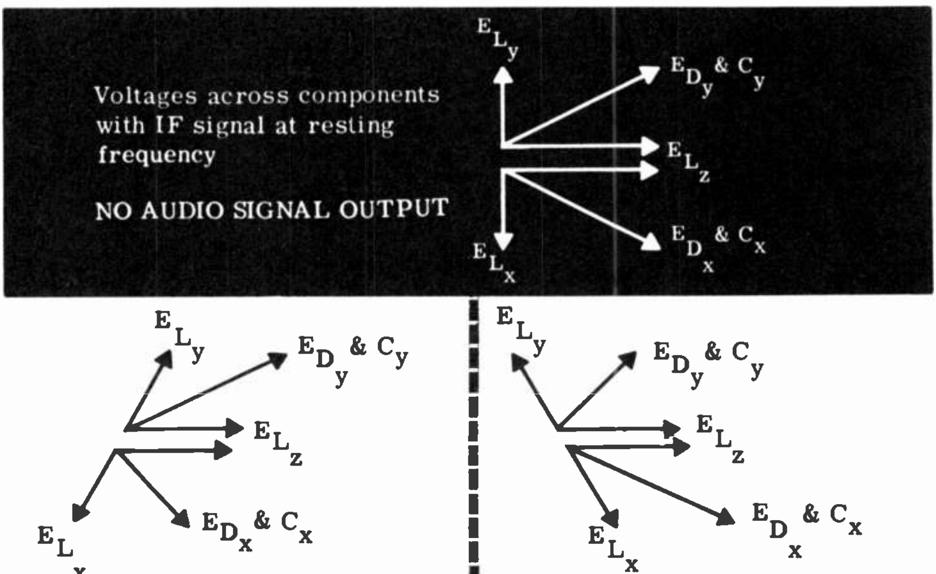
Ratio Detector (continued)

Establishing a stable voltage across the RC network is very important in the operation of this circuit. As the IF signal deviates about its resting frequency, the voltages across both Cx and Cy vary widely. However, the voltage across the RC network remains constant and the sum of the voltages across Cx and Cy is always equal to that across the RC combination.

The tuned circuits in the ratio detector operate in exactly the same manner as that described for the Foster-Seeley discriminator. The phase shift diagrams described in the discussion for that discriminator are also true for the ratio detector.

When the IF signal is at its resting frequency, the transformer secondary is at resonance. As a result the voltages across Cx and Cy are equal, but opposite in polarity. The voltage across output resistor R_3 is equal to zero, and there is zero volts applied to the audio amplifier.

When the IF signal rises in frequency, the phase shift diagram for the Foster-Seeley circuit holds true; and the resultant voltage across diode Y is higher than that across diode X. Similarly, when the IF signal frequency decreases, the resultant voltage across diode X is higher than that across diode Y. In both of these cases the voltage across Cx and Cy are unequal, although their sum always equals the voltage across R_3. The shifting inequality in the voltage across Cx and Cy causes the development of an audio frequency voltage across R_3, and this output signal is applied to the audio amplifier.



Voltages across components with IF signal above resting frequency Negative Voltage to Audio Amplifier

Voltages across components with IF signal below resting frequency Positive Voltage to Audio Amplifier

Evaluation of Detectors

Both the limiter-discriminator and ratio detector arrangements have advantages and disadvantages when compared with each other. Because of the nature of the relative advantages and disadvantages there can be no clear-cut statement as to which is the better. Each type is preferred by some designers, and it is essentially a matter of personal preference as to which is used.

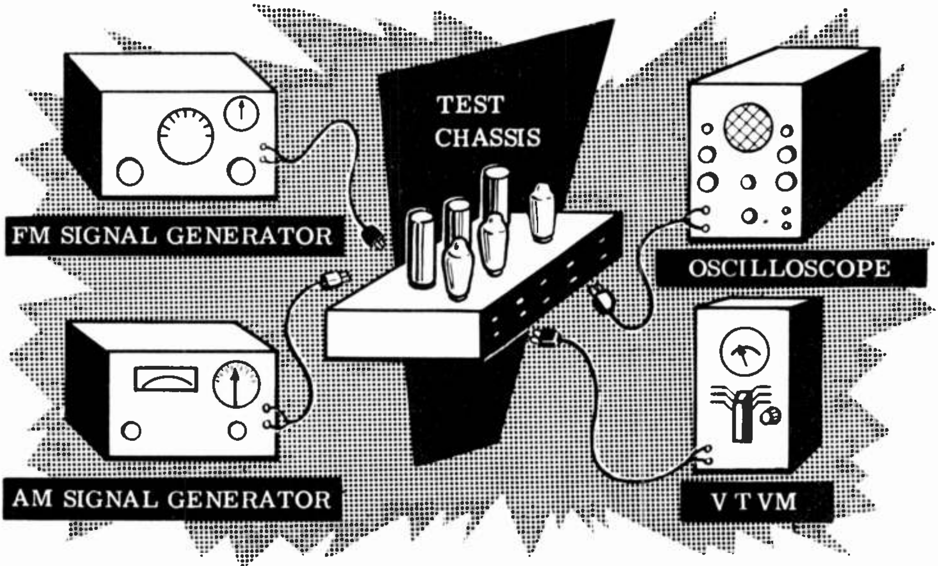
The important advantage of the limiter-discriminator arrangement is that it is relatively a simple matter to balance the two sides of the discriminator and obtain excellent reproduction of the audio-frequency signal. One disadvantage of this arrangement is that the limiter does not operate unless the incoming signal has sufficient amplitude to cause the limiting action to take place. When limiting action does not take place, the amplitude variations in the signal result in interfering noise and signal distortion. This means that high-gain RF and IF stages must be used to boost the signal amplitude into the limiter. In a number of FM receivers two limiters are used in a cascade arrangement to assure that adequate limiting action will take place. Even under these conditions signals of very low amplitude will result in amplitude modulation into the discriminator.

