Transistors

THIRD EDITION

MILTON S. KIVER

Editor, Electronic Packaging and Production Magazine

McGRAW-HILL BOOK COMPANY

New York Toronto London

Transistors

Library of Congress Catalog Card Number 62-15145.

Preface

TRANSISTOR development continues unabated, no less today than when these remarkable devices were first made available to the engineering fraternity somewhat more than a decade ago. So rapid has been the flow of new forms of transistors that a complete revision of this book was required. The revision has been extensive, in many instances to the extent of a virtual rewriting of entire chapters. All material deemed no longer applicable to present devices was deleted; all characteristics and circuits have been updated and a considerable amount of new material has been added.

The scope of transistor application is as wide as engineering ingenuity. Already, this remarkable speck of dust has usurped many of the roles of the vacuum tube in electronic equipment from miniature pocket radios to room-filling computers. There is no segment of the electronics field in which it cannot or will not be used, and it is necessary that everyone who has any interest in this field become intimately acquainted with this amazing device.

This book is written for electronics technicians and all other technical workers who desire to gain a working knowledge of transistors and transistor circuits. The discussion starts with modern electron theory and then progresses to the operation of diodes and transistors. Chapter 3 contains a fairly extensive examination of a variety of transistor types, showing the historical progression from the early transistors to the units that are now capable of operating far into the megacycle region. As this story unfolds, the reader is able to perceive clearly the complete dependence of product development on fabricating techniques and how laboriously obstacles must be surmounted step by step. Chapter 3 is followed by a series of chapters in which a variety of transistor circuits are analyzed. All circuit explanations employ the highly successful step-by-step approach, starting with the simplest facts and proceeding gradually to the more complex. No mathematics of any difficulty is used in the text.

Chapter 9 focuses attention on the growing number of transistorlike devices whose behavior is markedly different from the triode transistor. These include the tetrode transistor, the PNPN transistor, the unijunction transistor, and the now-famed tunnel diode, among others. A considerable section is devoted to the tunnel diode because of its potential range of applications. In Chap. 10, there is a discussion of the various precautions to observe when servicing transistor circuits and transistor devices. A method of approach to the servicing of all such units is also given. In Chap. 11, a series of simple and easily worked transistor experiments are included for readers who may wish to learn of transistor operation firsthand. To the practical man, this is valuable experience. In Chap. 12, the reader is introduced to transistor circuit design, and this chapter should make the behavior of transistors in various circuits and at different frequencies better understood. For those readers who wish to become generally familiar with circuit design, this chapter will provide all the information required; for those who intend to engage in circuit design, Chap. 12 will serve as a gateway to the more advanced texts available.

The book can be used in technical institutes, electronics schools, vocational-technical schools, armed-service schools, industrial-training programs in electronics, or for home study. Questions are included for each chapter to be used by an instructor to test the progress of a student or as a form of self-test for those studying alone.

The author wishes to extend his appreciation to the many companies in this field, among them the Radio Corporation of America, Philco Corporation, Western Electric Company, Admiral Corporation, and Howard W. Sams & Co., Inc., of Indianapolis, Indiana, for the data and material which they so graciously provided. The author is also indebted to Philip Thomas of the Lansdale Division of the Philco Corporation for his aid in supplying valuable transistor information and for his many critical comments.

MILTON S. KIVER

Contents

CHAPTER 1

Introduction to Modern Electron Theory

THE STORY of the transistor is, in large measure, the story of matter and how the scientists at the Bell Telephone Laboratories have been able to make that matter amplify electric currents. If the transistor served no commercial end, it would still be important for its contribution to our understanding of matter in the solid state. It demonstrated, for the first time in history, that man can achieve amplification in a solid, a feat which heretofore could be accomplished only by using vacuum tubes. If we consider the vacuum tube as man's first significant advance into the field of communications, then the transistor must certainly be heralded as man's second most important step. For here, surely, is as radical a departure from what has heretofore been done as the invention of the triode by Dr. Lee De Forest in 1906.

The invention of the transistor is officially credited to John Bardeen, William Shockley, and W. H. Brattain, three scientists working for the Bell Telephone Laboratories. The first public announcement of the transistor was made in June, 1948. Thus, in terms of time, the transistor is barely out of its infancy. In terms of application, however, it must be classed with the vacuum tube. And while there is no imminent prospect that it will completely replace the vacuum tube, the transistor has nevertheless made serious inroads in a field that was once exclusively the province of the vacuum tube.

The most obvious attractions of transistors lie in their higher operating efficiency and smaller size than comparable electron tubes. Not quite as obvious, but nonetheless as attractive, is the ability of transistors to operate at extremely high frequencies. (Kilomegacycle transistors have been developed by several manufacturers. This is far above the operating range of any conventional vacuum tube.) A transistor, being a solid, requires no special envelope surrounding a vacuum; furthermore, it requires no filament heating element to serve as the provider of electrons. The latter fact alone represents a considerable saving in power, since in most standard receiving tubes as much or more power is frequently expended in heating the filament as in drawing current through the tubes. For example, in a 6CB6, the filament requires a current of 0.3 amp at an applied voltage of 6.3 volts. This represents a power dissipation of 6.3×0.3, or 1.89 watts. For typical operation as an amplifier, plate current is about 9.0 ma when the plate voltage is 200 volts. The power dissipated in this circuit is equal to 200×0.0090, or 1.8 watts. Elimination of the filament would reduce the overall power needs of this amplifier by half. Actually, in most applications, the power saving is greater than this because of the higher efficiency of the transistor. For example, with a transistor operated at the conventional collector current level of 5 ma and a collector voltage of 6 volts, the power dissipated is 30 mw. This compares with the 3.7 watts indicated above for a 6CB6. Add to this a volume which is on the order of one-thousandth that of a vacuum tube and a weight which is reduced by a factor of 100 and the attractiveness of the transistor becomes quite evident.

Operation of vacuum tubes depends upon the flow of electrons from filament to plate and the control of this flow by intermediate grids. Operation of the transistor is also largely dependent upon an electron flow, although there are considerable differences between the two units. In order to appreciate these differences, it is best to review what we know concerning the structure of matter and the role the electron plays in that structure.

Atoms and Molecules

Every substance or material that we come in contact with or which is known to man can be divided into particles known as molecules. These are the smallest segments into which a substance can be divided and still retain all its individual characteristics. Molecular units are so minute that we have not been able to devise instruments which will enable us to see them, and it is doubtful that we ever shall. In order to see something which is extremely small in size, we must design an instrument, such as an optical microscope or electron microscope, which will detect this "something" and then enlarge it so that we, with our gross eyesight, shall be able to see it. Every instrument we use, however, is itself composed of molecular building blocks. How, then, could we distinguish between the molecules of the substance we are checking and the molecules of the instrument? Thus, although it is unlikely that we shall ever see a direct picture of a molecule, highly refined indirect methods have been developed for determining molecular structure so that we know quite a bit about it.

We know, for example, that the molecule of a substance may consist of a single element of nature or a complex association of a number of elements. Chemists and physicists have discovered over a period of hundreds of years that there are more than 100 different elements which, either singly or in combination, make up all the matter on this planet. This figure is not a static one but has gradually been raised as man's knowledge and scientific know-how have broadened. Although it is unlikely, the number of elements may expand much more. The most familiar elements are hydrogen, oxygen, gold, silver, nickel, copper, iron, etc.

Most substances are formed of two or more elements. Thus, water consists of hydrogen and oxygen; common table salt is formed of sodium and chlorine; alcohol is a combination of hydrogen, oxygen, and carbon. Substances can also consist of a single element. This is true of oxygen gas, gold, silver, copper, lead, etc.

A molecule, then, is the smallest portion of a substance which retains all of the properties of that substance. But what of the elements? What do they consist of? If we are considering the smallest portion of an element that is still identifiable as that element, then we get down to the atom. Each different element is represented by a different atom.

Atomic Structure

Thus, as we tread our way down the scale of size, we come first to molecules, then to the elements that compose the molecules, and finally to the atoms which represent the elements. When we investigate atoms, we find that they consist of a centrally situated nucleus with a net positive charge surrounded by a number of electrons which revolve about the nucleus. The central positive charge is said to be due to protons, while each of the electrons has a negative electrical charge. In a stable atom, the positive charge of the nucleus is exactly counterbalanced by the total negative charge of the externally revolving electrons. The net electrical charge is zero, and this is the conventional state of most atoms.

The atom possessing the simplest structure is hydrogen. It consists of a positive nucleus containing a single proton. Revolving around this proton is a single electron. The illustration most commonly employed for the hydrogen atom is shown in Fig. 1·1. Actually, we have learned enough about atomic structure to know that Fig. 1·1 is a highly simplified picture of the hydrogen atom. However, we would not gain any greater understanding of transistor action by modifying this illustration to conform to more recent theories, and its simplicity does impart an understanding that might not be obtained otherwise. Hence we shall remain with this, the more classic method of representation.

Helium follows next in order of complexity, and its atomic structure is indicated in Fig. 1·2. The central positive charge is 2, indicating two protons, and this charge is offset by having two electrons rotating about the nucleus. Each element then follows in numerical turn, with

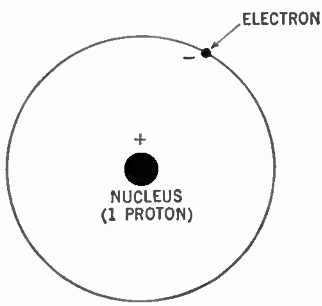


Fig. 1·1 The structure of a hydrogen atom.

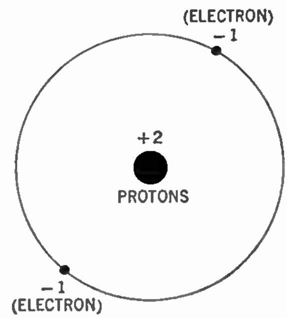


Fig. 1·2 The atomic structure of helium.

the central positive charge increasing in steps of 1 and being electrically counterbalanced by additional electrons revolving in paths or orbits about the nucleus. It can be appreciated that the nucleus and the associated electrons soon attain a highly complex structure.

The electrons which revolve about the nucleus do not follow random orbits; rather, they fall into definite energy levels. These levels may, for simplicity, be visualized as shells, each successive shell being spaced at a greater distance from the nucleus. The shell, or energy level, closest to the center carries a single electron (as in the hydrogen atom) or two electrons (as in helium and all other atoms). These electrons may rotate at any angle about the nucleus, but they are more or less bound to remain within the confines of the shell.

When we reach the third element, lithium, we find that the nucleus has a positive charge of 3, which is electrically counterbalanced by 3 negative electrons revolving around the nucleus, Fig. 1·3. Two of the electrons revolve about the center within the boundaries established by the energy level, or shell, just mentioned for the hydrogen and helium atoms. The orbit of the third electron, however, is much farther removed from the nucleus, and the third electron may be said to operate within another level which is entirely distinct from that of the first shell. This second level has been found capable of holding up to a maximum of 8 electrons, a condition that is achieved in the element neon. Neon, with a positive charge of 10, has 2 electrons in the first shell and 8 electrons in the second shell, Fig. 1·4. Thereafter, additional electrons start filling up a third level which can hold 18 electrons and then a fourth level which can hold 32 electrons. Beyond this, there are two additional shells, but these are never entirely filled

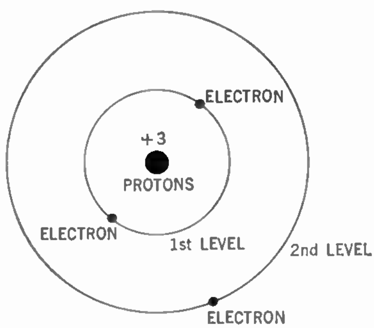


Fig. 1·3 The atomic structure of lithium.

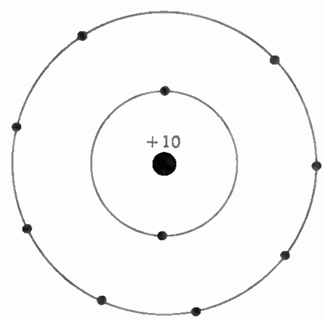


Fig. 1·4 Neon, with a positive nuclear charge of 10, has 2 electrons in the first shell and 8 electrons in the second shell.

because there are only 101 elements in all. Hence, we do not actually know just how many electrons they could hold.

Comparison of Atomic Structure and Solar System

The structural arrangement of an atom, with the central nucleus and the revolving electrons, has often been compared with our own solar system, Fig. 1·5. In this analogy, the atomic nucleus is equivalent to the sun and the revolving electrons are equivalent to the planets. Actually, many differences exist between the two systems, as the reader can appreciate. In the solar system, each planet follows its own path independently of the others; no two planets are equidistant from the sun. In the atom, the electrons revolve at certain specific distances for which we have used the name "shell." It must be recognized, of course, that there is nothing physical about these shells. They merely represent certain energy levels, and each level can carry a certain number of electrons.

It is also important to note that even in atoms containing large numbers of revolving electrons, there is still more empty space than there is solid material. The nucleus is rather closely packed, and it contains practically all the mass of the atom. Electrons provide only a negligible portion of the atomic weight, even though their combined electrical charge is equal to that of the far heavier proton.

To emphasize the relatively great distances that separate the nucleus and the first shell of electrons, it has been estimated that if we were to enlarge a helium atom so that the central nucleus attained the size of a golf ball, we would find each of the electrons to be as large as tennis

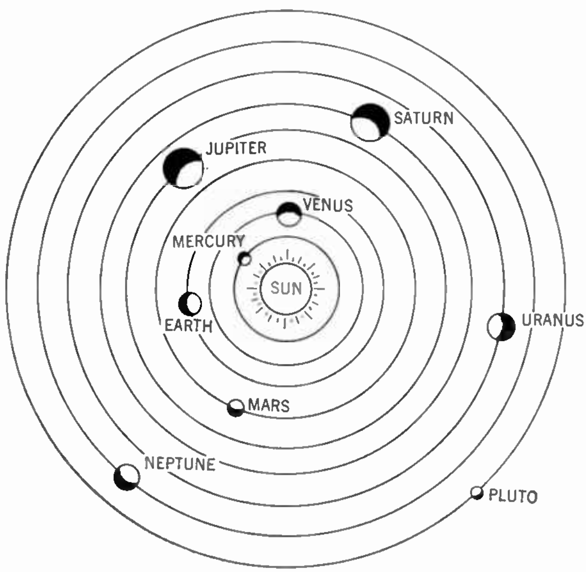


Fig. 1·5 The solar system. In a limited sense, the central nucleus of an atom may be compared to the sun and the electrons may be compared to the planets.

balls situated about 20 miles away. Thus, it is readily possible for two atoms to pass through each other's system without coming into physical contact with each other.

Composition of Nucleus

The central nucleus has been indicated to consist solely of protons. Actually, this is true only for the element hydrogen. As the elements increase in atomic number, it is found that the nucleus may contain electrons as well, with each such electron paired off with an offsetting proton. The net electrical charge of this combination is zero, and hence it does not influence the net positive charge of the nucleus. To achieve this charge, we must have unattached protons. Thus, when we say that a nucleus has a positive charge of 15, for example, we mean that it has 15 "free" protons. Actually, there may be more protons in this nucleus, but each of those beyond 15 has its charge neutralized by an electron which is bound to it within the nucleus.

This combination of a proton and a closely bound electron is called a neutron, the name stemming from the fact that the electrical charge is zero or neutral. In the lighter elements, those with relatively small numbers of revolving electrons, the number of nuclear neutrons is generally equal to the number of free protons. However, as the atomic number increases, the number of neutrons rises faster than the number of free protons. A neutron adds as much mass to an atom as a proton and hence is an important factor in the overall weight of an element. Actually, each neutron adds somewhat more mass than a proton because a neutron possesses an electron too.

It may be added that in recent years, a number of additional particles have been discovered in the nucleus. The meson and the neutrino are two examples of these particles. The meson may have a positive or negative charge, while the neutrino is devoid of charge. The mass of each particle is quite small, scarcely enough to be a significant factor in the overall atomic weight of the element.

The foregoing discussion has been given in some detail so that the reader will better understand transistor operation. The transistor is a solid-state substance composed essentially of atoms of germanium arranged in a definite geometric pattern or lattice. The flow of current through this substance depends upon our ability to dislodge electrons from the outer shells of the various germanium atoms. Hence, it was first necessary to understand how the electrons are distributed around the nucleus of an atom. Now, with this appreciation, we can turn our attention to additional data concerning electron behavior in atoms.

Electron Behavior in Atoms

It was noted in the preceding discussion that the electrons filled the first shell first, then the second shell, then the third shell, etc. In a stable atom, no electrons exist in an outer shell unless the inner shells are completely filled. (This is true of the first three shells; in the fourth and higher shells, there is less tendency to follow this rigorous pattern, probably because, with increasing distance, the influence of the nucleus on the outer electrons decreases rapidly.) Furthermore, a shell is in its most stable state when it carries a full complement of electrons. For the innermost level, this means 2 electrons; for the next shell, this means 8 electrons; etc. It has been found that elements whose outer ring is not complete are more chemically active than elements whose outer ring is complete. A good illustration of the former type of element is sodium. The atomic number of this element is 11, and the two inner rings or shells are each filled to capacity (i.e., 2 and 8 electrons, respectively). The third ring contains 1 electron, and since sodium would be more stable without any electrons in this third ring, we find that sodium is chemically active because it readily loses or gives up this sole electron. As a matter of fact, sodium is so anxious to give up its extra electron that it is never found by itself in nature. A suitable element that combines readily with sodium is fluorine to form sodium fluoride. These two elements react particularly readily with each other because fluorine (atomic number 9) has 7 electrons in its second ring and the addition of 1 electron completes this ring. Thus, one way of looking at this combination is to consider the sodium atom as giving up its lone third-ring electron to form a stable atom having two completed rings or shells while fluorine receives this extra electron and it, too, forms two complete shells.

Another explanation of this combination is to say that sodium and fluorine share these additional 8 electrons, permitting each to have a completed outer ring. That is, fluorine uses the 1 sodium electron to complete its second shell while sodium uses the 7 fluorine electrons to complete its third shell. The latter view is perhaps the more realistic one, since obviously the sodium and fluorine atoms (in the sodium fluoride salt that is formed) do not leave one another but rather coexist in a crystal structure. This view will also provide a better insight into transistor action, because electron sharing is an integral part of the germanium crystal.

From this "desire" on the part of an atom to attain a state in which its outer shells possess a complete complement of electrons, we can also surmise that atoms in which this is naturally true will be extremely stable and "well-satisfied" and will not enter *easily* into chemical combination with other elements. There are six such elements: helium, neon, argon, krypton, xenon, and radon. In helium, the first ring is complete with 2 electrons. In neon, the next heaviest of these atoms, there are 2 electrons in the first shell and 8 in the second shell. Argon, the third inert element, has 18 electrons divided into three shells. The first shell is complete with 2 electrons, the next shell has its full quota of 8, and the third shell has 8 electrons. Since the maximum capacity of the third shell was previously indicated to be 18, it would appear that in argon this shell is not complete. This is not true, however, because it was found that all shells beyond the first one can be divided into subshells, Fig. 1·6. For the second main shell we have two subshells, one holding a maximum of 2 electrons and one a total of 6 electrons. Note that the total previously given for this second shell, 8, still holds, but the 8 electrons are divided into two subshells of 2 and 6.

For the third main shell, the maximum number of electrons is 18, divided into three subshells of 2, 6, and 10. In argon, the first two subshells are complete, and this has the same stabilizing effect as though

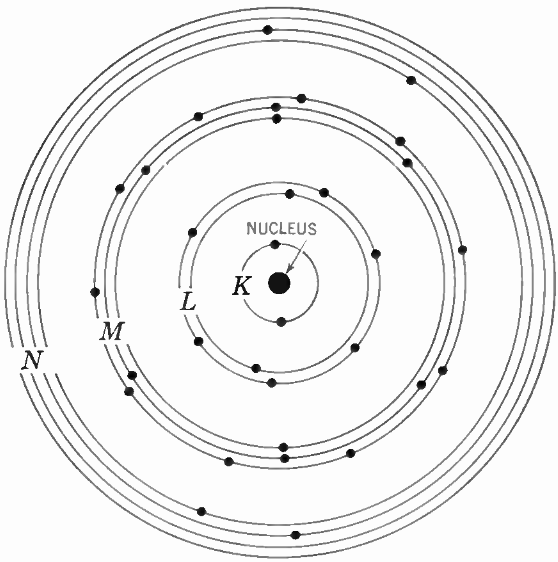


Fig. 1·6 All energy levels or shells beyond the first, the *K* shell, are divided into subshells. Each group of shells is designated by a different letter beginning with *K* and followed by *L* to Q. The first subshell of any group is given the letter *s*; the second subshell, *p*; the third subshell, *d*; and the fourth subshell, *f*. It is thought that no known element possesses more than four subshells. The arrangement shown is for the element germanium.

we had completely filled all three subshells. [This behavior also explains why sodium is "satisfied" when it shares its electron with fluorine's 7 electrons in its (sodium's) third shell. What happens is that the first and second sublevels of the sodium third shell are filled.]

The atomic number of inert element krypton is 36, and here all sublevels of the first three main shells are filled with electrons. In addition, two sublevels of the fourth main shell are filled with electrons.

The chemical behavior of any element, then, is directly related to the number of electrons contained in its outermost atomic shell and how close this shell comes to being filled. If the shell (or one or more of its sublevels) is filled, the atom needs no additional electrons and therefore has no tendency to enter into chemical combination with other elements. If the shell is not filled, the atom does seek other electrons and we say it is chemically active. Because of the importance of these outermost electrons, they are given the special name of valence electrons. Further, the shells or energy levels they occupy are called valence levels.

Chemical activity can also be related to electrical conductivity, because an atom whose outer ring is filled shows no tendency to part with any of its electrons. Since we need free electrons to obtain an electric current, the inert elements are insulators. On the other hand, atoms which part easily with an electron in order to end up with a complete shell make good conductors. Copper, for example, has an atomic number 29; this means that all subshells of the first three main shells are completely filled (i.e., 2, 8, and 18), leaving 1 electron for the first subshell of the fourth ring. It is fairly easy to take this electron away from copper, and hence copper is an excellent conductor of electric currents.

Note that if we were to attempt to take away more than 1 electron from an atom of copper, we would have to apply a considerably greater amount of energy. This is because the second electron would have to come from a complete ring, and since this is a particularly stable condition, the atom would resist the removal with tenacity. If we applied enough energy, however, the electron removal could be accomplished.

Electron Removal

When one or more electrons are removed from an atom, that atom is no longer electrically neutral in charge. If we remove 1 electron, the nucleus has one more positive charge than the outer electrons and the overall charge is +1. The atom has now become what is known as an ion; in the present illustration, a positive ion.

It is also possible for an atom to gain an electron, and when this occurs, the overall electric charge becomes −1. The atom in this instance becomes a negative *ion*. Situations such as this occur when an atom needs one more electron to complete a ring. It then attempts to obtain this additional electron from some other element, particularly one possessing a lone electron in its outermost ring.

Energy is required to remove an electron from an atom. That is, atoms do not part with any of their electrons unless they are forced to do so, and one way to pry an electron loose is to provide it with enough energy to escape from the attractive force of the positive nucleus. Commonly used forms of energy, particularly in electronic devices, are electric fields, heat, light, and bombardment by some other particle. In wires, for example, we force the copper atoms to give up an electron each by applying an emf across the ends of the conductor, Fig. 1·7*a*. In a vacuum tube, we heat a cathode until the outer-ring electrons have absorbed enough energy to escape from their respective atoms and leap into the interelectrode space, Fig. 1·7*b*. There they

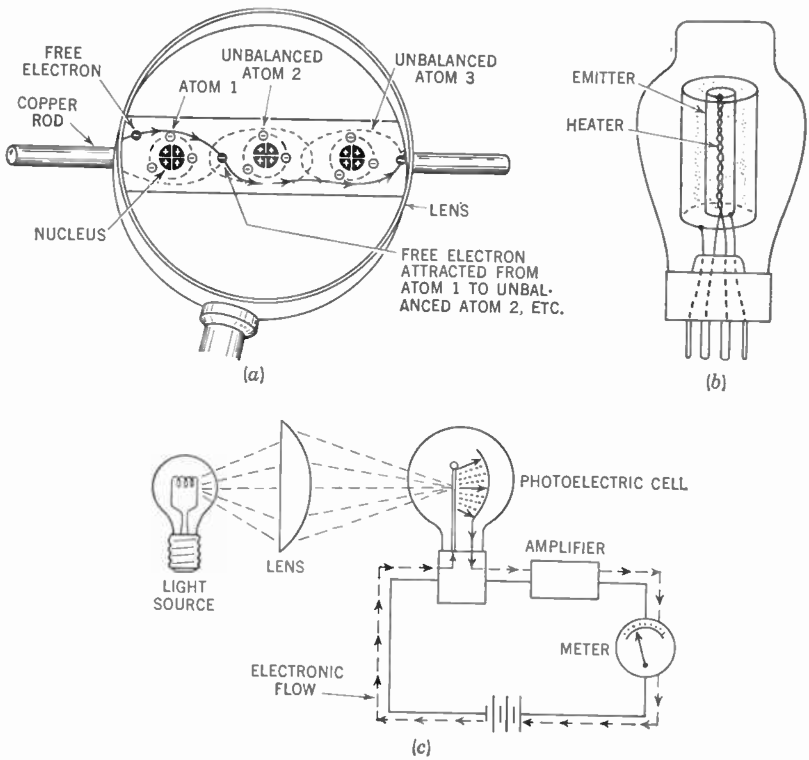


Fig. 1·7 Several common methods of separating electrons from their atoms.

are attracted by a positive emf on the plate, and thus we get a flow of current through the tube and the outer circuitry of the system. Light, as an activating agent, is employed in photoelectric tubes; the energy which light rays bring to the atoms of the photoelectric cell enables some of the electrons to escape from their atoms and again reach a positively charged anode, Fig. 1·7*c*.

Finally, bombardment to produce ions is the basis of operation of gaseous tubes such as the thyratron. Thus, ionization is important to the electronics industry. And ionization, no matter where it occurs, is due to a transfer of energy from one substance to another.

The Quantum Theory

The mechanism by which bound electrons are freed has been the subject of a considerable amount of investigation, and certain facts which are important to us in our study of transistors have been discovered. For example, it has been revealed that when we supply energy to an electron held in an atom, we must supply a definite amount of energy in order for it to have an effect on the electron. The various shells in an atom represent definite energy levels, and in order to move an electron from a lower shell (or subshell) to a higher shell, a certain amount of energy is required. Failure to provide enough energy to the electron will cause it to remain at its present level. This is true even if the energy provided is just barely shy of the required amount.

If more than enough energy is provided for the electron to leave its orbit and move to the next higher level, then the excess will be to no avail *unless* enough extra is provided to enable the electron to move to a still higher shell. In other words, energy is required in definite, discrete amounts called quanta, and the electrons can receive these quanta only in whole numbers, such as 1, 2, or 3 quanta.

Electrons can lose energy as well as receive it, and when an electron in an atom loses energy, it moves to a level which is closer to the nucleus. This lost energy may appear as heat, as in a conductor when current is passed through it, or as visible light, as in a gaseous tube. In the latter devices, the light emitted by the electrons of the gas molecules produces a visible haze. In fact, one of the ways of recognizing gas in a vacuum tube is by the bluish light which surrounds the filament. This merely represents electrons returning from some outer shell to an inner shell.

Different elements have different energy levels for their electrons, and, consequently, the amount of energy absorbed or released varies as the electrons move from level to level. This accounts for the different-color light emitted by various substances when they are excited. Heating sodium over an open flame produces a characteristic yellow light; neon gas, when activated in electric signs, emits an orange-red glow. The energy required to produce red light is less than the energy needed to produce blue light. This is because the energy in a quanta bundle depends on frequency and red has a lower frequency than blue.

In gases, electrons of one atom tend to act independently of the electrons of other atoms. In solids, however, the forces which bind atoms together greatly modify the behavior of the associated electrons, and here we are dealing with the aggregate action of many electrons rather than with individual electrons. One direct consequence of the close proximity of atoms in a solid is to cause the individual energy levels which exist for an isolated atom (depicted in Fig. 1·6) to break up to form bands of energy levels. Within the bands, discrete permissible energy levels still exist, but the act of bringing many atoms close together has produced many more permissible energy levels. It has also caused some energy levels to disappear.

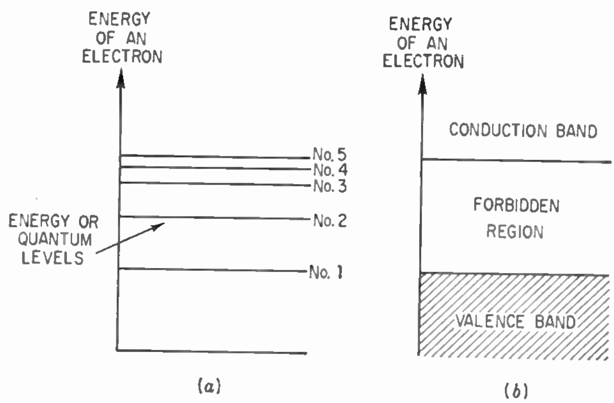


Fig. 1·8 (a) Energy levels in an isolated atom. (b) Energy bands in a solid.

Thus, where before, in an isolated atom, we had an energy-level arrangement such as that shown in Fig. 1·8*a*, in a solid material we find bands of energy as shown in Fig. 1·8*b*. As a matter of fact, only the three upper bands are shown in Fig. 1·8*b*. Additional energy bands exist below the valence band, but since these are not important to an understanding of semiconductor behavior, they will not be shown or discussed.

The uppermost band of Fig. 1·8*b* is the conduction band. When electrons are found there, they can be readily removed by the application of external electric fields (developed by the application of electric voltages). When a material has many electrons in the conduction band, it acts as a good conductor of electricity.

Below the conduction band is a series of energy levels that collectively form the forbidden band. Electrons are never found in this band. Electrons may jump back and forth from the bottom valence band to the top conduction band, but they never come to rest in the forbidden band.

The valence band is formed by a series of energy levels containing the valence electrons. These electrons are more or less bound to the individual atoms; hence, their range of movement is somewhat restricted, certainly far more so than the range of movement of electrons in the conduction band. It is possible to move electrons from the valence band to the conduction band by the application of energy, generally thermal energy. This movement happens to a certain extent in semiconductors, and we shall be covering this point more extensively in subsequent discussions. Of more immediate interest is the fact that the extent of the forbidden band, or the separation between the conduction and valence bands, will determine whether a substance is an insulator, semiconductor, or conductor. Figure 1·9 shows the difference between insulators, semiconductors, and conductors in terms of their three bands. Figure 1·9*a* is called an "insulator" because of the wide extent of the forbidden band. The wider this band, the greater the amount of energy which must be fed to any electron in the valence band in order to bring it up to the conduction band where it can be employed as a carrier of electricity. Obviously, in an insulator a lot of energy is required to get even a minute amount of current through the substance.

In a semiconductor, the extent of the forbidden band is smaller, Fig. 1·9*b*, which means that less energy need be fed to the electrons in the valence band in order to bring them through the forbidden band and into the conduction band. Hence, in semiconductors, more current will flow for a certain applied voltage, although this current will not be as large as we would obtain in a conductor.

The third illustration, Fig. 1·9*c*, is for a conductor. Here we see that the valence and conduction bands overlap. It now takes a very small amount of energy to move electrons into the conduction band, and, consequently, electricity is readily passed by conductors. All this, of course, is common knowledge; what is not so universally known is how the various energy levels within a molecule cause the molecule to act as an insulator, semiconductor, or conductor. That is where the quantum theory so admirably fills in the gaps.

The side axis of each of the three illustrations in Fig. 1·9 is labeled simply as "Energy." It is the generally accepted practice for physicists to use electron volts as a convenient measure of energy. One electron volt is the energy acquired by an electron in falling through a potential difference of 1 volt. If we use this method of measuring energy, then the width of the forbidden band in an insulator is 1 ev (electron volt) or more. For a semiconductor such as pure germanium, the width of this band is 0.7 ev; for silicon, another semiconductor, it is 1.1 ev; and for conductors, where the forbidden band is absent, we need perhaps 0.01 ev to bring an electron into the conduction band.

It should be understood, of course, that each of the three illustrations in Fig. 1·9 is representative of a class of materials and there will naturally exist many substances whose characteristics fall somewhere in between those of insulators and semiconductors on the one hand and semiconductors and conductors on the other. Germanium, for

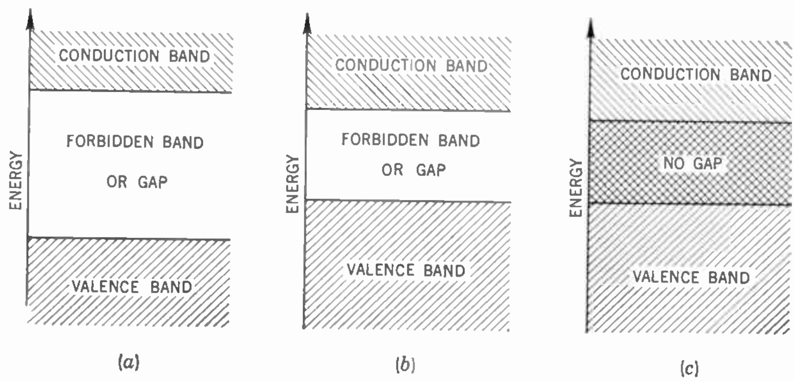


Fig. 1·9 The difference between insulators, semiconductors, and conductors in terms of their valence bands, forbidden bands, and conduction bands. (a) Insulator. (b) Semiconductor. (c) Conductor.

example, in the highly purified state, is a very poor conductor of electricity. As we add certain controlled amounts of impurities to germanium, however, we find that its conduction increases materially, indicating that while a substance may be basically an insulator, its properties can be altered. The alteration is actually what happens in germanium and silicon transistors and is responsible for the ability of these units to function usefully. In Chap. 2, we shall see just what is the effect of adding selected impurities to germanium and silicon.

QUESTIONS

1·1 What advantages that transistors possess make them especially attractive for communications applications?

1·2 Differentiate atoms, molecules, and elements. Name 15 elements that you have personally come in contact with, either singly or in combination.

1·3 Describe the structure of an atom in general terms. Compare this structure with that of the solar system.

1·4 How are the electrons arranged within an atom? Consider first a simple element, then a fairly complex element.

1·5 What differences exist between the nucleus of a simple atom, such as hydrogen, and the nucleus of a complex atom?

1·6 What causes an element to be chemically active? Stable?

1·7 Is there any apparent relationship between the chemical activity of an atom and its electrical conductivity? Explain.

1·8 What happens when an atom gains an additional electron? Loses an electron? What is the altered atom called?

1·9 What methods may be used to remove electrons from an atom? Describe one or two methods in detail.

1·10 What is the quantum theory with respect to electron removal?

1·11 Define the valence band, conduction band, and forbidden band.

1·12 Explain the difference between conductors, insulators, and semiconductors in terms of energy bands.

CHAPTER 2

Semiconductor Diodes and Transistors

IT WAS NOTED in the preceding chapter that the chemical activity of an atom is determined primarily by the number of electrons contained in the outermost ring of the atom. When this ring is filled, the atom is stable and shows little inclination to combine with an atom of any other element. The activity increases, however, when the number of electrons is less than the full number needed to complete a ring.

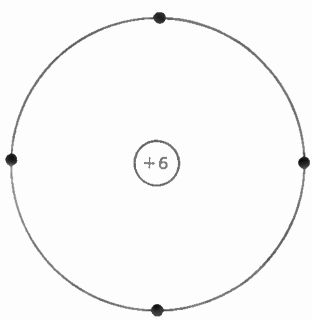


Fig. 2·1 A simplified illustration of the carbon atom. Only the valence electrons are shown.

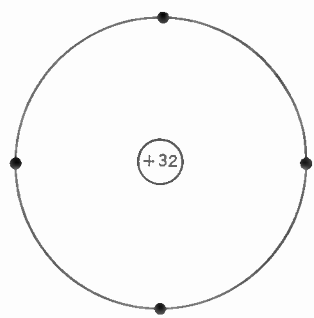


Fig. 2·2 The germanium atom, using the simplified method of presentation.

Because of the importance of these electrons, they are given the special name of valence electrons. Furthermore, it is common practice in illustrations of atoms to show only the valence electrons, Fig. 2·1. The carbon atom, with an atomic number 6, is shown here. The +6 at the center represents the nuclear charge. Since there are 4 valence electrons, we know that these electrons are in the second ring; the 2 electrons not shown would be in the first ring.

Of immediate interest in transistors is germanium, and this element, too, contains 4 electrons in its outermost ring. The atomic number of germanium is 32, giving us three completed shells of 2, 8, and 18 electrons each and 4 electrons in the fourth shell. The latter electrons are the valence electrons, and they are the ones represented in Fig. 2·2.

Lattice Structure and Crystals

Germanium in the solid state possesses a crystalline structure in which a group of germanium atoms combine, through their valence electrons, to form a repeated structure having a number of basic

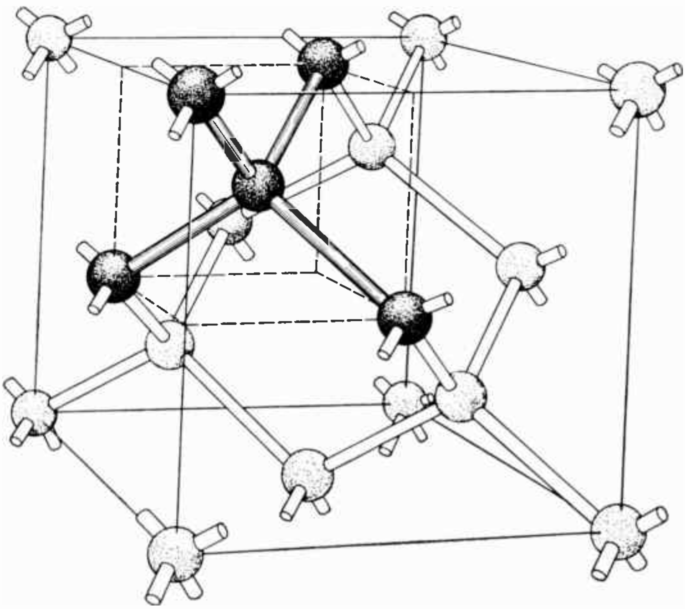


Fig. 2·3 The crystal lattice structure of germanium. Each sphere represents a nucleus surrounded by its inner shells. The spokes that join the atoms and support the structure represent the covalent bonds. (After W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Company, Inc., Princeton, N.J., 1950)

cubical lattices such as shown in Fig. 2·3. Each of the "balls" in the illustration represents a germanium atom; the rods between the balls represent the electronic forces binding each atom to its neighbors.

This cubical configuration, known as a diamond structure, is characteristic of the solid state of a number of elements, among them carbon, silicon, and germanium. A large, visible crystal of germanium would be composed of millions upon millions of these basic cubical lattices.

A two-dimensional illustration of the manner in which the germanium atoms are bound to one another is given in Fig. 2·4. Focusing our attention on any one of the central atoms, we see that each of its 4 electrons is shared by four other germanium atoms. This gives the central atom a total of 8 electrons in its outermost ring; 4 of these electrons are its own, and the other 4 it "borrows" from the surrounding atoms. Since 8 electrons in this subring provide for a stable arrangement, the germanium crystal forms a stable compound.

What is true of the central atom is true of all its neighbors: each shares its four outer electrons with four other germanium atoms. This, too, is shown in Fig. 2·4 for a limited number of atoms. All the valence electrons are tightly held together. Consequently, pure germanium is not a very good conductor of electricity. A good conductor

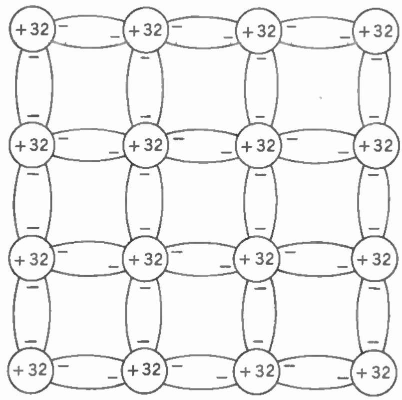


Fig. 2·4 Representation of electron-paired bonds in germanium.

would require an abundance of free electrons, and as we see, all electrons in a germanium crystal are held fairly tightly because, in combination, they tend to complete the outer subrings of the germanium atoms.

The sharing of the valence electrons between two or more atoms produces a shared or covalent bond between the atoms. It is this bond which is largely responsible for the cohesion which the crystal structure possesses and which actually keeps the crystal structure intact.

Note that by this sharing to form complete valence rings, germanium in its pure form is a fairly good insulator. That the substance is not a complete insulator, but rather a semiconductor, stems from the fact that thermal agitation, arising from the energy imparted to the electrons by the heat of its surroundings, causes an electron here and there to break away from its bond, move up into its conduction band, and wander through the crystal lattice structure in a more or less aimless manner.

Electrons and Holes

The bond from which an electron escapes is left with a deficiency of 1 electron, and hence we should find there a positive charge of 1. This electron deficiency has been given the rather descriptive name of "hole," as though a physical hole had actually been left by the removal of the electron.

We have been conditioned by our previous training to accept the fact that electrons are quite mobile and may be moved readily from point to point. Numerous tests have been performed in which this fact has been demonstrated, and in our own experience with electric circuits we have never encountered any action which would cause us to think otherwise. Hence, when the statement is made, as it was above, that the freed electron wanders aimlessly through the crystal, every reader will accept it without question.

It is, however, also valid in this case to state that the hole left by the electron will wander about within the crystal structure, and on this point we would run into a general raising of eyebrows. And yet it has been conclusively demonstrated that holes do travel through germanium crystals. In fact, the concept of hole travel is basic to an understanding of transistor operation and hence warrants a more detailed description.

When a bound electron departs, the charge deficiency, or hole, that it leaves behind is confined to the valence ring of the atom. If, now, a nearby electron held in a covalent bond acquires enough energy to leave its bond and jump into the waiting hole, then in essence what we have had is a shift in position of the positively charged hole from its first position to this new position, Fig. 2·5. This same action can occur a number of times, with successive changes in hole position, so we can very well state that a hole drifts about in a random manner in exactly the same fashion as the electron which left the hole originally.

The foregoing discussion has dealt with a single electron and a single hole, but in actual crystals there would be many such pairs. And with many negative electrons and positive holes present, a considerable number of recombinations will be taking place all the time. By the same token, the energy (be it heat or light or an electric field) being supplied to the crystal will constantly be breaking other bonds. Eventually, a dynamic equilibrium will be attained in which the number of bonds being broken will equal the number being re-formed.

If the energy supplied to the crystal is an electric field developed by the application of an emf across the germanium crystal, then the motion of the electrons and the holes will be less random and more directed in a direction determined by the voltage. Electrons will move toward the positive terminal of the battery, while the holes will drift toward the negative terminal of the battery. The opposite flows of these two charges do not, as one might suppose, cancel each other Rather they aid each other. This was demonstrated in an experiment performed in 1889 by the physicist H. A. Rowland. On an ebonite disk he placed negative charges of static electricity, separated by raised portions of the disk. When the disk was rotated at high speed, a magnetic field was produced, and the field was identical with what

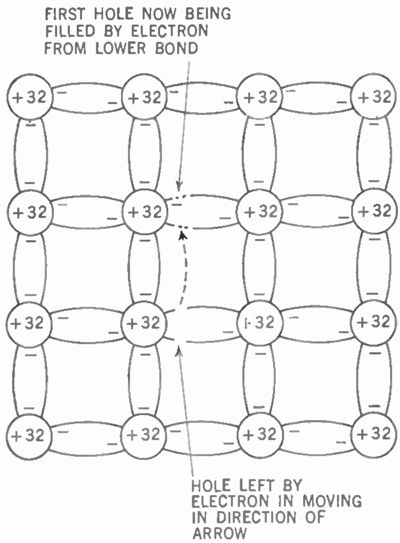


Fig. 2·5 Method by which holes travel through a germanium lattice structure.

would have been expected if a flow of electrons had taken place in a loop of wire in the same direction of rotation.

Rowland then removed the negative charges and replaced them by an equivalent number of positive charges. The disk now was rotated in the opposite direction, and the resulting magnetic field had exactly the same direction as its predecessor. Thus, we obtain the same electrical effect whether we have negative charges (i.e., electrons) moving in one direction or equivalent positive charges (i.e., holes) traveling in the opposite direction. Ohm's law or any other electrical law we know would yield identical results in either case. This is a significant fact to remember, because all our electrical studies have emphasized electrons and electron flow, and the idea of mobile positive charges will come as a surprise. It is particularly important to appreciate both types of current flow, since both occur in transistors.

It is interesting to note that the rate at which electrons and holes diffuse through fairly pure samples of germanium has been determined to be as follows: for electrons, 3,600 cm^2 per volt-sec; for holes, 1,900 cm^2 per volt-sec. These rates are at a temperature of 27°C and with no applied emf. When a voltage is applied, the rates increase considerably.

N-type Germanium

Externally applied heat and light may be used to produce free electrons and holes in a germanium crystal, but a much more efficient method of achieving the same result is to add exceedingly small

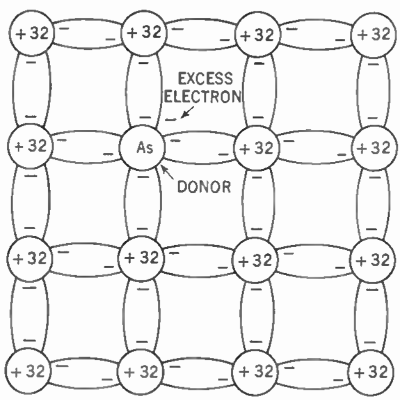


Fig. 2·6 The effect of an arsenic atom replacing a germanium atom.

amounts of selected impurities, generally to an extent no greater than 1 part in 10 million. Impurities frequently employed are arsenic and phosphorus. The impurity enters the crystalline structure of the germanium and takes the place of single germanium atoms at various points throughout the crystal lattice structure. A diagrammatic representation of this condition is shown in Fig. 2·6. Arsenic has 5 valence electrons, and 4 of them enter into covalent bonds with four surrounding germanium atoms. This is in accordance with the structural arrangement in a germanium atom. The fifth electron is simply held in place by the positive attractive power of the arsenic nucleus. However, since the arsenic atom is sharing its other 4 electrons with four other germanium atoms, in essence it possesses the equivalent of 8 electrons. Since all that the arsenic atom needs for a stable arrangement is 8 electrons, the force with which the extra (i.e., the fifth) electron is held is fairly weak and the bond between this electron and the atom is easily broken. Thus, by the addition of minute quantities of arsenic to the germanium structure we have, in effect, provided the germanium with a source of free electrons. Substances like arsenic or antimony which serve as sources of electrons are called donor impurities. Furthermore, the germanium crystal containing these donor atoms is known as N-type germanium. The N, of course, refers to the fact that the electrical conduction through the crystal is done by electrons, which possess a negative charge.

P-type Germanium

It is also possible to add impurities which possess 3 rather than 5 electrons in their outer orbit. Boron, gallium, and indium are ex-

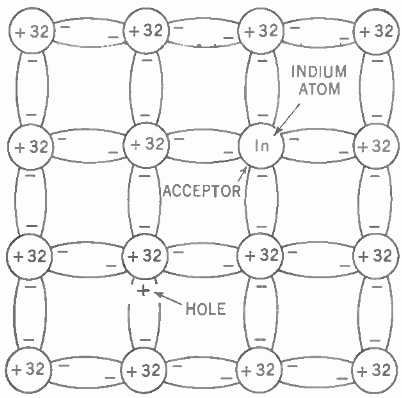


Fig. 2·7 **A hole is produced when an atom of a trivalent impurity such as indium replaces a germanium atom.**

amples of such substances. As with the arsenic, each trivalent impurity atom will replace a germanium atom in the lattice structure, Fig. 2·7. However, in this case, instead of having an excess of 1 electron, we now find ourselves with a deficiency of 1 electron. In order to complete the four electron-pair bonds, the trivalent atom "robs" an electron from a nearby germanium bond. The net result of this robbery is to leave a hole in the neighboring electron-pair band.

Thus, when the impurity added to the germanium crystal structure has only 3 valence electrons, a series of holes is produced. Under the stress of an applied emf, electrons from other nearby bonds will be attracted to these holes, thereby filling the gaps but creating a similar number of holes in their former bonds. Thus, we have the equivalent of a movement of holes through the crystalline structure, and conduction is said to take place by holes.

Trivalent impurities which create holes are known as acceptor impurities, and the germanium crystals which contain these substances are known as P-type germanium. Thus, by the careful selection of the impurity to be added, we can determine whether the germanium is of the N or P type. Both are employed in transistors, and it is important that the reader understand the differences between them and how electrical conduction occurs through each.

It should be noted that a number of holes are present in N-type germanium because of the normal breaking of bonds arising from the absorption of heat or light energy. This is exactly similar to the action in pure germanium. However, the electrons released because of the addition of arsenic or other pentavalent impurity atoms are, by far, the principal conductors of electricity in N-type germanium. By the same token, free electrons exist in P-type germanium, but again, it is the holes created by the addition of trivalent-impurity atoms that account for the major portion of the electrical conduction that takes place here. The holes in N-type germanium and the electrons in P-type germanium are called minority carriers.

The impurities must be added in carefully controlled amounts; otherwise, the germanium crystal structure is modified to such an extent that transistor action is not obtained.

The reader may wonder what would happen if both acceptor and donor impurities were added to a slab of germanium. The holes created by the acceptor atoms would be promptly filled by the extra electrons of the donor atoms. If both impurities were present in equal amounts, the excess electrons would just fill the excess holes and the germanium would act as pure germanium containing no impurities (provided the amount of impurities added were minute in quantity). On the other hand, if one impurity were present in greater amount, the electrons or holes it provided would become the principal carriers of electricity.

We might pause here for a moment and note what effect the addition of impurities has on the energy-band distribution in the atomic structure of a crystal lattice. Before the impurity addition, the energy bands were depicted as shown in Fig. 2·8*a*. These were the valence band, the forbidden band, and the top conduction band. If, now, we add minute amounts of a donor impurity, we find that extra energy levels which were not present in the pure semiconductor are developed. The extra electron added now establishes an energy level in the forbidden band just below the conduction band, Fig. 2·8*b*. The position would appear to be logical, since this extra electron is only loosely held by its parent atom and it should therefore be relatively easy to remove from that atom. That would bring its position close to the conduction band. It would not be completely in the conduction band, since that would assume it was completely free of any influence exerted by its parent atom.

The amount of energy required to raise the electrons provided by the impurity atoms is on the order of 0.05 ev. This is far less than the 0.7 ev needed to bring an electron from the valence band of a germanium atom to the conduction band. The difference is even greater in pure silicon, where the requisite amount of energy is 1.1 ev.

Note that the forbidden band of Fig. 2·8*a* exists only in pure semiconductor. When an impurity is added, the extent of this band is modified as indicated in Fig. 2·8*b*.