The important advantage of the ratio detector is that it is not sensitive to amplitude variations in the incoming signal, as was shown in the description of that circuit. In addition, the ratio detector circuit eliminates the need for the limiter stage, or stages, and does not depend upon the use of RF and IF stages of unusually high gain. The noise level rises with weak signals, but distortion does not take place until the signal amplitude is too low for reception. One disadvantage of the ratio detector is that special care must be taken to balance the two sides of the detector; otherwise some of the insensitivity to amplitude modulation will be lost, and noise will accompany weak signals. In addition, the ratio detector will produce more distortion in the audio signal output when the IF amplifier has an insufficiently broad frequency response range.

FM Detector Comparison

	ADVANTAGES	DISADVANTAGES
LIMITER DISCRIMINATOR	Simple balancing	Weak signals subject to AM interference
RATIO DETECTOR	Weak signals not subject to AM interference	Distortion and noise rise with detector unbalance and decreasing IF bandwidth

Demonstration—Operation of Basic Detector Circuits

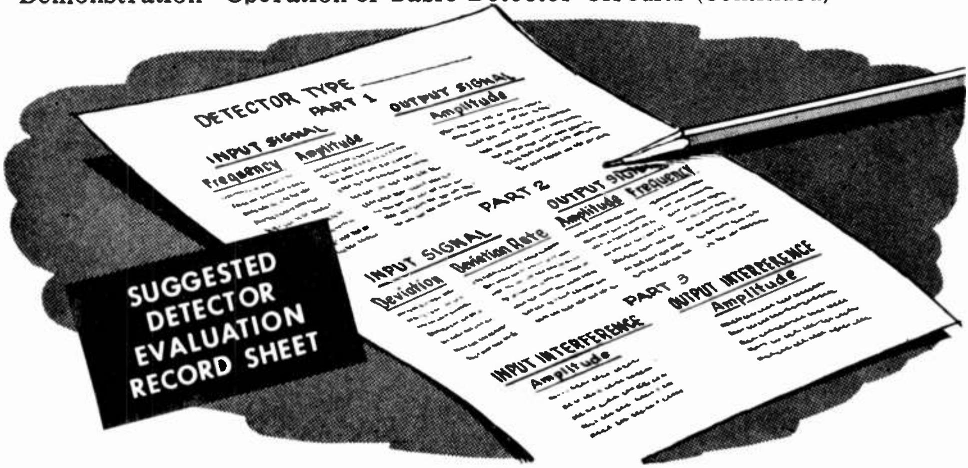


The purpose of this demonstration is to indicate to you tne outstanding characteristics of the basic types of FM receiver detector circuits. The equipment required is a chassis upon which are constructed a limiter and discriminator circuit, a limiter and Foster-Seeley discriminator, and a ratio detector. Test equipment required is an FM signal generator and a standard AM (RF) signal generator, with a built-in audio amplitude modulator, both of which are tunable to the 10.7 megacycle region. Also required is an oscilloscope and a vacuum-tube voltmeter.

The demonstration is begun by connecting both signal generators to the input of the limiter-discriminator circuit. Both the voltmeter and oscilloscope are connected to the output of the circuit.

To begin the demonstration all the equipment except the FM signal generator and the oscilloscope is turned on. The standard signal generator is set to 10.7 megacycles with the audio modulator off. The output on the voltmeter is equal to zero. As the frequency of the signal generator is slowly raised the voltmeter indicates an increasingly higher positive DC voltage. At various frequencies the amplitude of the signal generator output is raised and lowered. No essential change takes place in the voltmeter reading, showing that the circuit is insensitive to large changes in the amplitude of the input signal. Comparison of the results obtained will verify the conclusions indicated in the "Evaluation of Detectors." The voltage becomes a maximum when the frequency is slightly over 10.75 megacycles. As the frequency of the signal generator is slowly lowered, the voltmeter indicates an increasingly lower positive voltage which reaches zero at a frequency of 10.7 megacycles. Further decrease in frequency results in an increasingly negative voltage which reaches a maximum slightly below 10.625 megacycles. Returning the signal frequency to 10.7 megacycles returns the output voltage to zero.

Demonstration—Operation of Basic Detector Circuits (continued)



This first part of the demonstration indicates that the limiter-discriminator circuit increases and decreases its instantaneous DC output voltage in proportion to the deviation of the input signal from the center frequency. Also indicated is the fact that the circuit is insensitive to changes in the amplitude of the input signal.

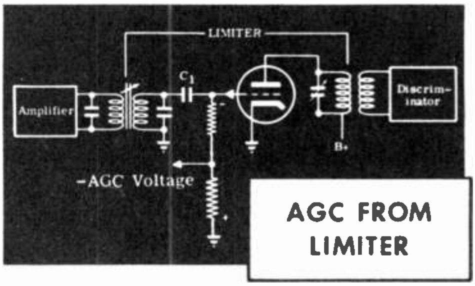
The second part of the demonstration shows that an audio signal is generated at the limiter-discriminator output when the frequency of the input signal deviates at an audio rate. To demonstrate this the standard signal generator is turned off and the FM signal generator and the oscilloscope are turned on. The tuning of the sweep signal generator is set to 10.7 megacycles.

With the frequency deviation limits of the signal generator set to about 15,000 cycles, the frequency of the audio modulating signal is varied over the entire range of adjustment. Under these conditions the oscilloscope display will show audio signals of constant amplitude which correspond in frequency to the frequency modulating audio signal. This indicates that the frequency of the limiter-discriminator output signal is determined by the rate of frequency deviation. Increasing and decreasing the extent of the frequency deviation raises and lowers the amplitude of the audio signal. Because of the limiting circuit raising and lowering the amplitude of the sweep generator output signal has little effect.

In part 3 of the demonstration the insensitivity of the limiter-discriminator to amplitude modulated signals is demonstrated by turning on the standard RF signal generator, tuning it to 10.7 megacycles, and turning on the audio modulator. When the oscilloscope screen is examined, only very small amounts of interfering audio signal will be visible. These interfering signals become most visible when the amplitude of the sweep generator output signal is reduced to below the level where limiting action takes place.

Now that the outstanding characteristics of the limiter-discriminator circuit have been indicated, the same procedure is repeated on the Foster-Seeley circuit and the ratio detector. Similar results are obtained, and the conclusions reviewed in the "Evaluation of Detectors" are verified.

FM Receiver Special Circuits—AGC

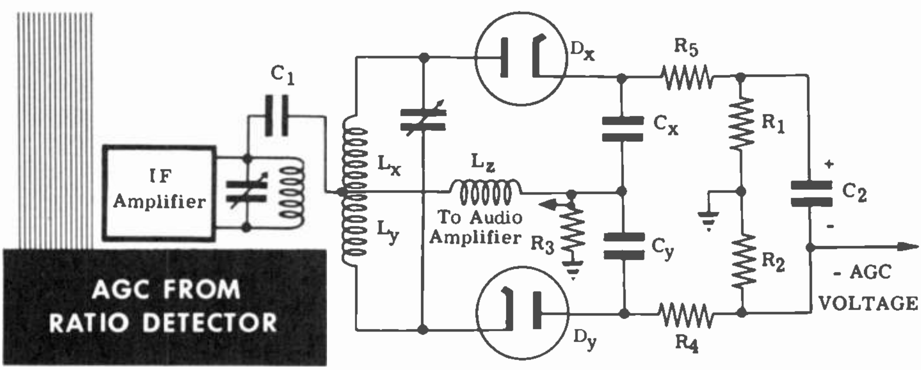


One of the few types of FM receiver special circuits that are important and frequently used are those designed to develop automatic gain control (AGC) voltage. Because of the transmission characteristics of the radio waves upon which the signal is modulated, FM signals from distant stations are subject to significant variations in signal strength. In addition there may be considerable difference in the signal level received from distant and nearby stations. Although FM receivers are largely immune to differences in signal amplitude, the signal level at the detector section should always be suitable for proper operation of the limiter-discriminator or the ratio detector stages.

As in AM superheterodyne receivers, the negative AGC voltage varies the gain of the IF stages in accordance with the amplitude of the received signal. The gain is reduced by the highly negative AGC voltage when there are strong incoming signals and the gain is reduced to a lesser degree in the case of weak incoming signals. This results in a stable signal amplitude at the detector, regardless of the incoming signal level.

When a limiter-discriminator combination is used, AGC voltage can be obtained from the resistor in the limiter grid to cathode circuit. Remember that the capacitor plate next to the grid retains a negative charge. If two resistors, rather than one, are placed in the grid-to-cathode circuit, a voltage divider arrangement is formed; and AGC voltage is available at the junction of the two resistors. Since the negative voltage on the grid increases with increasing IF signal level, the requirements for AGC voltage are met.

In the ratio detector circuit it was seen that the voltage across the R_1, R_2, C_2 network increases and decreases with corresponding changes in the IF signal level. Since a negative voltage is available this is a convenient source of AGC voltage.