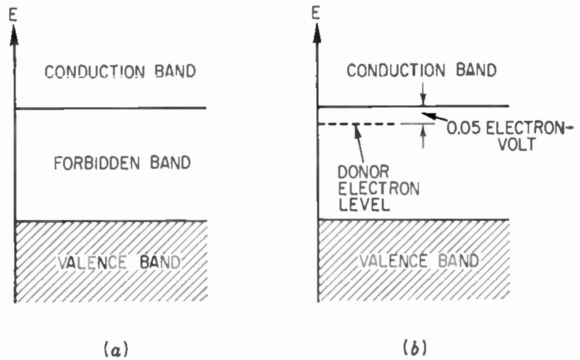
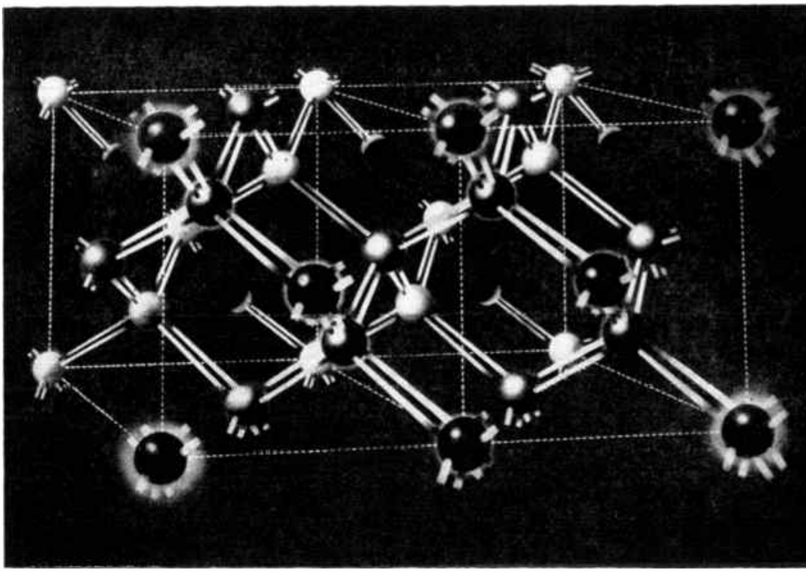


Fig. 2·8 (a) Energy levels in a pure semiconductor. (b) Energy levels in a semiconductor with a donor impurity.

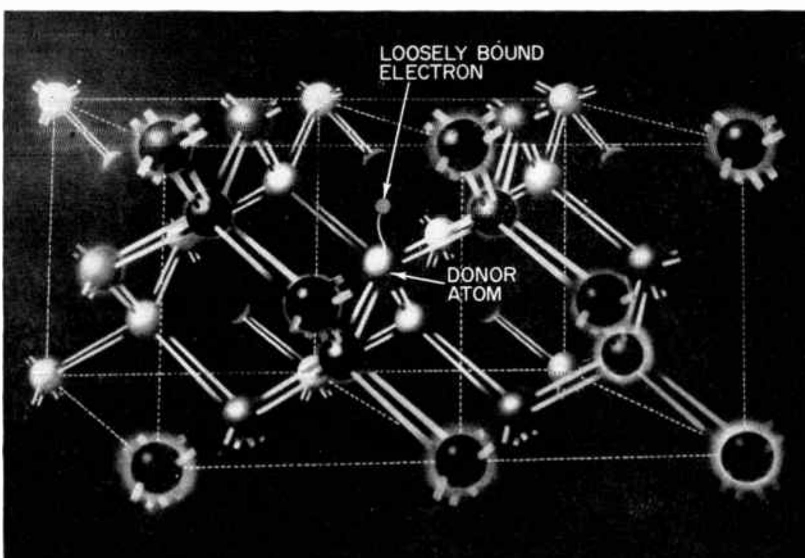
Through a similar process, addition of a P-type impurity will also modify the energy-level distribution of a semiconductor, again making it easier to bring about conduction. Thus, with impurities, the conduction provided by a semiconductor is overwhelmingly taken care of by the extra electrons or holes provided by the impurities. Very little current is contributed by the breaking of covalent bonds. When a semiconductor does not contain any impurities, it is said to be in the intrinsic state, as shown in Fig. 2·9*a*.

PN Junctions

If we take a section of N-type germanium and a similar section of P-type germanium and join the two together, as shown in Fig. 2·10*a*, we obtain a device which we know as a germanium diode. The N-type germanium is at the right, and the P-type is at the left. The circles at the right with the positive sign represent the donor atoms. They possess positive signs because their fifth electrons have been

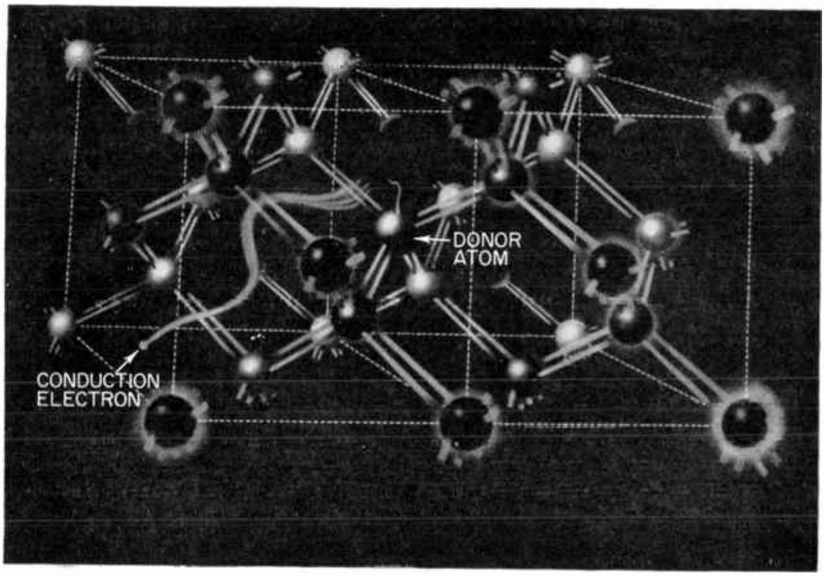


(a)

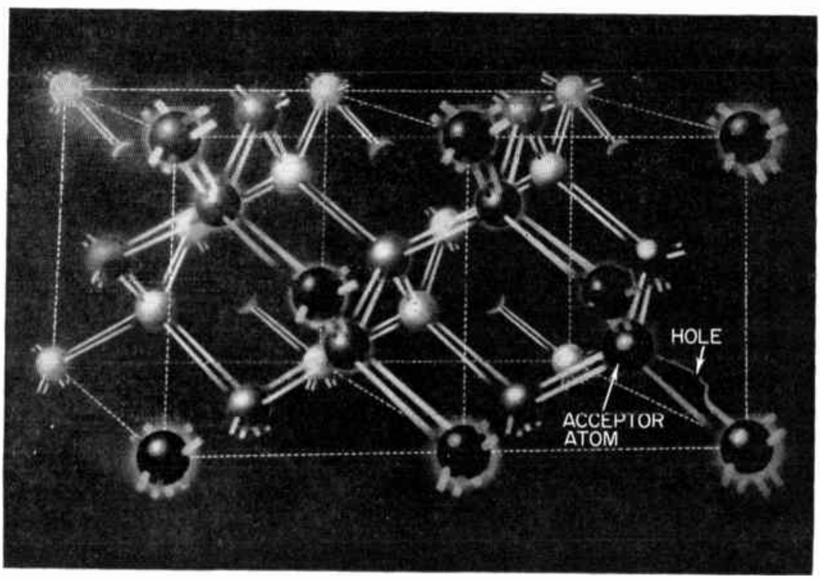


(b)

Fig. 2·9 A pictorial representation of germanium lattice showing shared electron bonds and the manner in which holes and electrons travel from point to point within the crystal structure. (a) Pure germanium crystal. (b) Donor atom substituted in germanium lattice. Note loosely bound extra

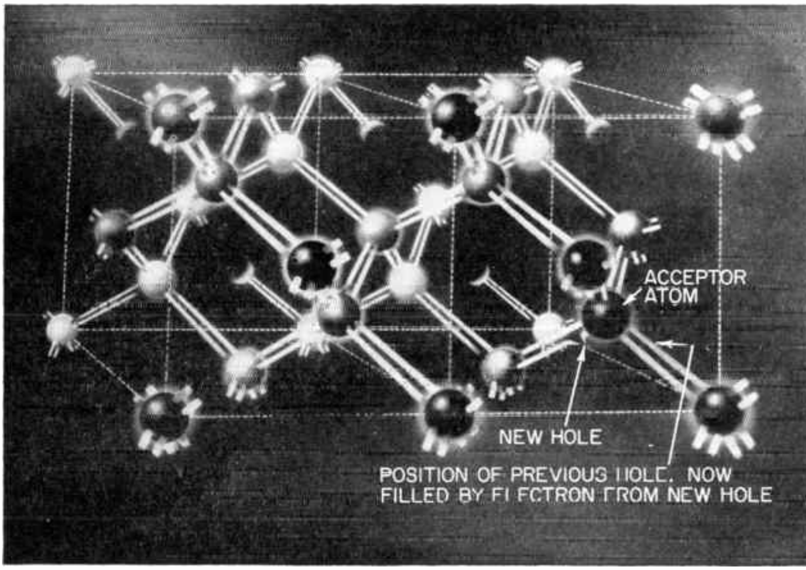


(c)

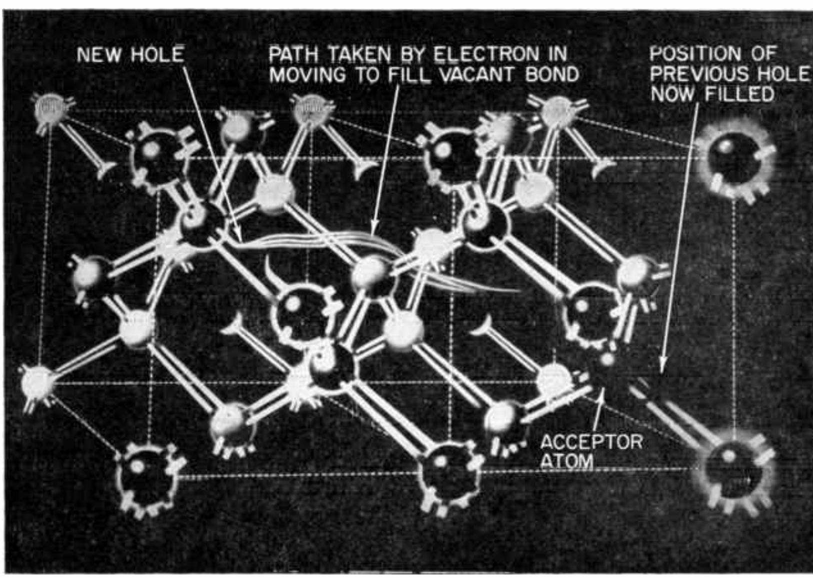


(d)

electron. (c) N-type conduction; electron free of donor atom. This may occur because of heat or electric energy. (d) Acceptor atom substituted in germanium lattice. Note unfilled band in lower right-hand corner of illustration.



(e)



(*f*)

Fig. 2·9 (*continued*) (e) Filling vacant bond by thermal excitation of electron from nearby bond. (*f*) P-type conduction. Hole migrating through lattice by excitation of electron from bond to bond. (General Electric Company and *Electrical Engineering*) removed, leaving each of the atoms with a +1 charge. The free electron is indicated by the negative sign.

By the same reasoning, an acceptor atom in the P-type germanium is represented by a circle with a negative sign, the latter being due to the presence of the additional electron which was "robbed" from a neighboring electron-pair bond. The hole left by this electron is represented by a small plus sign.

When these two germanium sections are joined together, one might suppose that all of the excess electrons on the right would immediately cross the junction and combine with the excess holes on the left. This action actually starts to occur, but, before it can progress very far, it is brought to a halt. Here is why it is. When the sections

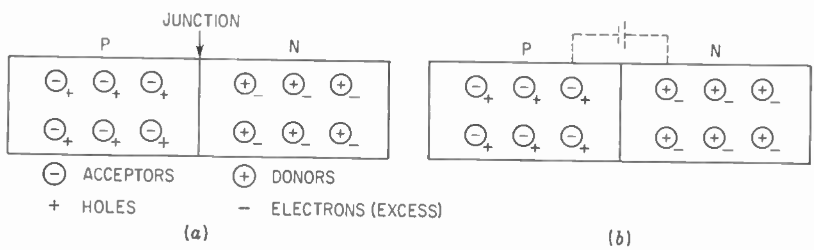


Fig. 2·10 (a) A PN junction forming a familiar germanium diode. (b) The battery drawn across the PN junction represents the restraining forces present at the junction.

are initially brought together, some of the electrons in the N-type germanium cross the junction and combine with a corresponding number of holes in the P section. Since the N section was initially electrically neutral, loss of some of its electrons leaves it with a net positive charge. This charge increases as the number of departed electrons increases, and a point at which no additional electrons are capable of overcoming this positive force is quickly reached.

By the same token, the P section itself was also electrically neutral at the start. When it loses some of its positive holes through combination with electrons from the N section, it develops a net negative charge. As additional electrons from the N section attempt to approach the junction, not only are they held back by the net positive charge existing in their own section, but they are also repelled by the net negative charge of the P section. Thus, the migratory action is halted.

The region in the immediate vicinity of the junction is called the depletion region because of the absence of any mobile charge there. In the N section, the free electrons initially present were lost, as described above, by combination with the free holes just on the other side of the junction in the P section. Beyond the depletion region, free electrons are present in the N section. These free carriers, however, are restrained from approaching the junction as explained above.

Several pictorial methods have been employed to indicate the restraining forces present at a junction. In one illustration, a small battery is placed across the junction in the manner shown in Fig. 2·10*b*. The negative terminal of the battery connects to the P-type germanium, while the positive terminal of the battery attaches to the N-germanium side. Electrons attempting to travel from the N-germanium side to the P side encounter the negative field of the battery and are repelled. By the same token, holes attempting to move from

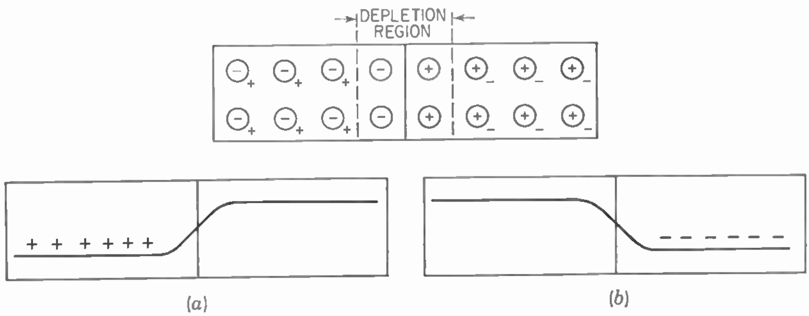


Fig. 2·11 A second method of indicating the forces present at the PN junction which prevent the combination of the holes at the left with the electrons at the right.

the P side to the N side see the positive battery terminal, and they, too, are repelled.

A second method of representation utilizes what are called potential "hills." The electrons on the N side have to climb a negative potential hill in order to reach the P side, Fig. 2·11*b*. The hill, of course, is the repelling force of the acceptor atoms. On the other side of the junction, the holes have to climb a positive potential hill in order to move to the right, Fig. 2·11*a*.

In order to produce a flow of current across the junction, we must reduce the potential hill that exists there. This can be done by applying an external potential across the ends of the two germanium crystals, Fig. 2·12. The negative terminal of the battery connects to the N-type section, and the positive terminal of the battery connects to the P section. The free electrons in the N section are repelled by the negative battery field and move toward the PN junction. At the same time, the positive holes in the P section are forced toward the junction by the repelling force of the positive battery field. If the battery is strong enough, it will ~~lower~~ the potential hill at the junction and enable the carriers to move across to the opposite side. Once the junction crossing is made, a number of electrons and holes will combine. For each hole that combines with an electron from the N-type germanium, an electron from an electron-pair bond in the crystal and near the positive terminal of the battery leaves the crystal and enters the positive terminal of the battery. This creates a new hole

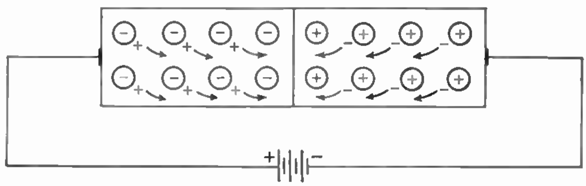


Fig. 2·12 The effect on the PN junction of opplicotion of forword bosis.

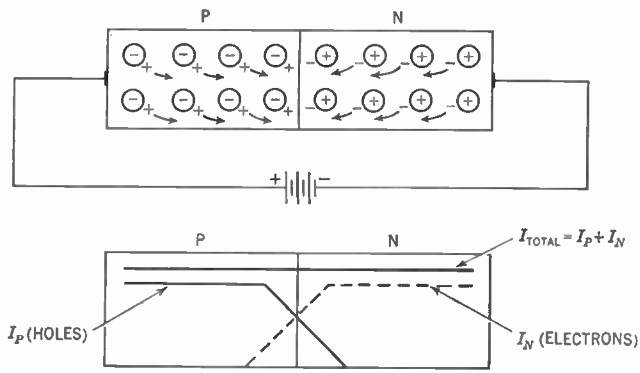


Fig. 2·13 The current flow in the N region is by electrons; in the P region, the current is corried by holes. In the vicinity of the junction, both types of corriers ore present.

which, under the force of the applied emf, moves to the junction. For each electron that combines with a hole, an electron enters the crystal from the negative terminal of the battery. In this way we maintain the continuity of current flow. Stoppage at any point immediately breaks the entire circuit, just as it does in any ordinary electric circuit If this were not so, then electrons would pile up at some point and result in a gradually increasing charge or potential at that point. Since this does not occur, we must treat the circuit operation in the manner which has just been indicated.

Note that current flow in the N region is by electrons; in the P region, the current is carried by holes, Fig. 2·13. As we approach closer to the PN junction, we find both types of carriers. The overall value of current, however, remains constant and is a function of the applied voltage.

As the external voltage is increased, it gradually overcomes the restraining forces present at the junction and the current rises. Once the restraining forces are completely overcome, the current rises sharply, as shown in Fig. 2·14. If the current flow is permitted to reach too high a value, the heat generated will permanently damage the junction and the unit will no longer function in the manner described above.

In the preceding discussion, the diode was biased in its forward, or low-resistance, direction. If, now, we reverse the polarity of the ap-

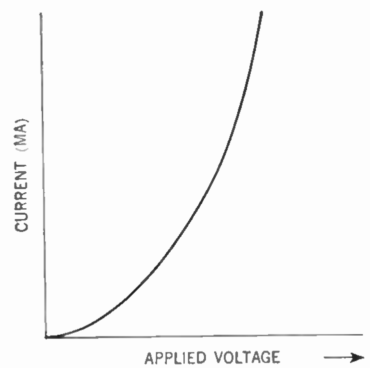


Fig. 2·14 The behavior of the current flow through a germanium diode as the forward voltage is increased.

plied voltage, we find that the battery acts in consort with the potential barrier at the junction and practically no current passes at all. The diode is biased in the reverse direction.

The reason for the current decrease is readily understood. With the negative battery terminal connected to the P-germanium section, the excess holes, with their positive charge, are attracted away from the junction, Fig. 2·15. At the same time, the positive terminal of the battery at the N side attracts the excess electrons away from the junction.

The overall characteristic curve of a germanium diode is shown in Fig. 2·16. The portion of the curve to the right of the zero line represents the forward current, and the reason for the sharp, upward swing of the curve has already been discussed. The swing upward occurs at relatively low voltages, possibly no more than 0.2 to 0.3 volt for germanium and 0.6 to 0.8 volt for silicon. To the left of the zero line we have the diode characteristic under reverse bias. We see that a current does flow, although it is usually on the order of microamperes and remains so until the reverse voltage is brought well above 20 to 30 volts for most diodes.

The sources of this reverse current are the minority electrons and holes which receive enough energy from heat and light reaching the crystal to break their covalent bonds. The reverse bias attracts these

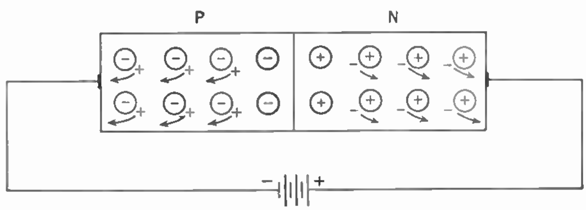


Fig. 2·15 When the battery connections are reversed, the electrons and holes are drawn away from the PN junction.

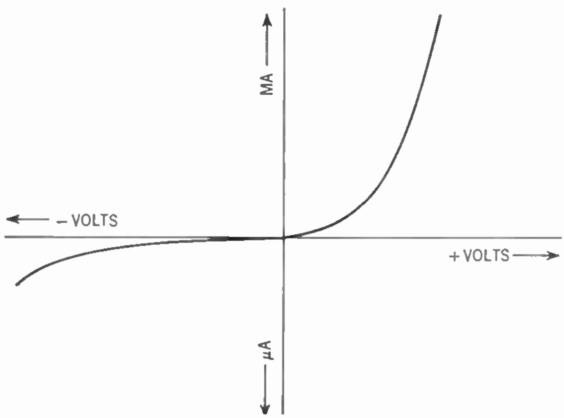


Fig. 2·16 The overall static characteristic curve of a germanium diode. The portion of the curve to the left of the vertical line is shown dropping faster than it should (by comparison with the curve at the right).

electrons and holes, and a minute current flows through the circuit. The electrons of the P-type material travel to the positive battery terminal, and as they enter this terminal (or the lead extending from the terminal), an equivalent number of electrons enter the germanium from the negative side of the battery. Similarly, holes of the N-type material move toward the negative terminal, and when they reach this point, they receive electrons with which they combine. For each such combination, one electron leaves a covalent bond in the crystal near the positive terminal of the battery and enters the battery. This creates a new hole which, under the force of the applied emf, moves across the crystal. The situation here is similar to what it was when the battery voltage was reversed and the majority carriers were involved. Now we are dealing with minority carriers, those electrons and holes that exist on the "wrong" side of the PN junction.

The reverse current is small because the number of minority carriers in each section is small. Actually, together with this reverse current is a minute leakage current that flows along the surface of the diode. This is essentially an Ohm's law phenomenon governed by $E = IR$. The R is the resistance offered by the diode surface to the applied voltage. Further, since the current obeys Ohm's law, it will be present whether the diode is forward- or reverse-biased. However, its effect is more noticeable under reverse-bias conditions because it more closely approaches the reverse current than the forward current in value.

As the value of the applied reverse voltage increases, a point at which there is a sharp increase in current is reached. This steep rise is due to a phenomenon known as avalanche breakdown, in which minority electrons, passing across the PN junction, gain sufficient energy to knock off valence electrons bound to the crystal lattice and raise them to the conduction band. One minority electron may thus free several valence electrons through collision. In turn, these valence electrons may each gain enough energy to free two or more additional electrons, until a considerable flow of current ensues. To sum up, then, germanium diodes offer a relatively low resistance when biased in the forward direction and a very high resistance when biased in the reverse direction.

Avalanche breakdown is sometimes called the Zener effect, after the American physicist Clarence Zener. About twenty years ago, Zener made theoretical investigations of the problem of electrical breakdown in insulators. He came to the conclusion that when the voltage across an insulator rose high enough, electrons could be torn from the valence band and raised to the conduction band fast enough to account for the large breakdown currents. When breakdown was observed in PN junctions, it was felt that it occurred by the same mechanism. Subsequent investigation proved this was not so, but rather that the action described above took place. In spite of this, the breakdown voltage of junctions is often called the Zener voltage rather than avalanche voltage.

Junction Transistors

An NPN junction transistor is formed by placing a narrow strip of P-type germanium between two relatively wide strips of N-type germanium. This is shown in Fig. 2·17. Low-resistance contact is then made to each strip for external circuit attachment. The N-type crystal at the left is called the emitter; the central P-type strip is known as the base; and the end germanium crystal is called the collector. Although these names have no particular significance as yet, they will tie in with the operation of the transistor. Some transistors, particularly those fabricated by the grown-junction method, actually take the physical form shown in Fig. 2·17. Most other junction transistors assume somewhat different forms, although the sequence of emitter, base, and collector remains unaltered. The form shown in Fig. 2·17 lends itself quite well to the present discussion and that is why it is being used.

As with the PN junction diode, the two end sections of the NPN transistor contain a number of free electrons, while the central P

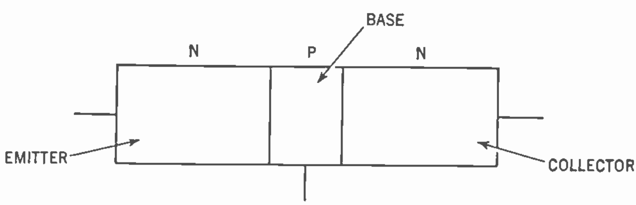


Fig. 2·17 An NPN transistor and the names of its three sections.

section possesses an excess of holes. At each junction, the action that takes place is the same as that previously described for a diode; i.e., depletion regions develop and small potential hills appear. These are then modified by the external voltages which are subsequently applied.

Transistor biasing. To employ the NPN transistor as an amplifier, we would bias the emitter and base sections in the forward, or low-resistance, direction. This is shown in Fig. 2·18*a*. At the same time we would bias the base and collector sections in the reverse, or high-resistance, direction, Fig. 2·18*b*. Both bias voltages are shown in Fig. 2·18*c*. Now let us see what happens under these conditions.

Since the emitter and base sections are biased in the forward direction, current will flow across their junction. Every time an electron from the emitter section crosses the junction and combines with a hole of the base section, an electron leaves the negative terminal of the battery and enters the emitter crystal. The battery cannot continue to supply electrons from the negative terminal without receiving an equivalent number at the positive terminal; hence, for each electron leaving the negative terminal, the positive side receives an electron from the base section. This loss of electrons in the base creates holes which then travel to the junction for eventual combination with electrons from the emitter.

Thus far, of course, we are on familiar ground—ground which was previously explored. The main carriers of electricity in P-type germanium are holes. And this is precisely the situation described above. If the center base section were made quite thick, then practically the entire current flow would occur in the manner described and would

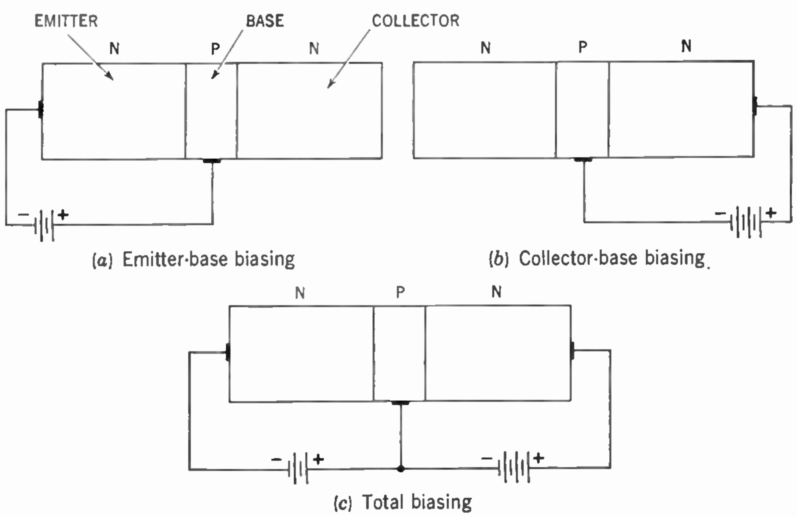


Fig. 2·18 The proper method of biasing an NPN transistor for use as an amplifier.

be confined entirely between emitter and base. There would be little current between base and collector because of the reverse biasing existing there.

When, however, the base strip is made exceedingly thin, transistor amplifying action is achieved. For with the base thin, electrons leaving the emitter pass right through the base section and into the collector region, where they see a positive attractive force that impels them on. Thus, they travel through the collector section and around the external circuit back to the emitter again, completing their path of travel.

At this point the reader may wonder why the emitter current flows through the collector when it was specifically stated that the collector was biased in the reverse, or high-resistance, direction. If we disregard the base for a moment and simply consider the path from the emitter to the collector internally and from the collector to the emitter externally, we see that the two bias batteries are connected series aiding. Thus, any emitter electrons that pass through the base region without combining with the holes present there will find the attracting force of the collector battery urging them on through the collector section. The reverse biasing between collector and base does not affect the emitter electrons that pass through the base and reach the collector.

With the base strip made very thin, the number of combinations between emitter electrons and base holes will be quite small, probably no more than 2 per cent of the total number of electrons leaving the emitter. The remaining 98 per cent of the electrons will reach the collector strip and travel through it. Thus, while the number of electrons leaving the emitter is a function solely of the emitter-base voltage, the element which receives most of this current is the collector. The analogy here to vacuum-tube behavior is very marked. In a tube, the number of electrons leaving the cathode (i.e., the emitter) is governed by the grid-to-cathode voltage. However, it is the plate (i.e., collector) which receives practically all these electrons. In a tube, the amount of current flowing is regulated by varying the grid-to-cathode voltage. In a transistor, the emitter-collector current is varied by changing the emitter-base voltage.

Note, too, that because the base current is very small, a change in emitter bias will have a far greater effect on the magnitude of the emitter-collector current than it will on base current. This also is desirable, since it is the current flowing through the collector that reaches the output circuit. (By the same token, it is the current flowing through the plate circuit in a tube that is important.)

Transistor gain. We achieve a voltage gain in the transistor because the input, or emitter-to-base, resistance is low (because of the forward biasing between the two elements), whereas a high load resistance can be used because the collector-to-base resistance is high (owing to the reverse bias there). A typical value for the emitter-to-base resistance is about 100 ohms, and a typical value for the load resistor is 10,000 ohms.

The current that reaches the collector is 98 per cent of the current leaving the emitter. If, now, we multiply the current gain (0.98) by the resistance gain 10,000/100, we shall obtain the voltage gain of the collector circuit over the emitter circuit. Numerically, this is

$$\text{Voltage gain} = \text{current gain} \times \text{resistance gain} = 0.98 \times \frac{10{,}000}{100} = 98$$

Thus we see that while the current gain here is actually a loss, this is more than made up by the extent to which the collector resistance exceeds the emitter resistance. Furthermore, this overwhelming differential in resistance will also provide a power gain. This means that with a small amount of power in the input, or emitter-to-base, circuit, we can control a much larger amount of power in the output, or collector-to-base, circuit. Both of these characteristics are important; without them the transistor would have only limited application in electronics.

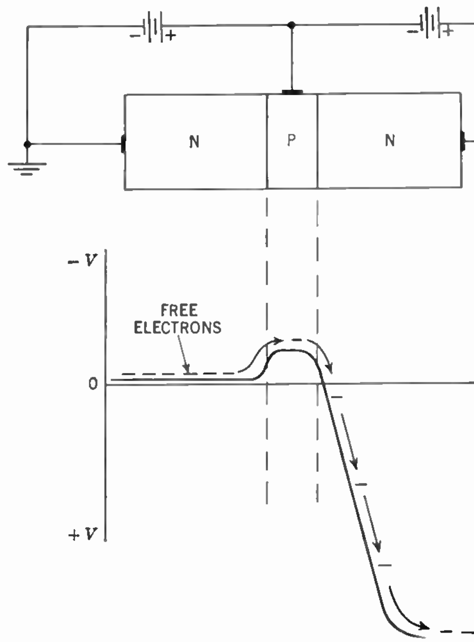


Fig. 2·19 The potential-hill diagram for an NPN junction transistor. (After H. K. Milward, Introduction to Transistor Electronics, *Wireless World*, March, 1955)

The voltage gain indicated above is that which would be obtained if the transistor operated into a very high impedance circuit. Actually, one of the problems which is encountered in cascaded transistor amplifiers is that of matching the relatively high output impedance of a prior stage with the low input impedance of the following stage. This point will be discussed in greater detail in Chap. 4.

Potential hills. Because a complete understanding of what happens within the transistor is so vital to future circuit application, still another approach, based on potential hills, is deemed desirable. This method, Fig. 2·19, reveals the effects of emitter, base, and collector voltages and presents a simplified visual picture of junction-transistor operation. (We continue to repeat the word "junction" because we shall presently consider another type of transistor.)

When the emitter is biased in the forward direction and the collector in the reverse direction, electrons leaving the N-type emitter will see only a small potential hill in front of them (at the NP junction), one that many of them can surmount. Once atop the hill, the "ground" levels off and the electrons move through the P layer of the base quite readily. When they reach the junction between the base and the N-type collector, the electrons come under the influence of the positive battery potential and surge forward strongly. In the voltage diagram, this attraction is represented as a downward slope which electrons (like human beings) find simple to traverse.

If the forward bias on the emitter is reduced, we are, in effect, raising the height of the base potential hill. Electrons leaving the emitter will find the higher hill more difficult to climb, and only those electrons possessing the greatest amount of energy will be able to reach its summit and, from there, move to the collector ahead. Current will consequently be reduced.

By the same token, increasing the forward bias on the emitter will reduce the height of the hill, thereby enabling more emitter electrons to enter the base region. Thus the biasing voltages used in a transistor have a very important effect on its operation. Another significant controlling factor is the width of the base. This is demonstrated by the next two illustrations.

Two three-dimensional representations of this potential diagram are given in Figs. 2·20 and 2·21. The difference between the two drawings lies in the width of the base sections. If the base section is wide, the tendency for emitter electrons (represented here by balls) to end up at the base electrode (because of combination with holes) is much greater than it is when the base section is narrow. In a physical model of these illustrations, the potential surfaces through the transistor are formed by a rubber membrane supported at several points. The holes in the base section are represented by a slight dip, or "valley," at the center of the base membrane. The wider the base section, the more difficult it is for the balls to roll through the base valley and over the edge of the precipice into the collector region without being trapped by the base dip.

On the other hand, if the base region is made quite narrow, any balls having enough energy to surmount the initial rise of the base hill possess enough energy to reach the far edge of the base and fall down into the collector.

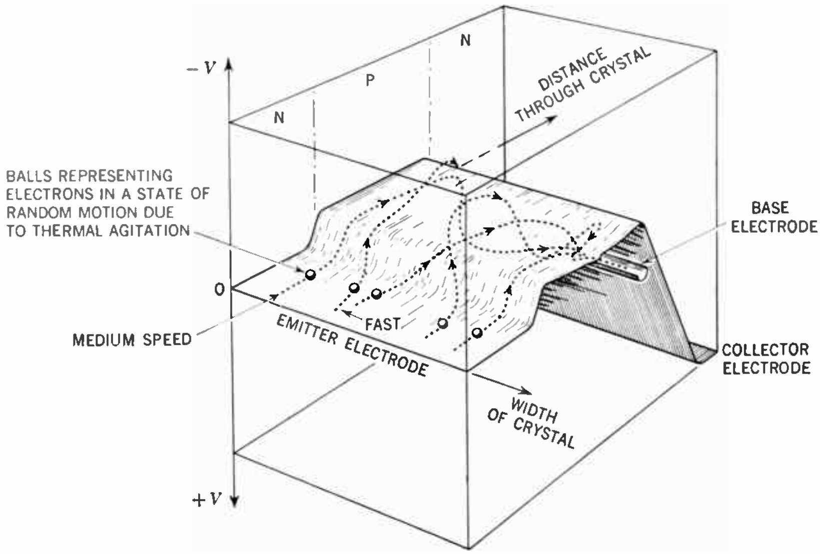


Fig. 2·20 A three-dimensional representation of the potential levels in an NPN germanium transistor. Suitable bias voltages are assumed to be applied to the various electrodes. In this illustration, the base is made wider than normal to demonstrate its effect. (*Wireless World*)

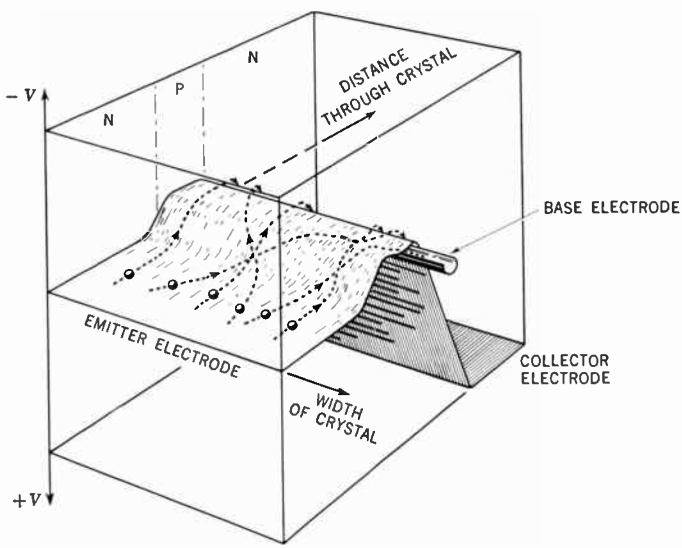


Fig. 2·21 The same illustration as Fig. 2·20 except that the base layer is narrow. (*Wireless World*)

Thus, the width of the base section has a very direct bearing on transistor gain, both voltage and power. For if the percentage of current reaching the collector decreases to very small values, it will reduce the voltage gain in the same proportion. Power, being proportional to the square of the current, will be adversely affected to an even greater extent.

From the foregoing discussion we can formulate two rules concerning the normal use of this and *all* transistors:

1. The emitter is biased in the forward, or low-resistance, direction.
2. The collector is biased in the reverse, or high-resistance, direction.

That these rules are always true can be seen if we consider their alternatives. If the emitter is biased in the reverse direction, it will not permit any electrons to reach the base region. And a reverse-biased emitter, with a reverse-biased collector, will produce a transistor in which current never passes. There are occasions, as with vacuum tubes, when it is desirable to bias the transistor to cutoff. In the case of a vacuum tube, this is done by increasing the negative bias on the grid with respect to the cathode. In the case of a transistor, cutoff is achieved by bringing the emitter-to-base bias to zero or even inserting a small amount of reverse biasing voltage. In the majority of applications, however, the statement indicated in (1) is true.

If the emitter and collector are both forward-biased, then the general tendency will be for the emitter electrons to flow between emitter and base and for the collector electrons to flow between collector and base. In essence we shall have two junction diodes possessing a common base. If the collector forward voltage is larger than the emitter voltage, some of the collector electrons will flow back to the collector via the emitter. But in any event, the desired amplification will not be obtained, and the purpose of the transistor will be defeated.

At this point, a note of caution regarding the application of reverse voltage to transistors is necessary. As we shall see later, the emitter bias voltage is quite small, on the order of 1 volt or less. The collector reverse voltage is generally much higher. If we should mistakenly connect the collector battery in the forward direction, the excessive current flowing through this section may develop enough heat to destroy the junctions and render the transistor worthless. Hence, always be certain the collector voltage polarity is correct *before* making connections.

It is interesting to note that a transistor possesses a bidirectional facility that is impossible to achieve in vacuum tubes. That is, we can forward-bias the collector with a low voltage, reverse-bias the emitter, and then feed the signal in at the collector. The current gain under these conditions will be somewhat less than it is when the unit is employed normally. As a matter of fact, in Chap. 7 there is described a television phase detector in which the emitter and collector sections are structurally identical and each takes turns sending current through the transistor.

Transistor amplifier. We are now ready to connect the NPN transistor into an actual amplifier circuit with the signal input at one end and the load resistor at the other, Fig. 2·22. The incoming signal is applied in series with the emitter-to-base bias, and the load resistor is inserted in series with the collector battery. When the signal voltage is zero, the number of electrons leaving the emitter and entering the base region is determined solely by the forward bias on the emitter. This situation can be represented by the potential-distribution diagram shown in Fig. 2·23*a*. When the signal goes negative, it adds to the forward bias, further reducing the height of the base hill and causing more electrons to flow through the transistor. This is shown in Fig. 2·23*b*. During the next half cycle, the signal goes positive, reducing the forward bias of the emitter and thereby reducing the number of electrons leaving the emitter and entering the base and, subsequently, the collector regions. This is shown in Fig. 2·23*c*, where the height of the base hill has been increased.

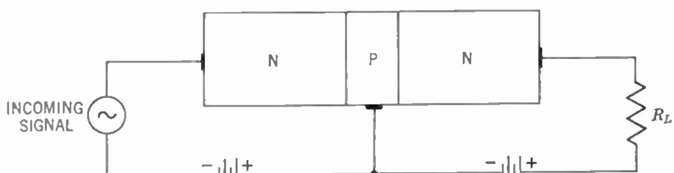
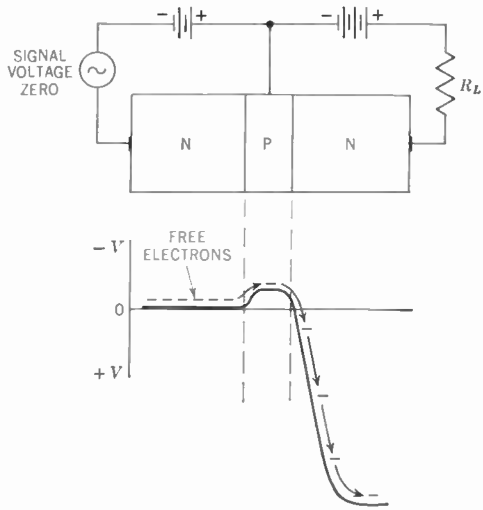


Fig. 2·22 The NPN transistor connected as an amplifier.

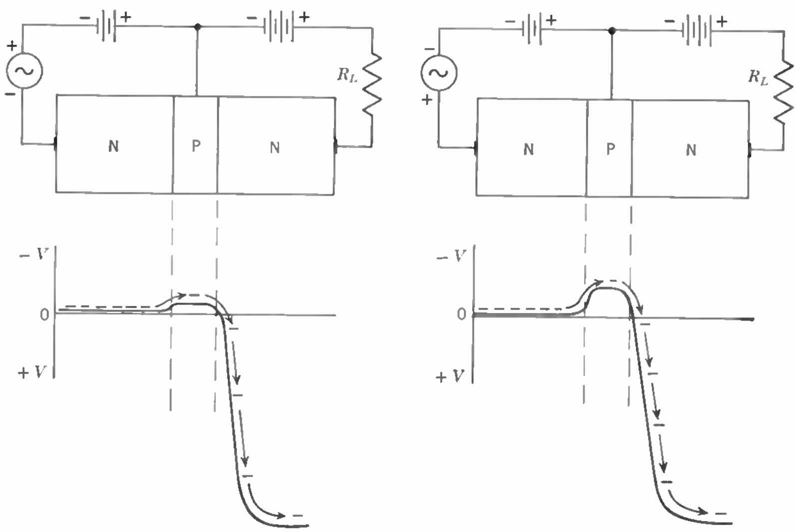
At the other end of the transistor, these current fluctuations produce corresponding voltage variations across R_L, the load resistor. When the input signal is negative and the current increases, the collector end of R_L becomes more negative. By the same reasoning, when the signal goes positive, current decreases and the collector end of R_L becomes relatively more positive.

Thus, through this transistor, amplification is achieved without the normal 180° phase shift we are accustomed to in vacuum tubes. This is not always true of transistors, and in our subsequent study we shall come across instances when signal phase reversal does occur.

Another point to note here is that with this transistor, an increase in



(a) Signal voltage zero



(b) Signal voltage aids emitter-base bias (c) Signal voltage opposes emitter-base bias

Fig. 2·23 Transistor-amplifier operation demonstrated by potential hills.

signal polarity (i.e., positive) causes the transistor current to decrease. This is in direct contrast to normal vacuum-tube amplifiers when the signal is applied to the grid. On the other hand, with the PNP transistors to be studied next, a positive increase in signal polarity causes the transistor current to increase.

All this is mentioned in an attempt to illustrate that although transistors and the more familiar vacuum tubes contain many points in common, they also differ in many respects, and it is suggested that the reader learn to think of each in terms of itself. We shall have occasion to make additional comparisons between these two devices. This, of course, is natural, since transistors do the same basic job that vacuum tubes do. However, transistors represent an entirely different approach to amplification, with many new features and characteristics. To attempt at each point to find an equivalent property in vacuum tubes for each characteristic in transistors will, in the long run, lead only to confused thinking. Learn to regard transistors in terms of their own operation and vacuum tubes in terms of theirs.

PNP Transistors

In the formation of the initial transistor from a PN junction, we added a second N section to evolve an NPN transistor. We can ap-

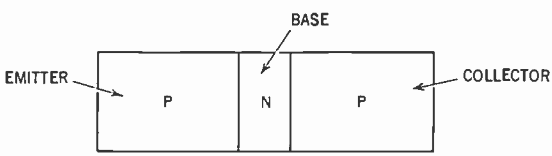


Fig. 2·24 A PNP transistor.

proach the same problem by adding another P section to produce a PNP transistor, Fig. 2·24. The emitter and collector sections are formed now of P-type germanium, while the base section consists of N-type germanium. Since this is actually the reverse—as far as material structure is concerned—of the NPN transistor, we should expect differences in the mode of operation and in the polarity of the voltages applied to the emitter and collector.

A typical bias setup with a PNP transistor is shown in Fig. 2·25. The positive side of the battery connects to the emitter, while the negative terminal of the battery goes to the base. The collector battery is attached in the reverse manner, with the negative terminal connecting to the collector and the positive terminal going to the base. Holes are the current carriers in the emitter and collector sections; in the N-type base section, electrons are the principal carriers. With the emitter-bias battery connected as shown in Fig. 2·25, the positive field of the battery repels the positive holes toward the base section. At the same time, the negative battery terminal at the base drives the base electrons toward the emitter. When an emitter hole and a base electron combine, another electron from the emitter section enters the positive battery terminal. This creates a hole, which then starts traveling toward the base. At the same instant, too, that the first hole and electron combine, another electron leaves the negative battery terminal and enters the base. In this way, current flows through the base-emitter circuit.

In the PNP transistor, the holes are carriers in the emitter section, and when they cross the junction into the base region, a number of them combine with the base electrons. However, well over 90 per cent of the holes do not combine with base electrons; rather, they pass through the base region and continue on to the collector. There they meet a negative attractive force and move toward the collector terminal. When the terminal is reached, an electron from the battery combines with a hole and effectively neutralizes it. At the same instant, an

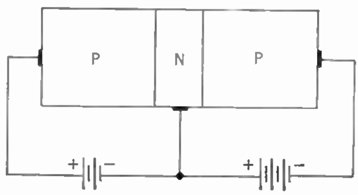


Fig. 2·25 Method of biasing a PNP transistor.

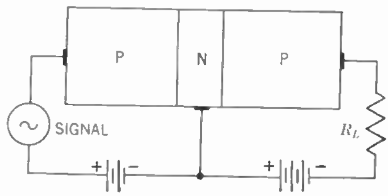


Fig. 2·26 A PNP transistor amplifier.

electron leaves the emitter region and starts on its way around the outer circuit to the collector battery.

Note, then, that although holes are the current carriers in P-type germanium, current conduction through the connecting wires of the external circuit is carried on entirely by electrons. This fits in with the current conduction that we are familiar with.

The incoming signal and the load resistor occupy the same positions in a PNP transistor amplifier that they do in an NPN transistor amplifier, Fig. 2·26. Only the polarity of the biasing voltage is reversed.

Point-contact Transistors

The discussion thus far has been concerned solely with junction transistors. These are the only types of transistor in use today. From the standpoint of discovery and initial development, however, the point-contact transistor comes first, and it is of historical interest to examine its mode of operation briefly. As with the junction transistor, it is best to start with a point-contact diode, such as we have had available for several years.

Point-contact diodes. A point-contact diode is shown in Fig. 2·27. One section consists of a small square rectangular slab of germanium to which a controlled amount of impurity has been added. Generally, a pentavalent impurity is added, giving us N-type germanium. The other half of the diode consists of a fine phosphor-bronze or beryllium-copper catwhisker wire which presses against the center of the germanium slab. The opposite end of this wire represents one terminal of the diode, while a metal plate deposited on the far side of the germanium crystal serves as the second terminal.

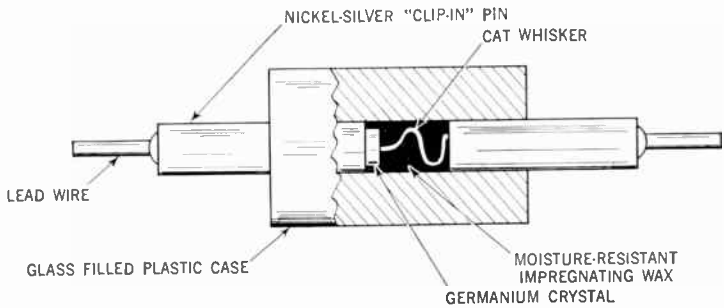


Fig. 2·27 Cross-sectional view of internal structure of a germanium diode. (CBS-Hytron)

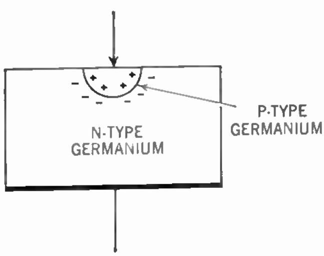


Fig. 2·28 A point-contact germanium diode showing the approximate distribution of the N- and P-type germanium.

An important step in the fabrication of this diode is the passage of a relatively large current from the catwhisker wire to the germanium slab. The purpose of this "forming" current is to produce a small area of P-type germanium in the region surrounding the point of contact of the phosphor-bronze tip. The germanium diode now consists of P- and N-type germanium, Fig. 2·28, and the explanation of its operation follows exactly along the lines previously indicated for a PN junction diode. Maximum current will flow when the base plate is made negative with respect to the catwhisker wire, and minimum current will flow when these voltages are reversed. The principal current carriers in the N section are electrons, while the principal carriers in the P section are holes. The boundary between the two can be considered as a PN junction, even though the contour of this junction differs considerably from the junctions previously discussed.

Point-contact-transistor operation. To form a point-contact transistor, two phosphor-bronze wires are mounted side by side as shown in Fig. 2·29. One wire forms the emitter of the transistor; the other wire, the collector. A third electrode, the base, is a large-area metal plate deposited on the underside of the germanium wafer. The final preparatory step is that known as electrical forming. In this process, relatively large surges of current are passed through the wires to the

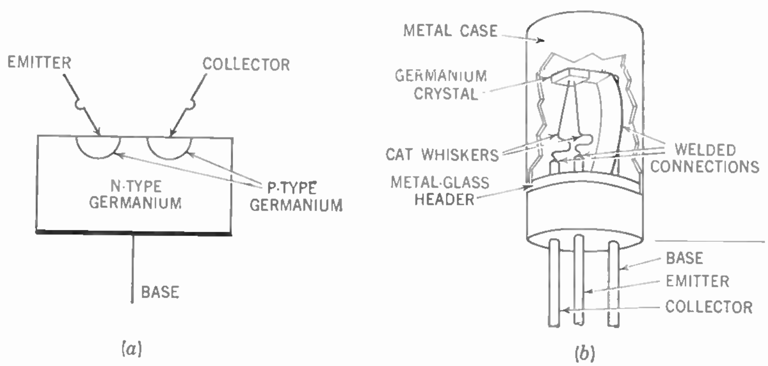


Fig. 2·29 (a) A point-contact transistor. The emitter and collector wires are each held tightly against the germanium block. (b) The physical construction of a point-contact transistor.

base. As in the preceding diode, this current serves to form small areas or islands of P-type germanium under each wire electrode. The area of each P section is extremely small, possibly no more than a few atomic layers thick.