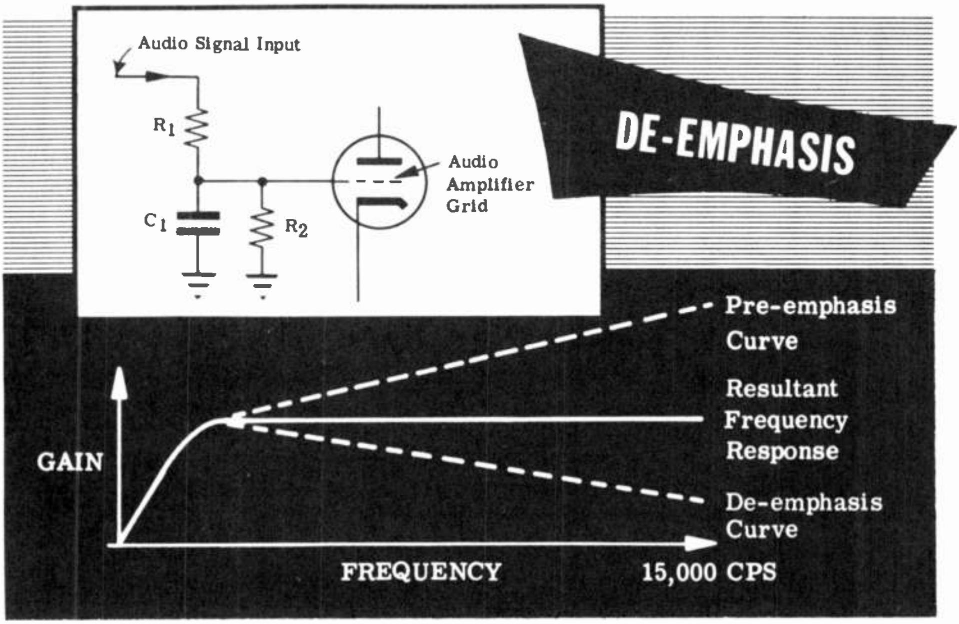
FM Receiver Special Circuits—De-emphasis

The purpose of pre-emphasis was explained earlier in the section concerning FM transmitters. It was indicated that pre-emphasis is a method of improving the signal-to-noise ratio of high frequency audio signals; and the method involves increasing the high frequency gain of the transmitter audio amplifier according to the frequency response curve shown here.

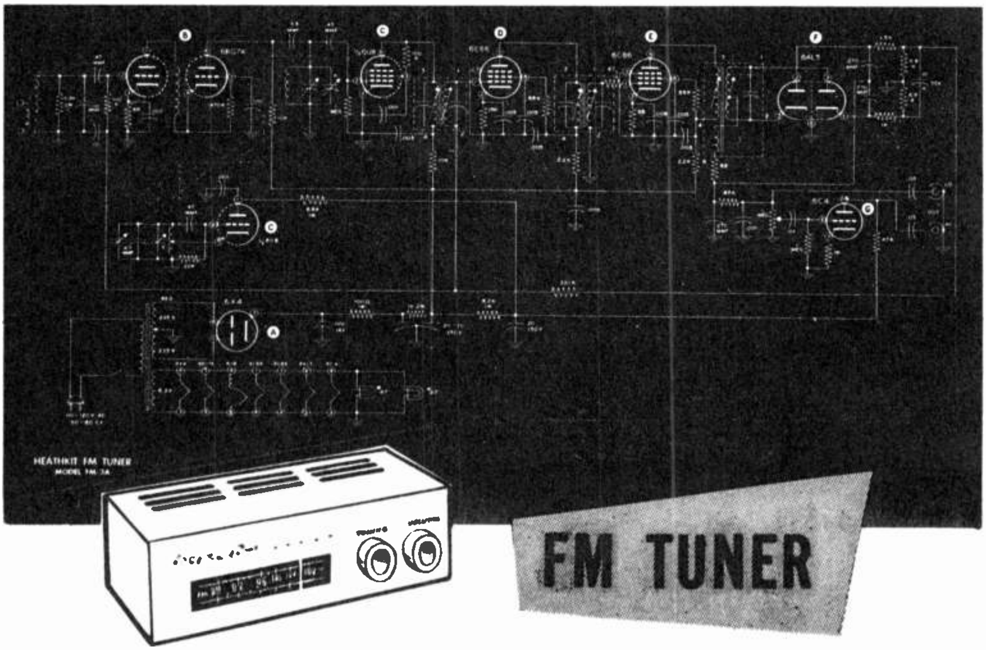
In the FM receiver arrangement described previously the high frequency audio signals would sound extremely unnatural, since, due to the transmitter pre-emphasis, these audio signals are amplified many times their normal amplitude. To bring the amplitude of the entire range of audio frequencies back to the distributuion normally found in voice and music, it is necessary to use "de-emphasis" to counteract the added high frequency gain introduced in the transmitter.

The method of accomplishing de-emphasis is to use an RC network which reduces the high frequency audio signal gain by exactly the same amount that it was increased in the transmitter. The basic arrangement for accomplishing this is to use the RC network shown in the diagram somewhere in the receiver audio amplifier — usually at the output of the detector.

The circuit is a low-pass filter which operates on the principle of a voltage divider. The capacitor is a small one, in the order of .001 microfarads. At low frequencies the capacitor has a very high impedance, and most of the audio signal voltage is applied to the grid. As the audio frequency rises, the impedance of the capacitor decreases; and increasingly less signal voltage is applied to the grid. The result is that the de-emphasis network reduces the high frequency gain of the amplifier and precisely counterbalances the effects of pre-emphasis.



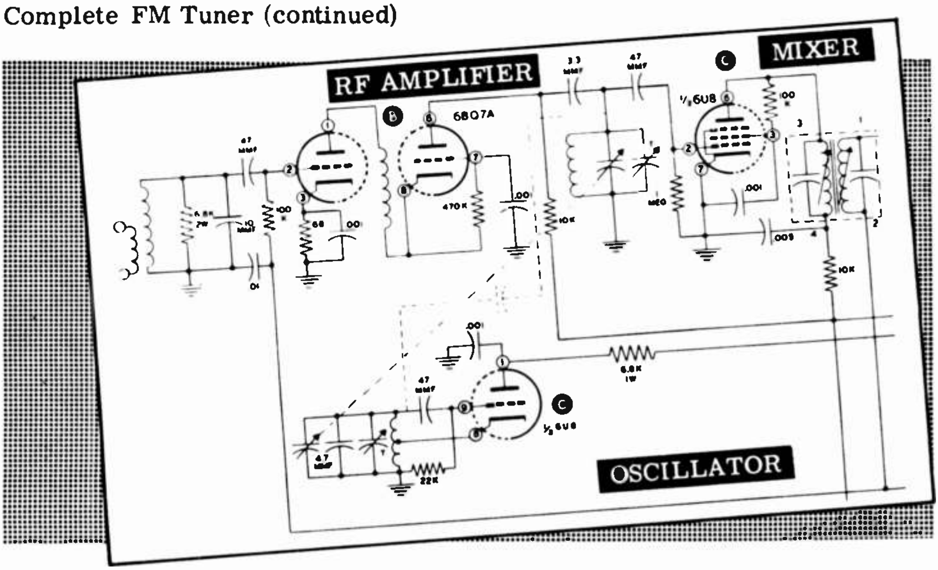
Complete FM Tuner



The equipment to be described now is a complete FM tuner, and its circuits are typical of those in widespread use today. An FM tuner is an FM receiver that is complete except for the audio amplifier. Many people, especially those interested in high-fidelity sound reproduction, have record players, tape recorders or AM superheterodyne receivers equipped with high quality audio amplifiers and loudspeakers. Such audio systems represent an investment of approximately $100 or more. Since an FM receiver requires an audio system of at least that quality to make full use of its capabilities, such a receiver would cost in the order of $175 - $200. By eliminating an audio amplifier of that quality, the cost of the receiver is reduced significantly, and the purchaser can make use of the audio amplifier equipment that he already has. If the FM tuner is bought in kit form (all components supplied but not assembled) the purchaser can do the construction and wiring and effect an additional savings.

The FM tuner to be described here contains 7 tubes, three of them dual-purpose tubes, making a total of 10 stages. The circuit consists of a cascade RF amplifier, oscillator, mixer, two-stage IF amplifier, ratio detector, audio voltage amplifier, and full-wave rectifier. Tuning range is from 88 to 108 megacycles, and the IF frequency is 10.7 megacycles. Maximum audio signal output is 15 volts, with a frequency response that is essentially flat over a range of from 15 cps to 17,000 cps. The purpose of the audio voltage amplifier is to make sure that there will be sufficient audio signal output to drive any type of audio power amplifier system.

Complete FM Tuner (continued)



Examination of the complete schematic diagram shows that there is very little in it that has not been considered previously. The paragraphs which follow review the major features of each section of the tuner, beginning with the RF stages and ending with the output to the audio amplifier.

The RF amplifier is a cascade amplifier. Incoming signals from the antenna are coupled to the 6BQ7A RF amplifier by means of the input transformer. Note that this transformer is not sharply tuned by means of a variable capacitor. Instead the transformer secondary is fixed-tuned by means of a fixed 10 mmf capacitor. To broaden the tuning of this coil and capacitor arrangement a 6,800 -ohm resistor is connected to the coil. Automatic gain control voltage is coupled from the detector stage, through the 100,000-ohm resistor to the input grid. Instead of using transformer coupling between the RF amplifier and the input of the mixer stage a 10,000-ohm load resistor and a coupling capacitor are used. This arrangement avoids the use of a sharply tuned circuit in the RF amplifier output and contributes to the broad-band amplification of that stage. Although gain is lost by these broad-band tuning techniques, this is compensated for by the high gain of the cascade amplifier and the stages which follow. The elimination of two variable tuning circuits has lowered cost without overall sacrifice of gain, has simplified the mechanical tuning arrangement, and has simplified the alignment procedure.