To understand how a point-contact transistor functions, let us connect the unit into the circuit shown in Fig. 2·30. The collector is biased in the reverse, or high-resistance, direction. In this case, since the collector wire is assumed to be the connecting terminal for a P layer, it receives the negative terminal of the battery. The base, then, is connected to the positive battery terminal. The emitter is biased in the forward direction; therefore, it connects to the positive terminal of the battery.

Now let us place current meters, switches, and resistors in each of the emitter and collector branches. The resistor in the emitter branch is more in the nature of a limiting resistor, since this portion of the transistor is biased in the forward direction. The resistor in the collector circuit is a load resistor across which the output signal will be developed.

As a start, let us open switch S and close switch S_2. Since the collector is biased in the reverse direction, very little current should flow under these conditions. Actually, with about 25 volts for battery B_2, a current of a few microamperes will be found to flow. This current, like the reverse current in a diode, stems from the presence of free electrons in P-germanium sections and holes in N germanium. These are the minority carriers, and they obtain their energy principally from thermal absorption, although impinging light photons may also be a factor. (The same behavior is displayed by the junction transistor.) If we continue to raise the value of the reverse voltage, we shall eventually reach a point where the stress of the applied electric field will be

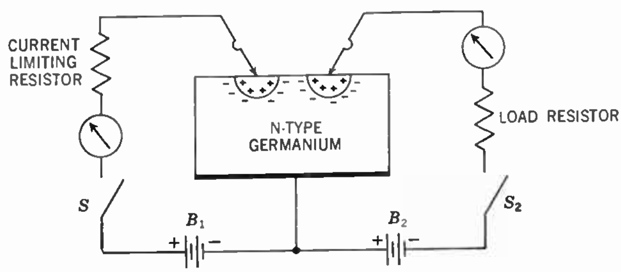


Fig. 2·30 A point-contact transistor with its bias voltages. These voltages follow the same rules of application as those indicated for junction transistors.

great enough to produce a cumulative buildup of electrons through collision and a large current will ensue. This occurs at the avalanche breakdown voltage. In the present discussion we are far from this level, and only the slight current mentioned above will be measured.

If, now, we permit the collector circuit to remain as it is and close switch S, two things will happen. First, current will flow in the emitter circuit. Second, the current in the collector circuit will rise. Furthermore, if an input signal is placed in series with the emitter bias, then as the total voltage increases (i.e., becomes more positive), the current in the collector circuit will increase also. On the negative half cycle of the signal, when its voltage is acting in opposition to the bias battery and reducing the overall positive voltage applied to the emitter, it will be found that the collector current will decrease too. In other words, variations in the input circuit will produce corresponding variations in the output circuit.

It is apparent from this behavior that the emitter circuit influences the collector circuit in point-contact transistors just as it does in junction transistors. Since the area around the emitter catwhiskers is P-type germanium and the base is N-type germanium, we can fall back on what we have already learned about junction transistors to attempt an explanation. With the emitter-to-base circuit biased in the forward direction, the holes in the emitter section move toward the base junction (the semicircular boundary of the P area). At the same time, the electrons in the base move toward the emitter junction. At the junction, a small number of the holes combine with base electrons. For every electron that the base loses in this manner, one electron leaves battery B_1 and moves into the base section. In the same way, every hole that the emitter loses by combination forces one electron to leave the emitter and enter the emitter wire on its way to the positive battery terminal. This loss of an electron produces a hole, which then starts migrating toward the base junction.

The majority of the holes which leave the emitter drift toward the collector, attracted by the negative collector voltage. On their way there, some of the holes combine with the free electrons present in the base. The holes arriving at the collector combine with electrons which the negative terminal of the collector bias battery provides. For every such combination an electron leaves the emitter section and enters the emitter wire. Electron travel is from emitter through both bias batteries to collector, with the base being completely bypassed. (The reader will note that a similar current exists in junction transistors.)

Now, if the foregoing represented the entire action in a point-contact transistor, then the current flowing in the emitter circuit would be larger than the current flowing in the collector circuit. The difference would be that portion of the emitter current diverted to the base. Actually, it is found that a change of 1 ma in the emitter circuit of a point-contact transistor produces a 2- to 3-ma change in the collector circuit. Here, obviously, is a significant departure from the effects observed in junction transistors, and the natural question to ask is: "Where does this additional current come from?"

It is believed that the additional current is due to the fact that the emitter holes, in traveling to the collector, form a positive space charge which attracts electrons from other sections of the germanium crystal and causes them to add to the collector current. These additional electrons (which are separate from the electrons that combine with the holes from the emitter) are confined solely to the collector circuit and travel in the path from the negative terminal of the battery through the collector and base sections of the germanium crystal and out through the base lead back to the battery again. In effect, what the holes do is reduce the internal resistance of the collector circuit, permitting a greater current to flow for the same applied voltage. Partial proof of this is the fact that whereas the internal resistance of the collector circuit in junction transistors is on the order of 500,000 ohms or more, in the point-contact transistor, it is typically about 20,000 ohms.

The foregoing behavior in point-contact transistors explains why the two catwhisker wires must be positioned close to each other. The greater their spacing, the less likelihood there is of emitter holes lasting long enough to reach the collector area. The separation distance of the two wires, then, is critical.

The reason for the marked difference in behavior of point-contact and junction transistors stems largely from the difference in construction. In the point-contact transistor, there is a large base area from which electrons may be drawn to enhance the normal collector current. The holes, traveling from emitter to collector, serve to attract these excess electrons through their strong positive field. In a comparable junction transistor, a PNP unit, the base section is quite narrow and can supply only a limited number of electrons to the collector current. Hence, the same current-gain effect is not observed.

The point-contact transistor is no longer being manufactured, because the junction transistor is superior to it in nearly all respects. Junction transistors are easier to manufacture; hence, they can be made more economically. A junction transistor will handle larger amounts of power, because each interelement junction possesses a larger area than the corresponding catwhisker wire and its junction. Finally, a junction transistor, operated in the common-emitter mode which we will study presently, can provide higher current and voltage gains than a point-contact transistor can.

Because of the almost complete disappearance of the point-contact transistor from the engineering scene, no further mention will be made of it.

QUESTIONS

2·1 Draw a simple diagram showing how the germanium atoms are bound to one another in a crystal.

2·2 Since all the electrons in a germanium crystal are held fairly tightly, explain why germanium is considered a semiconductor rather than an insulator.

2·3 What is a hole in a semiconductor, and how is it formed?

2·4 How do holes travel through a semiconductor?

2·5 When an electric field is applied to a semiconductor, what is the effect on holes and free electrons?

2·6 How is N-type germanium formed? List several substances which could be employed in this process.

2·7 How is P-type germanium formed? List several substances which could be employed in this process.

2·8 Are holes ever found in N-type germanium or electrons in P-type germanium? Explain.

2·9 Draw a PN junction showing the donor and acceptor atoms and the free holes and electrons on their respective sides.

2·10 What prevents the wholesale recombination of excess holes and electrons at a PN junction?

2·11 What do we mean by a "potential hill"?

2·12 How is a battery connected to a PN diode in order to initiate a flow of current? In order to prevent current flow?

2·13 Explain the mechanism of current flow through a crystal diode when the diode is biased in the forward direction.

2·14 Draw the characteristic curve of a germanium diode. Explain the reason for the reverse current.

2·15 What is the avalanche voltage in a crystal diode?

2·16 How is an NPN junction transistor formed? Sketch such a transistor and label each section.

2·17 Describe the rules that must be followed in biasing a transistor.

2·18 Why must the base section be made as thin as possible in a transistor? What happens when it is made too wide?

2·19 Explain how a transistor operates.

2·20 How is gain achieved in a junction transistor?

2·21 Why is there no signal reversal between input and output of the amplifier shown in Fig. 2·22?

2·22 Draw the diagram of an amplifier using a PNP transistor.

2·23 Describe the flow of current through a PNP transistor.

2·24 How do point-contact and junction transistors differ in structure?

2·25 Describe the operation of a point-contact transistor.

CHAPTER 3

Transistor Characteristics

WHEN TRANSISTORS were first developed, they could handle only milliwatts of power, and that at fairly low frequencies. Both of these limitations were serious drawbacks which severely limited transistor application. Since then, the frequency limitation has all but been removed, while the power-handling capabilities have been increased several thousand times so that we now possess many transistors capable of handling from 5 to 100 watts. Transistors still are subject to many limitations, and the more important of them will be covered in this chapter. Substantial manufacturing progress has taken place, however, and it is reasonable to expect that progress will not only continue at its present pace but even accelerate. In any event, the transistor is today a finely engineered device that possesses a high degree of stability capable of performing a wide variety of functions.

In the paragraphs to follow, we shall examine the more important characteristics of transistors, physical as well as electrical. In addition, we shall compare transistors with vacuum tubes to achieve a clear understanding of the differences and similarities.

Power

The power capability of a transistor is limited by three major factors. First, there is the maximum reverse voltage which the collector can withstand. When this reverse voltage is raised beyond a certain point, an electrical breakdown is created between the collector and base. This is generally the avalanche effect previously noted on page 34.

There is also another mechanism of voltage breakdown, called punch-through. This results from an expansion of the depletion region that exists on either side of the PN junction between the base and collector (see page 30). When a negative voltage is applied to the collector of a PNP transistor (positive voltage terminal to the base), the holes in the P-type collector will be attracted to the negative potential. Similarly, the electrons in the N-type base will be attracted by the positive potential applied to the base terminal. This attraction of the carriers away from the junction creates a depletion region in which there are no *free* carriers, Fig. 3·1*a*. As the voltage across the junction is increased, the depletion region expands and, at some particular voltage, will expand through the entire base region and actually come in contact with the emitter junction. This is shown in Fig. 3·1*b*. The voltage at which this effect takes place is called the punch-through voltage. When punch-through occurs, and at higher voltages,

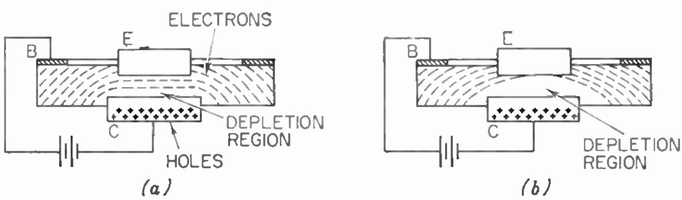


Fig. 3·1 Mechanism of punch-through voltage.

the normal transistor action ceases and the base can no longer control the current flow. As a practical matter, the resultant dynamic short circuit between collector and emitter permits a flow of current that is limited only by the resistance in the external circuit.

With either of the above two conditions, the transistor may be damaged if too much current flows. If the current flow is not excessive, however, the transistor may not be damaged.

A second factor that limits the maximum power capability of a transistor is the decrease in current gain with increased current. This will be examined more fully later because it involves a current factor (beta) which has not yet been discussed. With a decrease in current gain, however, the power gain similarly decreases and this reduces the effectiveness of the unit to function as an amplifier. (Power gain is proportional to the square of the current gain.)

The third factor which establishes a limit to the maximum power output of a transistor is the safe amount of heat which the material or junction can withstand. Another way of stating this is to indicate the maximum power dissipation of the transistor internally, although this is generally the collector dissipation since the greatest amount of heat is developed in the collector. Since the collector and emitter currents are nearly equal, differing only by the relatively small amount of current diverted by the base, the reader may wonder why more heat is generated at the collector. The answer lies in the fact that power is equal to I^2R and the collector, being reverse-biased, has a far higher internal resistance than the emitter has.

To assist transistors in achieving higher collector-dissipation ratings, special heat sinks, Fig. 3·2, in which these transistors are mounted

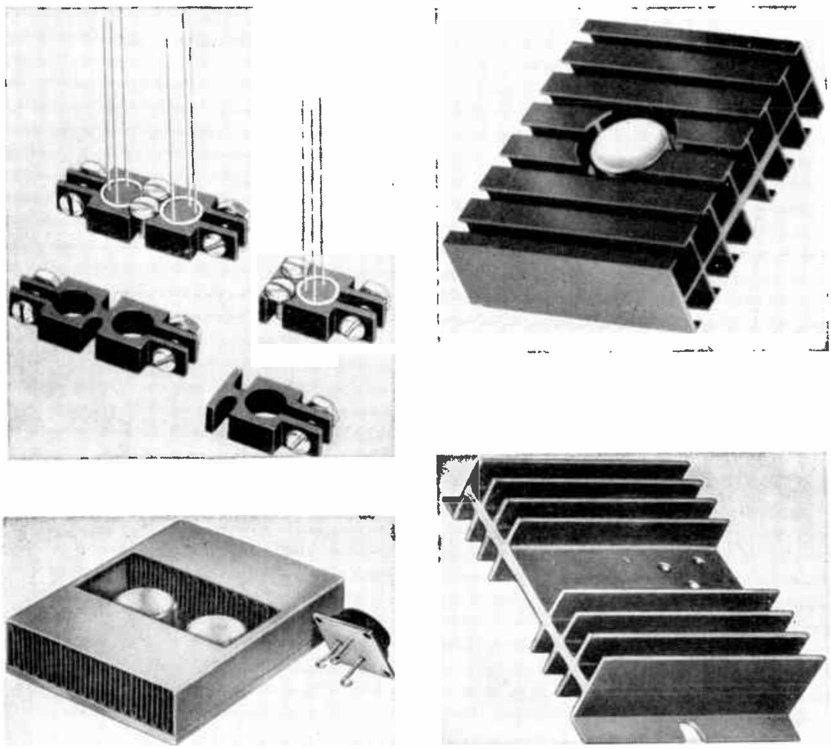


Fig. 3·2 Typical transistor heat sinks.

have been developed. These heat sinks, or heat dissipators, help conduct the heat away from the collector. Also useful for this purpose are metallic housings for transistors, Fig. 3·3.

It is customary, when the power dissipation of a transistor is given, to specify the temperature (usually 25°C) at which this dissipation can be obtained safely. Low-power transistors are rated in reference to the surrounding, or ambient, air temperature; medium- and high-power units are rated in reference to case or housing temperature. If it is desired to operate the transistor at a temperature higher than

25°C, the power-output rating must be lowered or derated. Either a specific derating factor, such as 1 watt per °C, is given or a derating curve is included with the specification sheet.

If a derating factor is given, then for each degree centigrade rise in temperature, the maximum power dissipation must be lowered by

Fig. 3·3 A power transistor housed in a metallic container. The unit shown will provide a maximum power output of 6 watts when operated class A. (Minneapolis-Honeywell Regulator Co.)

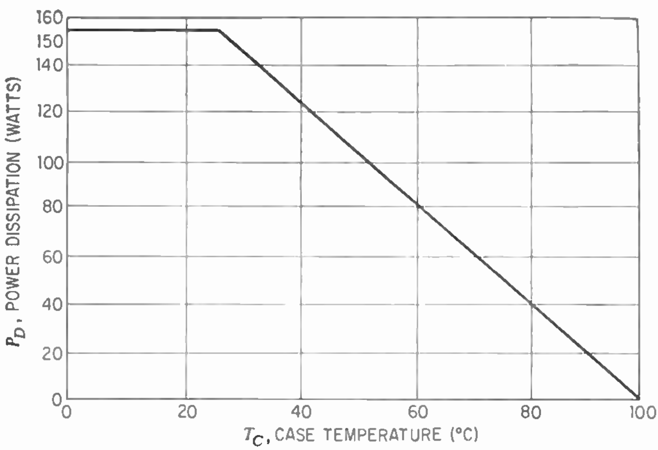


Fig. 3·4 A power-transistor derating curve.

the amount indicated, here 1 watt. If a derating curve is given, this information can be obtained directly. A typical curve is shown in Fig. 3·4 for a 2N174 transistor. Up to 25°C, the power dissipation is 150 watts. Beyond this temperature, the dissipation decreases linearly. Hence, at 60°C case temperature the power dissipation can only be 80 watts, while at 100°C it has been reduced essentially to zero. Failure to heed these derating factors will not only shorten the life of the transistor but will also frequently cause the transistor to burn out in short order.

Several types of power transistors are shown in Fig. 3·5. As a general rule, the higher the power rating, the larger the transistor housing.

Noise

Transistors, like vacuum tubes, develop a certain amount of internal noise which can be disturbing in low-signal circuits, such as we find

Fig. 3·5 Typical power transistors.

in the front end of a receiver or in the front end of a high-power audio amplifier. In the early days of their development, junction transistors tended to be considerably noisier than vacuum tubes. It was not unusual to find a transistor with a noise figure of 30 db compared to a 3-db noise figure for many vacuum tubes (at 1,000 cycles). Since then, transistor noise figures have been steadily lowered until, at the present time, they range between 4 and 20 db, with most values falling below 10 db.

Noise figure is a relative term designed to permit comparison between similar electronic devices or circuits. For example, if the noise figure of a transistor were 0 db, then the transistor would not be generating any internal noise. Such a transistor, inserted in any circuit, would add no noise to that already existing there and possibly stemming from any resistance present. When a transistor has a noise figure other than 0 db, however, then it contributes noise to the circuit. The higher the noise figure, the greater the noise being added.

The noise behavior of a typical transistor is shown in Fig. 3·6. The characteristic turns up at either end, but in between it remains quite flat and unchanging. At the low-frequency end, the noise figure is inversely proportional to frequency, dropping down to the flat base in

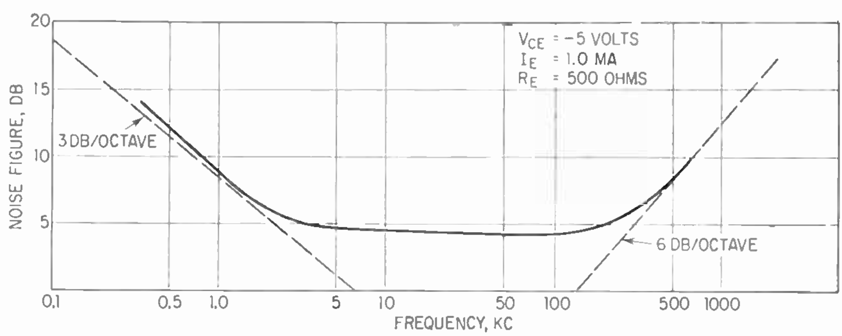


Fig. 3·6 Noise behavior of a typical transistor.

the frequency range between 100 and 1,000 cycles. The source of this noise has been traced to surface leakage and imperfections in manufacture. In recent transistors, special surface treatment has tended to reduce the extent of this disturbance and has extended the flat segment of the curve toward the low end.

The central section of Fig. 3·6 is developed by the diffusion of carriers within the transistor and from the thermal agitation of current carriers. The first of these effects is directly related to the shot noise in vacuum tubes, which arises from the physical flow of electrons from cathode to plate. The second, or thermal noise, is analogous to the noise voltages generated in any resistor or other conductor of current. Both of these noise voltages are completely random in their nature because their effect (i.e., polarity) cannot be predicted at any specific instant of time.

The second upturn of noise, at the right side of the curve, Fig. 3·6, is due to the drop in power gain of a transistor because of the frequency limitation of the device. Since the noise factor, by definition, depends on the ratio of signal to noise, anything that reduces the signal causes the noise figure to worsen.

Frequency Response

In any application of a transistor, its frequency limitations must be known not only for economic reasons (to avoid using a high-cost transistor with an extended high-frequency response where only a low-frequency response is required) but also because power gain falls off rapidly with frequency (above a certain point). There is no single factor which, by itself, completely controls frequency response. The end result is always a compromise of a number of considerations the more important of which will be discussed here.

An important characteristic in establishing the frequency behavior of a transistor is the time required for a signal to travel from emitter to collector. This, in turn, depends upon the mobility of the carriers within the germanium. Typical values are as follows: for electrons, about 3,600 cm^2 per sec per volt of potential difference; for holes, about 1,900 cm^2 per sec per volt of potential difference. We cannot apply signals whose frequency changes so rapidly that the carriers (holes or electrons) are unable to transport the changes from emitter to collector.

We have an analogous situation in vacuum tubes when the signal applied to the grid changes appreciably in the time it takes an electron to travel from the cathode to the plate. If the grid goes positive, attracting more electrons to the plate, and then turns negative before the electrons reach the plate, the change not only will reduce the peak number of electrons reaching the plate but will also cause the electrons to spread out over a greater interval. Operationally, this results in phase and amplitude distortion.

The mobility of the holes or electrons in a transistor is the velocity with which they move through the germanium when an electric field is applied. Since electrons move almost twice as fast as holes, we would expect those transistors in which electrons do most of the current carrying to have a higher frequency response than transistors which depend upon conduction by holes. Thus, consider an NPN transistor, Fig. 3·7*a*. Conduction from emitter to collector depends upon the diffusion of electrons from the emitter to the vicinity of the collector. On the other hand, in a PNP transistor, Fig. 3·7*b*, holes are the principal carriers from emitter to collector. Since holes travel slower than electrons, we would expect PNP transistors to have a lower cutoff frequency.

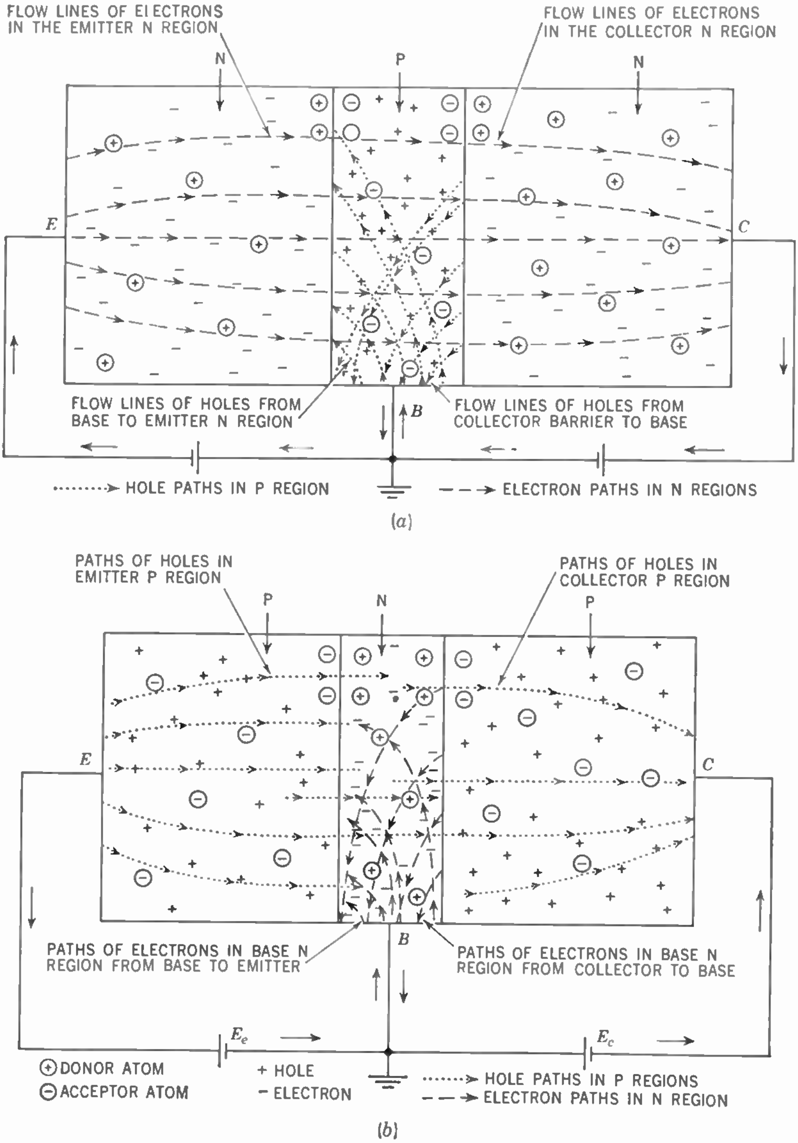


Fig. 3·7 (a) The chief carriers of current in an NPN junction transistor are electrons. (b) The flow paths of electrons and holes in a PNP junction transistor. Most of the current is carried by holes.

Another factor which limits the high-frequency response is the capacitance between sections of a transistor. The higher the frequency, the lower the impedance of the shunting capacitor and the greater its shunting effect on the applied signal. This is true in both the input and output circuits, as well as between input and output.

The frequency response of a junction transistor can be improved by making the central base-region layer as thin as possible. That is, the frequency f is inversely proportional to base width in accordance with the equation

$$f = \frac{2d}{w^2}$$

where d = diffusion constant, i.e., mobility, cm^2 per sec
w = base width, cm

One encounters manufacturing difficulties when the width is made too small, however. Moreover, decreasing base width serves to increase base resistance, and the latter will degrade the power gain as the frequency rises.

A higher base resistance also leads to a lower value of punch-through voltage, which is detrimental to the power capabilities of the unit. One way to avoid this is to make the base resistivity lower by adding more impurities during the manufacturing process. But emitter efficiency, or the ability of the emitter to inject its majority carriers into the base region, depends upon having the emitter resistivity much lower than the base resistivity (in other words, doping the emitter more heavily than the base). Thus, if we lower base resistivity, we lower the emitter injection efficiency, and this impairs current gain. Manufacturers get around this by adding a small amount of gallium to the indium they employ to fabricate a P-type emitter. For an N-type emitter, antimony is added to arsenic, the latter being the impurity that provides excess electrons. Another way to skirt this difficulty is provided by the drift transistor, which we shall study presently. In our subsequent study of the transistor other compromises will be apparent, so that the fabrication of a transistor to achieve a certain operating characteristic is not a simple, straightforward process.

There are a variety of transistor types on the market, each possessing certain desirable operating and economic features. Transistors have also passed through various stages of development, and some knowledge of these stages will not only assist the reader to better evaluate future progress but also provide him with a better understanding of transistor operation and application. In the paragraphs to follow, an attempt will be made to show this sequence of development.

Grown-junction transistors. Historically, the grown-junction transistor was the first junction type manufactured. Today, it is surpassed in performance by other transistor types and so finds very little application, although silicon transistors still utilize the grown-junction method. In time, it will undoubtedly be replaced here, too, by more advanced techniques.

The grown-junction transistor obtains its name from the fact that the junctions are produced by what is basically a growing process. A small crystal of germanium or silicon, called a seed, is dipped into molten germanium (or silicon) contained in a crucible (Ref. 3·1). As the seed is withdrawn slowly, the liquid freezes onto the seed, growing a crystal. The original melt contains a suitable impurity, perhaps arsenic for N type. After a sufficient amount of crystal has been grown on the seed with N-type impurity, a P-type impurity, in minute amount, is added to the melt to produce P-type germanium. The crystal is grown as P type for one-thousandth of an inch and then a second pellet, this time of N-type dope, is dropped into the melt. The amount is sufficient to overcome the P dope and the crystal is returned to N type for the balance of the crystal-growing process, Fig. 3·8.

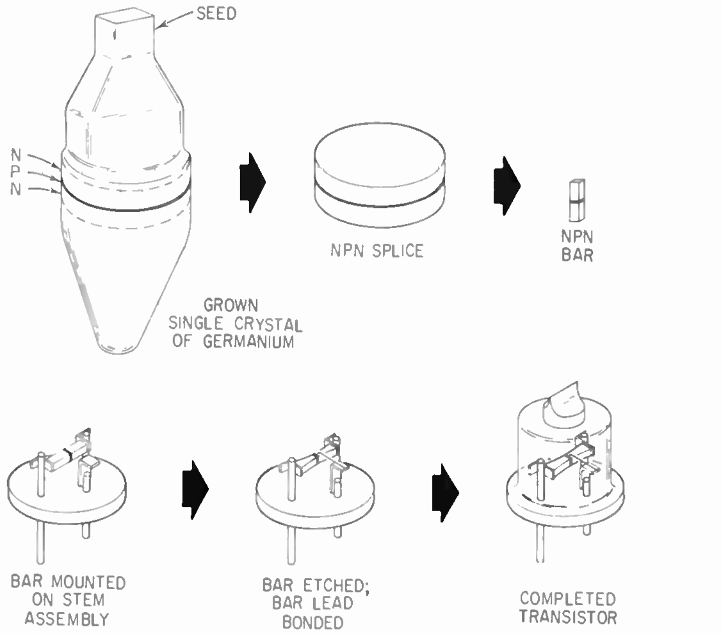


Fig. 3·8 **Manufacture of grown-junction NPN transistor from single crystal to final unit.** *(Western Electric Engineer)*

When the desired crystal has been grown, it is cut up into a number of wafers by a diamond saw, lapped smooth with an abrasive, cut up into little squares, and etched to remove any surface scratches. The final transistors have dimensions of 0.025 by 0.025 by 0.125 in. The end regions to which emitter and collector leads are attached are N type; the very thin base is P type.

The grown-junction transistor has several major limitations which prevent it from being employed more extensively than it is. Frequency is greatly dependent on base width, and for even nominal frequencies a very narrow base width is required. With the seed-growing technique it is difficult to achieve such narrow widths with sufficient accuracy to provide a suitable manufacturing process with a high degree of reproducibility.

A second limitation is base resistance. This tends to be high not only because the base is narrow, but also because contact between the base and the outer circuit occurs at a single point only. Much more desirable is a ring-type contact, but this is not feasible in this transistor. The foregoing limitations restrict the grown-junction transistor to frequencies below 20 Mc.

Alloy-junction transistors. The greater part of present-day transistor manufacture is devoted to the alloy-junction transistor. This process produces a wide variety of amplifier transistors (both low and moderately high frequency) as well as switching transistors in the low and medium speed ranges. The fabrication of alloy-junction transistors has been automated today to such an extent that units for the entertainment field can be sold for less than a dollar. This enables these transistors to be highly competitive with vacuum tubes also used in that range of frequencies.

The alloy-junction fabrication technique is quite different from the grown-junction technique. Whereas the latter is a batch-processing technique, the present alloy units are essentially made individually (although in large number at any one time). The starting point is a wafer of germanium about 0.080 in. square and 0.003 to 0.005 in. thick. (This wafer is grown similarly to that for a grown-junction transistor, except that the crystal is grown with uniform doping using one type of impurity only, either N or P. In the present discussion, N-type germanium will be assumed.) The crystal is cut into a multiplicity of wafers and then diced into the dimensions required for transistor production. This is indicated above to be about 0.080 in. square and 0.003 to 0.005 in. thick.

An impurity metal, usually indium, is then placed on opposite faces of the germanium and heated. The temperature is above the melting point of indium but below the melting point of germanium. While the indium is molten, it dissolves the germanium actually in contact with it. On cooling, the dissolved germanium recrystallizes onto the undissolved germanium. Since it is freezing from a melt containing indium, the recrystallized germanium is highly doped to P type. Thus we have a PNP transistor with the emitter and collector P type and the base N type, Fig. 3·9. Connections to the emitter and collector are made by wires soldered to the alloy. The other ends of these wires are then spot-welded to leads that make contact to the circuit in which the

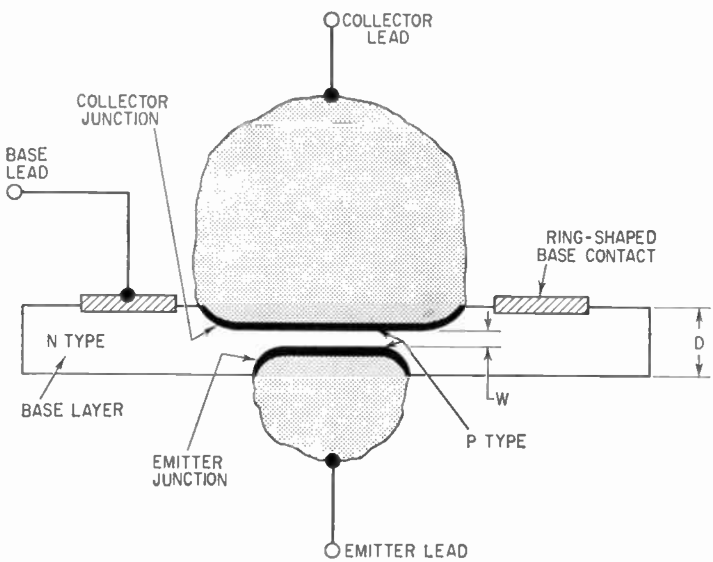


Fig. 3·9 Cross section of an alloy-junction PNP transistor. The collector and emitter elements are grossly exaggerated in size. (*Western Electric Engineer*)

transistor is placed. The base contact is usually made in the form of a ring that completely encircles the emitter. This permits a low-resistance connection to be achieved, Fig. 3·10.

With the alloy method of construction, several things have been accomplished. First, the separation between collector and emitter regions is on the order of only 0.0005 in. This permits a significant reduction in transit time between these two elements. Second, the base resistance can be made low by the use of a relatively thick germanium wafer at all points except in the small section between emitter and collector. Also, the emitter and collector diameters, 0.010 and 0.015 in., respectively, are kept small, thereby reducing the various capacitances which these elements introduce.

The alloy method is feasible for both PNP and NPN transistors. For an NPN unit, a P-type germanium wafer would be used and a pentavalent element (such as phosphorus) would be substituted for the indium.

The diffusion process. In the sequence of transistor development, the discovery of how to apply diffusion techniques to the fabrication of these devices represented a significant advance. Not only did it make commercially possible the manufacture of transistors capable of operating in the kilomegacycle region, but it also provided a means of accurately controlling the formation of a nearly unlimited junction area.

The process of diffusion is a process of mixing, on an atomic scale, two different sets of molecules through the random thermal motion of

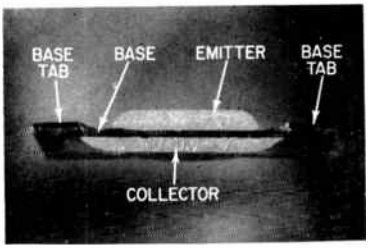


Fig. 3·10 An alloy-junction transistor.

molecules and atoms (Ref. 3·5). For example, if you place a drop of a colored dye in a glass of still water, you will find after several hours that the dye will be spread through the entire glass. The spreading has been accomplished by the random motion of the molecules of the dye and of the water.

The same process will take place in a gas or in a solid. In solids, however, diffusion is ordinarily ignored because it occurs so very slowly. For example, the colored dye mentioned above would, if dropped on a cake of ice, take millions of years to diffuse through the cake thoroughly.

All diffusion processes can be greatly speeded up by heat. In the fabrication of transistor junctions by the diffusion process, a temperature of about 1300°C is generally employed. Even at that temperature, several hours may be required to establish a junction only a few ten-thousandths of an inch below the surface of a semiconductor crystal. If it is desired to achieve a greater depth, the temperature can be raised.

Diffusion is well suited to the introduction of impurities into germanium or silicon crystals. Since the process is slow, the depth of penetration of the impurity can be accurately controlled. Different impurities move with different speeds; hence, their concentration and depth can be varied over a wide range. Also, since the impurity is added *after* the crystal is grown, the crystal itself can be developed under simpler growing conditions, i.e., from a melt possessing uniform characteristics.

The general approach in diffusion fabrication is as follows. After the crystal has been formed by the growing technique previously described, it is cut up into many small pieces and exposed in a furnace to an atmosphere containing one or more impurities in small but known concentrations. The gaseous molecules bombard the crystal surface, gradually forcing their way into the crystal interior. By regulating the amount of impurity present in the atmosphere, the temperature, and the exposure time, it is possible to control to a fine degree the penetration of the impurity into the crystal. If the crystal itself is N-type germanium and the impurity is indium, a PN junction can be formed in this manner.

One other feature concerning the diffusion process is of importance. When an impurity is diffused into a semiconductor crystal, the density of the impurity is greatest at the surface and gradually diminishes as we move into the interior of the crystal. This is useful in transistor operation and is actually taken advantage of, as we shall see.

The drift transistor. The drift transistor combines diffusion and alloy techniques. From this combination, we are able to achieve a significant improvement in the frequency limitation of transistors formed by the alloy method alone. The starting point is a wafer of N-type germanium having a fairly high resistivity. Whenever a semiconductor is said to possess a high resistivity, i.e., a high resistance per unit area, it is meant that the substance possesses only a small amount of impurity. In this case, it is N-type impurity. The wafer of N-type germanium is exposed, under controlled conditions, to an arsenic vapor at a high temperature. The arsenic impurity diffuses into the wafer, leaving the highest concentration at the surface.

This skin of graded arsenic is removed from one side of the wafer and a P-type collector junction is alloyed into the germanium there. The same P-type impurity is then alloyed into the other face of the wafer where the graded arsenic is still present. This produces the emitter junction. The result, as shown in Figs. 3·11 and 3·12, is a transistor in which there is a high density of impurity in the base end nearest the emitter, with a steady decrease until the germanium possesses very little impurity (i.e., reaches the original state of the wafer before the diffusion treatment) somewhat before the center of the base region. From there to the collector, the germanium remains uniform in characteristic. At the collector, the P-type condition appears.

A major difference, then, between the drift transistor and alloy unit resides in the varying impurity distribution that occurs in the base of the drift transistor. Just how this affects transistor operation

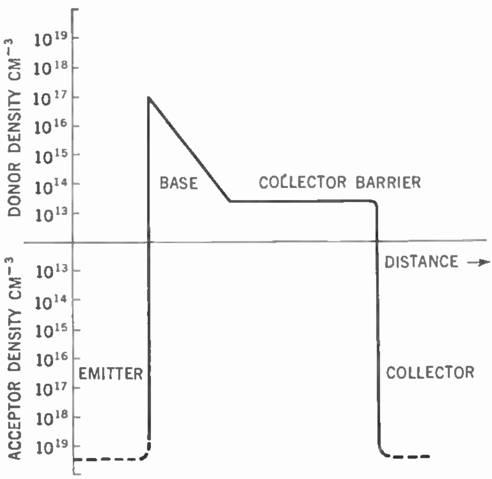


Fig. 3·11 Impurity distribution for a PNP transistor. The section labeled "collector barrier" has, relatively speaking, very few impurities and can be considered as intrinsic germanium.

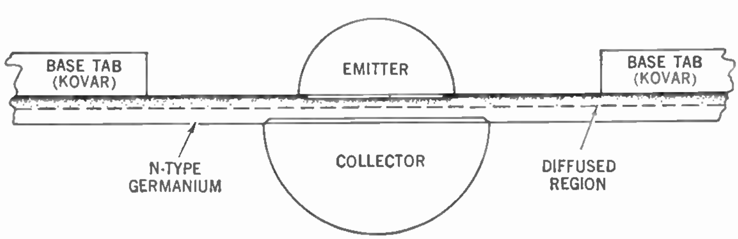


Fig. 3·12 Cross-sectional view of the drift-transistor structure. Note that the base tab is a circular ring.

can perhaps be better seen if we analyze in greater detail the movement of carriers from emitter to collector.

When a signal voltage (or any other voltage, for that matter) is applied between emitter and base, carriers from the emitter are injected into the base. These carriers must travel across the thin base layer and arrive at the collector junction. In a PNP transistor the carriers from the emitter are holes; in an NPN transistor the carriers are electrons. In either case, the carriers travel across the base by a process of diffusion. This motion, brought on by the thermal energy which the holes or electrons possess, consists of movement in random paths, as shown in Fig. 3·13*a*. The particles simply wander about aimlessly, colliding with each other or with germanium atoms and moving in all possible directions. Now, while the individual particles have a random motion, it is possible to obtain a flow of current across the base because the particles, being charged similarly, tend to move from a region of high particle concentration to a region of low concentration. This is shown in Fig. 3·13*b*. If, now, we apply a strong attractive electric field at the low-concentration end, the particles there are constantly being pulled out of the base, thereby encouraging other particles to take their place. This will produce a continuous flow of

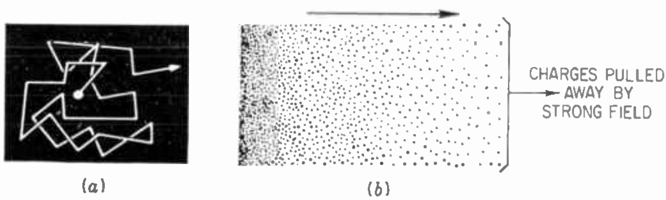


Fig. 3·13 (*a*) Random motion of holes and electrons implies no direction for diffusion, but (*b*) the holes or electrons will diffuse from a region of high concentration to a region of low concentration. (*Bell Laboratories Record*)

particles from the region of high concentration to the region of low concentration.

In a PNP junction transistor there is a high concentration of holes at the emitter end of the base created by the forward bias between emitter and base. These holes are injected by the emitter into the base. At the collector end of the base, there is a strong negative field that is due to the negative voltage applied to the collector and that pulls in all holes that reach this point. Thus, there is a steady current flow across the base region that is due to the diffusive action described above.

A small but measurable time is required for holes injected into the base by the emitter to reach the collector. Note that there are no electric fields *within* the base region. Whatever bias voltage is applied between emitter and base appears across the junction separating the sections. The same is true at the collector junction, where all of the voltage applied between base and collector appears. Once the injected carriers in the base reach the collector (in their aimless wandering), they then travel extremely fast because electrical forces are present there.

Now, if all the injected carriers required exactly the same travel time, the net effect would be simply to delay the output signal with respect to the input signal. In this random travel, however, not all the carriers take the same path, and consequently the carriers (holes or electrons) corresponding to a particular part of the input signal do not all arrive at the collector at the same time. When the signal frequency is low, this minute difference of arrival can be ignored. As we increase the signal frequency, however, some of the late-arriving carriers begin to interfere with the carriers representing the next portion of the signal, with resultant disturbance and cancellation effects. At this point the amplitude of the output signal begins to fall off. The dispersive effect becomes more and more pronounced as the signal frequency rises, and the frequency response continues to decrease.

To minimize this effect, the base section should be made very narrow. As we make the base thinner, however, we steadily decrease the reverse voltage which can be applied between it and the collector section. Also, with exceedingly thin base layers, we not only run into manufacturing difficulties but also encounter irregularities in thickness or in impurity distribution which can result in the collector-to-emitter short-circuit effect of punch-through. (There is also another effect, namely, increased base resistance, which is detrimental to high-frequency operation. This will be considered at a later point.)

In the drift transistor, we achieve the same effect as a thin base region without actually reducing the region to the same extent. By providing a graded impurity distribution in the base region, we establish an electric field there. Holes injected into the base region by the emitter, in a PNP transistor, are accelerated toward the collector. Thus, where previously they traveled aimlessly, they are now more or less directed toward the collector, arriving much sooner than if the base region possessed uniform doping. As the travel time decreases, the maximum operating frequency rises.

The impurity distribution for a PNP drift transistor is shown in Fig. 3·12. The emitter is doped fairly heavily with an acceptor impurity (such as indium). At the junction of emitter and base, the base impurity is at its highest level; thereafter it decreases until the germanium impurity level is quite low. This is somewhere around the center of the base. From there to the collector, the germanium purity remains constant. Since this section of germanium contains very few impurities, it is frequently called an intrinsic-region material; i.e., it is almost pure germanium. The electric field in this intrinsic region is quite strong, and the holes travel through the region quite rapidly. (In an NPN transistor, the base carriers would be electrons.)

Other names for the drift transistor are graded-base transistor and diffused-base transistor.

Meltback transistors. A similar type of graded-base impurity is achieved by the General Electric Company in their high-frequency

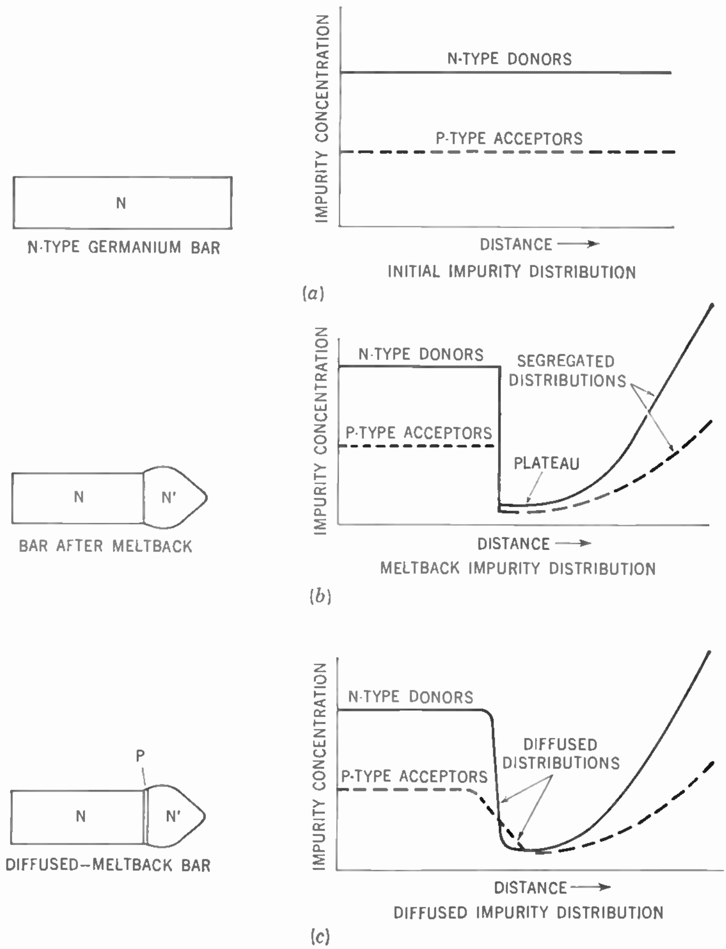


Fig. 3·14 Impurity distributions at each step of the diffused-meltback process.

meltback transistors (Ref. 3·3). The process starts with a germanium (or silicon) crystal which is intentionally doped with both N-type-donor and P-type-acceptor impurities. The doping concentrations are such that the N-type impurity predominates, as shown in Fig. 3·14*a*.

In the second step of the process, called *meltback*, one end of the bar is remelted, or melted back, and then solidified again. This forms the teardrop shape shown in Fig. 3·14*b*. Because the impurities are less soluble in the solid silicon than in the liquid, they will segregate as the melted portion freezes, thereby forming the distribution shown in Fig. 3·14*b*. Just at the dividing line between the portion of the silicon bar (at the left in Fig. 3·14*b*) that was not melted back and the end that was, the initial impurity concentrations drop to very low values. In this region, the impurity is still predominantly *N* type, but because there are fewer impurities, the resistivity here is higher. This high-resistivity section extends for only a few thousandths of an inch, and then the resistivity drops sharply as the impurity concentration rises. Note that after meltback, the base region has not been formed but is simply a junction between two N-type regions of different resistivity.

The last step of the diffused-meltback process is forming the base region, thereby developing the final NPN transistor structure, Fig. 3·14*c*. This is accomplished by subjecting the meltback bar to a long, high-temperature heating cycle lasting many hours. Under these conditions, the impurities on the high-concentration side of the meltback junction diffuse within the solid semiconductor to the "plateau region" of lower concentration. The chief feature of this action is that the P-type impurity has the property of diffusing almost 20 times faster than the N-type impurity in silicon. Therefore, on the plateau side of the junction, there results an excess of P-type impurities over the N type, corresponding to a thin P-type base region. By carefully controlling the entire manufacturing process, bases as thin as 2 microns can be obtained.

The final overall impurity distribution of the diffused-meltback bar is shown in Fig. 3·15, where the net impurity concentration is plotted as a function of distance. The emitter region corresponds to the undisturbed portion of the original silicon bar. Just at the emitter junction, the conductivity is slightly more N type because of the loss of P-type impurities that diffused from the region. The junction from emitter to base is quite abrupt, since the diffusion of the N-type impurities during the heating process is quite limited. Because of the predominant diffusion of the P-type acceptors, the base layer is characterized by a graded impurity, as shown. This introduces a built-in field which decreases carrier transit time in the base region, just as in the drift transistor.

Grown-diffused transistors. The diffusion process has also been applied to grown-junction transistors to produce a device which has a better high-frequency characteristic than ordinary grown-junction units have. This is true primarily because a thinner base can be developed with the grown-diffused method than by the grown approach itself.

The manner in which grown-diffused junction transistors are fabricated is as follows. There are three steps to the process. First, the collector is grown in the conventional manner, Fig. 3·16. Then the growing is stopped and the impurities required for the base

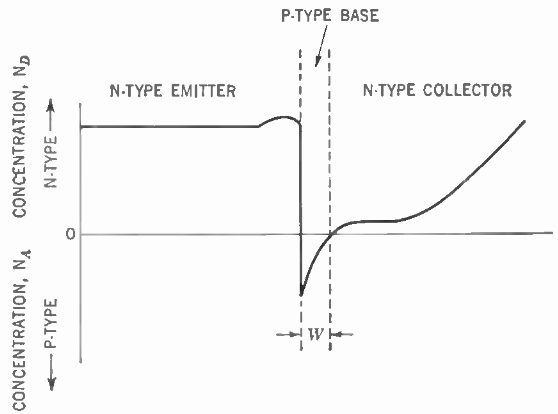


Fig. 3·15 Net impurity concentrations in diffused-meltback transistors.

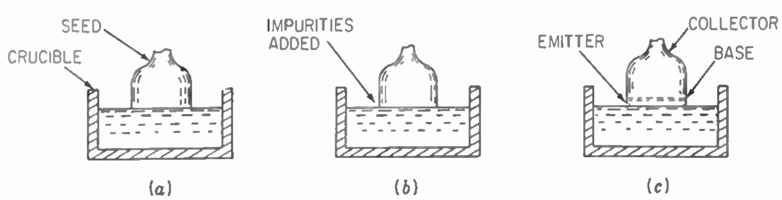


Fig. 3·16 The manufacturing steps of a grown-diffused transistor. (a) Collector region grown. (b) Base- and emitter-producing impurities added. (c) Emitter region grown. Base-producing impurities diffuse into collector region to produce base region.

and emitter are added; after this, the growing is resumed. During the growth that follows, the base-producing impurities diffuse up into the collector, thereby producing a narrow base region. The thickness of the base region depends on the relative doping levels, upon the impurities used, and upon the growth rate and time taken to grow the emitter region.

This approach is possible because donor impurities diffuse more rapidly than acceptor impurities in germanium and thus a PNP structure is produced. In silicon, the acceptor impurities diffuse more rapidly than the donor impurities and NPN structures are thereby produced. Note that because the base is formed by diffusion, the impurity distribution is graded, and this helps the high-frequency response.

The mesa transistor. The mesa transistor represents still another approach to the fabrication of high-frequency transistors. This method is being widely employed because it lends itself to a relatively simple fabrication technique and because its heat-dissipating characteristics are more favorable than those of alloy-junction transistors.

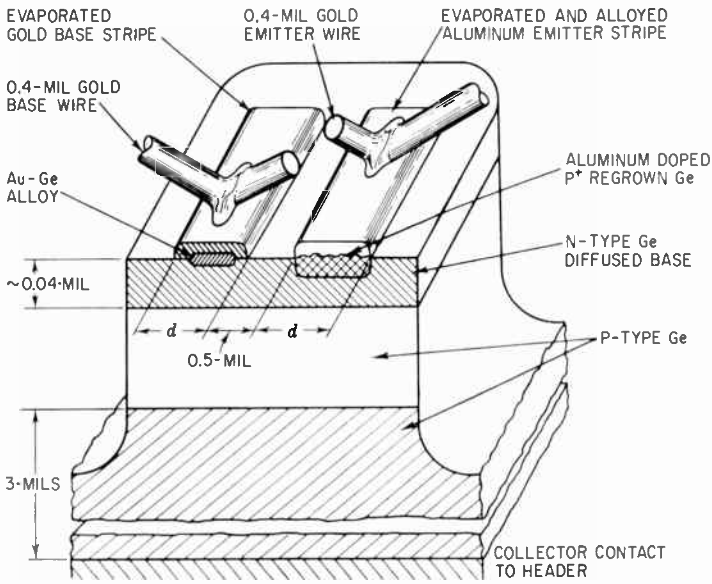


Fig. 3·17 A mesa transistor.