Output from the cascade RF amplifier is coupled to a variable tuning circuit and to grid number 2 of the 6U8 mixer-oscillator tube. The pentode section of this tube is connected as a mixer, while the triode section is a minor variation of the Hartley oscillator. Both of these sections contain tuned circuits which are adjusted by means of variable capacitors. The two variable capacitors are ganged together and are mechanically coupled to the tuning dial and its adjusting knob. The tuning is such that the oscillator frequency is always 10.7 megacycles higher than that of the RF amplifier, resulting in a 10.7-megacycle IF signal.

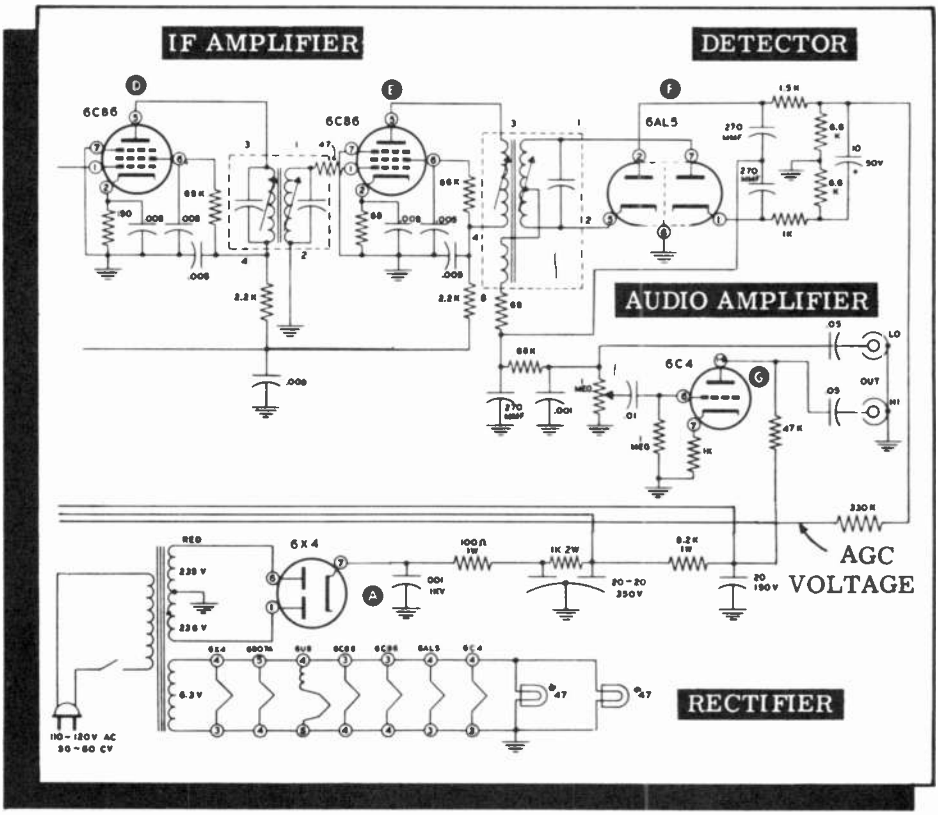
Complete FM Tuner (continued)

Two 6CB6 tubes are used to form two IF amplifier stages. Extremely high gain circuits are used, permitting very adequate performance with all transformers broad-band tuned to the same center frequency.

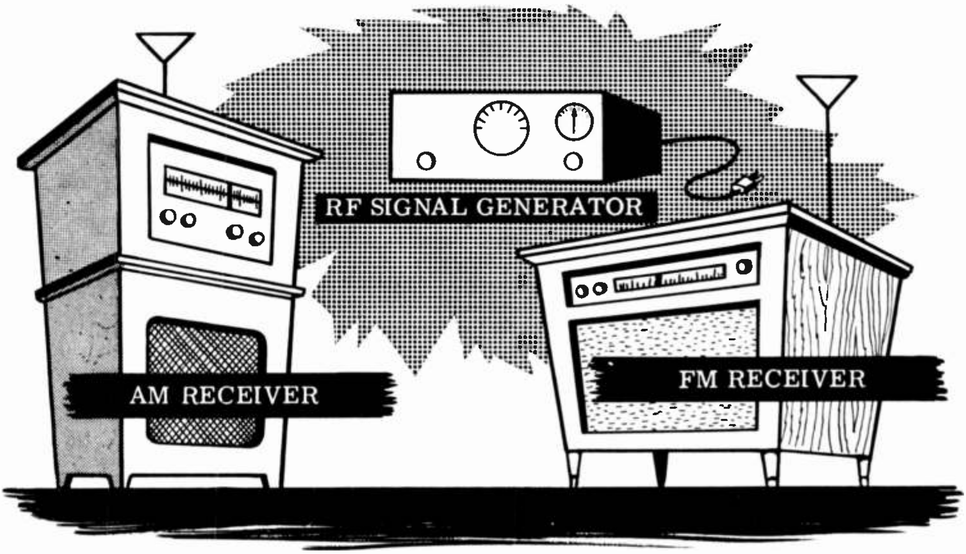
The output of the IF amplifier section is coupled to a 6AL5 ratio detector. The tube contains two diode sections connected in a manner which is identical to that considered earlier. Automatic gain control voltage is obtained from the detector and is used to vary the gain of the RF and first IF amplifier stages. A de-emphasis RC network is included between the ratio detector output and the variable volume control.