The internal structure of a mesa transistor is shown in Fig. 3·17. To start, a wafer of P-type germanium is placed in a hydrogen atmosphere containing antimony which diffuses into the germanium to a depth of about 0.04 mil. The next step is to evaporate gold and aluminum stripes onto the diffused surface. This is done through a rectangular hole in a molybdenum mask that is placed over the top of the germanium surface, Fig. 3·18. The aluminum and gold evaporation sources are located at different positions to produce the two stripes. An alloying operation follows the evaporation of the aluminum stripe. In this step a P-type emitter is formed. Thus, we have a P-type emitter, an N-type base, and a P-type collector. The gold stripe forms an ohmic connection to the base. After the foregoing operation, the transistor element is attached to the plated-metal header.

The transistor wafer, to this point, has a rectangular structure. In order to reduce the area of contact between the base and collector, corner sections are cut away on either side to achieve the physical form shown in Fig. 3·17. A mesa is formed, with the collector flaring out at the bottom so that it covers a greater area at its base than the area of contact between the collector and base. By developing this shape, the lowest possible collector capacitance is formed.

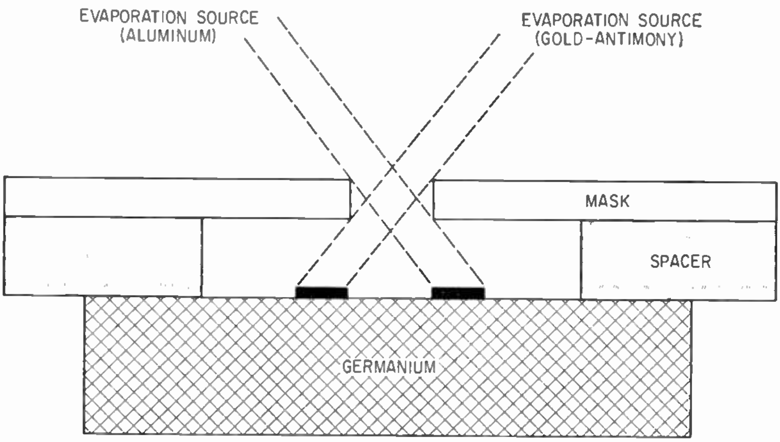


Fig. 3·18

The job is done by masking the hill, or mesa, area with wax and then stream-etching by using a mixture of nitric and hydrofluoric acids. The height of the mesa is approximately 0.5 mil. As a final step, gold leads are bonded to the base and emitter stripes for external connection to these elements (Ref. 3·7).

A variation of the mesa transistor is the planar transistor manufactured by Fairchild Semiconductor Corporation. Both units have basically the same structure; the difference between them lies in the manufacturing process. The planar transistor is fabricated by masking the entire surface of the germanium or silicon by an oxide layer except for those areas where diffusion is to occur. This results in an improvement in those characteristics which are sensitive to surface conditions, such as reverse leakage currents, breakdown voltages, noise figure, and current gain.

Both the mesa (and planar) constructions offer a wider collector which, when properly connected to a metallic header, provides a relatively large heat-dissipating surface. Since heat dissipation is always an important problem with transistors, any physical structure which facilitates heat removal is more desirable than structures which do not.

Epitaxial transistors. Recently, a new manufacturing technique has been developed for the fabrication of high-frequency transistors, particularly those to be employed for switching operations. The structure is again basically that of the mesa; however, the collector consists of two regions instead of one, Fig. 3·19. The larger portion of the collector is heavily doped to provide a very low resistivity. This is shown in Fig. 3·19 as N$^+$, indicating that there is a relatively high concentration of donor atoms. Over this is deposited a thin

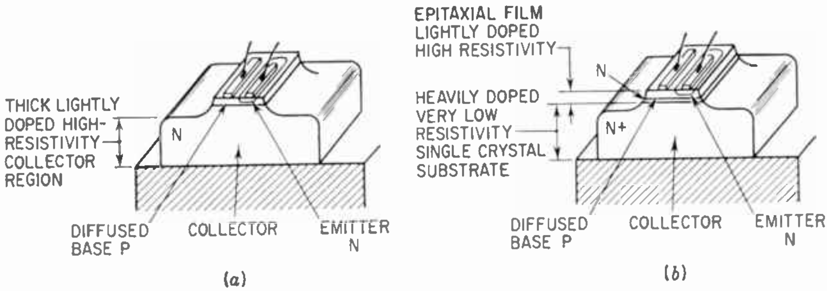


Fig. 3·19 (a) Mesa transistor. (b) Epitaxial transistor.

film, 0.1 mil thick, of very lightly doped semiconductor material. The thin film combines homogeneously with the crystalline structure below it (called the substrate) so that there is no discontinuity or break in the crystal structure. From this point, the standard techniques for fabricating mesas are used.

By this construction, lower collector resistance is obtained when the transistor is saturated (see page 106). The construction also results in a lower collector capacitance, and both of these features permit faster switching time, i.e., going from an "on" to an "off" condition and vice versa. (The use of a transistor as a switch will be discussed in Chap. 8.)

Electrochemical transistor. Another high-frequency transistor is the Philco surface-barrier transistor (Ref. 3·2). In cross-sectional appearance, Fig. 3·20, this unit closely resembles an alloy junction transistor. In mode of operation, however, it represents another approach to a solid-state amplifier. In all the junction transistors described thus far, two forms of germanium were employed, that is, P and N germanium. In the surface-barrier transistor, only one type of germanium, N germanium, is used. Electrodes serving as the emitter and collector are plated electrolytically on opposite faces of a germanium wafer. A metal contact is then soldered to one end of the crystal and serves as the base electrode. Unlike the situation in junction transistors, the emitter and collector electrodes remain coated to the surface of the germanium. There is no penetration of the germanium lattice structure by the atoms of the metals.

To understand the operation of the surface-barrier transistor, additional facts concerning the behavior of electrons inside a crystalline structure must be known. It has been found that energy levels,

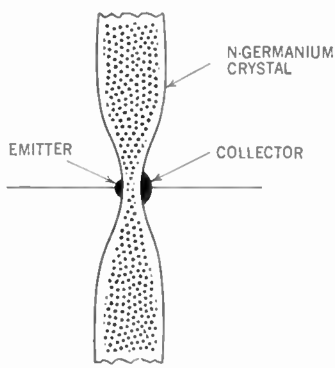


Fig. 3·20 Schematic cross section of a surface-barrier transistor.

or orbits, may exist on the surface of a crystal and not be found in the interior. It is believed that no orderly structure of energy bands (such as we have described in Chap. 1) exists on such a surface. Rather, the leftover bonds of germanium atoms, together with any atoms of other substances on the surface, form a two-dimensional solid with properties which are entirely different from those in the interior. It is felt that there are no forbidden bands among the surface atoms comparable to the forbidden bands found in the interior atoms. It is because of this absence that a number of free electrons move to the crystal surface and concentrate there in sufficient strength to produce a negative field which repels other electrons of the N germanium toward the interior, thereby creating a nearly insulating region containing a strong electric field just beneath the surface. This is the reason the N germanium in Fig. 3·20 is shown shaded in the interior but left unshaded in a narrow strip along the surfaces. The insulating strip is referred to as a surface barrier.

A metal electrode which is brought in contact with the germanium crystal, Fig. 3·20, can communicate with the main body of the crystal only through the surface-barrier region. If we apply a negative potential to the metal plate, it will further repel the interior electrons away from the surface and cause the intervening layer to become wider. Making the metal electrode positive will attract the interior electrons and reduce the width of the insulating layer. Current flow between the surface electrode and the inner portion of the crystal can thus be made smaller or larger as desired.

To form a transistor with the N germanium, we require an appropriate distribution of holes which will travel from emitter to collector as they do in a comparable PNP junction transistor. In the surface-barrier transistor it is found that a population of holes exists just under the germanium surface. These holes arise from the valence electrons that are thermally excited enough to leave their atoms and move into some of the energy levels at the surface which are intermediate between the conduction band and the valence band. The electrons come from the atoms located near the surface, and for every such electron departure, a hole is created. This action is confined to the atomic layer just below the surface; the rest of the germanium interior produces relatively few holes. Some metal contacts produce a denser hole population under the surface of the germanium than others. The most useful metals for this purpose are indium and zinc.

In review, then, we see that the surface-barrier transistor owes all its characteristics to the special conditions which exist at the surface of a crystalline structure. The strong electron field at the surface forces free electrons to remain in the interior. Also, because of the presence of intermediate energy levels at the surface, holes are found concentrated just below the surface. When a metal contact to the crystal is made positive, it repels these holes through the barrier. This would be the emitter electrode. The other electrode, the collector, is reverse-biased (i.e., biased negatively), and holes coming within its field after passage through the germanium body will be drawn to the surface.

The surface-barrier transistor thus consists of a germanium crystal which forms the base plus two metal electrodes, on opposite faces of the crystal, which serve as the emitter and collector electrodes. A positive emitter will drive the holes toward the collector, but at the same time it will attract the interior electrons. For efficient transistor operation, the electron current should be reduced as much as possible, since only the hole current is desired at the collector. This was achieved by bringing the collector electrode within 0.0002 in. of the emitter. The negative charge on the collector drives the germanium free electrons away from the emitter, while at the same time, it presents a greater attractive force for the holes.

To achieve the minute spacing required, a process of "electrolytic machining" is employed. Two tiny jets of a metal salt solution are directed from miniature glass nozzles toward opposite faces of a germanium wafer, with the latter serving as the anode and electrodes in the glass nozzles serving as the cathodes. This action etches away the germanium under each jet until the desired amount of material

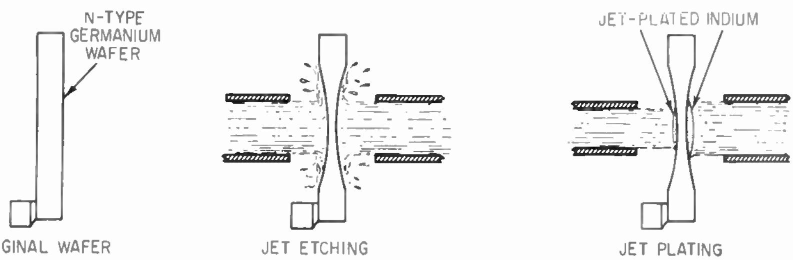


Fig. 3·21 Jet-etching and plating the surface-barrier transistor.

has been taken away. Then the voltage polarity is reversed, and now the same jets are made to electroplate the metal ions of the salt solution directly upon the freshly etched surface of the germanium, Fig. 3·21. This forms the desired emitter and collector electrodes. Thus, electrolytic machining is a very efficient and ingenious method of using a salt solution to accomplish both etching and electroplating. Diameter of the emitter is 0.005 in., and that of the collector is 0.007 in.

A later version of the surface-barrier transistor (commonly called SBT for short) is the microalloy transistor (MAT). This differs from the surface-barrier transistor in that it employs extremely shallow alloy contacts in place of the surface-barrier contacts. The result is a higher carrier injection efficiency (emitter to base) so that greater current gains can be achieved at higher frequencies.

The microalloy transistor is fabricated as follows (Ref. 3·4). A germanium blank is mounted on a base tab and jet-etched to a thickness of about 0.1 mil. Indium contacts of the order 3 to 8 mils in diameter are jet-plated in the bottom of the etch pits to a thickness of about 1 mil. A lead wire having a drop of solder composed of ¼ to 1 per cent gallium in indium is brought into contact with the jet-plated electrode, Fig. 3·22*a*. A hairpin-shaped heating element supplies the heat for the solder operation.

When the surface-barrier transistor is fabricated, the low-melting-point solder that is used prevents the indium from melting during the soldering of the lead. In the MAT, the indium-gallium solder melts at the same temperature as the jet-plated indium dot. The solder flows down into the jet-plated electrode and dissolves a small amount of germanium. Upon freezing, or solidifying, the germanium

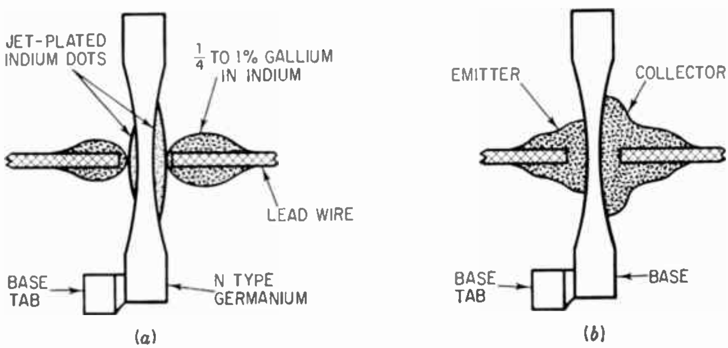


Fig. 3·22 **(a) Cross section of jet-etched jet-plated blank microalloy transistor with lead wires not soldered. (b) Lead wires soldered.**

is redeposited back onto the original germanium blank. This redeposited germanium is highly doped P type by the indium, Fig. 3·22*b*. Thus, we now more closely approach the PNP alloy-junction transistor.

A still later version is the so-called microalloy diffused transistor (MADT), in which the base has a graded or diffused impurity. This provides still higher frequency operation than either of the preceding two types. Prior to the jet-etching process, impurities are diffused into the starting material in a manner similar to that already described for the drift transistor. Thereafter, the processing continues as it does for the MAT.

The foregoing discussion has covered the most important types of transistors and their fabrication techniques. Variations of these methods wherein several methods have been combined to achieve a desired set of operating characteristics will be found. In time, newer techniques which will either modify existing methods or replace them completely may come along. At the present time, however, the frequency capabilities of transistors extend into the kilomegacycle range.

Cutoff Frequency

The frequency characteristic of a transistor is often given in terms of a cutoff frequency. One such frequency is that at which the ratio of collector or output current to emitter or input current drops to 0.707 (i.e., 3 db) of its value at 1,000 cycles. This ratio, incidentally, is designated by the Greek letter alpha (α).

It will be found that transistors will operate as oscillators at much higher frequencies than the indicated cutoff frequency. On the other hand, for amplification, a transistor having a 3-db drop in the current-amplification factor at, say, 3 Mc may have a significant reduction in voltage or power gain at 0.5 Mc or less. Thus, for any useful gain, it is frequently necessary to restrict the operation of a transistor well below the rated cutoff frequency.

There are other ways of specifying the frequency characteristic of a transistor, for example, f_T, f_{max}, or $f_{\alpha e}$. These will be discussed after we have examined the operation of transistors in various circuits.

Temperature Effects

The ability of properly processed germanium to serve as a transistor depends wholly on the electronic bonds and lattice structure existing within the germanium crystal. It was noted previously that too high a concentration of impurities will increase the conductivity of the germanium to such an extent that the effectiveness of the base in controlling emitter and collector current is destroyed. Conductivity will also rise with temperature. An increase in thermal energy will lead to more broken covalent bonds, more free electrons and holes, and a greater current flow in both input and output circuits for the same applied voltages. This, in turn, will reduce control of the collector current by the base and practically nullify the transistor action in the germanium. It is even possible for the thermal action to feed on itself and eventually destroy the transistor completely. The higher temperature results in more current, which raises the temperature even higher, which results in more current, etc., until the entire unit is permanently damaged. (This effect is frequently referred to as thermal runaway.)

Even if the increase in temperature does not prove detrimental to the transistor, it can be the cause of distortion because of a shift in operating point. When a transistor is employed in a circuit, say an amplifier, a suitable operating point is selected. This operating point represents a certain collector current at a certain collector voltage, perhaps 5 ma at 6 volts. When the signal enters the stage, it varies the collector current and the collector voltage follows suit. These variations in the collector circuit represent the output signal.

Suppose, now, that because of temperature, the operating collector current increases (in the absence of a signal) to 6 ma. This will cause a greater voltage drop across the load resistor and leave less voltage for the collector. Signals now passing through this stage can swing the collector voltage over a smaller range before cutoff occurs.

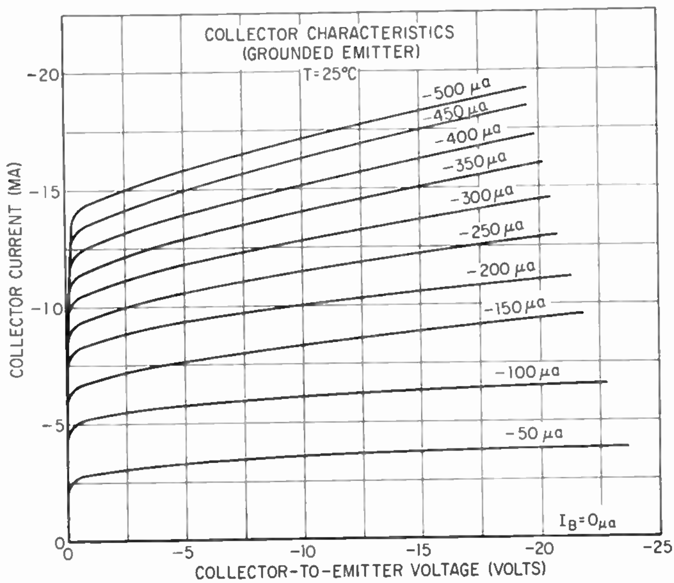


Fig. 3·23 Typical collector characteristic curves for a transistor.

Furthermore, as the collector voltage decreases, the transistor characteristic curves tend to become more nonlinear, Fig. 3·23, and this, too, is a source of distortion. Finally, by raising the collector-current operating-point value, more collector dissipation develops. This leaves less leeway for the signal before the maximum safe dissipation value is reached.

In transistor-characteristic charts, such as we shall study presently, the maximum collector dissipation is specified at a definite temperature (generally 25°C). If the operating temperature exceeds this value, then it becomes necessary to lower the collector-dissipation rating.

Sometimes, two maximum temperature ratings will be given: one, the lower one, in free air; the other when the transistor is mounted flush against a metallic surface (such as an aluminum chassis) which will conduct the heat away. These heat conductors are the heat sinks previously discussed on page 54, and they can make an appreciable difference in the maximum dissipation rating. For example, one Sylvania NPN transistor has a collector-dissipation rating in free air of 2.5 watts and a rating of 4.0 watts when mounted flush against an aluminum chassis. This is a significant point to keep in mind, particularly when the transistor is to be operated near its maximum rating.

In addition to an operating temperature, there is a maximum storage temperature. This is determined by life tests and is limited by the melting point of the materials within the transistor case or by parameter stability. Parameter instability (i.e., changes in operating characteristics) results because of the increased chemical action at elevated temperatures affecting the surface of the transistor. This, in turn, affects the transistor parameters or characteristics.

For example, the α cutoff frequency of a transistor will generally decrease with temperature. Another important parameter, I_{CBO}, increases with temperature, approximately doubling in value for every 10°C rise. I_{CBO} is an important characteristic that will be encountered again and again, and some explanation of it is in order. I_{CBO} is variously called the collector leakage current and the collector saturation current. Actually, saturation current is a preferable term, since leakage is only one component of its total value.

I_{CBO} is the current that flows through a transistor when the collector-base junction is reverse-biased and the emitter is open, Fig. 3·24. The notation I_{CBO} was selected to reflect this fact. Thus, the *CB* following the capital letter *I* stands for collector and base. They are the two elements that are reverse-biased. The condition of the third element, here the emitter, is open-circuited, as indicated by the letter *O*.

With the collector-base diode reverse-biased, one would expect zero current. In reality, however, it is impossible to fabricate P-type semiconductors without some free electrons or N-type semiconductors without some excess holes. These minority carriers account for a small current flow even though the collector-base diode is reverse-biased. Current flow takes place in the following manner. The minority carriers from each section, electrons in the P region and holes in the N region, are forced toward the junction under the influence of the applied battery voltage. Some of the electrons and holes combine and neutralize each other's electrostatic charge. But this loss of electrons from the P section upsets the electrical neutrality that originally existed there, leaving it with a net positive charge. The positive field attracts additional electrons from the negative terminal of the battery. At the same time, the loss of holes (due to the same action above) leaves the N region with a net negative charge. (Remember, both sections were electrically neutral *before* the battery was connected to them.) To satisfy this charge, additional holes are formed in the N section, while the electrons that were combined with these holes move to the positive terminal of the battery.

The foregoing provides a continuous flow of current, with electrons traveling from the negative battery terminal into the P section, while

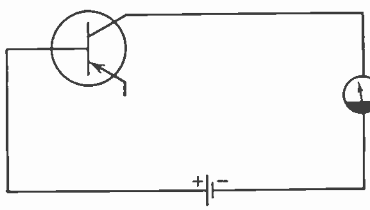


Fig. 3·24 The manner in which I_{CBO} is measured.

electrons from the N section enter the positive battery terminal. Since the number of minority carriers originally available was small, the current under reverse-bias conditions is also small.

The small current that flows when the collector-base diode (or any PN junction) is reverse-biased actually consists of two components. One is the current indicated above, i.e., the current produced by the minority carriers in each section. This current is the saturation current, and it doubles for approximately every 10°C rise in junction temperature. The heat sensitivity stems from the fact that as the temperature rises, thermal energy is imparted to the atoms forming the diode, enabling some of the previously held valence electrons to break loose from their covalent bonds. And, of course, for every electron that breaks loose, a corresponding hole is created. Thus, the supply of minority carriers rises and, with it, the reverse current rises.

The second component of I_{CBO} stems from the leakage of current across the outer surfaces of the transistor. This component is voltage dependent because it basically follows Ohm's law, $E = IR$. It is also somewhat temperature sensitive, however, probably because the resistance presented by the semiconductor surface itself is temperature sensitive.

Although I_{CBO} is ordinarily less than 10 μa for germanium at 25°C, it rises rapidly as the temperature goes up. Thus, at 65°C, it will attain a value of 160 μa, or 0.16 ma, a value too large to be ignored. I_{CBO} is detrimental to transistor operation because it represents an uncontrolled current which will, through its presence, upset the designated operating conditions for the circuit. Hence, I_{CBO} must be taken into account in any circuit design. I_{CBO} is the full notation for this current; frequently, however, it is shortened to I_{CO}.

The collector-current rise with temperature may also be caused by a change in the emitter-to-base resistance. This effect is discussed in Chap. 12.

Silicon Transistors

The emphasis in the preceding discussion has been on germanium as the semiconductor material from which transistors were made. It is also possible to employ silicon as the fundamental building block for a transistor, and this is indeed being done. Silicon is suitable for diode and transistor operation because its physical properties closely parallel those of germanium. Thus, silicon is a semiconductor with four valence electrons and, in the solid state, will form a cubic crystal lattice in which the various atoms are held together by the same mechanism of covalent bonds. It is possible to replace some of these atoms by impurities, of either the donor or acceptor variety, and form N-type or P-type silicon. By combining suitable P- and N-type sections of silicon, rectifier diodes or complete transistors can be fabricated.

In view of the physical similarities between silicon and germanium, it is only natural to investigate both substances to determine which is best suited for transistor operation. Actually, as we shall see, there are certain advantages and disadvantages to either, and it becomes a matter of selecting the unit that possesses the greatest suitability for a specific application.

Of the two substances, silicon is far more abundant in nature than germanium. As a matter of fact, silicon compounds form over 85 per cent of the earth's crust. All sand, for example, is silicon dioxide, while additional silicon compounds are present in many rocks. Unfortunately, silicon is never found in the free state, and in order to utilize it for transistor manufacture, extensive separation and refining methods must be employed. This represents a major obstacle, because silicon is not easily reduced to the pure state. It is an extremely difficult substance to melt or purify, and its processing requires high-powered, complex, and expensive furnaces. In spite of these difficulties, silicon is being mass-produced, and an increasing variety of silicon diodes and transistors is being marketed.

One of the most important advantages of silicon, which is in large measure responsible for much of the attention being devoted to this element, is the low collector saturation current I_{co} which silicon transistors exhibit. Table 3-1 contains a comparison between what might be termed typical junction-transistor characteristics for both silicon and germanium, and the great disparity in I_{co} values is readily observed. The ratio might be anywhere from 100:1 to 500:1. The rate of increase in I_{co} with temperature is about equal for both types of transistors. Since the value of I_{co} for silicon is so extremely small at room temperature, however, the unit can be used at much higher temperatures before it becomes troublesome.

Another property in which silicon excels is its collector resistance r_c. This value is higher than the comparable collector resistance of germanium transistors. As we raise the operating temperature, r_c in both types of transistors will decrease, but since we start with a higher value in silicon, it is possible to go to higher temperatures before r_c becomes too small to use. When the behavior of I_{co} and r_c is considered, it is seen why silicon transistors possess higher maximum dissipations and why they are useful as high as 150°C.

These characteristics are directly attributed to the larger energy gap that exists between the valence band and the conduction band in silicon atoms. It will be recalled, from Chap. 1, that in order for an electron to jump from the valence band to the conduction band, a certain amount of energy is required. More energy is needed to accomplish this jump in silicon than in germanium. This same factor also explains why more bias voltage is needed to produce a certain current in the emitter. This is indicated in Table 3·1, which also shows that the base resistance is higher in silicon transistors. When the silicon unit is employed as a grounded emitter, the higher base resistance requires more driving power from the preceding stage.

Both current and power gains of silicon units are lower than those of corresponding germanium transistors. The mobility of electrons in silicon is about one-seventh that of electrons in germanium. This tends to work against the frequency response of silicon transistors. There are, however, other factors that have to be taken into consideration when considering frequency response. For example, collector capacitance C_c is a very important frequency-determining factor (as we have already seen in this chapter), and in silicon units this value is lower than in germanium transistors. Another compensating factor is that higher collector voltages may be employed with silicon transistors; this serves to decrease further the effective value of C_c and thus aid the frequency response.

Silicon transistors are currently being fabricated by essentially the same methods as those employed for fabricating germanium transistors. These include grown-junction, alloy-junction, diffusion, and meltback techniques. The manufacturing processes are more difficult, and the best silicon crystals produced to date are still not so pure

Table 3·1 Comparative Characteristics of Silicon and Germanium Transistors

	Symbol	Silicon Grown	Germanium Grown	Germanium Alloy
Collector:				
Voltage (maximum), volts	$V_{c,\max}$	40	40	25
Dissipation (maximum), mw	. . . . .	150	50	50
Cutoff current, μa	I_{co}	0.02	2	10
Capacitance, μμf	C_c	7	14	40
Conductance—parallel, μmhos	g_c	0.3	0.2	1.0
Emitter:				
Current (minimum usable), ma	I_e	1	0.01	0.1
Reverse voltage (maximum), volts	$V_{e,\max}$	2	10	5
Bias voltage, mv	V_e	500	160	160
Resistance, ohms	r_e	100	25	25
Base, resistance, ohms	r'_b	500	150	300
Gain:				
Power, db	G_o	35	47	40
Current	B	26	35	40

SOURCE: *Electronic Design* magazine.

as comparable germanium crystals. In spite of this, silicon transistors which can operate above 100 Mc have been produced.

Life Expectancy

An important consideration in the application of any electronic device is its life expectancy. How long will this component last under normal operating conditions? In the case of transistors this is an especially pertinent question since they, like tubes, form vital links in circuits.

Failure of a transistor can be one of three kinds: mechanical failure, bulk failure (arising from some defect in the bulk of the material), or surface failure (Ref. 3·6). We are not concerned here with failure due to misuse, such as a current overload or the application of too high a voltage. Only failures arising from manufacturing difficulties will be considered.

Mechanical failure. Into this category fall such things as poorly made connections, excessive strain on sections of the transistor, the application of too much heat during the soldering operation (when wires are soldered to the transistor sections), or the differential expansion between adjoining parts. When a transistor is properly made mechanically, it will withstand centrifugal forces with accelerations as high as 31,000 times the force of gravity (i.e., 31,000 *g*) and impact tests with accelerations as high as 1,900 *g*. This is far in excess of the forces which will completely shatter most conventional vacuum tubes.

Bulk failure. Bulk failure arises from changes in the internal crystalline structure of the transistor body. For example, impurities that exist on the inside of the transistor housing or on the surface of the transistor can, in time, diffuse into the body of the transistor and alter the internal structure, particularly in the neighborhood of the junction. Fortunately, at normal operating temperatures, such diffusion is quite small and can ordinarily be neglected.

Of far greater importance is the effect of nuclear radiation which can penetrate into the transistor body and displace atoms from their normal positions, thereby causing them to appear at points where they were not intended to be. This disrupts the crystalline structure and modifies the electrical behavior of the unit. For example, the resistivity may decrease and thereby increase the leakage current for a particular applied voltage. Or the current gain may be affected, usually lowered, because the disrupture of the crystal structure in the base prevents as many minority carriers from reaching the collector.

The effects of nuclear radiation are generally cumulative, so that what may start out to be only a minor change in characteristics ends up by becoming a pronounced change. It is for this reason that transistors must be carefully shielded from all types of nuclear radiation.

Surface failure. By far the greatest number of transistor failures stem from changes at the surface of the transistor. Water vapor, for example, which condenses on a transistor surface will decrease the collector breakdown voltage and increase the collector current. Neither condition is desirable. A reduction in the collector breakdown voltage means that less collector voltage can be applied to the unit, and this limits output power. An increase in collector current means higher internal heat generation at the operating point and hence less leeway for the signal.

To avoid these changes, transistors are now hermetically sealed in a dry atmosphere. Even with hermetic sealing, surface changes may still occur. To minimize them, many manufacturers place a coating over the transistor surface which acts to protect the surface without reacting with it. Such transistors may be encapsulated with an inert gas such as hydrogen, argon, or helium, or the housing may be filled with an inert substance such as silicone grease.

Finally, it must be appreciated that in transistors we are dealing with extremely small dimensions, particularly at the junctions. Even a microscopic particle of dust falling across a junction can completely short-circuit it. Transistors must thus be fabricated under the most sterile conditions, and almost every transistor-fabrication plant has one or more "white" rooms where extensive air-cleansing equipment constantly filters out any dust particles that may be brought in.

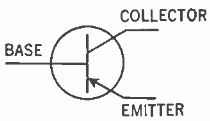


Fig. 3·25 The schematic symbol for a transistor.

These precautions extend even to the personnel, who are required to pass through special outer rooms where much of the dust from their clothing is removed. They then don white laboratory coats and even special head coverings so that any dust or dirt they may still carry on their bodies or clothing does not enter the atmosphere of the workroom. This is a far cry from the working conditions under which other components are fabricated, but the minuteness of the transistor structure and the sensitivity of the transistor to contamination make the precautions obligatory.

Transistors are too new a development to have enabled us to gather sufficient data concerning their full life span. It is, however, expected that they can be made to last more than 200,000 hr, either on the shelf or in operation. This means that a transistor might operate continuously for 20 years, a period greater than that we can now reasonably expect from all but a very few specially made vacuum tubes.

Transistor Symbols

Circuit symbols for transistors are still somewhat in a state of flux, although the point at which one will see fewer and fewer variations has been reached. The basic symbol for a transistor is shown in Fig. 3·25. The emitter element has an arrowhead, the base is a straight line, and the collector element is shaped like the emitter, but it possesses no arrowhead. This symbol is used for both point-contact and junction transistors and, like the tube symbol, may be placed in any position.

To distinguish between NPN and PNP transistors, the method shown in Fig. 3·26 is employed. If the emitter is a P-type germanium, the arrowhead is directed in toward the base. On the other hand, if N-type germanium is used for the emitter, the arrowhead points away from the base.

Three minor variations that have been employed are shown in Fig. 3·27. Their use is not extensive, and actually they would not cause any confusion. In addition, there are other types of transistors,

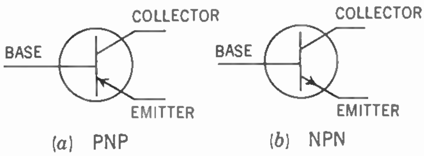


Fig. 3·26 To distinguish between PNP and NPN transistors, the method shown is frequently employed. This notation will be followed in this book.

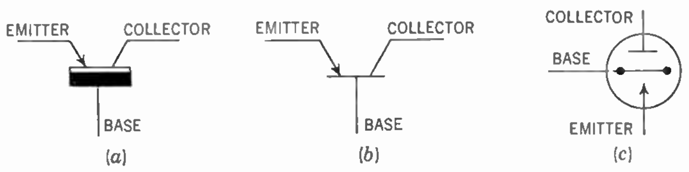


Fig. 3·27 Three additional symbols that have been employed to represent transistors.

and where their form differs significantly from the transistors discussed, other symbols are used. A number of these transistors are examined in Chap. 8, and appropriate symbols will be given there.

Comparison with Vacuum Tubes

Transistors are designed to perform the same functions as vacuum tubes, and it is therefore only natural to want to compare the two electrically to see wherein they differ and wherein they are similar. As a first step, let us consider these two components in the light of their internal operation. In a transistor, current flow through the various germanium sections is initiated by the flow of electrons or holes from the emitter section. In a vacuum tube, this initiation starts at the cathode. Thus we could say that the emitter in a transistor is equivalent to the cathode in a vacuum tube. (The word "emitter," of course, is a clue to the function of this element.)

The recipients of this current flow are the collector in the transistor and the plate in the vacuum tube; hence, these two elements can be considered to be equivalent in their actions. This still leaves the grid in the vacuum tube and the base in the transistor, and the equivalence of these elements is seen in the fact that whatever current flows from emitter (or cathode) to collector (or plate) must flow through the base (or grid) structure. Current flow in both devices is governed by the potential difference between emitter or cathode and base or grid. Figure 3·28 illustrates these analogies between transistor and vacuum tube.

The next step is to consider both devices in terms of the d-c voltages which are applied to their elements. In a vacuum tube the grid is practically always biased negatively with respect to the cathode. This makes the grid impedance very high (except at high

Fig. 3·28 Comparable elements in tubes and transistors: grid and base, cathode and emitter, and plate and collector.

frequencies, where other effects enter the picture). The plate, on the other hand, is always given a potential which is positive with respect to the cathode. The purpose of the plate is to attract the electrons emitted by the cathode, and since electrons possess a negative charge, a positive potential is needed to attract them.

In the transistor, conditions are somewhat different, though we wish to accomplish the same purpose. To initiate a flow of current, there must first be a flow of current between emitter and base and the bias battery must be connected to produce that current flow. This is what determines the polarity connections of the bias battery. If the emitter is formed by P-type germanium, the base will contain N-type germanium and current flow will occur between these sections when the positive battery terminal connects to the P-type emitter and the negative battery terminal connects to the base. We have spoken of this as forward-biasing, and under these conditions the impedance of the emitter circuit is low. Here, then, is a marked departure from conventional amplifier practice as we know it now.

When we employ N-type germanium for the emitter and P-type germanium for the base, we must reverse the battery connections if we are to obtain the desired current flow through the emitter. Thus, the guiding thought in the emitter circuit is current flow, and we alter the battery conditions to suit the type of germanium being used in order to achieve our objective. Here is a radical departure from anything we have known in vacuum-tube practice, and it points up something which we have hinted at throughout the preceding discussion. That is, transistors are current-operated devices, while vacuum tubes are voltage-operated. α, for example, is the symbol representing the ratio $\Delta I_C/\Delta I_E$, where ΔI_C is the change in collector current and ΔI_E is the change in emitter current. The counterpart of this symbol in the vacuum tube is μ, the ratio of a voltage change in the plate circuit produced by a voltage change in grid circuit. Again we see the emphasis on voltage in a tube as against current in the transistor.

In the collector circuit the proper battery biasing is such that the current flow is reduced to a minimum. (Note that it is not zero, although it is only on the order of microamperes.) This is known as reverse-biasing and is always true with collectors. To attain this condition, we must connect the battery in accordance with the type of germanium used in the transistor. If the collector is formed of P-type germanium and the base has N-type germanium (in a PNP junction transistor), then the negative terminal of the battery goes to the collector and the positive terminal to the base. Conversely, if the collector has N-type germanium and the base P type, the reverse is true. Great care must be observed when connecting the collector battery because biasing in the forward direction may cause so much current to flow through the collector-base sections that the resulting heat will permanently damage the transistor and render it unfit for further use. The excess current flow is due to the higher potential of collector batteries. While the applied emitter voltage is generally less than 1 volt, the applied collector voltage can be as high as 22½ volts. A junction biased in the forward direction with this high a voltage will receive enough heat from the ensuing current to be permanently affected. In a vacuum tube, no similar condition exists and we have never had to observe this precaution.

Basic Transistor and Vacuum-tube Amplifier Circuits

All vacuum-tube amplifiers can be divided into three classifications: grounded cathode, grounded grid, and grounded plate. A similar division exists for transistor amplifiers: grounded emitter, grounded base, and grounded collector. In the sections to follow, the basic differences among these groups will be examined. Further elaboration will then be made in the succeeding chapters dealing with applications.

Classifying vacuum-tube amplifiers in terms of grounded cathode, grounded grid, and grounded plate is a practice that has recently begun to gain favor among workers in the field. The term "grounded-grid amplifier" is not particularly new, but it has not been common practice to call the conventional amplifier a grounded-cathode ampli-

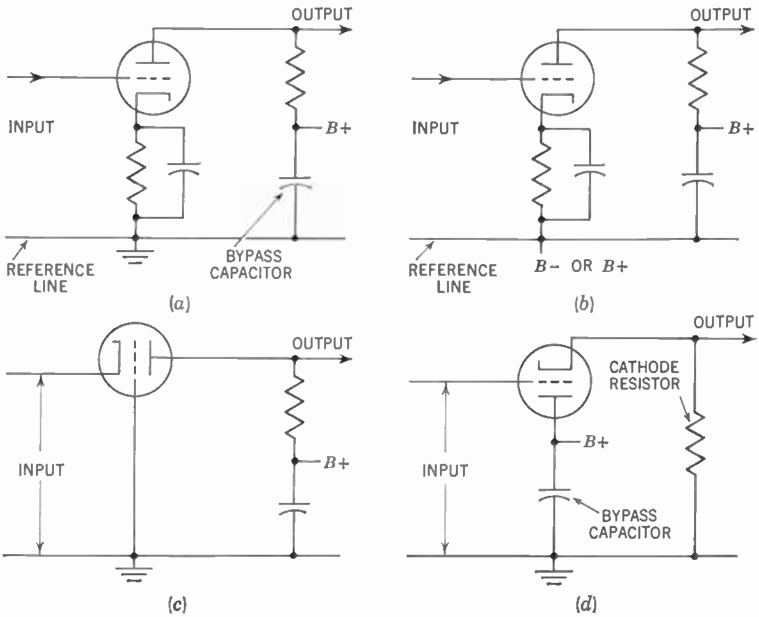


Fig. 3·29 (a) Grounded- or common-cathode amplifier. (b) Same amplifier with cathode returned to a d-c reference voltage instead of to ground. (c) Grounded- or common-grid amplifier, where grid is common to both input and output circuits. (d) Grounded- or common-plate amplifier. If the last stage appears strange, turn it around and the familiar cathode follower will be immediately evident. All the amplifiers have been drawn in the manner shown to bring out the reason for their indicated designations.

fier or the cathode follower a grounded-plate amplifier. Yet when you stop to think about it, that is precisely what these amplifiers actually are.

Consider, for example, the conventional amplifier circuit shown in Fig. 3·29*a*. The input signal is applied to the grid, while the output signal is taken from the plate. The cathode usually has a resistor in its circuit, but the resistor is bypassed in most applications by a capacitor that is frequently large enough to place the cathode at ground potential so far as the signal is concerned. Under these conditions, the cathode need not be at ground potential with respect to a d-c voltage and, indeed, usually has some positive voltage on it because of the voltage drop across the cathode resistor. Signalwise, however, the cathode is at ground potential. This, then, is a grounded-cathode amplifier.

Note the situation does not change if the cathode is returned to some positive or negative d-c voltage, as in Fig. 3·29*b*, instead of to ground. In this instance, we have simply changed our reference point from one d-c voltage (zero) to another d-c voltage which may be higher than zero (i.e., positive) or lower than zero (i.e., negative). Amplifier operation, however, remains the same. Whatever the polarity of the d-c voltage chosen, this is still the point from which the other d-c voltages on the tube are measured.

The above designation remains the same even when a cathode bypass capacitor is not employed. This simply has the effect of introducing some inverse feedback. The signal input and output points are unaltered, and the cathode is still common to both input and output circuits.

The key word in the last sentence is "common." The more general definition of a grounded-cathode amplifier is one in which the cathode is common to both input and output circuits. Hence, the name "common cathode" is interchangeable with "grounded cathode," and both names will be used throughout the book. The same applies to grounded-grid (or common-grid) and grounded-plate (or common-plate) amplifier. The word "ground," in nearly all its applications in electronics, should more frequently be considered in its general sense of being a reference point common to one or more circuits. The beginner in electronics is often led to believe that ground possesses special properties not found in other portions of the circuitry. A ground is best regarded as just another conductor which derives any special qualities it may have only by virtue of the fact that it is common to several circuits. The notation common-base, common-emitter, and common-collector amplifier is also widely used.

In a grounded- or common-grid amplifier, Fig. 3·29*c*, we place the grid at signal ground while the input signal is applied to the cathode and the output signal is obtained at the plate. Again, note that the grid may have some d-c voltage on it, for biasing purposes, without affecting the designation or operation of the stage.

The final arrangement, grounded- or common-plate, is shown in Fig. 3·29*d*. Here the plate is returned to signal ground, while the input signal is applied to the grid, and the output signal is obtained at the cathode. The best-known name for this amplifier is "cathode follower."

Each of these amplifiers has its own characteristics stemming from the method of connection. For example, the grounded-cathode amplifier will provide the greatest voltage and power amplification, while the grounded-plate amplifier will provide the least. On the other hand, the grounded-plate amplifier is best suited to provide a match between high- and low-impedance systems. Each has certain characteristics which make it the most desirable arrangement for certain applications. It will be found that the same is true of transistor amplifiers.

Grounded-base amplifier. It is convenient to start a detailed examination with a grounded-base transistor amplifier. This is shown in Fig. 3·30*a*. The input signal is applied to the emitter, and the

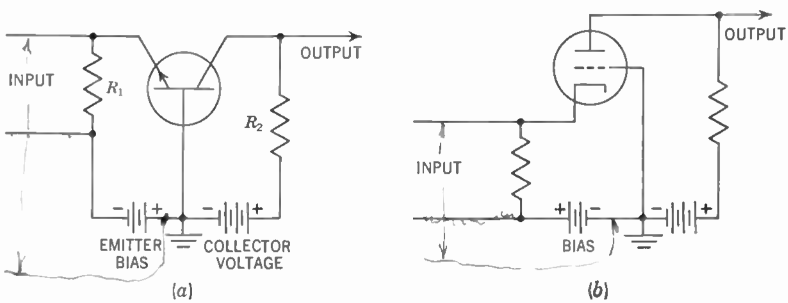


Fig. 3·30 (a) A grounded-base transistor amplifier and (b) the analogous grounded-grid vacuum-tube amplifier. The polarity of the voltages used in the transistor circuit will be governed by the doping of the semiconductor used for the various elements. Here an NPN transistor is shown.

output signal is obtained at the collector. The equivalent vacuum-tube amplifier is shown in Fig. 3·30*b*. The grid, being equivalent to the base, is grounded. The signal is then fed to the cathode; the plate is the output circuit.

In a grounded-grid amplifier, the input and output signals have the same polarity; i.e., passage through the tube does not alter the phase of the signal. In a grounded-base transistor amplifier, the same behavior is found. To illustrate this, the amplifier of Fig. 3·30*a* has been drawn using an NPN transistor, and the battery polarities have been chosen accordingly. Assume, now, that the incoming signal is positive at this instant. This positive voltage will counteract some of the normal negative bias between emitter and base and serve to reduce the current flowing through the transistor. This, in turn, will reduce the voltage drop across R_2, making the collector potential more positive. Thus, a positive-going input signal produces a positive-going output signal.

During the negative half cycle of the input signal, the emitter will be driven more negative than it normally is with respect to the base. This will increase the flow of electrons (here) from emitter to collector and cause the negative voltage drop across R_2 to increase. This will drive the collector more negative. Again we see that the polarity of the output signal is similar to that of the input signal.

For the output, a load resistor of 10,000 ohms is a common value. In so far as current is concerned, there is less at the output (i.e., the collector) than at the emitter. The difference is the 1 or 2 per cent that is diverted to the base. This ratio of output current, I_C, to input current, I_E, or I_C/I_E, is the α, or h_{FB}, of the transistor. (h_{FB} is the hybrid symbol for α. These hybrid characteristics or parameters are employed extensively on transistor data sheets and it is desirable for the reader to become familiar with them. They are explained at length in Chap. 12.) Thus, the current gain of this amplifier arrangement is less than 1, and this might lead one to believe that the circuit has little utility. This is not true; a sizable voltage gain may be obtained because the output load resistance value is so much higher than the input resistance. Thus, if we assume an input resistance of 50 ohms and a load resistance of 10,000 ohms, then the voltage gain (input to output) is

$$\text{Voltage gain} = \frac{E_{\text{OUT}}}{E_{\text{IN}}} = \frac{I_C R_L}{I_E R_{\text{IN}}}$$

$$\frac{I_C}{I_E} = 0.98 \qquad \text{for a typical value}$$

Hence,

$$\begin{aligned}\text{Voltage gain} = \frac{E_{\text{OUT}}}{E_{\text{IN}}} &= 0.98 \times \frac{10{,}000}{50} \\ &= 0.98 \times 200 \\ &= 196\end{aligned}$$

By the same token, a power gain is also possible:

$$\begin{aligned}\text{Power} &= I^2R \\ \text{Power gain} &= \frac{P_{\text{OUT}}}{P_{\text{IN}}} \\ &= \frac{I_C^2}{I_E^2}\frac{R_L}{R_{\text{IN}}} \\ &= 0.98^2 \times 200 \\ &= 0.96 \times 200 \\ \text{Power gain} &= 192\end{aligned}$$

We have indicated above that the input impedance of a transistor in the grounded-base configuration is quite low. The output impedance, if we remove the load resistor and look into the collector, is very high, on the order of 1 to 2 megohms. However, when we connect a load resistor of 10,000 ohms, then this is the value of the output impedance since it completely swamps the 1 to 2 megohms with which it is basically in parallel. It is well for the reader to keep this distinction in mind, because reference is often made in the literature to the high output impedance of the common-base arrangement, and this means, we indicated above, without the load resistor. Once a much smaller load resistance is connected to the collector, its value will essentially determine the output impedance.

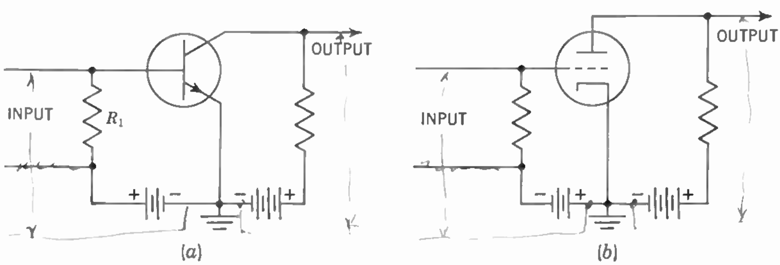


Fig. 3·31 (a) A grounded-emitter amplifier and (b) its vacuum-tube counterpart. An NPN transistor is used in a; for a PNP transistor, the polarities of the batteries would have to be reversed.

Grounded-emitter amplifier. The grounded-emitter amplifier, Fig. 3·31*a*, is the most popular of the three types. The input signal is applied to the base, and the output signal is obtained at the collector. The equivalent vacuum-tube amplifier is shown in Fig. 3·31*b*, and the reader will immediately recognize this as the most common amplifier in use today.

It turns out, when the mathematics of grounded-emitter circuitry is worked out, that this arrangement possesses a number of advantages for the junction transistor over the grounded-base approach. For one thing, the input impedance is higher, averaging between 700 and 2,000 ohms. The output impedance, looking into the collector before any load resistor has been connected, is about 500,000 ohms. This is less than the value presented by the common-base amplifier. The same 10,000 ohms of load resistance is usually employed here as well, however.

Since the input signal is applied to the base in the common-emitter arrangement, it is the variations in signal which control the collector current. Hence, current gain here is

$$\text{Current gain} = \frac{I_C}{I_B}$$
$$= \beta \quad \text{or} \quad h_{FE}$$

This ratio is called beta (β), or h_{FE} when the hybrid notation is employed. In a typical transistor, if the emitter current is 5 ma, the base will get 0.1 ma, while the collector will receive the rest, or 4.9 ma. Substituting these values into the equation, we obtain

$$\beta = \frac{4.9}{0.1}$$
$$= 49$$

Note that a sizable current gain is obtained, in contrast to the small loss occasioned in the preceding amplifier. The gain arises from the fact that very minute variations of the base current produce significant variations in the collector current.

A voltage gain is also obtained because there is not only a current gain but also a resistance gain. If we assume an input resistance of 2,000 ohms and a load resistance of 10,000 ohms, then the voltage gain is

$$\text{Voltage gain} = \text{current gain} \times \text{resistance gain}$$
$$= \frac{I_C}{I_B} \times \frac{R_L}{R_{IN}}$$
$$= 49 \times \frac{10{,}000}{2{,}000}$$
$$= 245$$

This is somewhat larger than the voltage gain achieved with the common-base circuit. The difference is not very much, however. Power gain, is considerably better:

$$\text{Power gain} = \frac{I_C^2}{I_B^2}\frac{R_L}{R_{IN}}$$
$$= 49^2 \times 5$$
$$= 12{,}005$$

It is because of its higher current and power gains that the common-emitter amplifier is the most popular arrangement employed in transistor circuits.

An interesting feature of the grounded-emitter form of connection is the phase reversal that occurs as the signal passes through the stage. In this the grounded-emitter amplifier is similar to its vacuum-tube prototype, the grounded-cathode amplifier. The reason for the reversal can be understood by considering the amplifier shown in Fig. 3·31*a*. The base-emitter circuit is biased in the forward direction, with the negative side of the bias battery connecting to the emitter and the positive side of the battery to the base. (In this way, the negative battery terminal repels the excess electrons in the N-type emitter toward the PN junction while the positive battery potential drives the holes in the base to the same junction.)

If, now, we apply a signal to the base, here is what will happen. When the signal is negative, it will tend to reduce the bias potential applied between emitter and base. This means that the electrons in

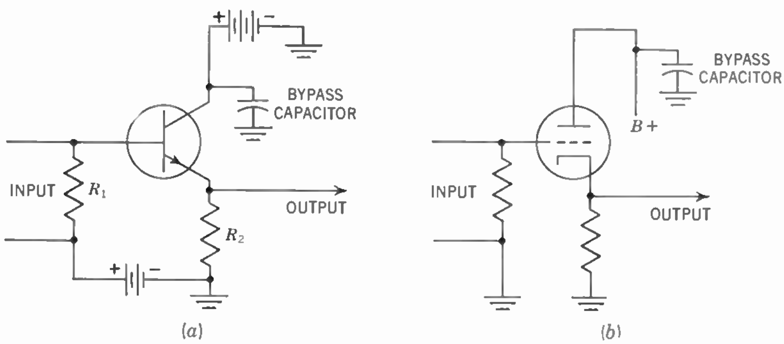


Fig. 3·32 (a) A grounded-collector amplifier and (b) its vacuum-tube counterpart. The cathode resistor in b would have a fairly high value.

the emitter and the holes in the base will have less compulsion to overcome the inherent separating force at the junction and less current will flow. This, in turn, will reduce the collector current and provide less voltage drop across the load resistor. As a result, potential at the collector will become more positive.

During the positive half cycle of the signal, the total voltage in the emitter-base circuit will rise. This will increase the flow of current through the emitter, the collector, and R_2. The increased voltage drop across the collector resistor will make the top end of this resistor more negative. Thus, in ground-emitter amplifiers, the output signal is 180° out of phase with the input signal.

Grounded-collector amplifiers. The final transistor amplifier circuit arrangement is the grounded-collector one. This is shown schematically in Fig. 3·32, together with its vacuum-tube counterpart. Note that the plate of the vacuum tube is not d-c grounded, since this ele ment still requires a positive potential (relative to the cathode) in order to attract electrons. However, the plate is at a-c ground by virtue of the large bypass capacitor, and that is actually what we are interested in.

The grounded-plate vacuum-tube amplifier will be recognized as the familiar cathode follower. It possesses a high input impedance between cathode and grid and a low output impedance. Voltage gain of this arrangement is always less than 1. In the grounded-collector arrangement we find many of the same characteristics. Thus, the input impedance, base to collector (which is here the common element), is quite high because of the reverse bias which exists between these two elements. (In contrast, the input impedance of the two preceding arrangements was low because the input circuit was between the base and emitter and the diode was forward-biased.) Of course, R_1 in Fig. 3·32 is hung across the input, and if this resistor is low-valued, it will cause the input impedance to be low. If R_1 has a high value, however, then we can obtain input impedances as high as 1 megohm in this circuit. The output impedance, on the other hand, is low, frequently falling below 100 ohms.

The current gain of a common-collector circuit is slightly higher than β; actually, it is $\beta + 1$. This can be shown quite simply as follows. The input current is the base current, I_B. The output current which flows through R_2, Fig. 3·32, is the emitter current, I_E, and this, we know, is equal to $I_B + I_C$. Hence,

$$\text{Current gain} = \frac{I_E}{I_B} = \frac{I_C + I_B}{I_B}$$

$$= \frac{I_C}{I_B} + 1$$

$$\text{Current gain} = \beta + 1$$

The voltage gain is always less than 1, although generally it is not much less than 1. This is true here for the same reason it is true in a cathode follower. The emitter (or cathode) resistor is fairly large and unbypassed. Hence, the signal voltage which develops here bucks the input signal at the base, so that only a small portion of the input signal is effective in producing an output voltage. In short, what we have is considerable degeneration.

Power gain is achieved in this stage because of the large current gain, but the gain is less here than it is in the other two configurations. Phase reversal of the signal does not occur in this stage. Any signal applied to input will appear at the output with the same phase. This, too, is like the cathode follower. And to complete the analogy, the common-collector circuit is frequently referred to as an emitter follower.

Table 3·2 summarizes the general characteristics of the three amplifier configurations.

Table 3·2

Characteristic	Common emitter	Common base	Common collector
Current gain	Large	1, approx	Large
Voltage gain	Large	Large	1, approx
Power gain	Largest	Large	Lowest
Input resistance	Low	Lowest	Highest
Output resistance	High	Highest	Lowest
Signal phase shift between output and input	180°	None	None

Transistor Characteristic Curves

The difference in operation between transistors and vacuum tubes, i.e., that one stresses current while the other voltage, is reflected in the characteristic curves of the two devices. In the characteristic curves

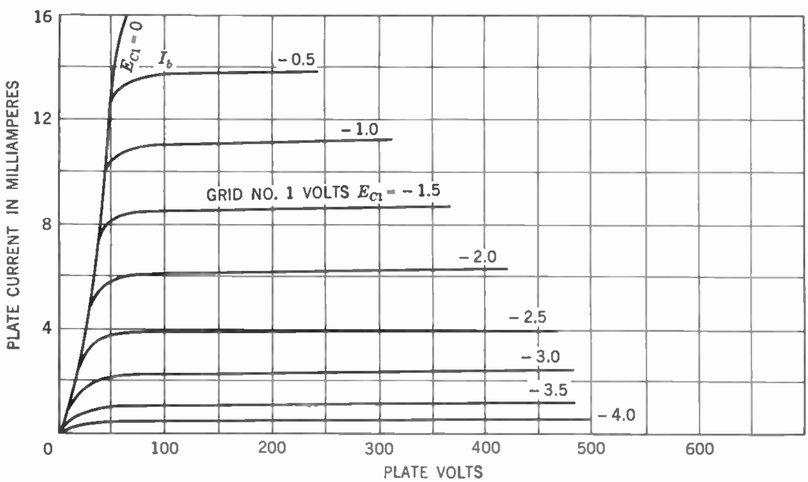


Fig. 3·33 The E_p-I_p characteristic curves for a 6AG5 pentode vacuum tube.

for a 6AG5 pentode, Fig. 3·33, the plate-current–plate-voltage curves are given for a series of grid-voltage values. The corresponding set of curves for a transistor is given in Fig. 3·23. Here, collector voltages and collector currents are plotted for various values of base current. Note the great similarity between these two sets of curves. In the case of the 6AG5 pentode, the plate current is relatively independent of the plate voltage from approximately 50 volts on. The only factor that determines plate current is grid voltage. For the transistor, collector current is likewise independent of collector voltage and is wholly a function of emitter current. This follows from the basic operation of a transistor, wherein the base-emitter potential determines how many minority carriers the emitter injects into the base. Because of its very low value, it is inconvenient to measure base-emitter potential. The base current, which is related to the base-emitter potential, is a more easily measured component and hence is employed as the governing parameter on the family of I_c-E_c curves. Furthermore, the input resistance is not linear. That is, the input resistance will change as the input current changes. If we attempted to use the base voltage as the running parameter, the characteristics would be made unnecessarily nonlinear.

An interesting feature of the transistor characteristic curves is the fact that when we reverse the collector voltage (as we do at the left-hand side of the chart), the collector current drops sharply to zero and then, if the chart had extended down, would have rapidly reversed itself and started flowing in the opposite direction (i.e., the forward direction). The latter condition is not desired, since it would quickly lead to excessive current flow with consequent overheating and permanent damage to the crystal.

In the plotting of graphs, it is customary to place the more important (i.e., the independent) variable along the horizontal axis. This should be done with the transistor curves; i.e., the collector-current axis should be placed horizontally and the collector-voltage axis vertically. In practice, both types of presentations will be found, with perhaps greater emphasis given to the form shown in Fig. 3·23 because of its correspondence to the more familiar vacuum-tube curves.

It is not uncommon to find collector-voltage values listed with negative values in these characteristic charts. This is to indicate that the applied collector voltage is negative in polarity, such as we would use with a PNP transistor. For an NPN transistor, the collector voltage would be positive.

Negative signs are also found in front of the current figures, and there their presence is associated with the direction of collector-current flow in relation to the conventional indication of current flow through a circuit. Conventionally, electric current is taken to flow from the positive side of a voltage source through the circuit to the negative side. Electron flow, with which most technicians are familiar, travels in the opposite direction. Because of this difference in treating current direction, it is suggested that any negative signs appearing in front of collector-current values be ignored. As long as the proper battery polarity is applied to the collector, the current will take care of itself.

Most manufacturers, when they give transistor characteristics, include the output characteristic curves for the common-emitter connection. Occasionally, the curves for the common-base connection will be given, although this practice is decreasing.

Alpha and beta. We have seen that the α of a transistor is the ratio of the collector current to the emitter current when the transistor is connected in a common-base arrangement. The collector current should be measured with no load resistor in the collector circuit because, by definition, α is the short-circuit current gain. The β of a transistor is defined as the ratio of the collector current to the base current when the transistor is connected as a common-emitter amplifier. Here, too, the collector current should be measured with no resistance in the collector circuit because, by definition, β is the short-circuit current gain of the transistor. (By defining α and β in this way, a definite frame of reference is established for all such measurements.)

Since the α and β are characteristics of the same transistor, they are related to each other. We know that the emitter current is equal to the sum of the base and collector currents. That is

$$I_E = I_B + I_C \tag{3·1}$$

Rearranging, we obtain

$$I_B = I_E - I_C \tag{3·2}$$

We also know that the collector current I_C is equal to αI_E. Hence, expression (3·2) can be written

$$I_B = I_E - \alpha I_E$$
$$I_B = I_E(1 - \alpha) \tag{3·3}$$

Now we divide $I_C = \alpha I_E$ by Eq. (3·3):

$$\beta = \frac{I_C}{I_B} = \frac{\alpha I_E}{(1 - \alpha)I_E}$$
$$\beta = \frac{\alpha}{1 - \alpha} \tag{3·4}$$

Conversely, we can express α in terms of β:

$$\alpha = \frac{\beta}{1 + \beta} \tag{3·5}$$

Equation (3·5) is derived by simply rearranging Eq. (3·4).

Typical Transistor Data

An indispensable tool for anyone dealing with the design, operation, or service of electronic equipment is a tube manual. Here we find the mechanical and electrical specifications for each type of tube, plus a set of characteristic curves. In similar fashion, equivalent data are published by transistor manufacturers for each of their products.

Transistor manufacturers' sheets contain the specifications of a particular transistor, including maximum ratings, characteristic curves, and physical outline. Some of the more important items described are:

1. Transistor number
2. Collector-junction voltage rating
3. Emitter-junction voltage rating
4. Current-handling capacities
5. Power rating
6. Temperature limitation
7. Thermal resistance
8. Transistor-case outline
9. A-c h parameters
10. I_{CO}
11. I_{EO}
12. α cutoff frequency
13. Static collector characteristic
14. Temperature variation of transistor parameter
15. D-c β and α
16. Saturation resistance

Each of these parameters may be used in circuit design and specification of a particular transistor. The location of the parameters on the data sheet need not be the same and the method of presentation may also vary. Furthermore, not all parameters may be given for any specific transistor; only those that are deemed necessary will be included.

A typical specification sheet (General Electric Company) is shown in Fig. 3·34. The various sections of this listing are numbered 1 to 10, and appropriate explanations of each are given below.

1. The lead paragraph is a general description of the device and usually contains three specific pieces of information: The kind of transistor, in this instance a silicon NPN triode; a few major application areas, here amplifier and switch; and general features such as electrical stability and a standard size hermetically sealed package.

2. The absolute maximum ratings are those ratings which must not be exceeded. To exceed them may cause device failure.

2N337, 2N338

Outline Drawing No. 4

(1) The General Electric Types 2N337 and 2N338 are high-frequency silicon NPN transistors intended for amplifier applications in the audio and radio frequency range and for high-speed switching circuits. They are grown junction devices with a diffused base and are manufactured in the Fixed-Bed Mounting design for extremely high mechanical reliability under severe conditions of shock, vibration, centrifugal force, and temperature. For electrical reliability and parameter stability, all transistors are subjected to a minimum 160 hour 200°C cycled aging operation included in the manufacturing process. These transistors are hermetically sealed in welded cases. The case dimensions and lead configuration conform to JEDEC standards and are suitable for insertion in printed boards by automatic assembly equipment.

SPECIFICATIONS

(2) **ABSOLUTE MAXIMUM RATINGS: (25°C)**

	Voltage			
	Collector to Base	V_{CBO}	45	volts
	Emitter to Base	V_{EBO}	1	volt
	Current			
	Collector	I_C	20	ma
(3)	**Power**			
	Collector Dissipation*	P_C	125	mw
	Temperature			
	Storage	T_{STG}	−65 to 200	°C
	Operating	T_A	−65 to 150	°C

(4) **ELECTRICAL CHARACTERISTICS: (25°C)**
(Unless otherwise specified; $V_{CB} = 20v$; $I_E = -1$ ma; $f = 1$ kc)

			2N337			2N338			
	Small-Signal Characteristics		**Min.**	**Typ.**	**Max.**	**Min.**	**Typ.**	**Max.**	
(5)	Current Transfer Ratio	h_{fe}	19	55		39	99		
	Input Impedance	h_{ib}	30	47	80	30	47	80	ohms
	Reverse Voltage Transfer Ratio	h_{rb}		180	2000		200	2000	$\times 10^{-6}$
	Output Admittance	h_{ob}		.1	1		.1	1	µmho
	High-Frequency Characteristics								
(6)	Alpha Cutoff Frequency	f_{ab}	10	30		20	45		mc
	Collector Capacitance ($f = 1$ mc)	C_{ob}		1.4	3		1.4	3	µµf
	Common Emitter Current Gain ($f = 2.5$ mc)	h_{fe}	14	24		20	26		
	D-C Characteristics								
(7)	Common Emitter Current Gain ($V_{CE} = 5v$; $I_C = 10$ ma)	h_{FE}	20	35	55	45	75	150	
	Collector Breakdown Voltage ($I_{CBO} = 50$ µa; $I_E = 0$)	BV_{CBO}	45			45			volts
	Emitter Breakdown Voltage ($I_{EBO} = -50$ µa; $I_C = 0$)	BV_{EBO}	1			1			volt
(8)	Collector Saturation Resistance ($I_B = 1$ ma; $I_C = 10$ ma)	R_{SC}		75	150				ohms
	($I_B = .5$ ma; $I_C = 10$ ma)	R_{SC}					75	150	ohms
	Cutoff Characteristics								
(9)	Collector Current ($V_{CB} = 20v$; $I_E = 0$; $T_A = 25°C$)	I_{CBO}		.002	1		.002	1	µa
	Collector Current ($V_{CB} = 20v$; $I_E = 0$; $T_A = 150°C$)	I_{CBO}			100			100	µa
	Switching Characteristics								
(10)	Rise Time	t_r		.02			.06		µsecs
	Storage Time	t_s		.02			.02		µsecs
	Fall Time	t_f		.04			.14		µsecs

*Derate 1 mw/°C increase in ambient temperature over 25°C

Fig. 3·34 A transistor specification sheet. *(General Electric Co.)*

3. The power dissipation of a transistor is generally limited by the junction temperature. Therefore, the higher the temperature of the air surrounding the transistor (ambient temperature), the less power the device dissipates. A factor which indicates how much the transistor must be derated for each degree of increase in ambient temperature in degrees centigrade is usually given. Note that the 2N337 (given on this specification sheet) can dissipate 125 mw at 25°C. By applying the given derating factor of 1 mw for each degree increase in ambient temperature, we find that the power dissipation will drop to 0 mw at 150°C. This, then, is the maximum operating temperature of this transistor.

4. All of the remaining ratings define what the device is capable of under specified test conditions. These characteristics are needed by the design engineer to develop matching networks and to calculate exact circuit performance.

There is one important difference between the absolute maximum rating and the design characteristics listed on specification sheets. The absolute maximum rating must not be exceeded under any circumstance and to exceed it automatically releases the transistor manufacturer from any warranty he may give with the unit. Characteristics, on the other hand, although they may entail some guarantee, are presented primarily as a guide to the user. Some of the parameters, for example, I_{co}, h_{fe}, V_{BE}, etc., have their maximum values guaranteed but not at end of life (i.e., usually after 1,000 hr). Other parameters, such as breakdown voltage, are rated on an end-of-life basis. However, none of the typical values listed are guaranteed.

5. "Current-transfer ratio" is another name for β. In this case we are talking about an a-c characteristic, so the symbol is h_{fe}. If the d-c beta is meant, the symbol is h_{FE}. β is partially dependent on frequency, so some specifications list it for more than one frequency. The remaining h parameters in this section deal with items necessary for transistor circuit design. h_i is the hybrid symbol for input impedance. If a small b subscript is added, that is, h_{ib}, then the symbol stands for the input impedance of a common-base amplifier, generally given at zero a-c collector current. By the same token, h_{ie} is the input impedance of a common-emitter amplifier.

h_{rb} is the reverse voltage transfer ratio for the common-base amplifier. It is the ratio of the a-c voltage appearing at the emitter to an a-c voltage applied to the collector for essentially an open circuit to alternating current at the emitter. h_{re} is the reverse voltage transfer ratio for the common-emitter configuration. In this instance, it is the ratio of a-c voltage at the base to the a-c voltage at the collector.

h_{ob} is the output circuit conductance for the common-base configuration with the input open-circuited. Conductance is the reciprocal of resistance, so that $1/h_{ob}$ is the output impedance of the transistor. It is defined as the ratio of the a-c collector current to the a-c collector voltage. h_{oe} is the output circuit conductance with the input open-circuited for the grounded-emitter configuration. It too is the ratio of the a-c collector current to the a-c collector voltage.

The h parameters are measured at 1,000 cps and are useful in ampifier design since they can be used to determine the input impedance, the output impedance, and the current gain of a transistor.

At this point in the book, the hybrid parameters will probably prove more confusing than enlightening. It is suggested that the reader wait until he reaches Chap. 12, where a full discussion of hybrid and other parameters is given. The only reason for mentioning these parameters at this time is that they appear on the data sheets.

6. The frequency cutoff f_{α_b} or f_{hfb} of a transistor is that frequency at which the grounded-base current gain drops to 0.707 of the 1,000-cycle value. It gives a rough indication of the useful frequency range of the device.

7. The collector breakdown voltage BV_{CBO} is the inverse voltage between the collector and base with the emitter open at which there is a sharp increase in current flow between the collector and base. This point is known as the avalanche breakdown, in which minority electrons, passing the PN junction, gain sufficient energy to knock off valence electrons bound to the crystal lattice and raise them to the conduction band (see page 13). BV_{CBO} is usually specified at some value of reverse leakage current.

Emitter breakdown voltage BV_{EBO} is the maximum voltage which can be safely applied between emitter and base when these elements are reverse-biased (with the collector open) at some specified value of reverse leakage current. This value is given in specification sheets in order to indicate how large a reverse voltage may be applied to the input of a common-emitter amplifier before the input circuit will break down.

Application of voltages in excess of the maximum breakdown values may or may not damage a transistor. In most instances, the transistor can be made to operate again satisfactorily after the excessive voltage has been removed. In some instances, however, sufficient damage is wrought that the transistor becomes unusable. The emitter-base breakdown voltage value will depend on how the transistor was fabricated. For alloy transistors, it is comparable to the collector-base breakdown voltage. With diffused transistors, however, it is much less than the collector-base breakdown voltage. In the specification sheet of Fig. 3·34, the emitter breakdown voltage is 1 volt, whereas the collector breakdown voltage is 45 volts. In normal circuit design all voltages should be kept well below the maximum ratings, because transients or other voltage variations which drive the transistor into the breakdown region can cause an immediate and permanent damage.

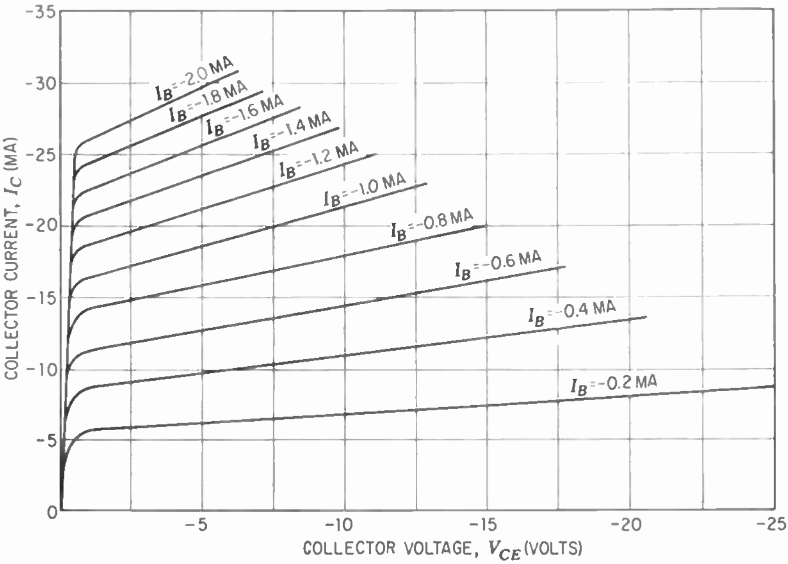
8. Collector saturation resistance. A transistor is saturated when both junctions are forward-biased. The saturation resistance for this condition is equal to the collector-to-emitter voltage divided by the collector current. It consists of two components. The first component is the bulk resistance of the material from the collector and emitter terminals to their respective junctions. The second component is due to the transistor action of the device; it decreases as the base current is increased for any given value of collector current. Thus, overdriving the transistor will reduce the saturation resistance.

Collector saturation resistance is generally of considerable importance to engineers who are designing logic circuits in which the transistor itself acts as a switch, going from a very high impedance condition, when it is essentially cut off, to a very low impedance condition, when it is saturated. In such design work it is important to know the saturation resistance, or the resistance of the transistor when it is in the low-impedance condition.

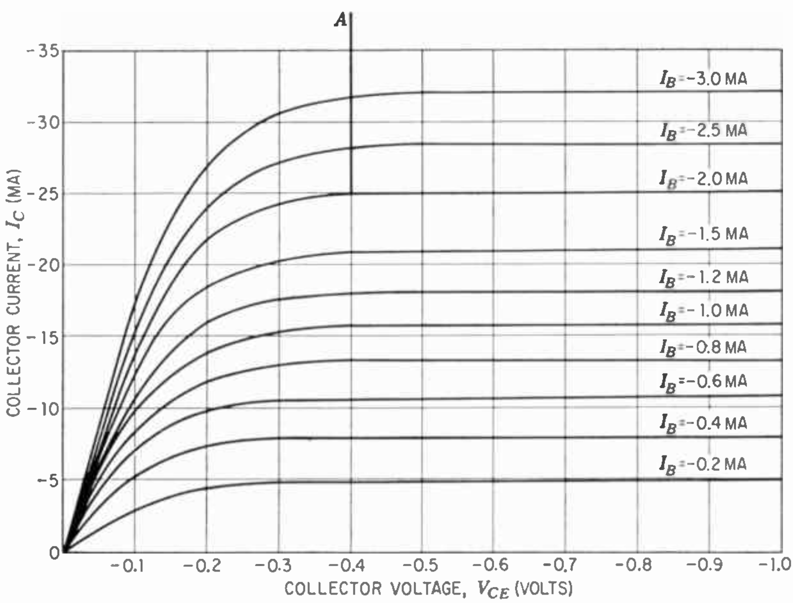
The saturation region on the characteristic curves of a transistor is at the extreme left-hand side where the curves appear to come together, Fig. 3·35*a*. Actually, when this portion of the graph is enlarged, Fig. 3·35*b*, it can be seen that each curve is separate and distinct. In this region the curves slope downward in a straight line and, if the ratio of collector voltage to collector current at any point is computed, the saturation resistance value is obtained. (Another way of stating the same thing is to say that the reciprocal of the slope of a curve in this region is the saturation resistance.)

Many manufacturers will list a collector saturation voltage $V_{CE(sat)}$. This voltage is essentially the minimum voltage necessary, at a particular collector current, to sustain normal transistor action, and it occurs when the emitter-base voltage equals the emitter-collector voltage. At lower collector voltages, the base-collector diode becomes forward-biased and the current-voltage relationship changes abruptly. This is the region where the curve lines slope sharply downward.

It might be noted, in passing, that there is also a cutoff region on the characteristic-curve plot. This occurs below the curve marked $I_B = 0$. In Fig. 3·35*a* the cutoff region is somewhat below the curve marked $I_B = 0.02$ ma.



(a)



(b)

Fig. 3·35 (a) The saturation region on the characteristic curves of a transistor is at the extreme left-hand side where the curves appear to come together. (b) Enlargement of saturation region of (a). This region is to the left of line A.

9. The collector cutoff current is the current from collector to base when no emitter current is being applied. This is the I_{co} which has been mentioned previously. It varies with temperature changes and must be taken into account whenever any semiconductor device is designed into equipment which is used over a wide range of ambient temperatures.

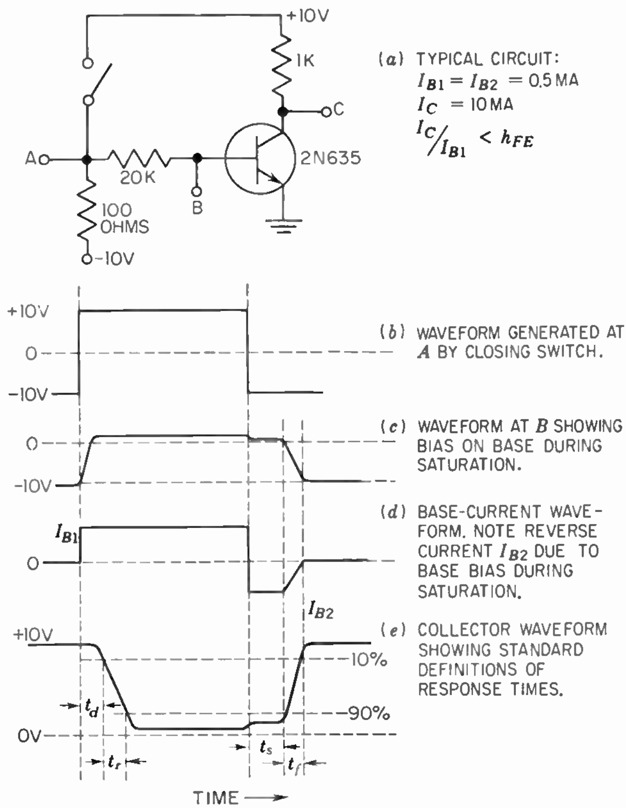


Fig. 3·36

10. The switching characteristics show how the device responds to an input pulse under the specified driving conditions. These response times are very dependent on the circuit used. The terms used in this section of the specification sheets are explained in Fig. 3·36.

t_d is the delay time, or the time it takes from the application of the input voltage at point A, Fig. 3·36, until the output voltage has reached 10 per cent of its final value.

t_r is the rise time or the time interval required for the output to go from 10 to 90 per cent of its saturation value.

t_s is the storage time, or the time it takes from the removal of the input signal for the output to go from its saturation value to 90 per cent of that value.

t_f is the fall time, or the time interval required for the output to go from 90 to 10 per cent of its saturation value.

The delay time is partly due to the time required to discharge the emitter-base capacitance which has been charged to the reverse space emitter bias voltage (-10 volts in Fig. 3·36) through the base resistance. Secondly, time must be allowed for the emitter current to diffuse through the base region. The rise time refers to the turn-on of the collector current. The storage time is due to the length of time required to sweep out the stored charge carriers in the base region which resulted from the collector-base region being forward-biased during saturation. (During saturation, both the emitter *and* collector inject carriers into the base region. The emitter normally does this under all circuit conditions; the collector only when it is forward-biased, as it is during saturation.) This is true for alloy transistors. For grown-diffused and mesa transistors, the primary charge storage takes place in the collector region rather than in the base region. As soon as the carriers have been swept out of the base region or the collector regions, the transistor begins to turn off.

In reading symbols, it is necessary to note carefully what upper- and lowercase letters are used in both the major letters and their subscripts. Lowercase letters are used as major letters to represent instantaneous (or alternating) values of current, voltage, power, or whatever. Examples are i, v, and p. Uppercase or capital letters are used as major letters to represent d-c or rms values. Examples are I, V, and P.

In subscripts, d-c and instantaneous values are indicated by uppercase letters. Examples are I_{CO}, V_{EB}, h_{FE}. Lowercase subscripts, such as those in V_{ac}, i_c, and h_{fe}, indicate a-c component values. (Readers wishing to learn more about the IRE standards for semiconductor notation are referred to the July, 1956, issue of the *Proceedings of the IRE*, pages 934 to 937.)

The final item which is found on transistor specification sheets is an outline of the transistor housing and an indication of the positioning of the transistor leads. Transistor cases are assigned so-called TO numbers such as TO-1, TO-5, and TO-9. The letters "TO" stand for Transistor Outline.

Lead placement varies. Sometimes leads are arranged in a straight line, sometimes they are arranged around a circle, and sometimes they are bunched together at a single point. The illustrations in Fig. 3·37 are representative of the more common arrangements. In

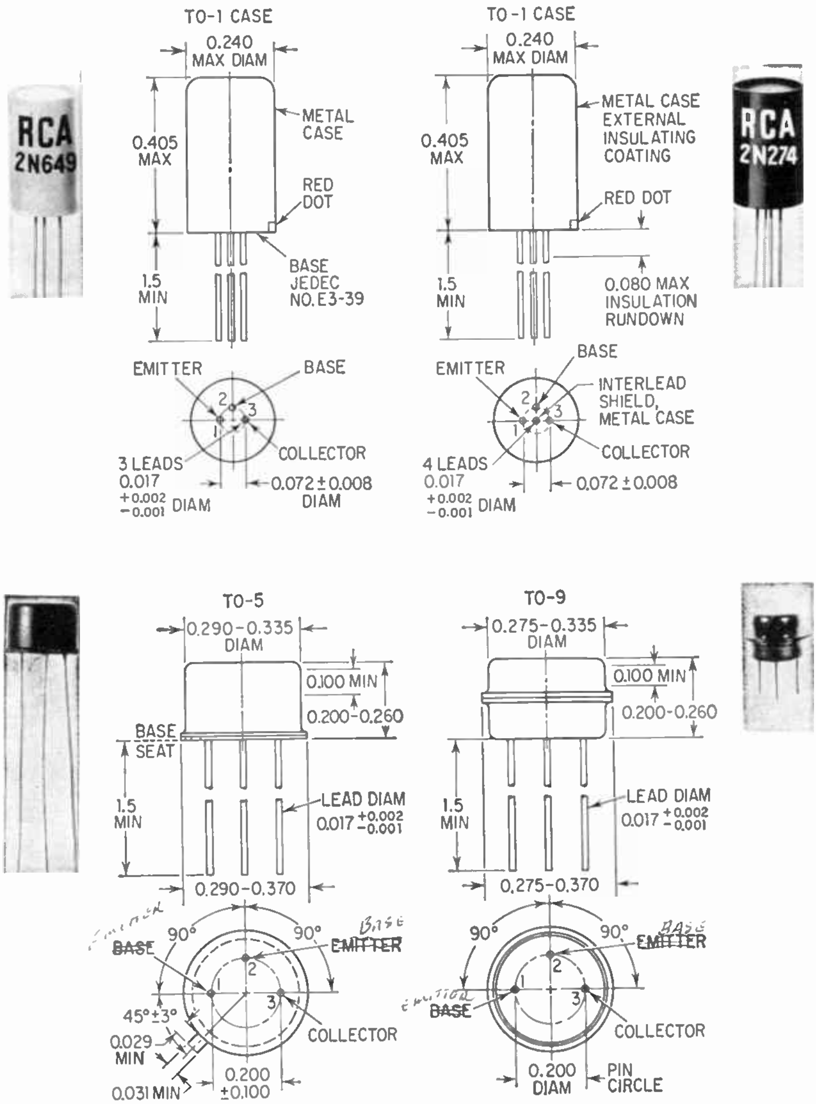


Fig. 3·37 Above and on the two pages following are shown typical transistor housings and lead arrangements. In instances when four leads are found, one is usually connected to the case and should be grounded (generally) in the circuit. In the larger power transistors, where only two leads or terminals are found, one is the base and the other is the emitter. The transistor case then serves as the collector connection, being internally connected to the collector element.

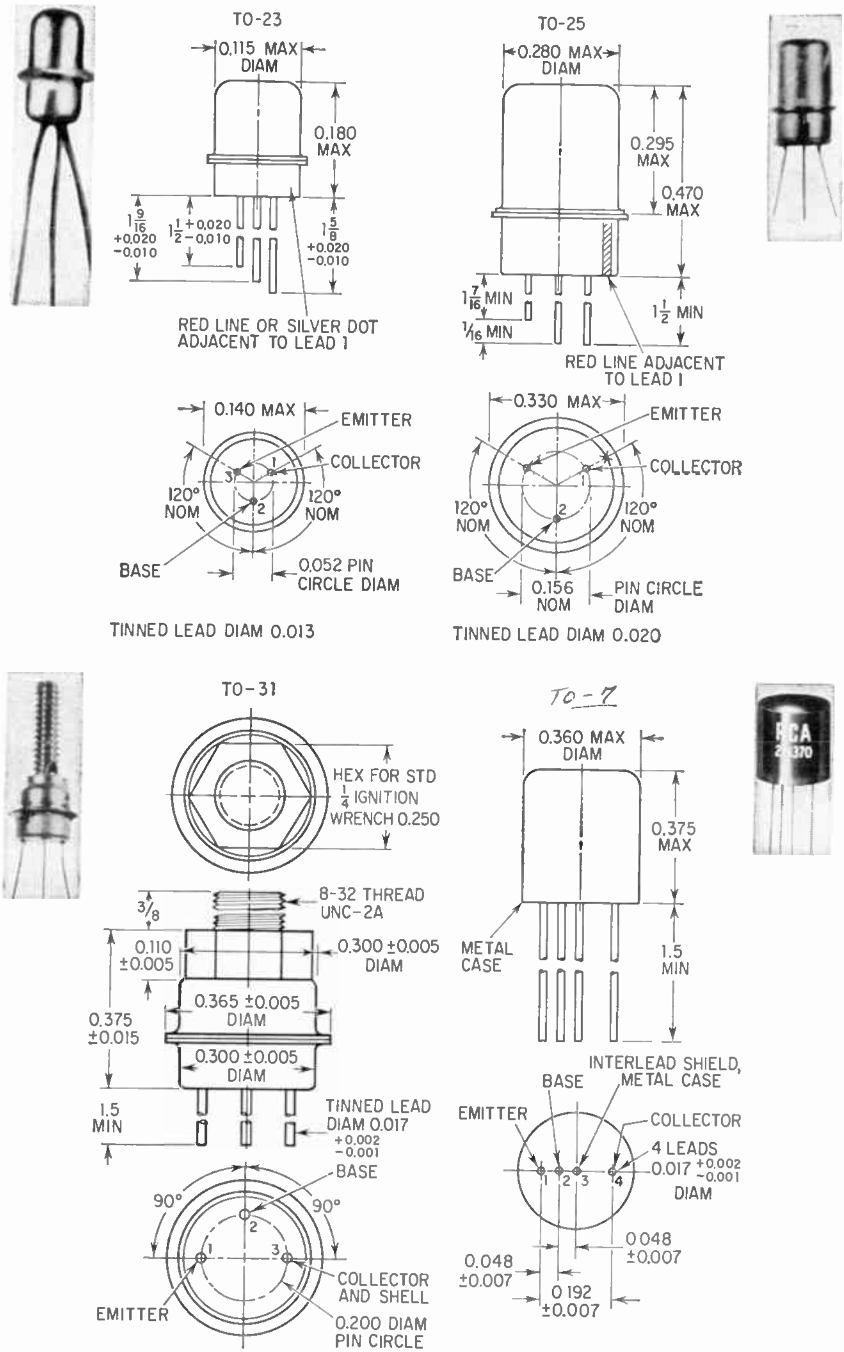


Fig. 3·37 (Continued)

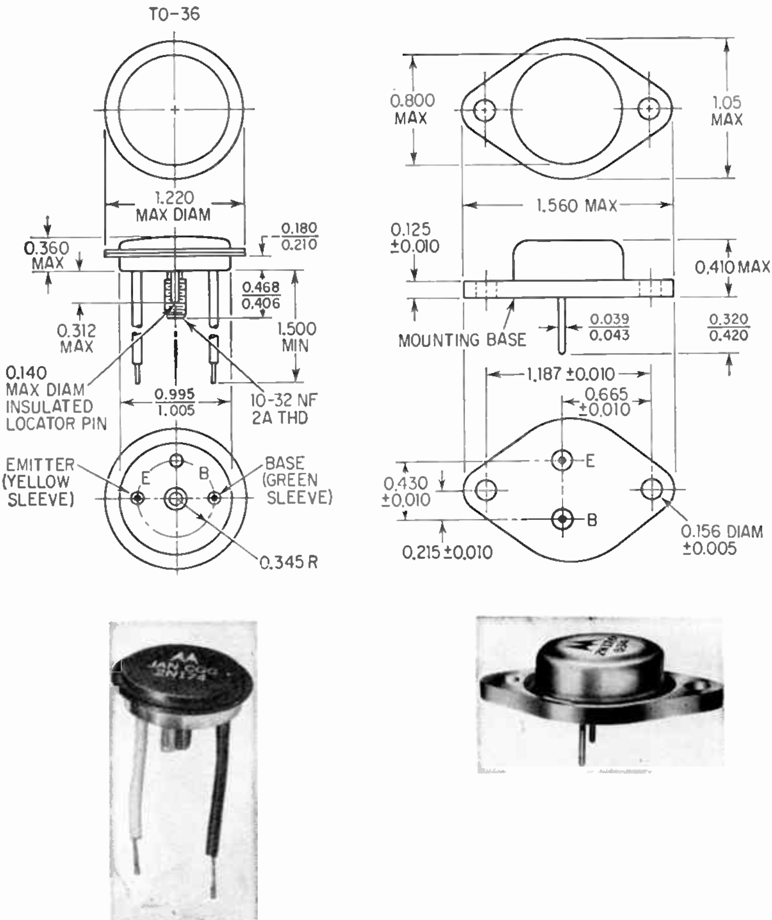


Fig. 3·37 (Continued)

all instances, the manufacturer's specification sheets should be checked before any connections are made to transistor leads.

Transistor Equivalent Circuits

As one works with transistors, one finds that there is a considerable degree of dependence between the input and output circuits. This is in direct contrast to vacuum tubes, where the input and output circuits are relatively independent of each other. (Feedback effects which may occur can generally be counteracted by suitable means.)

In the common-emitter arrangement, for example, the variation in input resistance with load resistance is as shown in Fig. 3·38. Note how the input resistance decreases with increase in load resistance, eventually leveling off to a value of about 500 ohms when the load resistance becomes inordinately large. This is for a typical junction transistor.

Similar curves showing the effect on the output resistance for different input resistances could be drawn. The reason for these interactions can perhaps be better understood when the equivalent electrical circuit of a transistor is examined. Equivalent circuits are convenient devices that enable an engineer to develop a relatively sim-

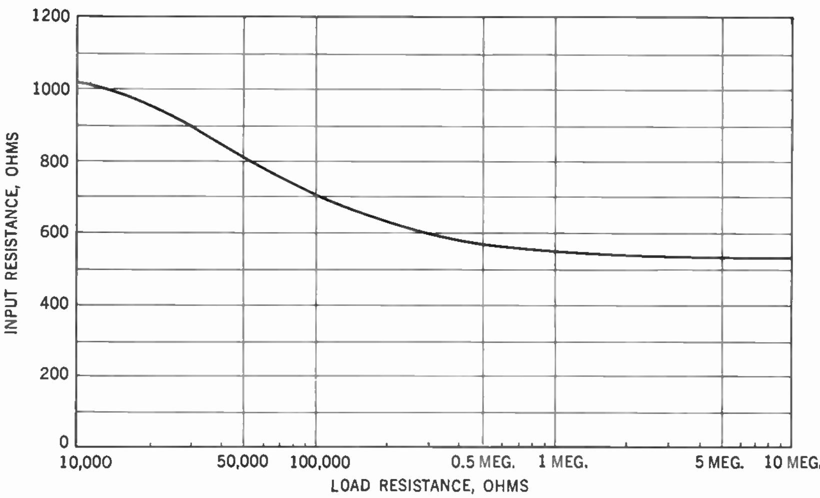


Fig. 3·38 Variation of input impedance with load resistance in a junction transistor.

ple electrical network which will function in the same manner electrically as some complex circuit which he may be investigating. It is interesting to study the engineer's approach to equivalent circuits because it will give the reader a better appreciation of the value and purpose of these circuits. The start is made with a little black box in which the circuit or system to be analyzed is contained. Access to the box is prohibited, and all we have from the box are four terminals, two representing the input and two the output. The procedure then is to take this black box and perform a series of four measurements on it. One measurement is to apply a signal to the input terminals 1 and 2 and record the voltage that is applied and the current that flows in the input circuit with the output circuit open, Fig. 3·39*a*. This will give us, when V_1 is divided by I_1, the input resistance R_1.

A second measurement is made by applying the signal to the input terminals of the black box and recording the current flowing in the input circuit and the voltage developed across the output circuit. This is illustrated in Fig. 3·39*b*. This measurement indicates what effect the input circuit has on the output circuit.

The third test is made with the signal generator connected across the output terminals and the voltage and current meters recording these respective quantities in the output circuit, Fig. 3·39*c*. The input circuit is open.

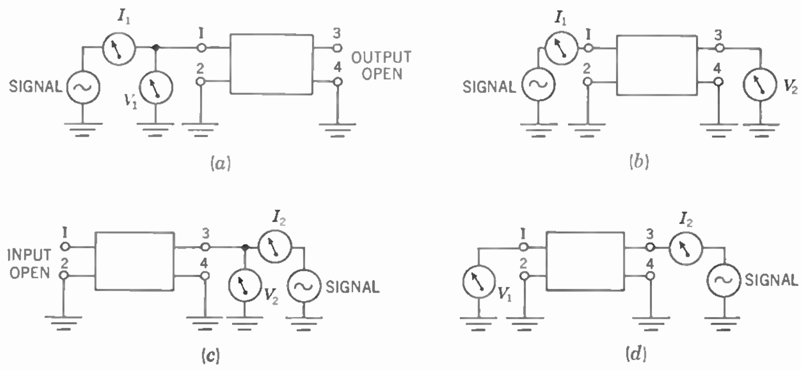


Fig. 3·39 Voltage and current measurements made on a system in order to determine its equivalent circuit. (*a*) Signal is applied to input. Voltage and current measurements are made on input with output terminals open. (*b*) Signal is applied to input. Current flowing in input is measured, together with voltage across output. (*c*) Signal is applied to output with input open; voltage and current in output are measured. (*d*) Signal is applied to output. Current flowing in output is measured, together with voltage across input.

The final check is made under the conditions indicated in Fig. 3·39*d*. Here we apply the signal to the output circuit and measure the voltage it produces across the input circuit.

The results of these four measurements are then used to draw a simple network which will give exactly the same results when the measurements indicated above are made. If such a network can be found, then we know that it will act, under all conditions, as the circuit or system in the black box acts, and we can call this latter network the equivalent of the box system and deal with it rather than the generally more complex system it replaces.

Using the foregoing method, one equivalent network obtained for a transistor is as shown in Fig. 3·40*a*, where r_e is the internal resistance of the emitter, r_b is the internal resistance of the base, and r_c is the internal resistance of the collector. Note that the base resistance is common to both the emitter and collector circuits, a fact that we discovered previously when studying the manner in which current is conducted through the transistor.

Now, if all we had in our equivalent circuit were these three resistances, then we would have a simple resistive network in which signals (or voltages) could pass from input to output or from output to input with equal ease. This, we know, is not true of transistors. Furthermore, a simple resistive network could introduce only attenuation, not amplification, and transistors do amplify. Obviously, something more is needed, and that something is the small generator placed in series with r_c. For mathematical reasons which are related to the design equations of transistors, this generator is given a value of $r_m i_e$, where i_e is the current flowing through the emitter resistance

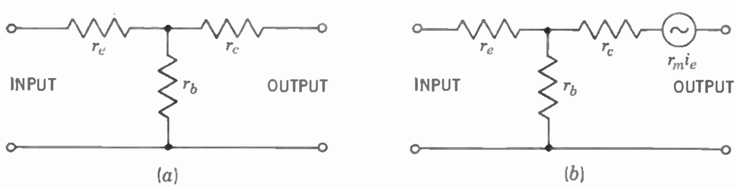


Fig. 3·40 Steps in the development of an equivalent circuit for a transistor. (*a*) Preliminary equivalent circuit of a transistor. (*b*) A more nearly complete equivalent circuit of a transistor

r_e and r_m is a mutual resistance of the system. For our purpose here, we need simply regard this generator as adding its voltage to that of the input signal to produce a greater (i.e., an amplified) signal at the output. In this way we achieve an equivalent circuit which reveals how a signal applied to a transistor is amplified and just what that amplification will be under various types of load resistances.

For those readers who find this added generator strange or confusing, attention is directed to the equivalent circuit for a triode vacuum tube, Fig. 3·41. We note that a voltage e_1 applied between grid and cathode produces the same effect as a voltage in the plate circuit which is μ times greater. μ, of course, is the amplification factor of the tube.

Returning to Fig. 3·40, we begin to see why the input and output circuits of a transistor are so dependent on each other. Any current flowing in the collector circuit will also flow through r_b, and the voltage developed here will directly influence the current flowing in the input circuit (containing r_e and r_b). And, of course, anything that happens in the input circuit will be immediately felt in the output circuit. In a vacuum tube, where the grid is negative and the frequency is not very high, the equivalent circuit of Fig. 3·41 shows quite plainly that the grid and plate circuits are isolated from each other and we do not have the same dependence between the impedances in each circuit that we have in a transistor.

A more extended discussion of transistor equivalent circuits will be found in Chap. 12. It is recommended, however, that the reader not

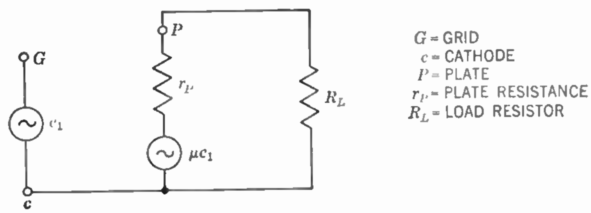


Fig. 3·41 The equivalent circuit of a vacuum tube The incoming signal e_1 appears in the plate circuit as a greater voltage μe_1

turn to this chapter immediately, but wait until he reaches it in the normal sequence of reading.

REFERENCES

3·1 Anderson, A. Eugene: Transistor Technology Evolution, *Western Electric Engineer*, July, October, 1959, January, 1960.

3·2 Philco Research Division: The Surface-barrier Transistor, *Proceedings of the IRE*, December, 1953.

3·3 Phillips, A. B., and A. M. Intrator: A New Frequency NPN Silicon Transistor, *Annual Convention of the IRE*, March, 1957.

3·4 Rittmann, A. D., and others: Microalloy Transistor, *IRE Transactions on Electron Devices*, April, 1958.

3·5 Sparks, M., and W. J. Pietenpol: A Breakthrough in Semiconductor Device Fabrication, *Bell Laboratories Record*, December, 1956.

3·6 Stevenson, I. R.: Transistor Ratings and Reliability, *Proceedings of the Australia IRE*, March, 1960.

3·7 Talley, H. E.: A Family of Diffused-base Germanium Transistors, *IRE Wescon Convention Record*, part 3, p. 115, 1958.

QUESTIONS

3·1 What factors limit the power-handling ability of transistors? Describe each briefly.

3·2 How does the noise figure of transistors vary with frequency?

3·3 List some of the factors which govern the frequency response of transistors.

3·4 Why does an NPN transistor have a higher frequency response than a comparable PNP transistor?

3·5 Describe briefly how grown-junction transistors are fabricated.

3·6 What major limitations do grown-junction transistors have?

3·7 Why is it possible to manufacture alloy-junction transistors with higher cutoff frequencies than grown-junction transistors?

3·8 Describe briefly how alloy-junction transistors are manufactured.

3·9 In what respects is the diffusion technique of transistor manufacture superior to either alloy or grown methods?

3·10 Can the alloy and diffusion techniques be combined for transistor maufacture? Explain your answer.

3·11 What does a drift transistor possess that is not present in an alloy-junction transistor? In what way is this helpful?

3·12 Describe how a grown-diffused transistor is manufactured.

3·13 What is the difference between a mesa transistor and an epitaxial mesa?

3·14 How does the surface-barrier transistor operate?

3·15 What is the difference between an SBT and an MAT unit?

3·16 Define I_{CBO}, I_{EO}, V_{CBO}, BV_{EBO}.

3·17 What differences exist between silicon and germanium in their use in transistors?

3·18 Why is I_{CO} important in transistor operation?

3·19 Why must the collector-dissipation rating of a transistor be reduced when the unit is employed above a certain temperature?

3·20 Name some of the ways in which the collector-dissipation rating can be increased.

3·21 How are PNP and NPN transistors differentiated schematically? What other conventions are employed in drawing transistor symbols?

3·22 Compare the elements in a triode vacuum tube with the sections of an NPN transistor. Do the same thing with respect to the d-c voltages which each device receives.

3·23 Draw the circuit of a grounded-base transistor amplifier, complete with d-c biasing voltages and input and output terminals. Draw the vacuum-tube counterpart of this circuit.

3·24 Follow the same procedure as in Question 3·23 for a grounded-emitter transistor amplifier. Draw the vacuum-tube counterpart of this circuit.

3·25 Answer Question 3·24 for a grounded-collector transistor amplifier.

3·26 Which of the three types of amplifier is best suited for a high-input-impedance low-output-impedance application? Which arrangement provides the best voltage and power gains? Would the same results be obtained if vacuum-tube amplifiers were employed? Explain.

3·27 Differentiate between the α and β values of a transistor.

3·28 Describe how the β value of a transistor may be determined from its characteristic curves. For your illustration, use Fig. 3·22.

3·29 How can you identify the various element leads of a transistor? Indicate two methods.

3·30 What characteristics are generally given for a transistor in the manufacturer's listings?

3·31 Why is there greater dependence between the input and output circuits of a transistor than of a vacuum tube?

3·32 Draw the equivalent circuit of a transistor.

CHAPTER 4

Transistor Amplifiers

IN PRECEDING CHAPTERS we noted how a transistor functions internally and how it achieves the desired goal of amplification. Several simple amplifier circuits were touched on in the course of this discussion; they were, however, incidental to the main discussion of transistor characteristics. In the present chapter we shall turn our attention completely to transistor amplifiers to see what forms they take and how they operate.

Transistors, like tubes, can be employed in three different configurations designated as grounded, or common, base; grounded, or common, emitter; and grounded, or common, collector. By way of review, it will be recalled that the common-base arrangement provides less than unity current gain, a very low input impedance, a high output impedance, and no phase reversal of the amplified signal. In the common-emitter amplifier, the current gain β is quite large, the input impedance is relatively low, the output impedance is moderate, and the signal suffers a phase reversal in going from input to output. In the final configuration, the common collector, the input impedance is high, the output impedance is low, and there is no signal phase reversal. The last arrangement is the direct counterpart of the vacuum-tube cathode follower.

Common-base Amplifier

Practical circuits employing each of the three transistor amplifier arrangements are shown in the illustrations on the pages to follow. In Fig. 4·1 a 2N104 transistor is used. If we did not know offhand whether this was an NPN or a PNP transistor, we could use the emitter and collector battery polarities as our clue. The emitter must be biased in the forward direction. This means that the positive terminal of B_1

must drive the emitter carriers to the emitter-base junction. A positive field, produced by the positive battery terminal, will repel holes. Therefore, we know that the emitter section is formed of P-type germanium. It follows then that the base has N-type germanium and the collector has P-type germanium. In short, the 2N104 is a PNP transistor.