Output from the volume control is connected to the grid of a 6C4 tube connected as a triode amplifier. An unbypassed cathode resistor is used to introduce negative feedback, thus reducing distortion in the audio signal and assuring a broad frequency response.

The power supply of the tuner is completely conventional. A 6X4 vacuum tube is connected to a power transformer in a full-wave rectifier circuit. A two-section RC filter is used to filter the rectifier output, so that the B+ output voltage is quite free of ripple.



Demonstration - Operation of an FM Receiver



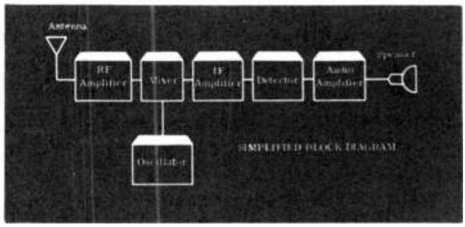
The purpose of this demonstration is to indicate the outstanding characteristics of an FM receiver and to compare these characteristics with those of an AM receiver. The equipment required is an FM receiver complete with antenna and suitable loudspeaker, plus an AM receiver similarly equiped and of comparative quality. Test equipment required is a standard RF signal generator, with a built-in audio amplitude modulator, which can be tuned to both the AM and FM broadcast bands.

In the first part of the demonstration the quality of AM and FM reception is compared. To do this each of the receivers is tuned to a local station which is broadcasting music. A better comparison can be obtained by tuning to a local station which is broadcasting the same program on both AM and FM. By alternately tuning the volume controls so that first one receiver and then the other is heard it will quickly become obvious that a much wider range of audio signals can be heard on the FM receiver.

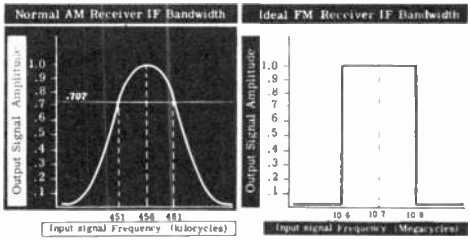
The second part of the demonstration shows that the FM receiver is many times more insensitive to interference than is the AM receiver. To do this the signal generator is connected to the antenna of the AM receiver. Then the signal generator, with the audio modulation on, is tuned until it is heard most loudly from the loudspeaker. Then the amplitude of the signal generator output is increased until the program being broadcast is practically inaudible because of the interference. Without changing the setting of the signal generator amplitude control, the generator is connected to the FM receiver antenna and tuned to the same frequency as the FM station being received. There will be little or no interfering signal audible from the FM receiver output. If available, other sources of interference such as buzzers, spark coils, defective neon tubes or motors, etc., may be tested to show the relative interference they produce in both receivers. In all cases the FM receiver will show great superiority in rejecting interference.

Review of FM Receivers

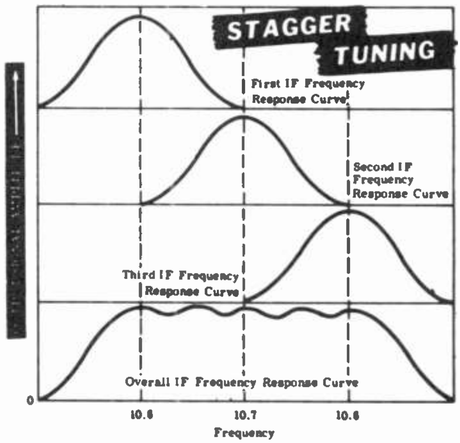
FM RECEIVER BLOCK DIAGRAM—There is a close similarity between the block diagrams of an FM receiver and an AM superheterodyne receiver. There are actual differences in all stages, with the most important differences in the IF amplifier and detector.



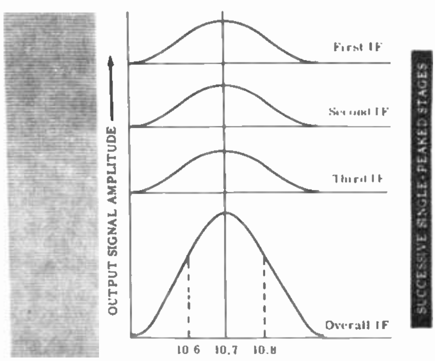
IF AMPLIFIER BANDWIDTH—To effectively amplify the received carrier and sideband signals without attenuation it is desirable for the IF amplifier to have a flat bandpass for 100,000 cycles above and below the center frequency.



STAGGER-TUNED IF—This type of IF amplifier contains three stages which are tuned to 10.6, 10.7 and 10.8 megacycles, respectively. Important characteristics are an excellent frequency response curve, an overall gain equal to that of a single stage, and the requirement for special alignment techniques.