The input signal is *RC*-coupled to the emitter. The emitter bias is established by battery B_1, 6 volts. Current flow through the emitter is governed by R_1, a 10,000-ohm resistor. By using Ohm's law,

$$E = IR$$

we find that

$$6 = I \times 10{,}000$$

$$I = 0.6 \text{ ma}$$

Actually, by this reasoning we completely neglect the resistance of the emitter-base section. The latter is so small with the indicated

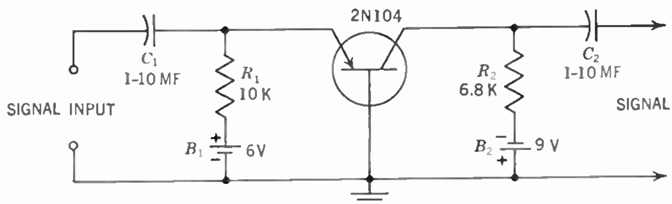


Fig. 4·1 A common-base transistor amplifier.

battery arrangement, however, that it scarcely alters the total current flow. Actually the emitter-base potential is on the order of 0.1 volt or possibly less.

The output, or collector, circuit possesses a 9-volt battery and a 6,800-ohm load resistor. Signal voltages developed across R_2 are then capacitively coupled to the next stage or output device.

The two coupling capacitors C_1 and C_2 are shown with capacitance ranges from 1 to 10 μf. The use of such high values is dictated by the relatively low input impedance of this stage (and for C_2, the relatively low input impedance of the following stage). With an input impedance on the order of 200 to 300 ohms, it is desirable that the impedance offered by C_1 to the lowest operating frequency be no more than 20 ohms. To achieve this, large values of capacitance must be used. Fortunately, the working voltage requirements are extremely low (here on the order of a volt or two), so that high-valued electrolytic capacitors can be manufactured at reasonable cost and with considerable compactness.

It is, of course, desirable to make C_1 (and C_2) as large as possible. When 1-μf values are used, the frequency response is such that the gain at 100 cycles is 16 per cent of the gain at 1,000 cycles. When we change to 10 μf, the 100-cycle gain rises to 46 per cent of the 1,000-cycle value, thus showing a marked improvement. The extent of the high-frequency end of the curve is governed by the capacitances shunting the circuit and the manner in which α drops off with frequency.

The use of two batteries is somewhat of a disadvantage, however, and this can be remedied by a voltage divider, as shown in Fig. 4·2. Since resistors R_3 and R_4 have values in the ratio of 2:3, the voltages developed across them will have the same ratio. Thus, 6 volts will develop across R_3 and 9 volts will develop across R_4. Across R_3 the

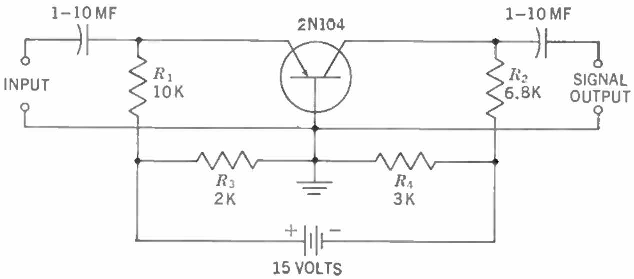


Fig. 4·2 A common-bose omplifier using only one bios bottery.

ungrounded end is positive, while across R_4 it is negative. This will provide the proper polarity voltages for the emitter and collector and enable the circuit to function in the same manner as that of Fig. 4·1.

In using the voltage-divider arrangement of Fig. 4·2, it is desirable to make R_3 and R_4 as low as possible so that variations in emitter or collector currents will not have any appreciable effect in altering the current through R_3 and R_4 and, in consequence of this, their voltage drops. On the other hand, if R_3 and R_4 are made too small, the current drain on the battery will become excessive. The values shown in Fig. 4·2 represent a compromise between these two conflicting goals.

Common-emitter Amplifier

Much more widespread than the use of common-base transistor amplifiers is the use of common-emitter amplifiers. With the common-emitter arrangement we obtain greater current and power gain. A circuit using the 2N104 in a common-emitter arrangement is shown in Fig. 4·3. Connection of the input and output resistors and capacitors remains the same as in the preceding amplifier. Note, however, the use of a single battery for both circuits. This is possible because the emitter is common to both input and output circuits and both collector and base require voltages which possess the same relative polarity with respect to the emitter. (In the present discussion, the emphasis is on circuit form rather than the selection of the most stable circuits. An analytical method of determining the sensitivity of an amplifier circuit to temperature and other factors is given in Chap. 12.)

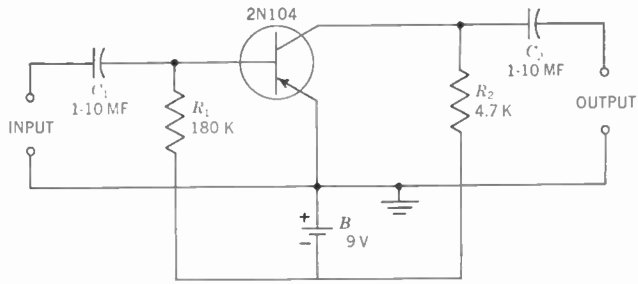


Fig. 4·3 A single-stage common-emitter amplifier.

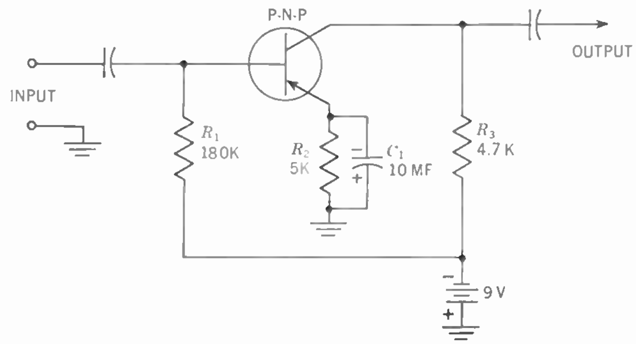


Fig. 4·4 A common-emitter amplifier that employs a stabilizing resistor R_2.

A form of common-emitter amplifier that is frequently seen is shown in Fig. 4·4. The chief difference between this circuit and that of Fig. 4·3 is the 5,000-ohm resistor R_2 and filter bypass capacitor C_1 which have been inserted in the emitter lead. Resistor R_2 serves to d-c stabilize the circuit by compensating for differences between transistors and by reducing the effects caused by temperature drift. Capacitor C_1 is shunted across R_2 to prevent degeneration with its reduction in gain. In some instances, the added stability provided by a-c degeneration may be desired, in which case C_1 would be omitted.

The problem of amplifier stability being affected by temperature changes is more serious in common-emitter and common-collector circuits than it is in common-base circuits because of the presence of a cutoff current I_{co}, which was mentioned briefly in Chap. 3. I_{co} is the current that flows through the collector-base diodes when the emitter current is zero. It stems from the presence of minority carriers in the base and collector sections, and it gives rise to a small current when the collector is reverse-biased. I_{co} is generally below 10 μa, and it is independent of the emitter current. Its value is determined chiefly by the particular transistor being used and by the temperature. It is the latter dependence which is particularly significant.

When a transistor is connected with the base common to both input and output circuits, as in Fig. 4·1, then the total collector current that flows is made up of two components; these are

$$I_C = \alpha I_E + I_{CO}$$

That is, I_c, in a junction transistor, is equal to 98 per cent (or so) of the emitter current I_E plus the collector cutoff current I_{co}. Since I_{co} is in microamperes and I_E is in the milliampere range or higher, changes in I_{co}, unless they are drastic, will not seriously increase the heat dissipated at the collector or significantly change the operating point. Hence, we need not take other than the normal precautions with common-base amplifiers.

Consider, now, the common-emitter circuit, Fig. 4·3. I_{CO} still flows between the base and collector sections, but now the base current *determines* the amount of collector current flowing. This, too, we noted in Chap. 3, and it was because of this relationship that we developed a second current-gain factor β, which is equal to $\alpha/(1 - \alpha)$, and values of β of 50 or more are not unusual. The total collector current now flowing is given by

$$I_C = \beta I_B + (1 + \beta)I_{CO}$$

Previously, the factor $(1 + \beta)I_{co}$ was ignored. When the transistor is subjected to fairly wide ranges in temperature, however, it is possible for this factor to develop values high enough to affect transistor operation seriously. I_{co} is extremely sensitive to temperature, and any increase in this current will be magnified 50 or more times because of the presence of $(1 + \beta)$. This can have a marked effect not only on the collector current but—what is equally disturbing to the circuit—on the bias or operating point. What it will do is shift this point out of the linear region, giving rise to an increase in distortion.

The insertion of a series resistor in the emitter leg is designed to prevent the foregoing action from occurring. If we use the circuit of Fig. 4·4 as an illustration, all of the collector current will flow through the emitter resistor. The voltage drop produced across R_2 serves to make the emitter negative with respect to ground. Note, however, that the base is also negative with respect to ground, and hence the base-emitter voltage will be the difference between the battery voltage drop across R_1 and the smaller voltage drop across R_2. Now let us say that the collector current rises because of a temperature-induced rise in I_{CO}. This will cause the voltage drop across R_2 to increase, making the overall base-to-emitter voltage less negative than it was before. This is actually working against the forward biasing voltage of the base-emitter circuit, resulting in *less* emitter current. Hence, we are counteracting the rise in I_C by decreasing I_B and I_E. In this way we achieve stabilization of our amplifier circuit.

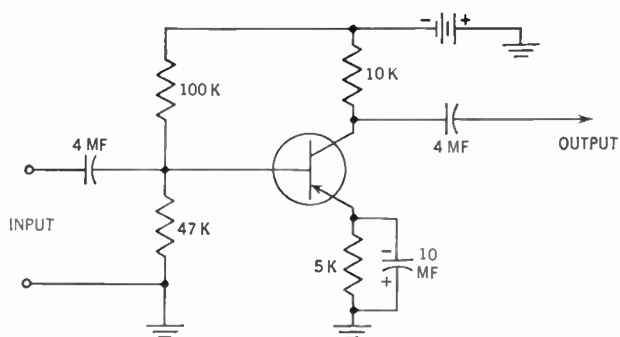


Fig. 4·5 A variation of the stabilized amplifier for Fig. 4·4.

A variation of this stabilization circuit is that shown in Fig. 4·5. Here the base is connected to a voltage divider. This arrangement provides greater stabilization than its predecessor, but the additional resistor does absorb more power from the battery and from the signal source. In this sense, then, this circuit is less efficient.

The reader will recognize that every transistor circuit has two kinds of stability, d-c and a-c (or signal). In d-c stability, we desire to maintain the same operating point irrespective of any changes in the transistor or in the values of any of the other components in the circuit. Generally, it is the changes in the transistor that cause the most trouble, but variations in the values of circuit resistors can also have a marked effect on circuit operation.

Alternating-current stability refers to the ability of the circuit to treat any signals that pass through in the same way. The most common variation occurs in the amplification accorded different-frequency signals, but different-amplitude signals can also be treated differently. For example, a strong signal may receive a greater distortion than a smaller signal. In Fig. 4·5, removal of the 10-μf capacitor shunting the 5,000-ohm emitter resistor will provide more uniform amplification to signals over a wider frequency and amplitude range than if this capacitor were permitted to remain. This is called current feedback. Another method is voltage feedback, and this will be illustrated presently.

Common-collector Amplifier

A single-stage common-collector amplifier is shown in Fig. 4·6. The input signal is applied between base and ground. Since the collector is at a-c ground potential because of C_1, however, we can say that

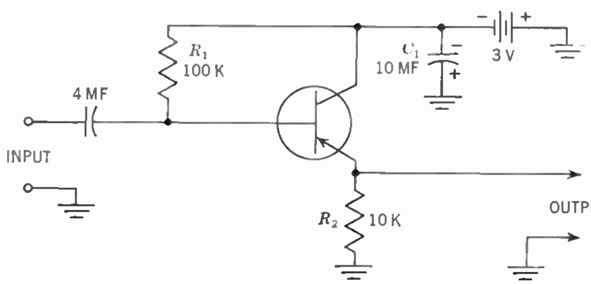


Fig. 4·6 A single-stage common-collector amplifier.

the signal is effectively being applied between base and collector. The output is taken from a load resistor between emitter and ground or, what is the same thing, between emitter and collector. This circuit has a high input impedance, on the order of 100,000 ohms or more, and an output impedance of 200 ohms. The voltage gain of a common-collector amplifier cannot exceed 1, and in the circuit shown it is about 0.9. Power gain, however, is 15. The voltage gain is relatively independent of frequency, but the current gain falls off with frequency exactly as it does for a common-emitter amplifier. The fall-off of power gain, with frequency, is therefore about midway between that of a common-base amplifier and that of a common-emitter amplifier.

Still another variation of the common-collector amplifier is shown in Fig. 4·7*a*. The load resistor is the 47,000-ohm resistor in the emitter leg. In addition, there is feedback (furnished by C_2) between the emitter and base. The latter is designed to decrease the shunting effect of the base voltage divider.

The frequency response of this circuit for two different bias voltages is shown in Fig. 4·7*b*. Note that in neither case does this gain exceed 1. (The change in circuit presentation is purposely being made to help the reader become familiar with the different methods of illustration that he will encounter. The important items to look for are the point of application of the input signal and the takeoff point of the output signal.)

Another name for this circuit is bootstrap circuit.

It is relatively simple to derive the current gain of a common-collector amplifier. The input current is I_B, since the incoming signal

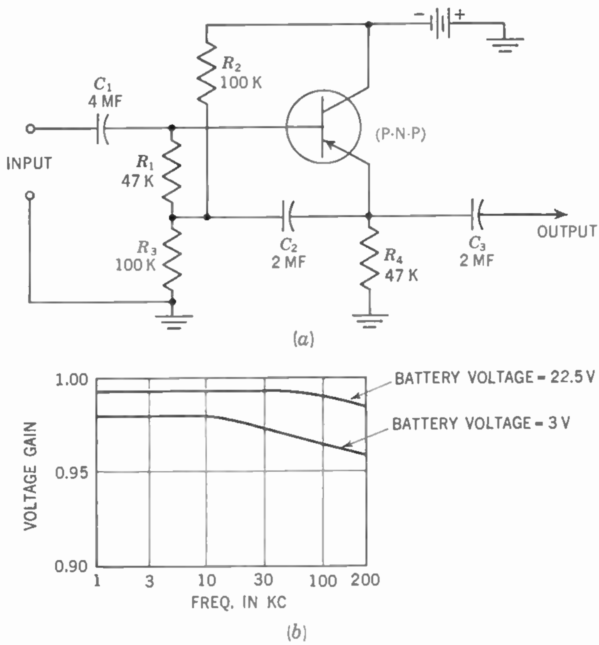


Fig. 4·7 (a) Another common-collector circuit. (b) The frequency response of the circuit shown in (a) for two different bias voltages. (After P. G. Sulzer, Junction Transistor Circuit Applications, *Electronics*, August, 1953)

is fed to the base. The output current is I_E, since the load resistor is in the emitter lead. Hence

$$\text{Current gain} = \frac{I_E}{I_B}$$

Further,

$$I_E = I_B + I_C$$

Thus,

$$\text{Current gain} = \frac{I_B + I_C}{I_B}$$

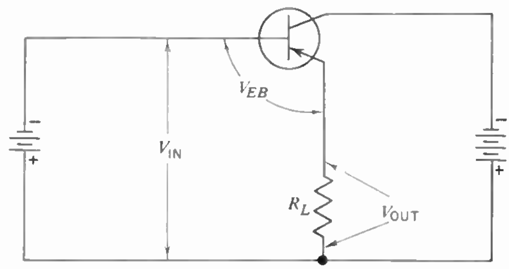
$$= 1 + \frac{I_C}{I_B}$$

$$\text{Current gain} = 1 + \beta$$

To all intents and purposes, this is equal to β, because beta is normally greater than 30. Thus, the common-emitter and the common-collector amplifiers provide the same current gain.

The input impedance of a common-collector stage is approximately equal to β times R_L, or βR_L. This can be shown as follows. The input impedance of the stage R_{IN} is equal to the input voltage divided by the input current. From Fig. 4·8, the input voltage is equal to

Fig. 4·8



$V_{EB} + V_{OUT}$. (These voltages can be added directly because they are always in phase for a resistive load. Thus,

$$R_{IN} = \frac{V_{EB} + V_{OUT}}{I_B}$$

$$R_{IN} = \frac{V_{EB}}{I_B} + \frac{V_{OUT}}{I_B} \tag{4·1}$$

Now, V_{EB}/I_B is the internal resistance between base and emitter r_{be}, and since this junction is forward-biased, the value is very low. On the other hand, V_{OUT}/I_B depends on the circuit as well as the transistor. We noted just above that I_E/I_B is equal to $\beta + 1$ or, essentially, β. From this we may state

$$\frac{V_{OUT}}{I_B} = \frac{\beta V_{OUT}}{I_E}$$

Since $V_{OUT}/I_E = R_L$, we have

$$\frac{V_{OUT}}{I_B} = \beta R_L$$

Substituting this into Eq. (4·1) above, we obtain

$$R_{IN} = r_{be} + \beta R_L$$
$$\cong \beta R_L$$

This demonstrates that it is desirable to use as high a value of R_L as practical in order to obtain a high input impedance. In turn, this means that for common-collector amplifiers, the operating bias current should be as low as possible; otherwise, the d-c voltage dropped across R_L would be so high that an uneconomically large bias supply voltage would be required.

One word of caution. R_{IN} is governed not only by R_L and β but also by any resistor which connects to the base. Thus, if the base bias current reaches this element through a low-valued resistor, the input impedance will be lowered accordingly because the resistor will shunt the input circuit. Also, circuit capacitance across the base will affect the input impedance, particularly as the operating frequency rises. This, of course, is just as true of transistor amplifiers as it is of vacuum-tube amplifiers. For example, at 16 kc, a 10-$\mu\mu$f capacitor presents an impedance of 1 megohm. At 100 kc, the capacitive impedance has decreased to about 167,000 ohms. The drop is significant, and it must be remembered that 10 $\mu\mu$f is quite small and readily developed through stray circuit-wiring capacitance alone. Hence, considerable care must be exercised when designing and constructing such stages.

By way of contrast, common-base and comon-emitter amplifiers have quite low input impedances and much more shunt capacitance would be required before the input impedance is affected.

Since β can also be represented by h_{FE} or h_{fe}, the above formulas are frequently shown using the latter symbols. h_{FE} is the d-c beta and h_{fe} is the a-c beta.

Cascaded Amplifiers

Transistor amplifiers, like vacuum-tube amplifiers, are seldom used singly. Rather, it is more common to find them in groups, with two, three, or more stages following each other in order, i.e., in cascade. When vacuum-tube amplifiers are used, it is a relatively simple matter to connect them one after the other, because a conventional vacuum-tube amplifier has a much higher input impedance than output impedance. Hence, when we attach the input of one stage to the output of the preceding stage, we do not ordinarily affect the preceding stage.

Consider, however, a transistor amplifier, say one designed with a common emitter. The input impedance is on the order of 1,000 ohms. The operating output impedance is more likely to be between 10,000 and 20,000 ohms. Obviously, a direct connection between two stages will result in a significant loss in gain due to the mismatch. If we accept this reduced gain, then it becomes necessary to use more stages in order to obtain a desired amplification. Another solution would be to insert a device (i.e., a step-down transformer) which will match the higher output impedance of one stage to the lower input impedance of the following stage. This solution has been used, and special miniature transformers, Fig. 4·9, have been devised for the purpose. Transformers, however, do not ordinarily possess the same flat frequency response that can be obtained from *RC* networks. Also, transformers are more costly, and hence it is often more desirable, from an economic standpoint, to add an extra amplifier stage and use *RC* coupling than to revert to transformer coupling. Both methods are used, however, and typical amplifiers of both types will be examined.

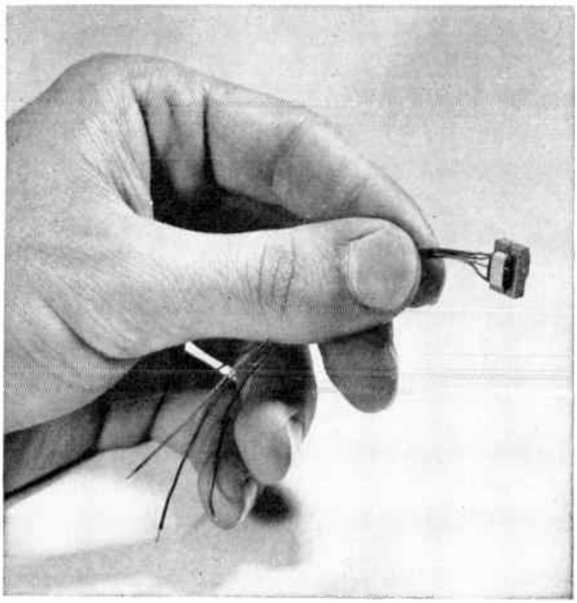


Fig. 4·9 A special miniature transformer designed for transistor application.

A two-stage transformer-coupled grounded-emitter amplifier is shown in Fig. 4·10. The interstage transformers have primary impedances of 20,000 ohms each and secondary impedances of 1,000 ohms each. Capacitors C_1 and C_2 are 10 μf in value, and resistors R_1 and R_2 are 150,000 ohms each. The two resistors are needed to establish the proper forward bias for the base-emitter circuits, and the two capacitors are inserted to prevent grounding of the base bias through the low d-c resistance of the transformer secondary windings. Overall power gain of this particular combination is approximately 50 db.

A resistance-coupled grounded-emitter amplifier that will provide approximately the same amount of overall power gain is shown in Fig.

4·11. Note that three stages are required because of the mismatch between the output of one stage and the input of the following stage.

A two-stage amplifier with high input impedance and d-c stabilization is shown in Fig. 4·12. The higher input impedance is achieved by

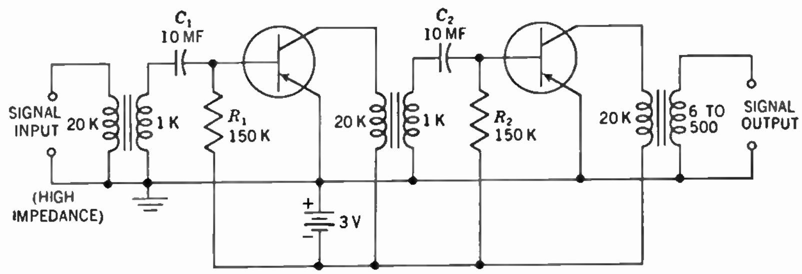


Fig. 4·10 A two-stage transformer-coupled grounded-emitter amplifier. Two 2N464 transistors are used.

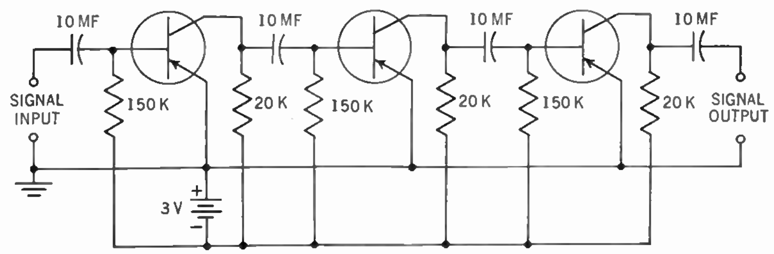


Fig. 4·11 A resistance-coupled grounded-emitter amplifier that will provide approximately the same amount of overall power gain as the amplifier of Fig. 4·10. Three 2N464 transistors are used.

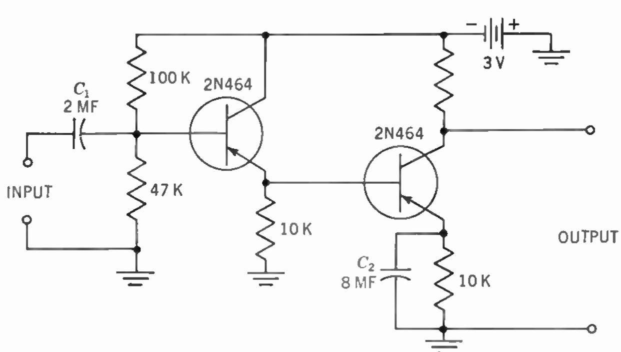


Fig. 4·12 A two-stage amplifier with high input impedance and d-c stabilization.

the use of a grounded-collector stage, the signal of which is forwarded to a grounded-emitter amplifier. Insertion of 10,000-ohm resistors in the emitter leads of both transistors provides amplifier stabilization against temperature changes. The first 10,000-ohm resistor cannot, of course, be bypassed, since the signal is obtained from this point. In the second stage, however, an 8-μf bypassing capacitor is employed.

The frequency-response behavior of this two-stage amplifier, at two different bias voltages, is shown in Fig. 4·13. In transistor circuits, as in vacuum-tube circuits, we obtain more gain for higher voltages. The only precautions to observe are those dictated by the maximum

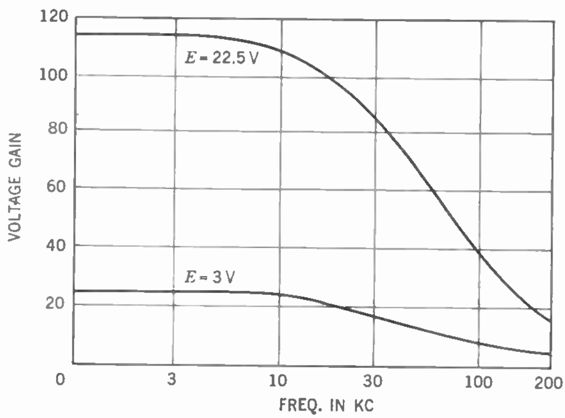


Fig. 4·13 The frequency-response behavior of the amplifier shown in Fig. 4·12 at two different bias voltages.

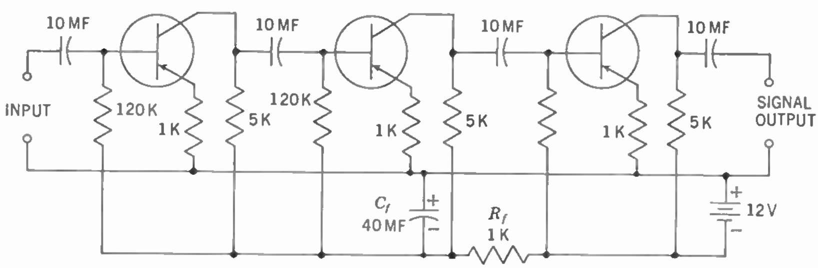


Fig. 4·14 A resistance-coupled amplifier with a decoupling filter C_f and R_f. It might be desirable to add additional filter sections if motorboating is encountered. The transistors are 2N191's.

safe operating currents, the temperature, the breakdown-voltage values, and collector dissipation.

In cascaded amplifier circuits, it is frequently desirable to employ a decoupling filter across the battery or power supply. This is shown in Fig. 4·14, where R_f and C_f serve this function. The need for these components stems from the impedance of the power source (be it battery or an a-c supply) and the necessity of preventing positive feedback from a later stage where the signal level is high to a prior stage where it is low.

In choosing values for R_f and C_f, it is best to restrict R_f to fairly small values so that the operating voltage to the stages situated prior to R_f is not reduced to a value which will materially affect their gain. Remember, however, that the smaller R_f becomes, the larger C_f must be made in order to obtain effective filtering action. In general, the time constant of $R_f \times C_f$ should be greater than $1/f$ for the lowest frequency passed by the amplifier. In computing this time constant, C_f is expressed in farads, R_f in ohms, and f in cycles per second. The values indicated for R_f and C_f are typical.

Negative Feedback in Transistor Amplifiers

Negative feedback can be employed in transistor amplifiers for the same reason and in the same manner as in vacuum-tube amplifiers. Negative feedback will improve amplifier stability, reduce distortion, increase input impedances, and reduce the variations in gain caused by different transistors (or tubes). The last feature is particularly important in transistor amplifiers because of the fairly wide range in characteristics that one encounters among transistors of the same type. Fortunately, this situation is being steadily improved, and one can depend upon a greater uniformity among transistors today than, say, two or three years ago. However, a wide latitude will still be found among similar units, and the use of negative feedback can often reduce the variations in amplifier performance caused by these differences to a considerable extent.

All the advantages of negative feedback are not obtained without some penalty, this being the loss in gain. The loss is not a serious one, however, because of the higher and higher voltage amplifications which transistors are providing. We seldom lack sufficient gain; usually, we have more than we actually require.

A simple form of negative feedback is obtained by leaving the d-c stabilization resistor in the emitter lead unbypassed. This is a single-stage approach to be used or not as desired by each of the various stages.

Another form of negative feedback is shown in the two-stage amplifier of Fig. 4·15. The feedback loop here extends from the output circuit of X_2 to the emitter circuit of X_1. Involved in this feedback are two resistors R_1 and R_7 and one capacitor C_4. R_1 is needed to provide a means of inserting the feedback energy into the emitter circuit of X_1; hence, it was left unbypassed. R_2, in the same emitter circuit, is bypassed by C_1, and no feedback voltage is developed across these two parallel components. R_2, however, does provide d-c stabilization for X_1. R_5 does the same for X_2.

In the negative feedback of voltage, we know that the phase of the signal fed back must be 180° from the phase of the incoming signal. It may be instructive to check the signal polarities in the circuit of Fig. 4·15 to see if this condition holds true. (The procedure will also help the reader become familiar with the methods of checking signal polarities in transistor circuits.) If we assume that the incoming signal, applied to the base of X_1, is positive at some instant, then the signal voltage at the collector of this transistor is negative. This stems from the 180° phase reversal that occurs in a common-emitter amplifier.

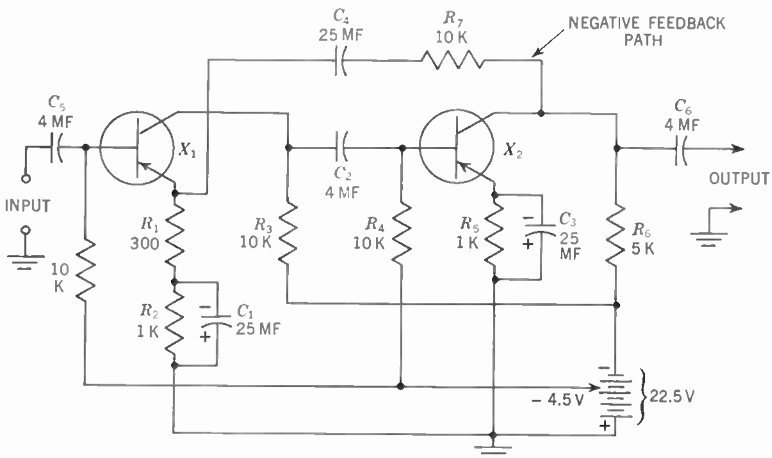


Fig. 4·15 A two-stage amplifier employing negative feedback. Two 2N464 transistors or their equivalent are employed.

The negative signal at the collector of X_1 is also negative at the base of X_2. This produces a positive signal at the collector of X_2, and a portion of this signal is fed back to the emitter of X_1. Thus, we have a positive signal at the base and a smaller positive signal at the emitter. Since these two voltages will work in opposition to each other in forcing current through the emitter base, we obtain negative feedback.

The effect of negative feedback on the frequency response is shown in Fig. 4·16. Note how much flatter the curve is with the feedback. Also instructive are the two curves in Fig. 4·17. The left-hand curve shows how the overall gain will vary with different transistors when there is no feedback. Note how much better the action becomes when feedback is employed.

The point of feedback return is governed by the phase conditions in the circuit around which the feedback is sent. Consider, for example, the two-stage amplifier circuit shown in Fig. 4·18. The first stage is a grounded collector, and it does not introduce any phase reversal in the signal. The second stage is a grounded emitter, and it causes a 180° reversal. Under these conditions, the feedback line from the collector of the second stage can be brought back to any signal point prior to the base of this stage. The point chosen in this particular circuit is the base of the grounded-collector stage, but it could just

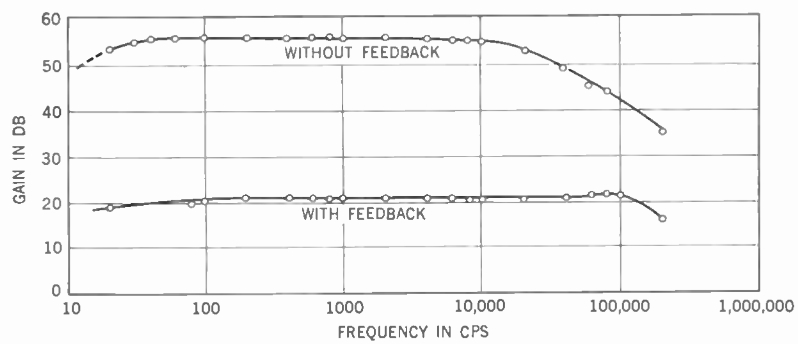


Fig. 4 16 The effect of negative feedback on the frequency response of the amplifier shown in Fig. 4·15. (*Electronics*)

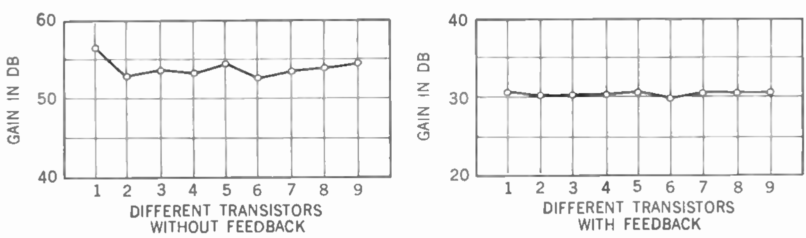


Fig. 4·17 Negative feedback in an amplifier (such as the one shown in Fig. 4·15) will serve to reduce the variations in overall gain that different transistors will introduce owing to varying characteristics.

as easily have been the emitter of this stage or the base of the second stage. At all these points, signal polarity is the same. The feedback in Fig. 4·18 is voltage feedback, in contrast to the current feedback obtained when the emitter resistor is left unbypassed.

Transistor-circuit Considerations

In dealing with transistors and transistor circuits, a careful distinction must be drawn between the limitations of the transistor as a device and the circuit in which it is placed. Perhaps the most obvious example of this occurs at the low-frequency end of an amplifier response characteristic. Here it is the external circuit elements—capacitors and resistors—which are responsible for a drop-off in gain. The transistor itself is capable of amplifying down to direct current, and once the signal is brought to the transistor, it will receive as much amplification at 30 cycles as at 300 or 3,000 cycles. It is in bringing the signal to the transistor through the coupling network that fall-off occurs.

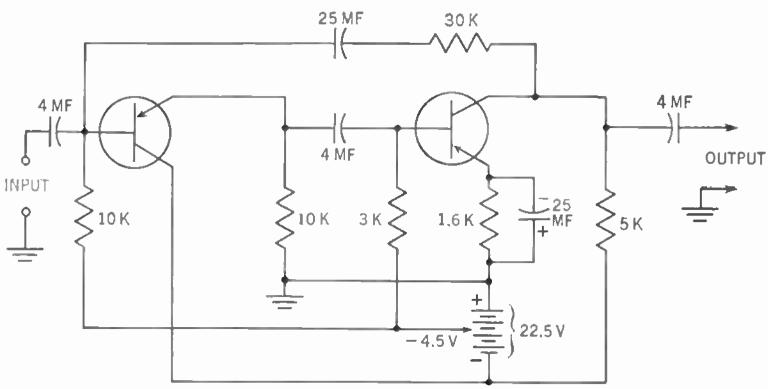


Fig. 4·18 Another negative-feedback arrangement. (*Electronics*)

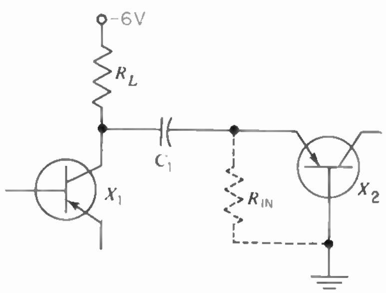


Fig. 4·19 Simplified diagram of interstage coupling network between two transistor stages.

At this point, still another factor must be recognized. That is, while the transistor itself has uniform amplification at the low frequencies, it can *influence* the external circuitry by its input and output impedances. For example, the common-base amplifier has a very low input impedance. That means that any coupling capacitor connected to the transistor input must itself possess a very small impedance so that it does not rob any signal from the transistor. This is shown in simplified form in Fig. 4·19. The current leaving the collector of the preceding stage sees two paths to follow. It can either flow through the collector load resistor R_L or it can go through C_1 and the input impedance of the following transistor. In both instances it can return to the emitter of its own transistor, thereby completing the circuit.

Now, in order to bring as much signal to X_2 as possible, it is desirable to divert as much collector current through C_1 and the input impedance of X_2 as possible. Since the input impedance of X_2 is very low, C_1 must be made as large as possible so that its reactance will be small.

Consider, now, a common-emitter circuit at X_2. Its input impedance will be larger than for a common-base circuit. Consequently, it will not be necessary to employ as large a capacitor C_1 as before because, with the input impedance larger, lowering the impedance of C_1 below a certain point will not measurably alter the total series impedance of the input circuit. Finally, with a common-collector arrangement at X_2, we will see a large input impedance. This means that C_1 will have even *less* effect on the total series impedance and so a smaller value of C_1 can be employed than in either of the two preceding circuits.

The reader will recognize, of course, that what we are dealing with here is the relative impedance of C_1 and the input to the transistor. If the transistor input is lower, C_1 will need to be larger before its effect can be considered negligible. As the input impedance rises, the best value for C_1 can change accordingly. In this sense, the transistor can affect its connecting circuit. But given a certain transistor, we can then look to the external circuit to establish the low-frequency end of the network.

At the high-frequency end of the amplifier response, it is the transistor and the drop-off in its β that govern circuit behavior. This will be discussed more fully presently. While the transistor is usually the primary cause, poor external-circuit design can also affect the high-frequency cutoff. For example, too large a shunting capacitor will lower the upper cutoff frequency. This, too, was noted previously. Also, too small a coupling capacitor can reduce the signal voltage reaching the next stage. But if the circuit has been properly designed, then in most instances it is the transistor which establishes the upper frequency limit.

Bias considerations. Another factor to consider when comparing the three transistor-amplifier arrangements is the precaution to observe when biasing these circuits. A very common biasing method, one which helps to maintain amplifier stability, is the circuit shown in Fig. 4·20. R_1 and R_2 form a voltage divider which provides the base with the necessary potential to forward-bias the base-emitter diode. With a given battery voltage, R_1 and R_2 can assume a wide range of values to achieve the proper voltage division for the base. The smaller

R_1 and R_2 are, however, the more stable the circuit. Now, let us see what complications this brings with the three transistor configurations.

For a common-base amplifier, R_1 and R_2 can have their lowest values because the input impedance of the common-base arrangement is low. Hence, very little difficulty is encountered here.

In a common-emitter circuit, the input impedance is higher and we must be careful not to make R_1 and R_2 so low that they bypass the signal current arriving from the preceding stage. Also, if R_1 and R_2 are low, they will load the preceding stage more than the input impedance of the transistor they are attached to would, and this will lead to a lower overall circuit gain. In selecting higher values for R_1 and R_2, however, we reduce the temperature stability of the circuit.

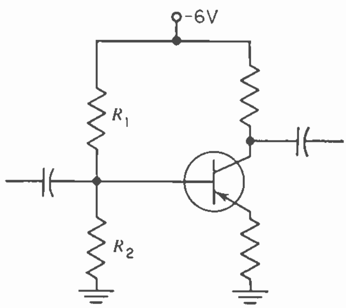


Fig. 4·20 A widely used biasing network, R_1 and R_2.

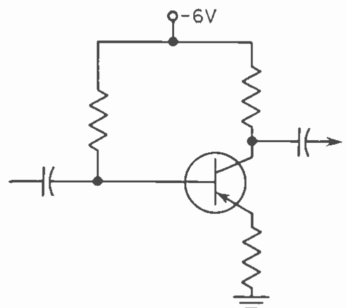


Fig. 4·21 Another method of biasing a transistor.

Finally, in a common-collector stage, where the input impedance is very high, we would try to avoid using the bias approach of Fig. 4·20 and go to something like the method shown in Fig. 4·21. The temperature stability of the stage with this bias network of Fig. 4·21 is much poorer than it would be by using the bias circuit of Fig. 4·20, but R_1 and R_2 of Fig. 4·20 would act to reduce the input impedance and thus undercut one of the major features of the common collector.

Direct-coupled Amplifiers

We have spoken of and demonstrated RC- and transformer-coupled amplifiers. Another type of amplifier that is extensively used is the direct-coupled amplifier. In this circuit, a d-c path exists from the output of one stage to the input of the next stage. In its simplest form, a direct-coupled stage would appear as shown in Fig. 4·22. Here, the output device, a pair of headphones, is directly connected to the collector element of the amplifier stage. In order to employ the phones in this manner, their impedance should match the amplifier output, their operation should not be affected by the collector current flowing through them, and their d-c resistance should not be too high or the resulting voltage drop will reduce the collector voltage to too low a value. In place of phones, we might use a relay, a meter, or any one of a number of devices.

Another direct-coupled amplifier is illustrated in Fig. 4·12, where a direct path exists between the emitter of the grounded-collector stage and the base of the following grounded emitter. Any decrease in low-frequency response in this circuit would be due entirely to the input capacitor C_1 and the second emitter bypass capacitor C_2.

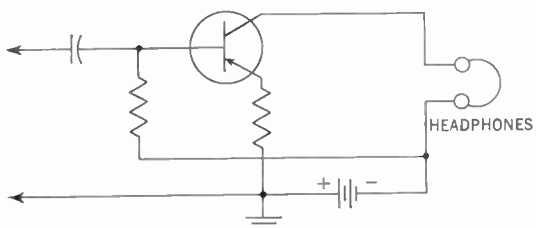


Fig. 4·22 A simple direct-coupled transistor amplifier.

Another type of direct-coupled transistor amplifier takes advantage of the fact that there are two basic kinds of junction transistors: NPN and PNP units. Each is the symmetrical counterpart of the other, and the polarity of an input signal necessary to increase conduction in a PNP transistor is the opposite of that necessary to increase conduction in an NPN transistor.

A direct-coupled amplifier that makes use of this symmetry is shown in Fig. 4·23. The first transistor is an NPN unit; the second, a PNP type. The first stage is set up so that the collector current flowing through its load resistor R_4 develops just enough voltage there to make the base of the PNP transistor negative with respect to its emitter. This establishes the proper conditions in the emitter-base circuit of the PNP unit to bias it in the forward direction. Thus, by the proper choice of resistor values and battery potential, both stages will operate as class A amplifiers.

The application of a signal to the base of the NPN stage will result in an amplified signal appearing across R_6. For example, when the signal at the base of the NPN transistor goes positive, an amplified negative voltage will appear across R_4 (collector end negative with respect to battery end). This increasing negative voltage will provide an even greater forward bias for the base-emitter circuit of the PNP transistor and cause an increased flow of current through this unit. Electrons will flow up through R_6, making the top end positive with respect to the bottom end.

Thus, the positive signal applied to the input of this amplifier appears with the same polarity, but in amplified form, at the output.

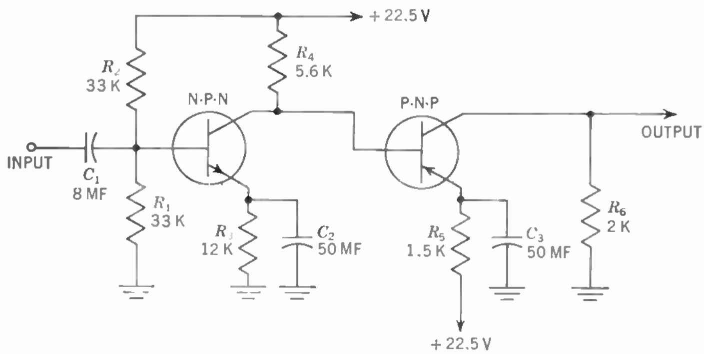


Fig. 4·23 A direct-coupled amplifier that makes use of the complementary nature of NPN and PNP transistors. Voltage gain of this system is 660; power gain is 53 db. (After R. D. Lohmen, Complementary Symmetry Transistor Circuits, *Electronics*, September, 1953.)

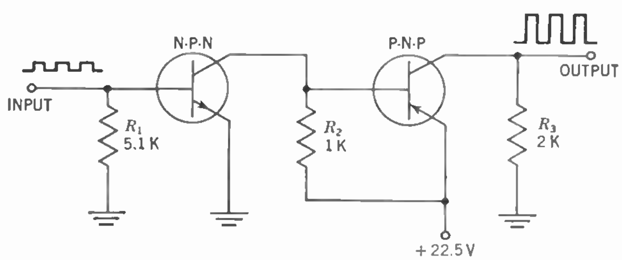


Fig. 4·24 A direct-coupled pulse amplifier.

Note the simplicity of this arrangement, requiring as it does no coupling capacitors and only one battery supply. (Both 22½-volt potentials shown would come from one source.)

Another two-stage amplifier designed along somewhat similar lines is shown in Fig. 4·24. This system has for its sole purpose the amplification of pulses, and its mode of operation is therefore modified accordingly. For example, if you examine the base-to-emitter circuits of both stages, you will note that no forward bias is employed. The characteristic curves for grounded-emitter operation (such as we have here) reveal that when the base current is zero (that is, $I_b = 0$), the collector current is quite small. In terms of operation, this means that the transistor is biased close to cutoff. This is true in both amplifiers of Fig. 4·24, although the second stage is not so close to cutoff as the first stage. This is because the small collector current that flows from the first transistor passes through R_2 and the small forward biasing voltage that is developed shifts the operating point of the second transistor away from cutoff.

In the circuit of Fig. 4·24, a positive pulse of 0.25 volt input to the first stage is amplified to a 20-volt peak at the output of the second stage. Conduction is required only when the pulses are applied, hence the reason for the cutoff biasing.

Direct-coupled amplifiers are not used as extensively as they might be because of the stability problems they present. Any change in the d-c operating point of one stage immediately alters the d-c operating points of all succeeding stages. This in itself might not be so bad if it were not for the fact that most stages are connected common-emitter and a change in input base current produces β times that change in the collector circuit. Since common values of β range from 50 to 100 or more, it will be readily recognized that in a direct-coupled string of stages, an infinitesimal change in an early stage will quickly build up, after a few stages, to substantial proportions. To prevent this buildup, considerable d-c feedback must be employed, and this will not only add to the expense of the circuit but also generally act to affect the gain adversely.

The principal offender, in most instances, is the change in I_{co} with temperature. Hence, any arrangement of direct-coupled amplifiers that may be devised must somehow minimize the effect of changes in I_{co}. This has led to circuit designs such as the one shown in Fig. 4·25. Two 2N1428 silicon transistors are directly coupled, with the signal output from X_1 proceeding directly through R_4 to the base of X_2. The two transistors share a common-emitter resistor R_E, and this has the effect of eliminating the degeneration that normally results from an unbypassed emitter resistor. Degeneration is effectively eliminated because the a-c signals through R_E from each transistor are 180° out of phase.

The effect of the I_{co_1} of X_1 on X_2 can be seen from the following analysis. Any increase in I_{co_1} will cause a change in the collector current of X_1 by a factor of βI_{co_1}. This increased current flowing through R_3 will produce a greater voltage drop, with the collector end of R_3 becoming more positive. Since X_2 is a PNP transistor, this increase in positive voltage across R_3 will drive it closer to cutoff;

hence, this will alter the operating point of X_2 and certainly modify its effect on any signals passing through the circuit.

In this arrangement, the I_{co_2} of X_2 and the I_{co_1} of X_1 are made to offset each other's effects. I_{co_1} flows in the path indicated by the solid arrows in Fig. 4·25, and I_{co_2} flows in the path indicated by the dotted arrows. By adjusting the value of R_4, we can regulate the additional current that the voltage across R_3 sends into the base of X_2. (Consider the additional voltage drop across R_3 as a small battery. Then the current fed to the base of X_2 will be proportional to this voltage divided by

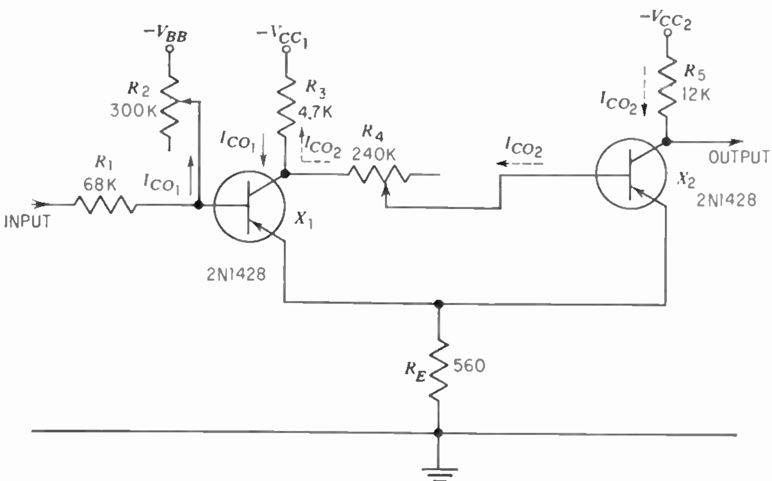


Fig. 4·25 A low-drift direct-coupled amplifier. (Philco Corp.)

R_4. This, of course, is Ohm's law.) R_4 is selected to have this current equal I_{co_2}. Since the two currents flow in opposite directions, changes in I_{co_1} and I_{co_2} can be minimized. By using similar transistors for X_1 and X_2, we are more likely to have fairly identical variations in I_{co_1} and I_{co_2}, and this will permit compensation over a wide range of temperatures. Figure 4·26 shows the low drift in voltage gain with temperature for this arrangement.

Two bias supplies are required in the amplifier. This is the case because R_3 is small compared to R_5, yet each transistor should have approximately the same operating point. Since the emitter voltage is developed across a common-emitter resistor, the collector supply voltage is necessarily smaller for the transistor with the smaller load resistor.

Power Amplifiers

Since all transistors are current-operated devices, all serve essentially as power amplifiers. In this section, however, we are concerned with high-power units, those which are capable of dissipating powers in excess of 1 watt.

A high-power transistor means, in essence, one that develops considerable amounts of current, generally in the ampere range. The initial approach to this goal is to simply enlarge the dimensions of the emitter, base, and collector elements until the desired amount of current is achieved at the specified heat-dissipation rating. For the lower-power transistors, this is all that is required.

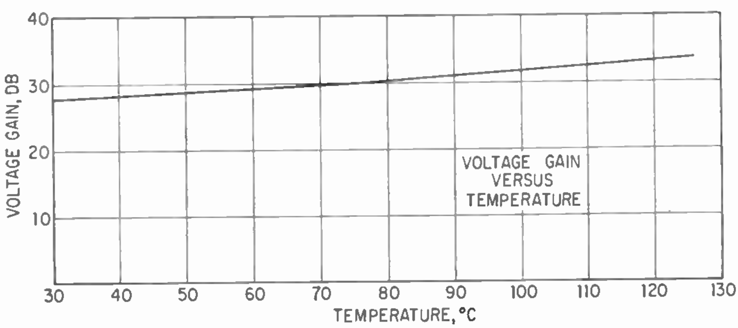


Fig. 4·26 The variation in voltage gain vs. temperature for the direct-coupled amplifier of Fig. 4·25.

As the transistor dimensions are increased, however, it is found that the emitter efficiency (or the emitter ability to inject carriers into the base) decreases. The reason for this stems from the fact that in a very large transistor, the base current is fairly substantial and, in flowing through the base layer, produces an ohmic voltage drop which reduces the forward bias on parts of the emitter distant from the base electrode. As a result, most of the emitter current is obtained from a very narrow strip around the edge of the emitter, while the central emitter region is less effective.

To get around this difficulty, a number of manufacturers use a ring-shaped emitter, Fig. 4·27. By making this ring large enough, sufficient current is obtained; at the same time, because of its shape, efficiency is maintained.

A variation of this construction is shown in Fig. 4·28. A small second base contact is placed at the center of the emitter ring and connected to the outer base ring. This additional base segment develops an electric field which helps to pull carriers from the emitter inward toward the collector and thereby assists in directing the flow of current from emitter to collector.

Still another high-power transistor is shown in Fig. 4·29. This transistor employs a mesa-like construction, using a long narrow emitter (in the center) surrounded on three sides by the base electrode. The collector, which cannot be seen, is underneath.

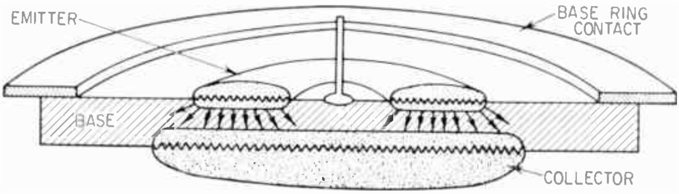


Fig. 4·27 A power transistor utilizing a ring-shaped emitter.

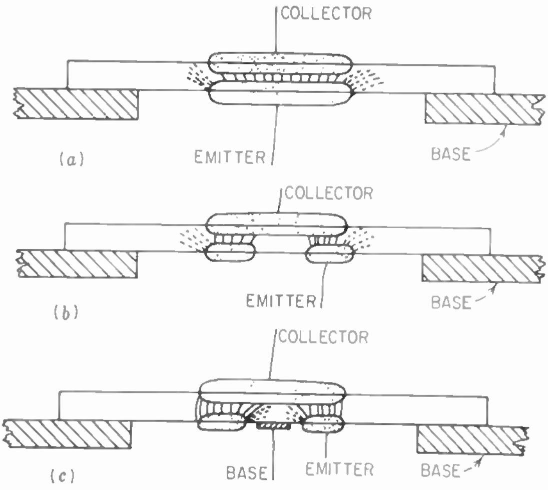


Fig. 4·28

Alloy-junction transistors, as well as many of the other types previously discussed, have been employed for power application. The power transistor shown in Fig. 4·30, for example, combines alloying and diffusion techniques. An alloyed junction is employed for the emitter, while the collector has a diffused structure. This makes it possible to eliminate the indium or lead alloy at the collector, where its high thermal resistance between the collector junction and the heat sink impedes the removal of the heat generated at the junction. With the diffused construction, a lower-thermal-resistance connection can be made to the heat sink, thereby permitting higher allowable dissipation.

Thermal resistance impedes the flow of heat just as ohmic resistance impedes the flow of current. By employing materials having low thermal resistance, we are able to draw the heat away from the collector junction more quickly and prevent a buildup of heat with its consequent deleterious effect on the transistor.

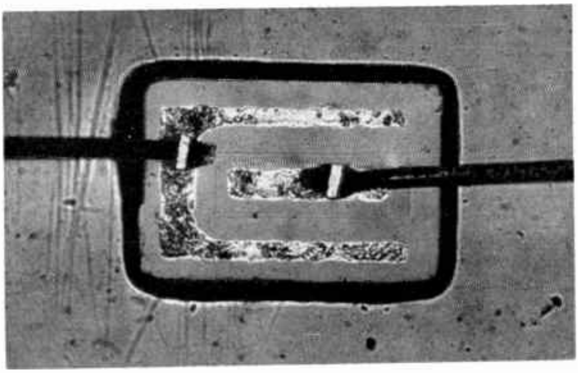


Fig. 4·29 A power transistor using a mesa-like construction.

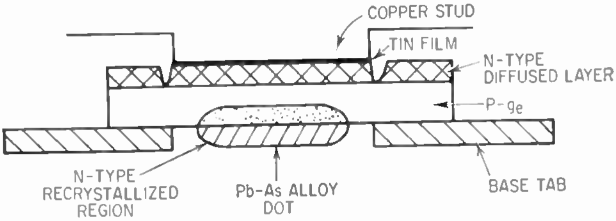


Fig. 4·30 Cross section of an alloy-emitter diffused-collector power transistor. (*RCA Review*)

In order to keep the junction temperature at a safe value, it is usually necessary to provide a means for rapidly transferring heat from the junction to the surrounding air or other medium. Copper and aluminum are suitable materials for the construction of heat sinks which have low thermal resistance. The efficiency of a heat sink increases with an increase in the surface area which is exposed to the cooling medium. (Application Note 1-A, Transistor Heat Sinks, Delco Radio Division, Kokomo, Indiana.) When a transistor is used in a low-dissipation application, the use of a copper or aluminum chassis will provide sufficient cooling if there are no other sources of heat on the same chassis.

High-dissipation applications will generate more junction heat than can safely be dissipated by the chassis itself. If unlimited space is available, plain sheets of copper or aluminum may be used as heat sinks. Figure 4·31 indicates the variation which may be expected in the heat-radiating efficiency of ⅛-in. aluminum as its area and orientation

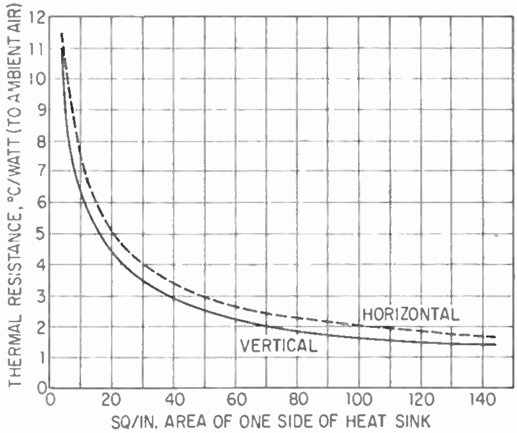


Fig. 4·31 The variation in heat-transfer characteristics of ⅛-in.-thick aluminum as its area and method of mounting are varied.

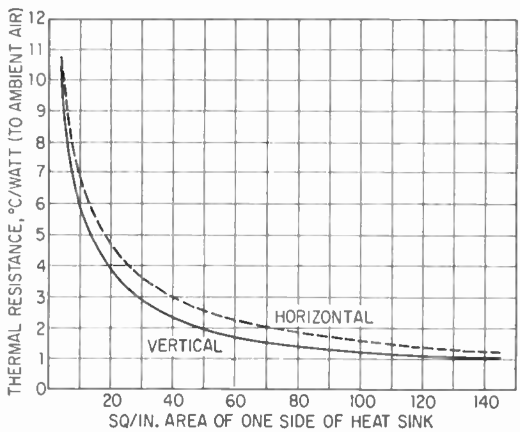


Fig. 4·32 The variation in heat-radiating efficiency of ⅛-in.-thick copper as its area and method of mounting are varied.

are varied. Figure 4·32 gives corresponding data for ⅛-in. sheet copper.

When space limitations forbid the use of large sheets of metal as heat sinks, it is possible to use finned heat sinks of a design which permits the concentration of large surface areas within a small space. The heat sink shown in Fig. 4·33 compresses 80 in.^{2} of radiating surface into an overall volume which measures 4⅝ by 1⅜ by 3 in.

Made of extruded aluminum, the heat sink is painted flat black to facilitate the radiation of heat.

The graph in Fig. 4·34 shows the temperature differential between the collector mounting stud and the ambient air at various collector-dissipation levels, with the heat sink in the horizontal and vertical positions.

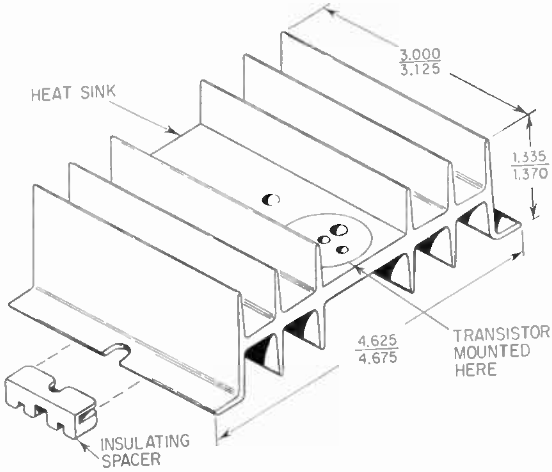


Fig. 4·33 A heat sink for limited space.

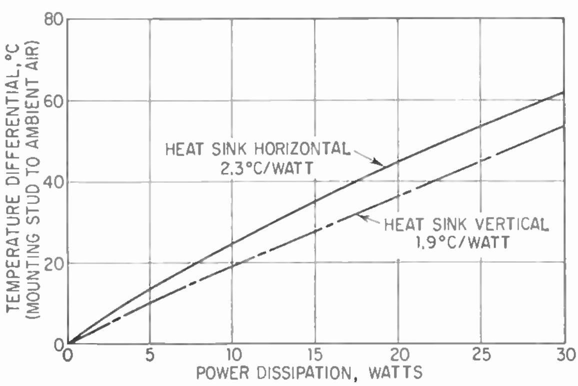


Fig. 4·34 Thermal characteristics of the heat sink shown in Fig. 4·33.

Figure 4·33 also shows an insulating spacer which may be used for electrically isolating the heat sink from the chassis. Electrical isolation of the transistor from the heat sink can be accomplished by special transistor-mounting kits shown in Figs. 4·35 and 4·36. The mica insulators will, however, have a thermal resistance which must be added to the thermal resistance of the heat sink. Best heat dissipation

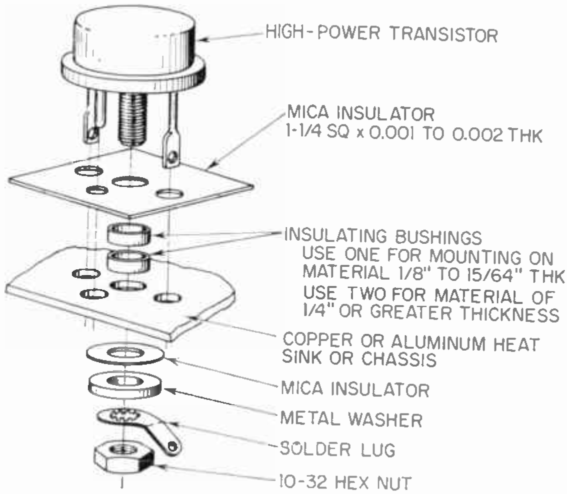


Fig. 4·35 The method of mounting a transistor on a heat sink.

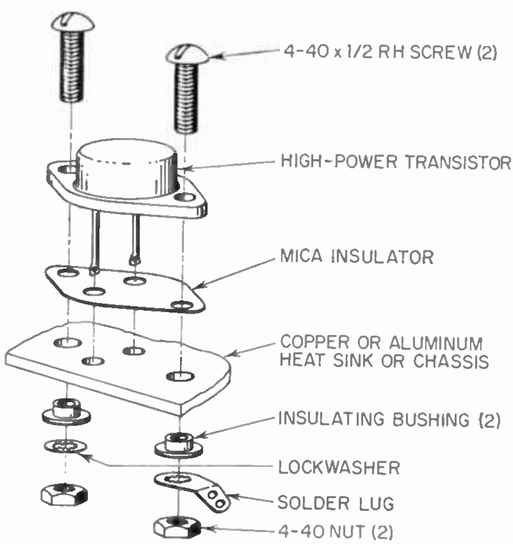


Fig. 4·36 Heat sinking another type of transistor.

is achieved when the transistor is mounted near the lower edge of the vertically mounted heat sink. This heat sink is prepunched to allow mounting either of the illustrated transistors—JETEC Type TO-3 (diamond) and JETEC Type TO-6 (doorknob)—directly to the heat sink. If the transistor is to be insulated from the heat sink, the user may enlarge the mounting holes to allow the use of the appropriate mounting kit.

A variety of heat sinks are available, and some typical units are shown in Fig. 3·2.

Any consideration of transistor heat dissipation should include an explanation of the factors which affect the generation of heat in a transistor junction. As a generalization it may be said that heating is determined by the mode of operation (switching service, audio amplifier, etc.), the level of bias and signal applied to the transistor, and the waveform of the applied signal.

Operation of a transistor as a square-wave oscillator or as a class B amplifier of square waves results in less junction heating than does the generation of sine or complex waves or operation as a class A amplifier.

In the generation or amplification of square waves, the transistor is alternately cut off and saturated. When cut off, no heating occurs because there is practically no current flowing through the junction. When saturated, the resistance of the junction is so low that there is practically no voltage developed across the junction which would cause power losses.

Efficient square-wave operation requires the application of relatively high levels of signal to the control element of the transistor in order to assure complete cutoff and saturation. Any rounding of the waveform indicates power losses which are dissipated as heat.

Class A operation of transistors for the amplification or generation of a sine or complex wave results in the dissipation of power in the transistor junction. This is true because the transistor is operating essentially as a variable resistor rather than as a switch. In class A operation the transistor bias is adjusted to a value which will keep the transistor operating on the linear portion of its characteristic curve. Thus there is flowing a no-signal collector current which is equal to about one-half of the peak value of collector current. The power dissipated in the junction by this no-signal current is in the form of heat.

Power transistors are physically larger than low-level transistors and are mounted in metal cases which offer low thermal resistance. Several typical units are shown in Fig. 3·5.

Class A power amplifiers. Circuit arrangements of single-ended power amplifiers do not differ to any marked degree from those of corresponding voltage amplifiers. Figure 4·37 illustrates two class A power amplifiers designed to drive the loudspeaker of an audio amplifier (or a radio or television receiver). In one instance, two batteries are employed; in the other, a single battery is employed. The output transformer would be designed to match the impedance of the collector on one hand and that of the loudspeaker on the other. The amount of power that may be obtained from this arrangement will be governed by the size of the battery and the permissible dissipation in the transistor itself. As in vacuum-tube practice, a single-ended power amplifier can be operated only class A.

Power amplifiers can also be operated in push-pull. A typical illustration of an audio amplifier using a single driver stage and a class A

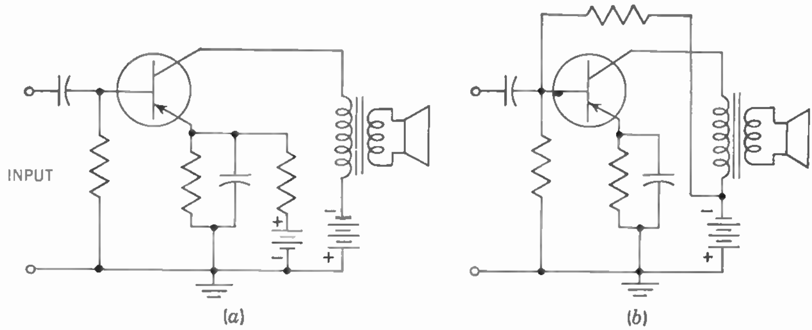


Fig. 4·37 Two class A power amplifiers.

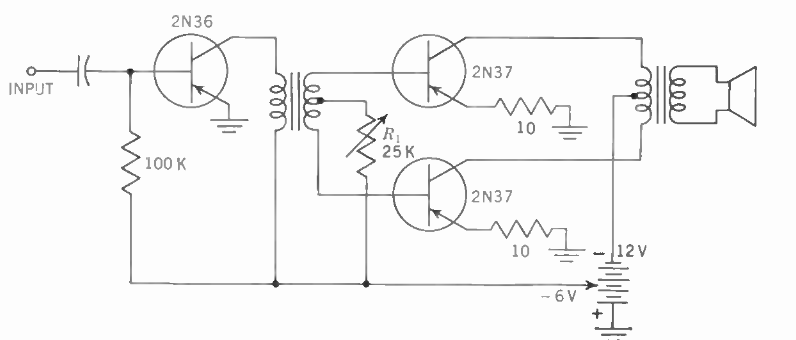


Fig. 4·38 An amplifier using a 2N36 driver and two 2N37's in push-pull. All these transistors are PNP units. The 10-ohm resistors in the emitter leads are designed to stabilize the transistor against temperature changes.

push-pull output stage is shown in Fig. 4·38. All transistors are operated with common emitters, and transformer coupling is employed between the driver and output stages and between the output amplifiers and the loudspeaker. The resistance R_1 is variable and is adjusted for a total collector current of 8 ma.

Push-pull amplifier operation results in the cancellation of second harmonics within the stage. For the same amount of distortion, then, a class A push-pull amplifier can be driven harder, thereby providing greater output. This also means that we can obtain more output with push-pull operation than we can get by using two similar transistors as single-ended amplifiers.

Class B power amplifiers. In class A push-pull operation, the average current that flows remains steady, whether or not a signal is being applied to the stage. More efficient operation can be achieved with class B operation, where each transistor is biased to cutoff. When no signal is applied, practically no current flows and no power is being dissipated.

The circuit of a class B push-pull amplifier is shown in Fig. 4·39. Three power transistors are employed; the first one serves as a class A

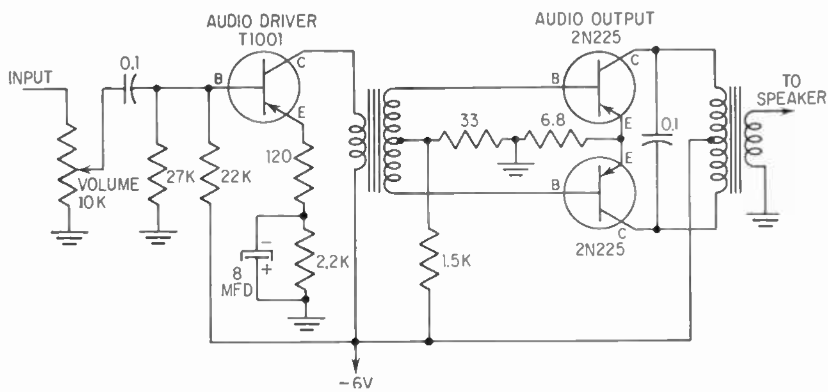


Fig. 4·39 A two-stage audio amplifier. The output stage is operated class B.

driver amplifier and the remaining two serve as a class B output stage. Efficiency of the class B stage is close to 75 per cent. This is achieved because, with no signal, the total class B collector current is extremely low, since the stage is biased near cutoff. In a class A amplifier, the efficiency is perhaps half this amount or less because a fairly sizable collector current always flows, signal or no.

Input signals are applied to the base of the first audio amplifier. A 2,200-ohm bypassed resistor in the emitter circuit of this driver stage provides thermal stabilization only; it does not introduce signal degeneration. However, just above it is an unbypassed 120-ohm resistor, and this does provide signal degeneration.

The output of the driver stage is transformer-coupled to the class B amplifier not only for proper impedance matching but also to provide two signals 180° out of phase with each other (as required by the class B amplifier). A 6.8-ohm resistor in the emitter circuit of this output stage introduces a small amount of signal degeneration to improve the stability of the circuit. The base circuit contains a 33-ohm resistor, across which a small voltage from the negative 6-volt supply is developed. The purpose of this voltage is to reduce crossover distortion. (This will be explained presently.) The collector elements of the two output transistors connect to opposite ends of the output transformer, and the d-c voltage is brought in at the center tap. The 0.1-μf capacitor across the primary of the output transformer removes the highs from the output signal for a more mellow output tone. The speaker is an extremely small one which, because of its dimensions, naturally tends to emphasize the higher frequencies. This is counteracted somewhat by the 0.1-μf capacitor.

Class B audio amplifiers are favored in many transistor receivers not only for their greater power and reduced distortion but also because their current drain is practically zero when no signal is being received. If two class A output amplifiers were connected in push-pull, an average current would always flow and impose a constant drain on the battery. Since these are power transistors, their current requirements are fairly large and a significant amount of power would be dissipated.

Now let us examine the reason for the 33-ohm resistor in the base circuit of the output amplifier. Its purpose is not so much to provide base-emitter bias as it is to minimize a condition known as crossover distortion. When transistors are connected back to back in push-pull arrangements and the bias is zero, there is a region near cutoff where their respective characteristic curves tend to become nonlinear, Fig. 4·40*a* and *b*. Note the jog in both curves near the origin. If we now introduce a sinusoidal base current into the input circuit, Fig. 4·40*c*, we will obtain the distorted collector current indicated to the right of the characteristic curve. This distortion becomes more severe as the signal level decreases.

To prevent crossover distortion, a small amount of forward bias is introduced between the base and emitter of each transistor. In Fig. 4·41*a*, the transfer characteristics of the transistors are shown back to back for zero base bias. These curves are not combined. The dashed lines indicate the base-current values when forward bias is applied to provide the overall dynamic operating curve of the amplifier. With this forward bias, the two curves must be shifted until the dashed lines are aligned with each other, Fig. 4·41*b*. Note that now the jog at the center of the curve has disappeared. If we apply a sine-wave signal, the undistorted output shown in Fig. 4·41*c* is obtained. The 33-ohm resistor in Fig. 4·39 provides this forward bias for the output stage. The output of the audio system in Fig. 4·39 is approximately 150 mw to a 3-in. speaker.

A second audio system is shown in Fig. 4·42. There is one small circuit variation in Fig. 4·42 that the reader should become familiar with. This occurs in the class B output stage, where two PNP transis-

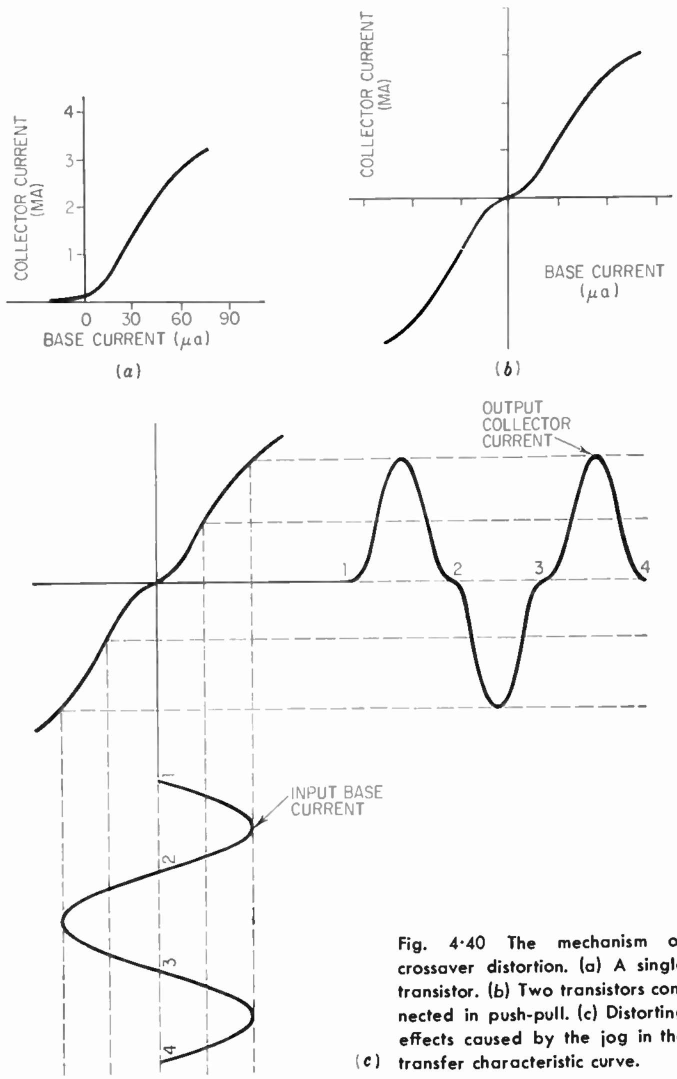


Fig. 4·40 The mechanism of crossover distortion. (a) A single transistor. (b) Two transistors connected in push-pull. (c) Distorting effects caused by the jog in the transfer characteristic curve.

tors are employed. The voltage from the battery is applied to the emitter and base from the +12-volt line. The collectors, however, are returned to ground (negative side of the battery in this circuit). In a vacuum-tube amplifier, this would be equivalent to placing a large negative voltage on the cathode and returning the plate to ground.

Since it is the relative potential between the two elements that produces current flow through the device, it makes little difference whether the cathode is made negative or the plate is made positive. The same type of reasoning applies to transistors. Note, however,

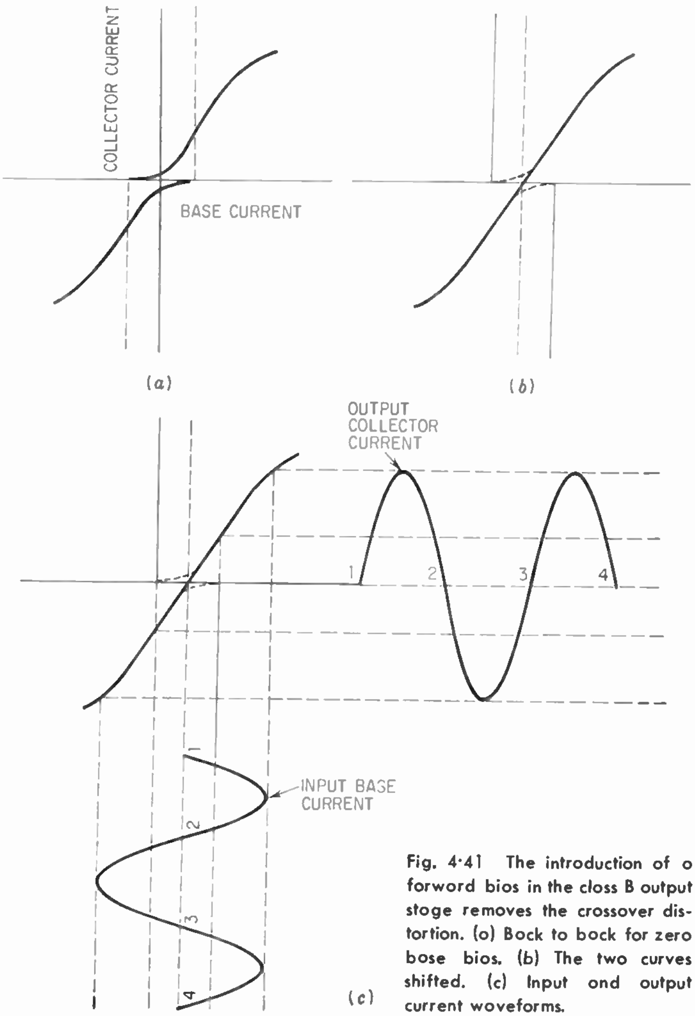


Fig. 4·41 The introduction of a forward bias in the class B output stage removes the crossover distortion. (a) Back to back for zero base bias. (b) The two curves shifted. (c) Input and output current waveforms.

that because a large positive voltage is applied to the emitter, the same +12-volt line must also be directly connected to the base circuit. If the base were similarly grounded and the large positive voltage were applied to the emitter, excessive current would flow through the base-emitter circuit and destroy the transistors.

A similar voltage arrangement is employed in the driver stage preceding the class B output amplifier.

Complementary Push-Pull Amplifiers

The complementary symmetry of PNP and NPN transistors was employed previously in direct-coupled amplifiers. These same features may also be utilized to obtain push-pull operation without any input or output transformers. This is possible because the collector currents of NPN and PNP transistors react in opposite ways when subjected to the same applied signal.

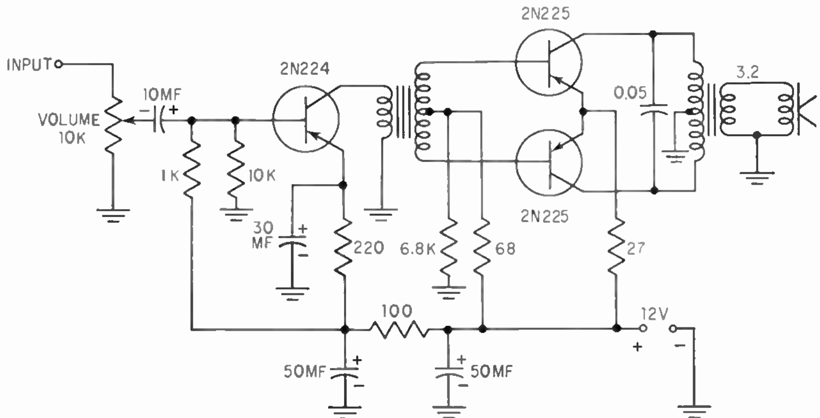


Fig. 4·42 A second audio amplifier system.

To see this more explicitly, consider the class A push-pull amplifier shown in Fig. 4·43. At the top of the illustration we have a PNP transistor; below, an NPN unit. Both are suitably biased with a 22.5-volt battery (one for each transistor). R_1 and R_2 for the PNP transistor and R_3 and R_4 for the NPN transistor serve to establish the base-to-emitter bias suitable for class A operation. The 560-ohm emitter resistors provide d-c stabilization. Each of these resistors is suitably bypassed to prevent a-c degeneration, which would reduce the gain of the amplifier. The load is a 500-ohm voice coil of a loudspeaker, and it is directly connected to the collectors of the two transistors. A single input line is provided, with the base of each transistor connected to this line.

Assume, now, that a sine wave is being amplified and, at the moment in question, the positive half of the sine wave is active. This means that both bases will be driven positive simultaneously. In X_1, this will cause the base-emitter current, and with it the collector current, to decrease. Since the collector current flows up through R_L (i.e., the speaker voice coil), it will serve to make the top, or collector, end of this load impedance less positive or more negative.

Now let us turn to the NPN transistor. The path for its output current is from the collector to R_L. Hence, the voltage drop across R_L due to this transistor is such as to make the top end of R_L negative.

When the positive half of the applied signal reaches the base of the NPN transistor, it acts to increase the forward bias there, thereby increasing the base-emitter current. This, in turn, increases the collector

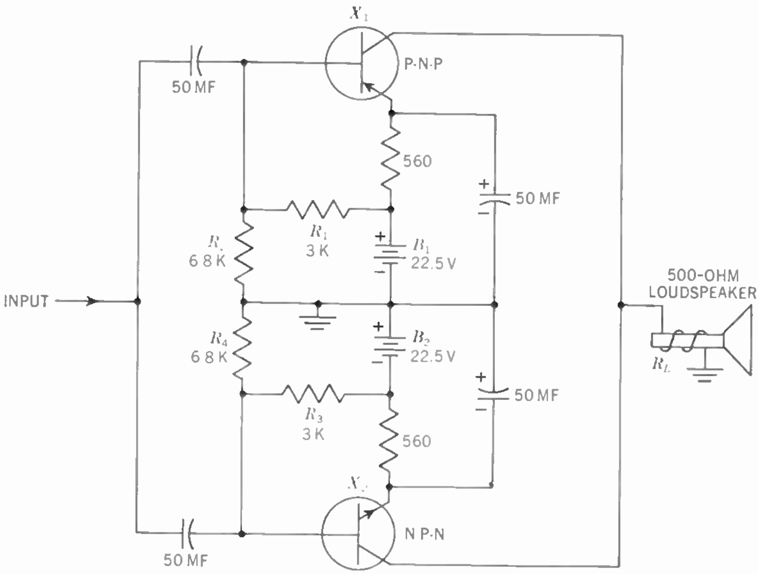


Fig. 4·43 A push-pull class A amplifier using neither input nor output transformers.

current, causes more of a voltage drop across R_L, and raises the negative potential present at the top of the load. This serves to work with or strengthen the voltage drop produced by the PNP transistor.

During the next half cycle, when the negative half of the signal is active, the reverse set of conditions occurs. That is, the current through the PNP transistor increases, producing more of a positive voltage across R_L. At the same time, the current through the NPN transistor decreases, lowering its negative voltage drop across R_L, which, in essence, is equivalent to a positive increase.

Thus, both sections of this circuit work in unison with each other, producing a larger output than either one could by itself. This is demonstrated in Fig. 4·44, where the individual output waveforms of each transistor are shown, together with the combined waveform. Note the differences in relative sizes.

A class B push-pull transistor amplifier with complementary symmetry that can feed its output directly to the 16-ohm voice coil of a loudspeaker is shown in Fig. 4·45. Across the top section of the diagram we have a PNP transistor directly coupled to an NPN transistor. Across the bottom section we have the reverse situation. Both halves

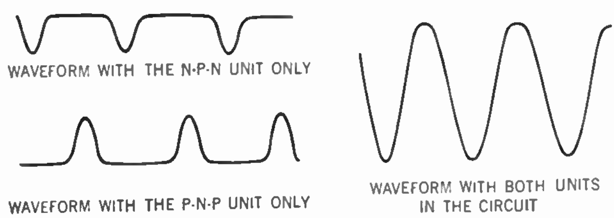


Fig. 4·44 Waveform in the push-pull amplifier of Fig. 4·43.

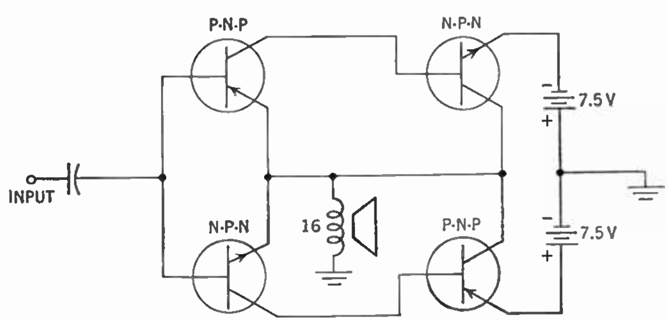


Fig. 4·45 Another push-pull amplifier using complementary symmetry. (*Proc. IRE*)

are similar to the direct-coupled amplifier of Fig. 4·23 and operate in the same manner. In addition, the two sections form a push-pull arrangement. Power gains on the order of 30 db (i.e., 1,000:1 ratio) have been obtained in this manner.

Phase-inverter Circuits

Oppositely phased signals of closely similar amplitudes can be obtained from a single transistor stage as shown in Fig. 4·46. One output voltage is taken from across the emitter resistor, while the second output voltage (of opposite phase) is obtained from the collector load resistor. While perfect balance cannot be obtained, because the current gain α is not equal to 1, the voltages can be made to approach each other quite closely. Typical values of voltage gain to both outputs are shown in the diagram.

In a vacuum tube, the grid does not ordinarily draw any current. Hence, whatever current passes through the plate circuit also flows entirely through the cathode resistor, Fig. 4·47. By having equal-valued resistors in the plate and cathode legs of the tube, equal output voltages will be obtained. In the transistor circuit, Fig. 4·46, a portion

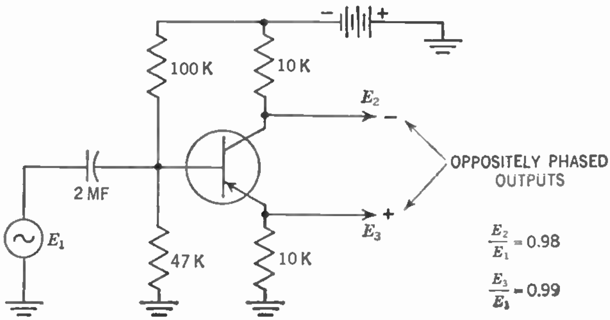


Fig. 4·46 A transistor phase inverter. (*Electronics*)

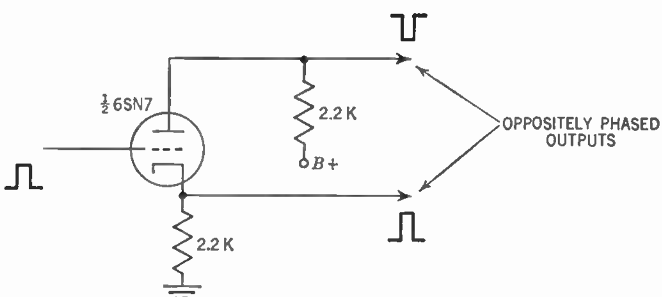


Fig. 4·47 A vacuum-tube phase inverter.

of the emitter current does not reach the collector. Hence, equal-valued collector and emitter resistors will not produce equal output voltages. If we alter the resistances to achieve better balance, we change the circuit operating conditions, including the various currents that flow. Thus, while we may come close, we shall not attain a perfect balance.

It is, of course, possible to obtain balanced signals of opposite polarity by using two transistor stages. One stage will serve to provide one polarity signal, while the other stage will take a portion of this signal, amplify it, and invert it and thereby provide the required second signal.

Volume-control Placement

In the amplifiers that have been shown and discussed thus far, volume controls were omitted in order to keep the circuitry down to its essentials. However, volume controls are normally found in amplifiers, and it is important that we understand the do's and don't's of volume-control application.

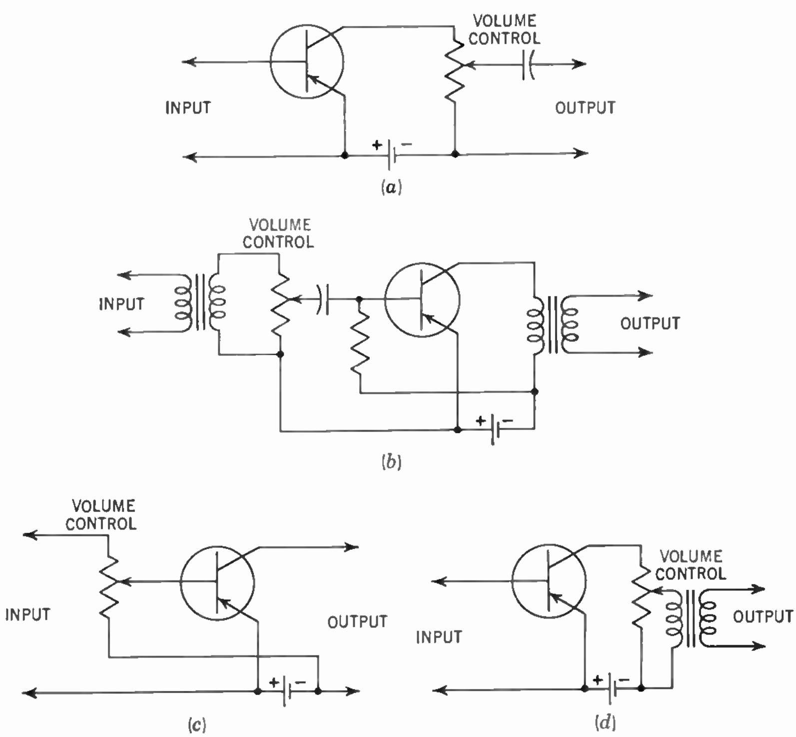


Fig. 4·48 (*a*) and (*b*) satisfactory and (*c*) and (*d*) unsatisfactory methods of connecting a volume control into a transistor amplifier circuit.

Four examples of how volume controls may be wired into a circuit are shown in Fig. 4·48. In Fig. 4·48*a* and *b* variation of the volume-control arm does not vary the base current, the collector current, or the value of the load as seen by the transistor. In the first case, Fig. 4·48*a*, we are varying the amount of voltage being taken from the load resistor; in the second instance, Fig. 4·48*b*, we vary the amount of signal voltage being fed to a transistor.

In the third and fourth illustrations of Fig. 4·48, rotation of the control arm will alter the operating conditions in the transistor amplifier. For example, in Fig. 4·48*c* rotating the volume-control arm will vary the base current and, with it, the collector current. In Fig. 4·48*d* the volume control will vary both the collector current and load impedance. Hence, neither of the two latter arrangements would be desirable in actual circuits.

Tone controls present similar problems, and they, too, must be so inserted that they do not affect either the direct operating currents or the load impedance of the transistor.

Radio- and Intermediate-frequency Amplifiers

With the development of diffusion methods for fabricating transistors, units capable of operating with a fair amount of gain at hundreds of megacycles became possible, and this barrier to transistor utilization has all but disappeared. As a matter of fact, transistors which will function in the kilomegacycle region (i.e., thousands of megacycles) can be built today. This is far higher than any conventional vacuum tube can attain.

In dealing with transistors in low-frequency circuits, any decrease in stage gain is due not to the transistor, but rather to the external circuitry. For example, in an audio amplifier, the gain drop-off as the frequency is decreased stems from the rising impedance of the coupling capacitors such as the 10-μf units shown in Fig. 4·14. Eventually, so much signal voltage is dropped across these coupling capacitors that very little reaches the transistor. Hence, very little can appear in the output.

At high frequencies, things are quite different because now the current gain β enters the picture. Current gain, or β, can be defined in two ways. The first, d-c β, is given as

$$\text{d-c } \beta = \frac{I_C}{I_B}$$

where I_C = d-c collector current
I_B = d-c base current needed to produce this collector current

The second, small-signal or a-c β, is defined as

$$\text{Small-signal } \beta = \frac{i_c}{i_b}$$

where i_c = a-c collector current
i_b = a-c base current needed to produce this collector current

In addition to the letter β, another symbol is employed extensively for current gain. This symbol, for d-c β, is h_{FE}; for a-c β, it is h_{fe}.

In order to specify and measure small-signal β (h_{fe}), it is necessary to specify also the frequency at which it is being measured. A typical variation of small-signal β as a function of frequency is shown in Fig. 4·49. Specifically, this graph is for an output that is short-circuited for alternating current (i.e., the a-c load impedance is zero). With an actual load resistor, the same shape curve is obtained, but with lower β values.

In Fig. 4·49, β_0 at the left-hand side of the chart represents the value of h_{fe} at some low frequency, perhaps several hundred cycles.

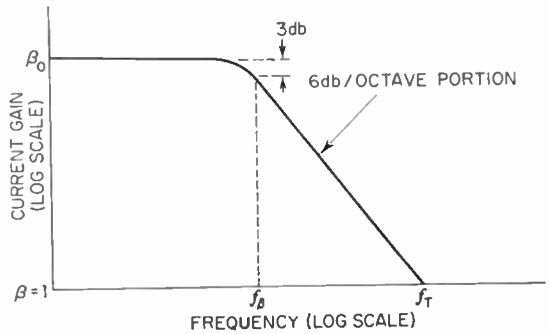


Fig. 4·49 Typical variation of β with frequency.

This value remains fairly steady over a limited range of frequencies and then starts to decrease. f_β in Fig. 4·49 is the frequency at which the current gain is 3 db (i.e., 0.707 β_0) down from its β_0 value. Beyond f_β, the curve slopes down at a constant rate of 6 db per octave. This simply means that each time the frequency is doubled, h_{fe} decreases by 6 db. Eventually, the value of h_{fe} reaches 1, and this frequency is labeled f_T. This latter value is listed frequently in transistor data sheets.

It should be noted that while frequency determines the value of small-signal β, the collector-to-emitter voltage and collector current also affect this characteristic (agc control of stage gain is based on this behavior).

Interstage coupling networks. The most widely used amplifier connection is the common emitter, and with this arrangement the input impedance is fairly low (on the order of 1,000 ohms or so) while the output impedance is in the neighborhood of 10,000 to 20,000 ohms. It is the purpose of the interstage coupling network not only to provide whatever frequency selectivity is desired but also to match these input and output impedances.

There are a number of interstage coupling networks possible, and the more important of these are shown in Fig. 4·50. In the first group, Fig. 4·50*a*, the second amplifier is connected directly into the parallel-

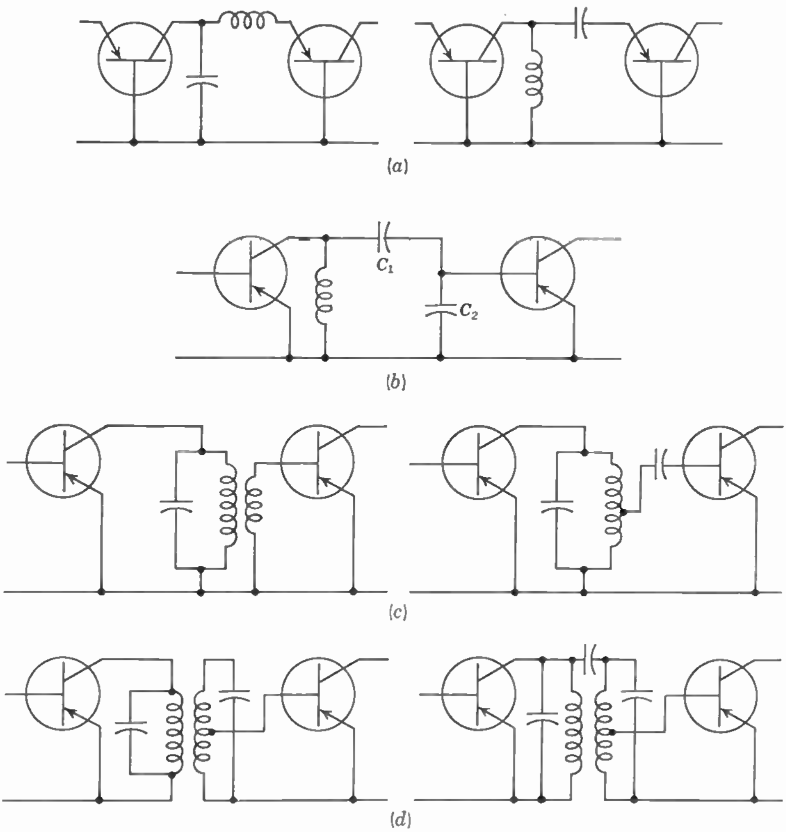


Fig. 4·50 Various methods of coupling transistor stages together. (After W. F. Chow, High frequency Transistor Amplifiers, *Electronics*, April, 1954)

resonant circuit, either in series with the inductance or in series with the capacitance.

In the second group, Fig. 4·50*b*, the second stage is connected to the junction of two capacitors C_1 and C_2, which resonate with the inductance. By properly proportioning the values of C_1 and C_2, we can use the network to match the high output impedance of the first transistor to the much lower input impedance of the second transistor.

The third group, Fig. 4·50*c*, employs inductive coupling between stages. In the first illustration of Fig. 4·50*c*, the primary circuit is tuned, offering a high impedance to the first transistor stage. The signal is then transferred to an untuned secondary containing fewer turns. This step-down action enables the low-impedance input of the second transistor to match the output of the first stage. In the second illustration of Fig. 4·50*c* we obtain the same electrical action by dispensing with the secondary winding and tapping directly into the primary inductance. In this arrangement, a coupling capacitor is needed to prevent the higher collector bias of the preceding transistor from reaching the base of the second unit.

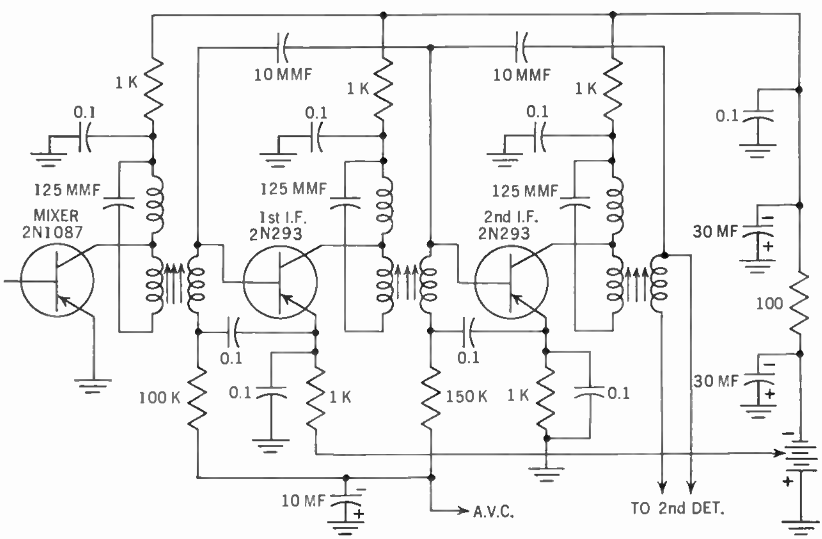


Fig. 4·51 A two-stage i-f system operating at 465 kc. (Raytheon Mfg. Co.)

In the final group of coupling networks, Fig. 4·50*d*, double tuning is employed. Again note how the second stage must be tapped down in order to achieve the proper impedance match.

Intermediate-frequency amplifiers. A two-stage i-f system suitable for a broadcast receiver is shown in Fig. 4·51. The circuit uses two 2N293 high-frequency PNP transistors in a grounded-emitter configuration. Operating frequency is 465 kc, and the overall gain is at least 90 db. The i-f transformers have 155 total turns on the primary, tapped at 55 turns, with an 18-turn secondary. The coils are bifilar wound and enclosed in an adjustable ferrite cup. They are tuned by a fixed 125-$\mu\mu$f capacitor across the primary.

Each emitter possesses a 1,000-ohm d-c stabilizing resistor. Alternating-current or signal degeneration is prevented by the use of 0.1-μf bypass capacitors across these resistors.

It will be noted that each i-f stage is neutralized by connecting a 10-$\mu\mu$f capacitor from the base of the following stage to the base of the preceding stage. (These two points are 180° out of phase because of the grounded-emitter arrangement.) Neutralization was deemed necessary because enough internal capacitance existed in the two 2N293 transistors to lead to oscillation. There are transistors in which

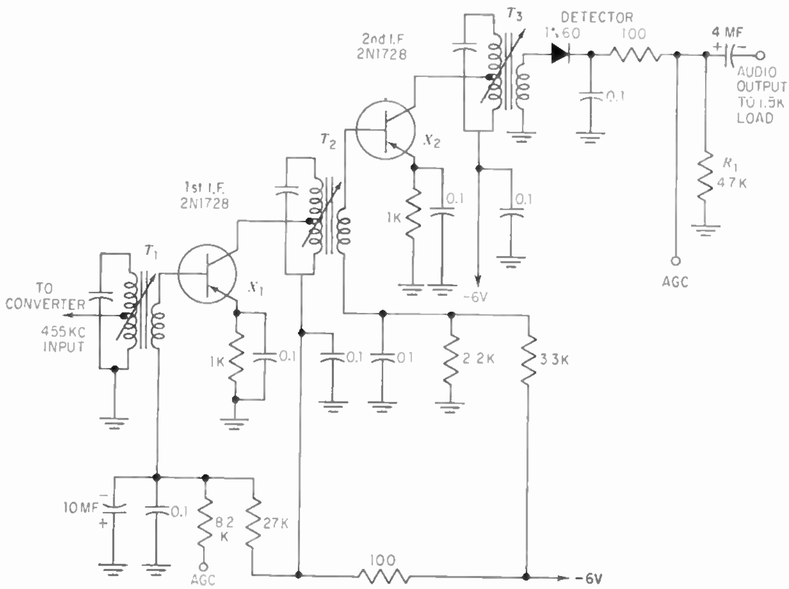


Fig. 4·52 A second i-f system designed for a broadcast receiver.

the signal feedback is so small that special neutralizing networks are not required and hence not used.

The bases of both i-f stages connect into an agc line. Operation of agc systems will be explained in Chap. 6.