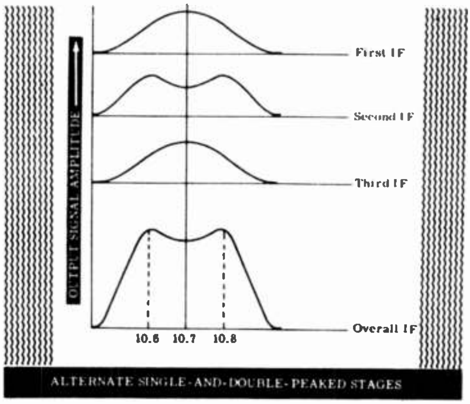


CENTER-TUNED BROAD BAND IF—In this type of IF amplifier all stages employ less than critical coupling in the IF transformer, have low Q transformer coils, and are all tuned to the center frequency. Adequate gain can be achieved, frequency response is acceptable, and the alignment procedure is simple.



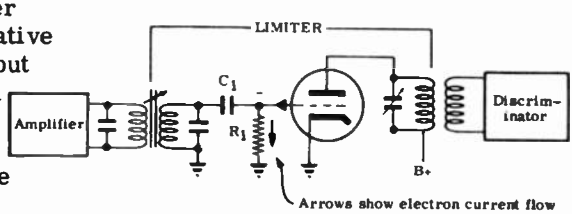
Review of FM Receivers (continued)

DOUBLE-PEAKED IF STAGE—The frequency response curve obtainable with center-frequency tuning can be improved by using an over-coupled IF transformer in one stage. The double-peaked response curve produced broadens the overall response of the complete IF amplifier. The alignment procedure becomes only slightly more complex.

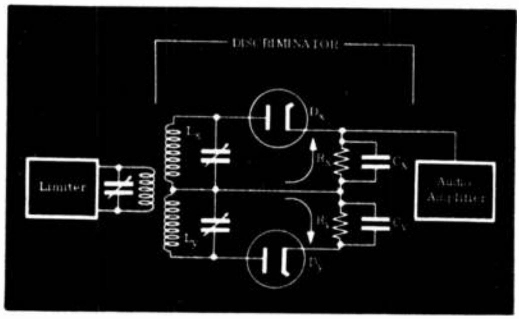


ALTERNATE SINGLE-AND-DOUBLE-PEAKED STAGES

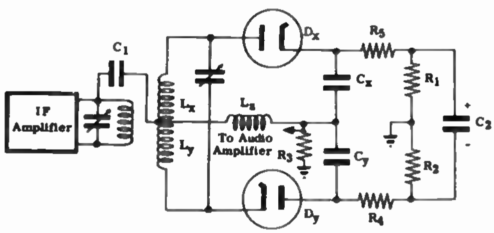
LIMITER STAGE—The limiter clips all the positive and negative peaks of the IF amplifier output signal and thus eliminates all amplitude variations. One or more limiting stages are required if a discriminator-type of detector is used.



DISCRIMINATOR—The discriminator converts the limiter output signal to an audio signal. The frequency of the audio signal is equal to the frequency of the carrier signal swing, and the amplitude is in proportion to the amplitude of the frequency variation. The discriminator is sensitive to amplitude variations and must be used with a limiter before it.



RATIO DETECTOR—The ratio detector provides FM detection without the need for a limiter. Operation is similar to that of the discriminator, except for the diode connections and the addition of an RC network to provide limiting functions.



CUMULATIVE INDEX

(Note: The first number in each entry identifies the *volume* in which the information is to be found; the second number identifies the *page.)*

CUMULATIVE INDEX

CUMULATIVE INDEX

CUMULATIVE INDEX

CUMULATIVE INDEX

$2.90

"BASIC ELECTRICITY"
by Van Valkenburgh, Nooger & Neville, Inc.

"BASIC SYNCHROS & SERVOMECHANISMS"
by Van Valkenburgh, Nooger & Neville, Inc.

These are civilian versions of the eminently successful illustrated training courses used by the United States Navy to turn out trained naval technicians in record time. Many of the nation's leading high schools are now using these courses for training. In industry, most large companies now conducting training programs in the field of electricity use these courses for personnel up-grading. Vast numbers of the technically-minded public have found that these "picture-books" teach the subjects faster and easier than was ever believed possible.

"BASIC ELECTRICITY"

TABLE OF CONTENTS:

#169-1,-2,-3,-4,-5, Available Individually.
#169 Soft Cover: 5 Volumes, 624 pp., 6x9".
#169-H, Cloth Bound: 5 Volumes in a single binding.

"BASIC SYNCHROS & SERVOMECHANISMS"

TABLE OF CONTENTS:

#180,-1,-2, Available Individually.
#180 Soft Cover: 2 volumes, 270 pp. 6x9".
#180-H, Cloth Bound: Both volumes in a single binding.

JOHN F. RIDER PUBLISHER, INC., NEW YORK
a division of HAYDEN PUBLISHING COMPANY, INC.