Another radio-receiver i-f system is shown in Fig. 4·52. The transistors here, too, are connected with the emitters common to both input and output circuits. Transformers T_1, T_2, and T_3 comprise three bifilar circuits which serve as interstage coupling networks, with essentially unity coupling between primary and secondary windings.

The first i-f transformer T_1 transfers the received signal to the base of the first i-f transistor X_1. This stage is also provided with agc bias.

The control action is accomplished by varying the base current of X_1 in step with the signal level at the second detector and this, in turn, varies the collector current. Figure 4·53 demonstrates what effect this variation has on the gain and on the input and output impedances of the transistor. Note that transistor gain decreases rapidly as the collector current drops below 0.25 ma. Observe, too, that the input and output impedances rise with emitter-current decrease, causing mismatching in the input and output circuits and further reducing gain.

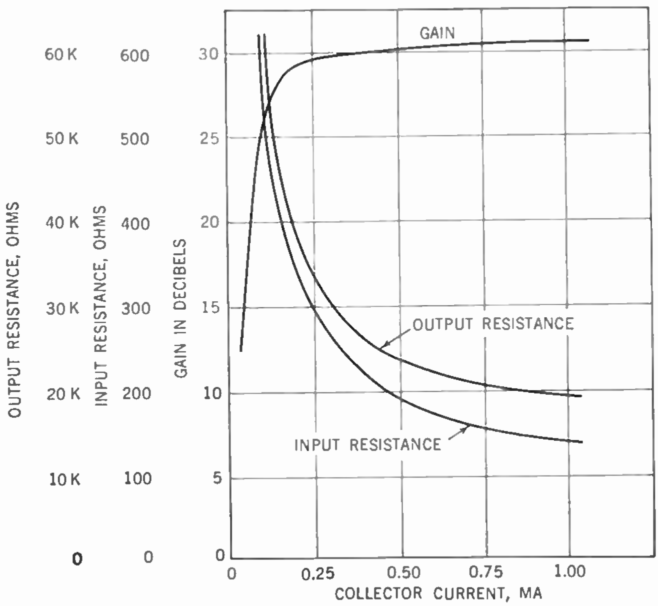


Fig. 4·53 Variation in input and output resistance of the first i-f stage in Fig. 4·52 with changes in collector current. Variation in stage gain is also indicated.

In the second i-f stage we have essentially the same circuit arrangement, although agc is not applied here. Because of this, the base bias voltage (hence, current) is different. Beyond this stage, the signal goes to a 1N60 germanium rectifier where it is demodulated. It is then ready for the audio amplifier stages. The agc voltage is developed across R_1. This resistor is also the load for the 1N60.

Radio-frequency amplifiers. The considerations which govern the design of the i-f amplifiers also hold true for r-f amplification. However, because r-f amplifiers operate at higher frequencies, we can expect lower gain.

The r-f stage of a transistor automobile radio is shown in Fig. 4·54. The input transformer T_1 is slug-tuned, with its slug mechanically ganged to the slugs of the converter and the local oscillator (not shown) coils. The Q of T_1 varies from 70 to 50 across the tuning band, 550 to 1,600 kc. The base of X_1 is returned to the junction point of R_2 and R_3, where the d-c potential is approximately 1.5 volts. A d-c stabilizing resistor R_4 is placed in the emitter leg of X_1 to make the stage relatively insensitive to changes in ambient temperature. A small 680-ohm resistor R_5 brings an agc voltage to the emitter. If no agc

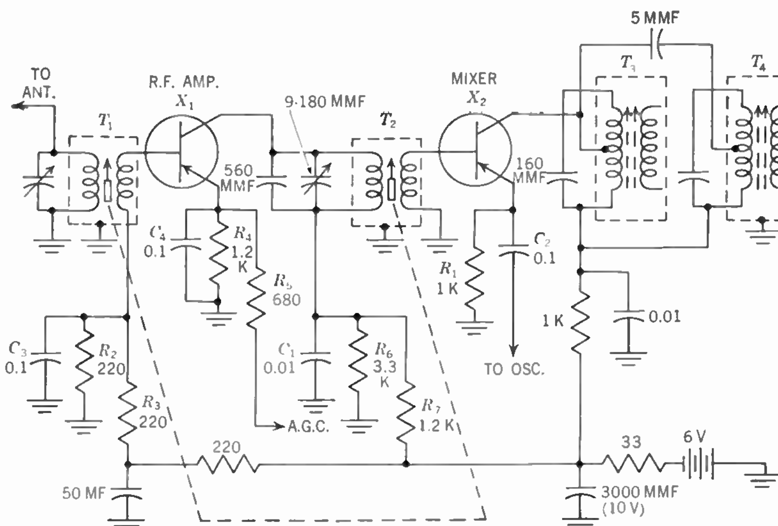


Fig. 4·54 The r-f and mixer stages of a transistor automobile radio.

control is desired, the connection between R_4 and R_5 can be severed. C_4, across R_4, serves to place the emitter at r-f ground. C_3, at the junction of R_2 and R_3, serves also as a low-impedance path to ground for radio frequency.

Transformer T_2 couples the signal from the r-f stage to the mixer. In the mid-frequency range of the broadcast band, the output impedance of X_1 is 10,000 to 15,000 ohms and the mixer input impedance is about 500 ohms. These are the two impedances that must be matched by T_2.

The r-f signal is applied to the base of X_2, while the locally generated oscillator signal is brought into the circuit by C_2 and developed across R_1. The latter resistor, incidentally, also serves to provide bias stability in the same manner as R_4 in the r-f amplifier stage. Approximately

0.4 volt rms of oscillator voltage is injected into the converter stage, this value having been found to provide optimum conversion gain of X_2. If the oscillator voltage is reduced below this level, the conversion gain drops rapidly, which means that we obtain a smaller i-f signal for a given amount of incoming r-f signal. On the other hand, if the oscillator signal is made larger than this optimum value, conversion gain will again decrease, although this time more slowly.

A fairly elaborate interstage coupling network is employed between the mixer and first intermediate frequency. This is designed to achieve the desired signal bandpass, with a fairly rapid fall-off on either side. Output impedance of the converter is in the neighborhood of 50,000 ohms, and it is not affected by signal frequency. Capacitor C_2 and the oscillator circuit that it ties into offer very low impedance to signals of intermediate frequency, so that for i-f signals R_1 is effectively bypassed and no degeneration results.

Interstage coupling networks other than the network shown in Fig. 4·54 could be employed between r-f stages, and these will follow closely the patterns indicated in Fig. 4·50.

QUESTIONS

4·1 How can you determine by looking at a schematic diagram whether a PNP or an NPN transistor is being employed? (Assume that this information is not indicated.)

4·2 Why do the coupling capacitors in transistor audio amplifiers possess high values?

4·3 Explain the purpose of C_2 and R_2 in Fig. 4·3.

4·4 Why must I_{co} be watched more closely in common-emitter amplifiers than in common-base amplifiers?

4·5 How is the effect of I_{co} minimized in common-emitter amplifiers?

4·6 Why is it more difficult to cascade transistor amplifiers than vacuum-tube amplifiers?

4·7 Why can more gain be obtained by using transformer coupling rather than RC coupling between transistor amplifiers?

4·8 Draw the diagram of a two-stage transformer-coupled transistor amplifier.

4·9 Explain the purpose of each component in the circuit of Question 4·8.

4·10 Draw the diagram of a three-stage RC-coupled transistor amplifier.

4·11 What considerations govern the choice of values for R_f and C_f in Fig. 4·14?

4·12 What advantages does negative feedback offer in transistor applications?

4·13 Illustrate a simple method of obtaining negative feedback.

4·14 Explain how negative feedback is obtained in the circuit of Fig. 4·15. Show that the feedback voltage is actually 180° out of phase with the voltage existing at the feedback point.

4·15 Would the operation of the circuit in Fig. 4·15 be altered if the feedback line terminated at the base of X_1 rather than at the emitter? Explain.

4·16 What do we mean by complementary symmetry in transistors?

4·17 Explain how the circuit in Fig. 4·23 operates.

4·18 Draw the diagram of a push-pull class A transistor power amplifier.

4·19 What advantages does class B operation offer over class A operation in the audio range?

4·20 How can the principle of complementary symmetry be used advantageously in push-pull amplifiers?

4·21 What precautions must be observed when incorporating a volume control into a transistor amplifier circuit?

4·22 Draw a transistor phase-inverter circuit. Explain how it operates.

4·23 Illustrate several suitable interstage coupling networks for transistor r-f or i-f amplifiers.

4·24 What is the purpose of the neutralizing circuits sometimes found in transistor i-f or r-f amplifiers?

4·25 How is neutralization achieved?

4·26 What is the difference between a-c and d-c β? Why is this difference important in high-frequency transistor amplifiers?

4·27 Define f_β, f_T, and β_0.

4·28 What do we mean by crossover distortion in a class B amplifier? How is it minimized?

4·29 What is thermal resistance? Contrast it to ohmic resistance.

4·30 What internal differences exist between low-power and high-power transistors?

4·31 What is a heat sink? How is it used? Describe several types of heat sinks.

CHAPTER 5

Transistor Oscillators

TRANSISTORS WILL FUNCTION as oscillators as readily as they will as amplifiers. For every vacuum-tube oscillator, there is a transistor counterpart. In addition, transistors can be designed to produce oscillations in a way that cannot be duplicated with tubes.

Oscillations in vacuum-tube circuits are normally produced by feeding a portion of the amplified signal in the plate, or output circuit, back to the grid, or input circuit. The phase of this feedback signal must be the same as the instantaneous phase of the grid signal in order that the two will add and reinforce each other. This is in distinction to degeneration or negative feedback, where the returned signal is 180° out of phase with the existing grid signal.

Low-frequency Oscillators

A simple audio-frequency oscillator, using a vacuum tube, is shown in Fig. 5·1. The transfer of energy from the plate to the grid is achieved through transformer T_1. The primary and secondary sections are so connected that the field set up by the secondary winding establishes an induced voltage in the primary which tends to maintain oscillations in the circuit. To appreciate the importance of winding polarity, all one has to do is reverse the connections to either T_1 winding and the oscillations cease. (Of course, if the connections to both windings are reversed, circuit behavior will be unaffected.)

The cathode of the oscillator tube is grounded, and grid bias is developed by the combination of C_1 and R_1. With this type of biasing, the circuit operates class C, which means that plate current flows in pulses for only a short time during each signal cycle. If desired, the tube can be made to function as a class A oscillator by moving C_1 and R_1 into the cathode circuit.

An equivalent transistor oscillator, using the same general arrangement of components, is shown in Fig. 5·2. An output control R_1 with an output coupling capacitor C_1 has been added to indicate one way in which an output signal could be obtained from this circuit. One battery is employed to bias both collector and base circuits. R_2 is inserted in the base lead to limit the current in the base circuit to the proper value dictated by this transistor.

In the design of this oscillator, the windings on the transformer T_1 must match the low impedance of the base-emitter circuit on the one

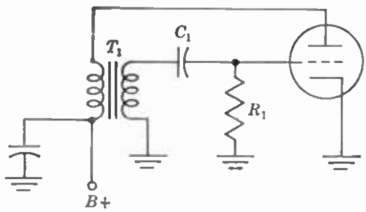


Fig. 5·1 A simple vacuum-tube audio-frequency oscillator.

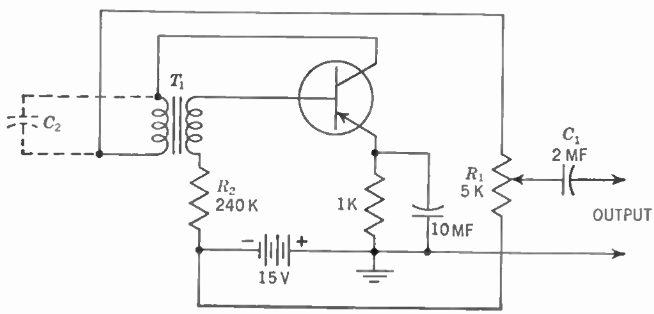


Fig. 5·2 A transistor audio-frequency oscillator. A 2N464 transistor or its equivalent can be used.

hand and the high impedance of the collector circuit on the other. The problem, in this respect, remains similar to what it is in amplifiers. Frequency of oscillation will depend upon the inductance of the windings and their distributed capacitance, and it may be lowered by shunting capacitors (such as C_2) across the high-impedance collector winding.

The same oscillator, using the transistor in a common-base arrangement, is shown in Fig. 5·3. Two batteries are required effectively to bias input and output circuits with the proper polarity.

In the common-base amplifier, it will be recalled that the signal polarity on the emitter (normally the input element) is in phase with the signal on the collector (the output element). Since these two elements are fairly close together, it is relatively simple for some of the collector signal to return to the emitter. This sets up positive feedback between these two elements, and if enough signal voltage is fed back, the stage will oscillate. In amplifiers, this feedback is discouraged, but in oscillators it is encouraged by connecting a small external capacitance between collector and emitter, or by transformer coupling as shown in Fig. 5·3.

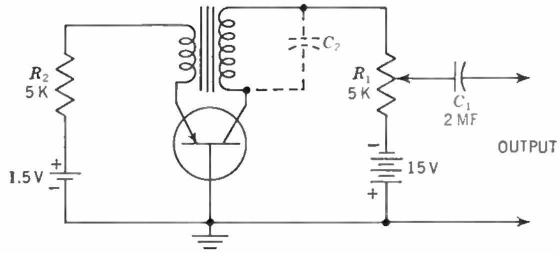


Fig. 5·3 The same oscillator as in Fig. 5·2 in a common-base arrangement.

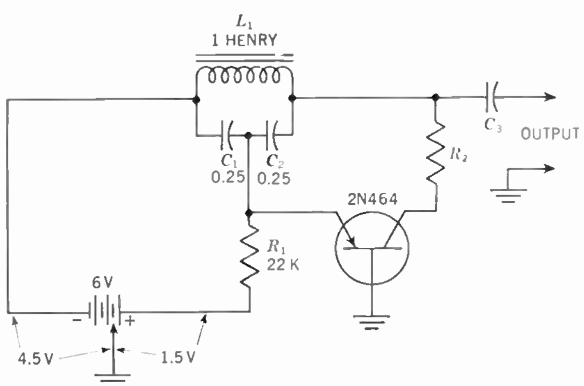


Fig. 5·4 A transistor oscillator using a Colpitts circuit arrangement. (After L. Fleming, Transistor Oscillator Circuit, *Electronics*, June, 1953)

The common-base oscillator is a widely used circuit, particularly at high frequencies.

Another low-frequency transistor oscillator, this time using an inductor having a single winding, is shown in Fig. 5·4. A 2N464 PNP transistor is used. Feedback of in-phase energy from the collector to the emitter is achieved by capacitors C_1 and C_2 in what is essentially a Colpitts circuit arrangement.

Resistor R_1 determines the bias between the emitter and base, and therefore it will govern the extent of current flow through this portion of the circuit during each cycle of oscillation. Any value between

5,000 and 100,000 ohms will work. Resistor R_2 is designed to limit the reverse collector current flow during that part of the half cycle when the collector is driven positive. If R_2 is made zero, the positive peaks of the voltage wave will have flat tops because the collector is driven to overload. The waveform is found to improve as R_2 is increased to a thousand ohms or so, and thereafter improvement occurs more slowly as the resistance is raised. When the value of R_2 reaches about 40,000 ohms, the collector voltage is reduced to such an extent that oscillations cease.

The values of C_1, C_2, and L_1 shown in Fig. 5·4 will produce a 1,000-cycle signal. Battery drain is less than 50 μa. The upper frequency limit for this transistor in this circuit is in the neighborhood of 50 kc.

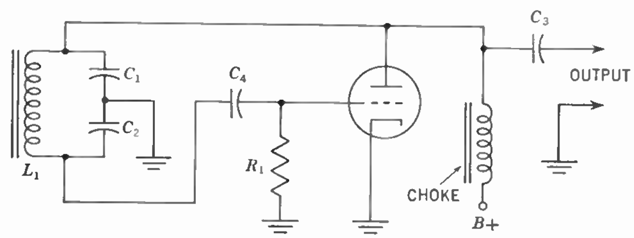


Fig. 5·5 A vacuum-tube Colpitts oscillator.

For those readers who might find that a comparison of a vacuum-tube Colpitts oscillator makes this circuit easier to understand, the circuit of Fig. 5·5 is shown. The main feature to look for in both circuits is the way the frequency-determining components (that is, L_1, C_1, and C_2) are connected. The voltages in the vacuum-tube circuit would obviously be applied differently from the bias voltages of the transistor. This difference, therefore, must be discounted. (Strictly speaking, the circuit of Fig. 5·5, to be directly equivalent to the oscillator in Fig. 5·4, should have its grid grounded. However, the arrangement in Fig. 5·5 is the one most frequently used and undoubtedly the one most familiar to the reader.)

The foregoing are all sine-wave oscillators. Important, too, are relaxation oscillators developing pulses, square waves, sawtooth waves, and other nonsinusoidal forms. One such unit is the blocking oscillator shown in Fig. 5·6. This circuit is closely similar to that shown in Fig. 5·2 except for the addition of a small coupling capacitor C_1 between the base and T_1. The frequency of the circuit is variable between 3 and 60 kc and is inversely proportional to $R_1R_2C_1$. The blocking oscillator can be synchronized to a pulse or a sine-wave input by coupling the signal to the base through capacitor C_2. Note the extreme simplicity of this circuit, including the manner in which one battery furnishes power to the entire circuit.

Another common vacuum-tube oscillator, particularly in television receivers, is the multivibrator. The basic form of this oscillator, shown in Fig. 5·7, is seen to consist of two resistance-capacitance-coupled

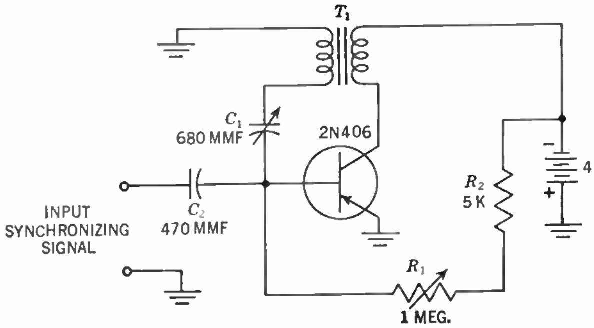


Fig. 5·6 A blocking oscillator. Capacitor C_1 is made variable to permit the pulse width to be changed.

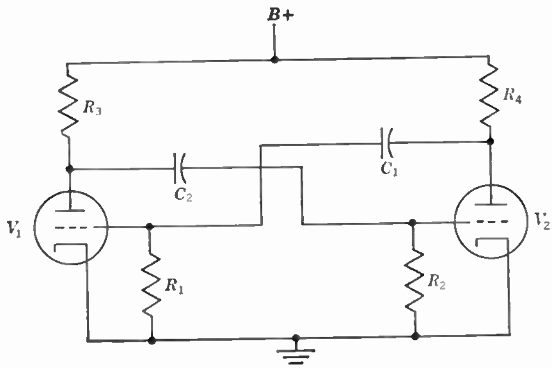


Fig. 5·7 The basic multivibrator circuit using vacuum tubes.

amplifiers with the output of the second stage fed back to the input of the first stage. Oscillations occur in this system because each tube reverses the voltage applied to its grid by 180° and two such reversals produce a signal at the plate of V_2 which is in phase with the voltage at the input of V_1.

In a multivibrator, one tube is cut off while the other one is conducting. How long this condition persists is determined largely by the values of the grid resistors and capacitors R_1, R_2, C_1, and C_2. To see how the shift in conduction is made from tube to tube, let us briefly follow one cycle in the operation of a multivibrator. Assume that the power supply has just been connected across the circuit. Owing perhaps to some slight disturbance in the circuit, the plate current of V_1 increases. This produces an increase in the voltage across R_3, with the plate end of the resistor becoming more negative. Capacitor C_2, which is connected to R_3 at this point, likewise attempts to become more negative, and the grid of V_2 also assumes the same potential. The net result is a lowering of the current through V_2 and R_4.

The lowered voltage across R_4 means that the plate end of this resistor becomes less negative, or relatively positive to its previous value. Capacitor C_1 transmits this positive increase to the grid of V_1, and, consequently, even more plate current flows through R_3. The process continues in this manner, with the grid of V_1 becoming more and more positive and driving the grid of V_2 increasingly negative by the large negative charge built up across R_2 and C_2. The plate current of V_2 is rapidly brought to zero by this sequence of events.

Tube V_2 remains inactive until the negative accumulation of charge on C_2 discharges and removes some of the large negative potential at the grid of V_2. The path of discharge of C_2 is through the relatively low resistance of r_p of V_1 and the relatively high resistance R_2. When C_2 has discharged sufficiently, plate current starts to flow through R_4, causing the plate end of the resistor to become increasingly negative. This now places a negative charge on the grid of V_1, and the plate current through R_3 decreases. The reduction in the voltage drop across R_3 causes the plate end of the resistor to increase positively, and the grid of V_2 (through C_2) receives this positive voltage. The increased current through R_4 quickly raises the negative grid voltage on V_1 (through C_1) and drives this tube to cutoff. When the excess charge on C_1 leaks off, the process starts all over again. C_1 loses its accumulated negative charge by discharge through r_p of V_2 and R_1.

The entire operation may be summed up by stating that first the plate current of one tube rises rapidly, driving the second tube to cutoff. This condition remains until the second tube is released from its cutoff state and starts to conduct. It is now the first tube which is cut off. When the first tube is again permitted to conduct, the second tube is driven into nonconduction. The switching continues in this manner, with the rapidity of turnover (i.e., frequency) determined by the grid resistors and capacitors.

An equivalent multivibrator, using transistors, is shown in Fig. 5·8. Each transistor must be connected as a grounded-emitter amplifier in order to provide the necessary 180° phase reversal in each stage. For the values of the components as shown, the repetition frequency is on the order of 2 kc. It is possible by altering these values to reach frequencies in the megacycle range.

Cutoff of a single stage is achieved in this multivibrator in exactly the same way that it is in the vacuum-tube multivibrator. For example, X_2 in Fig. 5·8 is cut off when the emitter-base voltage is such that the base is positive with respect to the emitter. Normally, the base should be negative with respect to the emitter because the 2N501 transistor is a PNP unit. The positive potential required by the base for effective cutoff is developed across resistor R_3. And X_2 remains

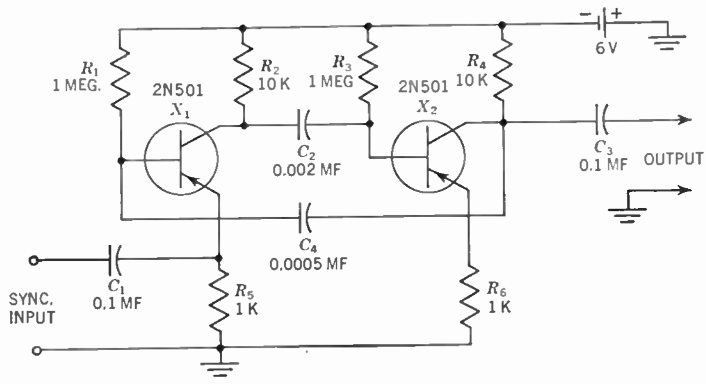


Fig. 5·8 A transistor multivibrator. The 2N501 transistors are of the PNP variety.

cut off until the positive charge held by C_2 (and applied across R_3) has been reduced through discharge. For transistor X_1, the controlling components are C_4 and R_1.

If a sawtooth wave instead of a rectangular wave is desired across the output terminals, it can be obtained by connecting a capacitor from the collector of X_2 to ground. This is satisfactory for low-frequency operation. It will not work too well at high frequencies because of the loading effect of the capacitor.

An emitter-coupled multivibrator using a minimum of parts is shown in Fig. 5·9. This is closely related to the vacuum-tube multivibrator of Fig. 5·10, a circuit that is employed extensively in television receivers.

Sine-wave *RC* oscillators. A sine-wave output may be obtained from an oscillator using an *RC* network in place of the inductance-capacitance network. The *RC* network is positioned between the output and input circuits which, in the common-emitter arrangement, means between collector and base. Furthermore, since the output signal is 180° out of phase with the input signal, the *RC*-coupling network must introduce an additional 180° phase shift in order to provide the base with an in-phase signal to sustain oscillations.

In order to achieve this 180° phase shift, three identical *RC* sections are generally employed, each shifting the signal by 60°. Each section consists of a series coupling capacitor and a shunt resistor.

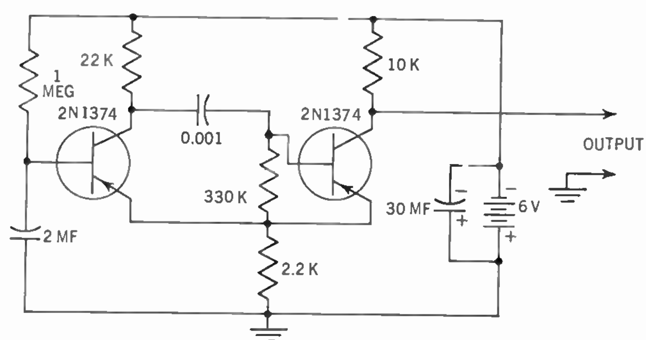


Fig. 5·9 An emitter-coupled multivibrator using a minimum of parts.

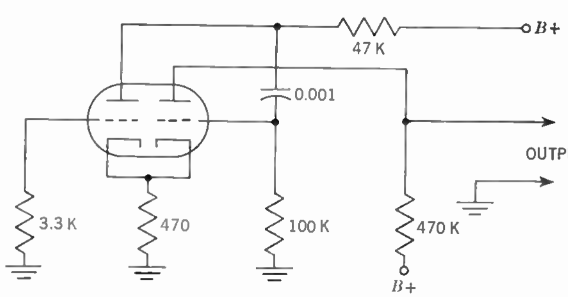


Fig. 5·10 A vacuum-tube cathode-coupled multivibrator.

In Fig. 5·11, these are R_1 and C_1, R_2 and C_2, and R_3 and C_3. The phase shift comes about because *R* and *C* in series produce a current which leads the applied voltage by a certain angle. This angle is determined by the numerical relationship of resistance and capacitance. The smaller the capacitance, the more the current will lead the voltage for a given resistance. By properly selecting *R* and *C*, a 60° phase shift per section can be achieved. Note that this shift will occur at only one frequency. If, however, either *R* or *C* is made variable, a range of frequencies can be obtained.

Transistor oscillator circuits can be designed simply, as illustrated by the preceding arrangements, or they can be incorporated in more complex designs. A low-distortion audio oscillator employing three transistors is shown in Fig. 5·13. To appreciate the operation of this circuit, it may be instructive to commence with a simplified version of its vacuum-tube predecessor, Fig. 5·12*a*. The first tube V_1 is a vacuum-tube amplifier; the second tube is a cathode follower. Positive feedback, necessary to maintain oscillations, is fed from the cathode of V_2 to the cathode of V_1 through a lamp. (A lamp is used, instead of a

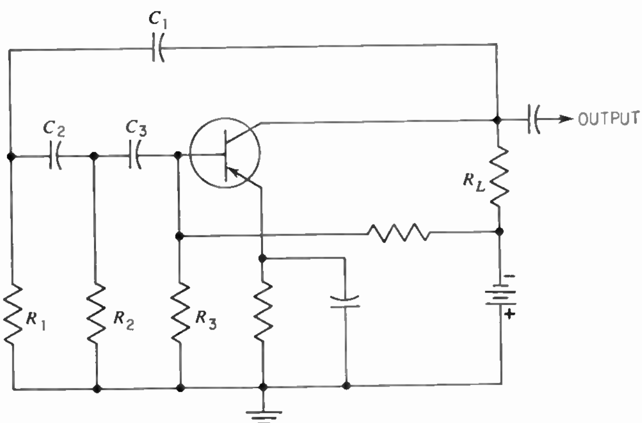


Fig. 5·11 An RC oscillator that generates sine waves.

conventional resistor, to stabilize the amplitude of the generated oscillations. Lamp-filament resistance rises with current, tending to maintain a constant positive-feedback voltage.) For stability, negative feedback is transferred from the same point on the cathode of V_2 to the control grid of V_1. A bridged-T network in this path produces a voltage minimum and zero phase shift at the operating frequency. At all other frequencies, more voltage passes through the bridged-T network, thereby increasing the degeneration and discouraging oscillation. In essence, then, what we have here is a bridge arrangement in which oscillations will develop when the positive feedback exceeds the negative feedback, and the frequency at which this occurs is determined by the bridged-T network.

The first step in the development of the transistor circuit is shown in Fig. 5·12*b*. The first stage is a grounded-emitter amplifier. This is a direct counterpart of vacuum tube V_1. The second transistor is operated as a grounded-collector stage, and it is equivalent in its action to the cathode follower V_2.

An improved version of the oscillator is shown in Fig. 5·12c. The emitter load resistor of the second junction transistor X_2 was replaced by a third transistor X_3. This change was made because the high dynamic collector resistance of X_3 permits more efficient operation of X_2.

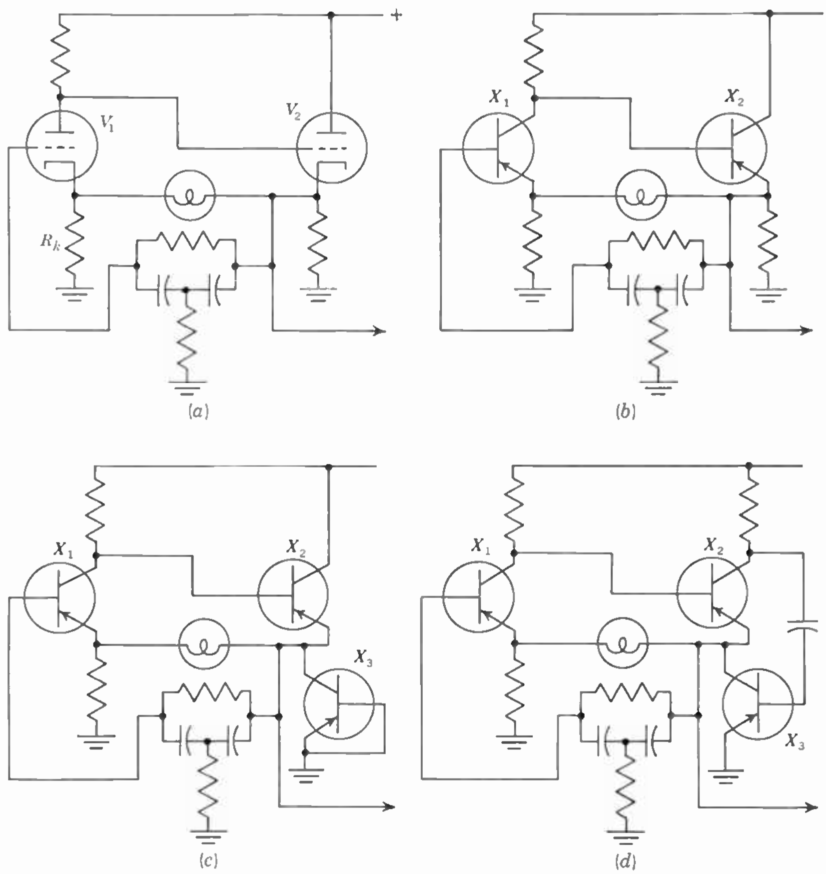


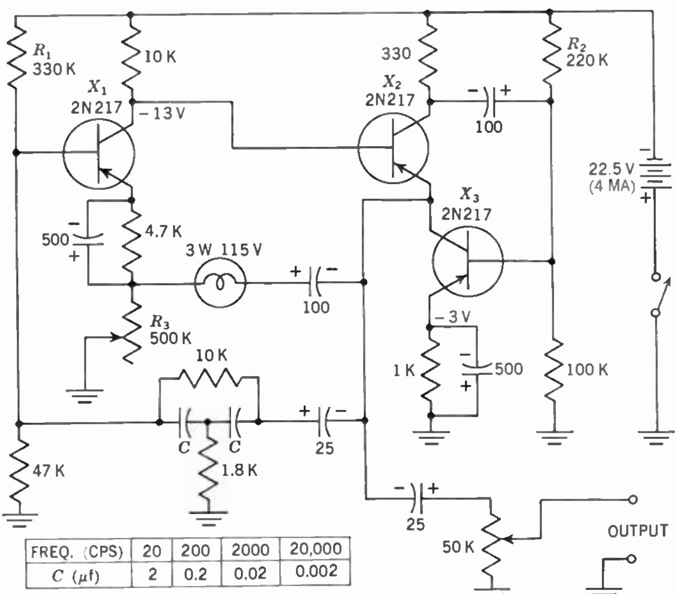
Fig. 5·12 The various steps in the development of a transistor oscillator from its vacuum-tube counterpart. (*Electronics*)

The third step in the development of this circuit, Fig. 5·12d, was made to reduce the amount of even harmonics present in the oscillator output signal. The base of X_3 is connected to the load resistor of X_2, effectively placing the two transistors in push-pull. When the collector current of one transistor increases, the collector current of the other decreases. The signal currents of both transistors flow through the same load resistor, and this provides for a high output current.

In the final version of this oscillator, Fig. 5·13, d-c stabilizing resistors were placed in series with the emitters of X_1 and X_3. Additional resistors, such as R_1 and R_2, establish the proper base operating voltages.

Radio-frequency Oscillators

The use of transistors in r-f oscillators is governed by the same considerations as for low-frequency oscillators, with the addition of a frequency imposed by the transistor. Frequently, the manufacturer



FREQ. (CPS)	20	200	2000	20,000
C (μf)	2	0.2	0.02	0.002

Fig. 5·13 Completed circuit of low-distortion transistor a-f oscillator. (*Electronics*)

will employ the term f_{max} on his characteristic sheets. This term is known as the maximum frequency of oscillation or the frequency at which the transistor power gain is unity. Theoretically, these two points are the same because the device must have at least a unity power gain to oscillate.

Transistors which will operate into the kilomegacycle range are available. Hence, frequency as such is not a basic limitation in designing an oscillator. What is important is power output, which can be more difficult to achieve as the frequency rises, and also circuit stability. These factors will be governed by circuit design and the transistor selected.

An r-f oscillator which is a direct outgrowth of the audio oscillator of Fig. 5·2 is shown in Fig. 5·14. L_1 and L_2 are two tightly wound coils which provide for the transfer of energy between output and input circuits. L_3 is wound close to L_2, and the energy it absorbs is transferred to whatever external circuit is connected to the oscillator.

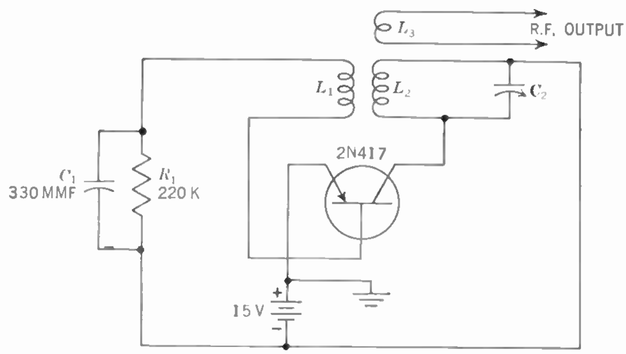


Fig. 5·14 An r-f transistor oscillator. The 2N417 is an alloy-junction transistor with a cutoff frequency of 20 Mc.

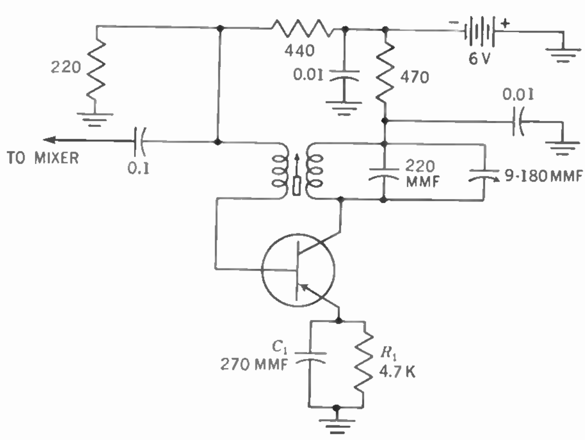


Fig. 5·15 An oscillator designed for a transistorized automobile receiver.

Capacitor C_2 tunes L_2 and enables the generated frequency to be varied. R_1 serves to limit the emitter current to a safe value; C_1 across R_1 assists in the oscillating action.

The oscillator shown in Fig. 5·15 is similar to the preceding circuit except that a 4,700-ohm resistor and a 270-$\mu\mu$f capacitor are inserted in the emitter lead. This particular oscillator is employed as the local oscillator in a broadcast receiver, and the R_1C_1 network is used to introduce a limited amount of degeneration into the circuit. This offsets some of the positive, regenerative feedback and serves to reduce the loading effect of the transistor input circuit on the oscillator tuned circuit. It was found that when this was done, oscillator tuning became relatively independent of the transistor input impedance. A further aid

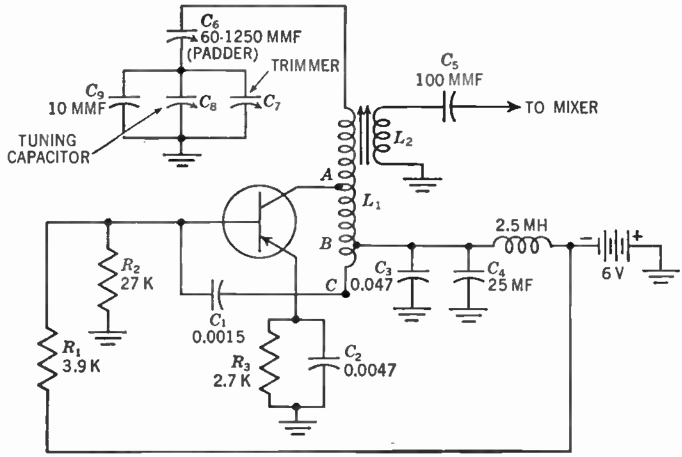


Fig. 5·16 An r-f oscillator employed in a radio receiver. An r-f PNP transistor would be used.

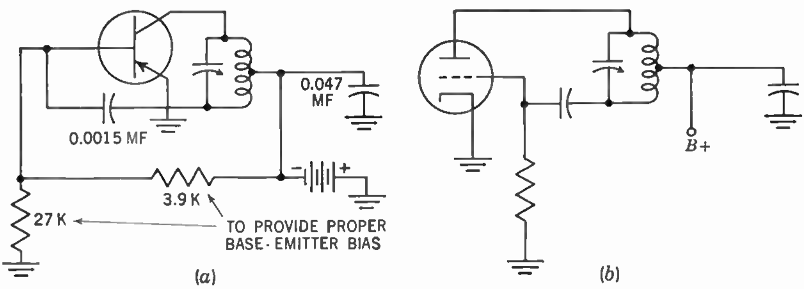


Fig. 5·17 A Hartley oscillator using (a) a transistor and (b) a vacuum tube.

to stability is the use of a relatively high tank capacitance across the collector-tuned circuit.

Another oscillator which has been employed in a radio-broadcast receiver is shown in Fig. 5·16. If we ignore, for the moment, the turns on L_1 which extend above point A, and if we also disregard winding L_2, then what we have here is a Hartley oscillator, Fig. 5·17. The voltage which is developed between points B and C of L_1 represents the energy which is fed back to the base input circuit via coupling capacitor C_1.

The collector is tapped down on L_1 to decrease the effect of its (i.e., the collector) capacitance, to provide a better impedance match between the transistor and the tuned circuit, and to improve frequency stability and tracking. (The last feature stems from the application of this oscillator in a radio receiver.) Tracking also explains the reason for the presence of capacitors C_6 (600 to 1,250 $\mu\mu$f) and C_7. In this particular design three-point tracking between oscillator and mixer was

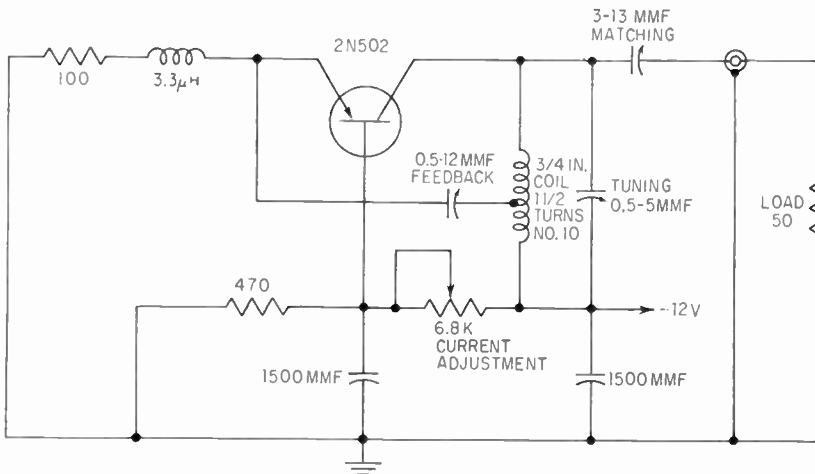


Fig. 5·18 A common-base oscillator designed to operate at 200 Mc.

obtained by using a slug in the oscillator coil, a padder capacitor C_6, and a gang capacitor trimmer C_7. The slug takes care of the central portion of the band, the padder provides an adjustment at the low end of the band, and the trimmer is employed to make the high end of the tracking curve fall into line.

Energy from the oscillator is transferred into the mixer circuit (not shown here) by a combination of inductive and capacitive coupling. The initial transfer from L_1 to L_2 is inductive; the second transfer, from L_2 through C_5 to the mixer, is capacitive. The designers of this circuit felt this arrangement would provide a more nearly constant oscillator injection voltage at the mixer.

The necessary biasing voltage for the transistor collector is brought in through coil L_1. A similar biasing voltage for the base is brought in via R_1 and R_2. The resistor R_3 and capicator C_2 in the emitter leg serve approximately the same purpose here that they did in the preceding oscillator circuit.

A 200-Mc oscillator using a 2N502 transistor in a common-base arrangement is shown in Fig. 5·18. Ten milliwatts of output power is provided. The current-adjust resistor is set for a collector current of

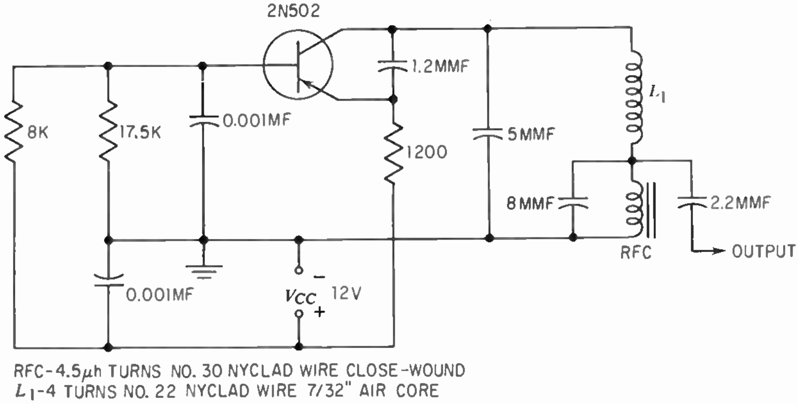


Fig. 5·19 An oscillator operating at 200 Mc.

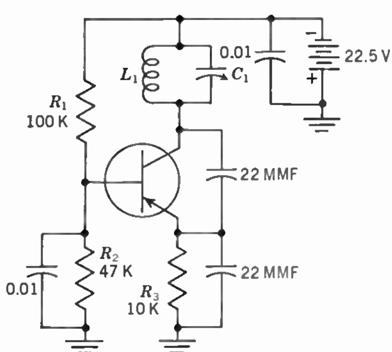


Fig. 5·20 A Colpitts oscillator using a transistor. A 2N414 PNP transistor equivalent would be employed. (After P. G. Sulzer, Junction Transistor Circuit Applications, *Electronics*, August, 1953)

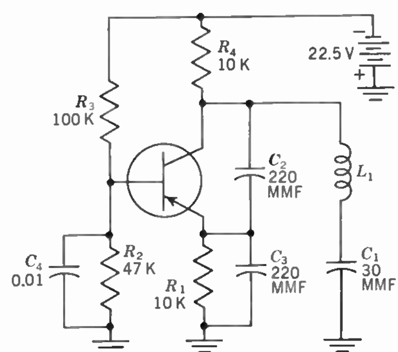


Fig. 5·21 A Clapp transistor oscillator similar to the oscillator shown in Fig. 5·20.

2 ma. The matching capacitor serves as an impedance transformer to match the oscillator to the 50-ohm load section. Figure 5·19 shows another oscillator arrangement operating at essentially the same frequency.

A transistorized r-f Colpitts oscillator is shown in Fig. 5·20. The single tuning circuit L_1 and C_1 is connected in the collector circuit between the collector element and the battery. To provide the proper feedback to sustain oscillations, 22-$\mu\mu$f capacitors are connected between collector and emitter and between emitter and ground. R_1 and R_2 provide the proper voltage for the base-emitter circuit, while R_3 functions as a d-c stabilizing resistor (to minimize the effects of temperature variations).

Another oscillator which is somewhat similar to the Colpitts oscillator but more stable is shown in Fig. 5·21. This is the Clapp oscillator, named after its originator J. K. Clapp. The tuning circuit C_1 and L_1 is series resonant, and it is connected between collector and ground. Feedback voltage between the collector, emitter, and base is provided

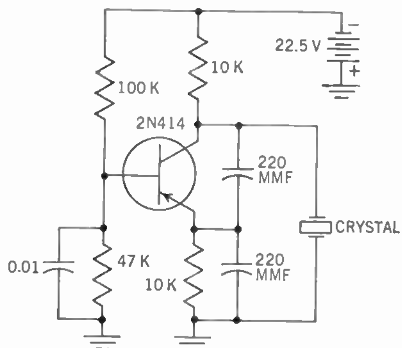


Fig. 5·22 A crystal-controlled Clapp oscillator.

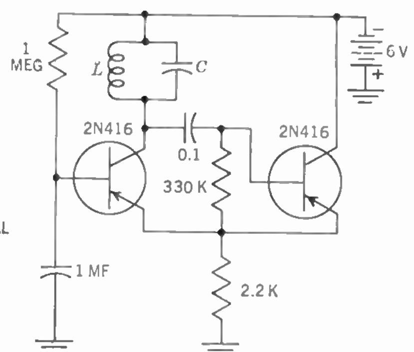


Fig. 5·23 An emitter-coupled sine-wave oscillator.

by C_2 and C_3. The remaining components of this circuit serve the same function as in the preceding oscillator.

The same Clapp oscillator with a crystal substituted for the series-resonant circuit is shown in Fig. 5·22.

An emitter-coupled multivibrator was previously shown in Fig. 5·9. The same circuit modified for sine-wave oscillation is shown in Fig. 5·23. Here the resistor in the collector circuit of the first transistor is replaced by a tuned circuit. With the choice of suitable junction transistors, oscillations above 10 Mc can be readily obtained. The vacuum-tube counterpart of this circuit will oscillate up to 80 Mc just as readily as it will at 1,000 cycles.

Crystal control can be incorporated into this oscillator by connecting a series-resonant crystal between the emitters of the two transistors, Fig. 5·24.

Considerable work has been done with transistorized crystal oscillators because of their inherent frequency stability. A crystal oscillator that was developed at the National Bureau of Standards is shown in Fig. 5·25. A PNP junction transistor is used, together with a 100-kc crystal. Bias battery voltage is only 1.5 volts. Current drain is 100 μa.

The 0.01-μf capacitors connected to ground from each side of the crystal serve to maintain a constant phase shift in the crystal feedback loop. The crystal itself is placed in the path between collector and base. The transistor is operated as a grounded emitter, and it develops

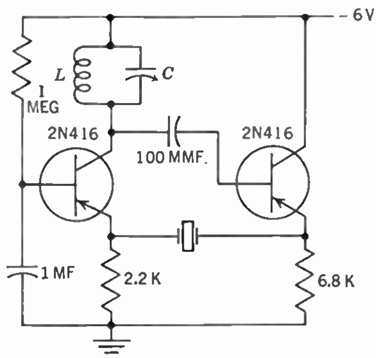


Fig. 5·24 An emitter-coupled sine-wave oscillator that is crystal-controlled. Resonant frequency of the crystal should be the same as the resonant frequency of the parallel-resonant network *LC*.

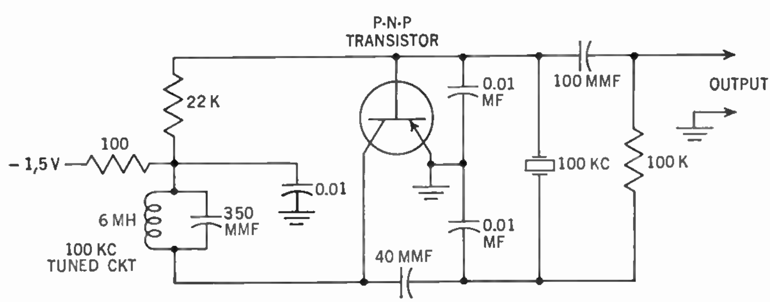


Fig. 5·25 A highly stable transistor oscillator developed at the Bureau of Standards.

0.8 volt across the 100-kc tuned circuit connected to the collector. Driving current for the crystal is obtained from a capacitive voltage divider consisting of the 40-$\mu\mu$f and 0.01-μf capacitors connected in series between the collector and ground.

The stability of this oscillator is excellent. Measurements of frequency with changes in temperature and voltage indicate that the frequency varies approximately 1 part in 100 million per °C and 1 part in 100 million per 0.1 volt. Short-time variations are about +3 parts in 10,000 million, and long-interval drift indicates changes of about 3 parts in 1,000 million (i.e., 1 billion) per 24 hr.

PNP-NPN oscillator. Before we end this discussion, mention might be made of an oscillator that combines NPN and PNP transistors, Fig. 5·26. The crystal between the two emitters serves as the feedback path, just as it does in Fig. 5·24. Direct coupling is employed between the two transistors, and although the voltage applied to the base of the second one is negative, it is less negative (hence, positive) with respect to its emitter.

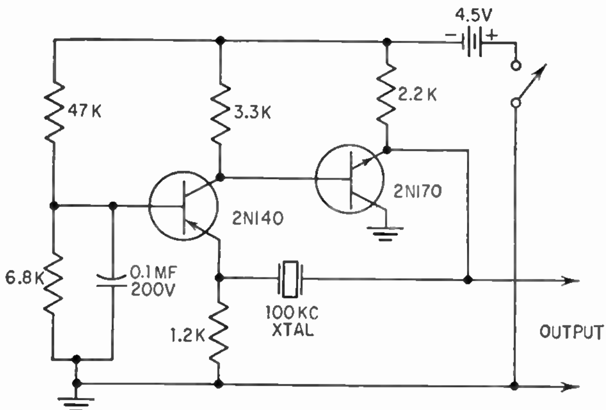


Fig. 5·26 A PNP-NPN oscillator. (After F. T. Merkler, PNP–NPN Transistor Oscillator, *Radio-Electronics*, November, 1960)

It is interesting to note that the first transistor is connected as a common-base amplifier, while the second transistor is a common-collector circuit. This is needed in order for direct coupling to be employed.

QUESTIONS

5·1 How are oscillations normally produced in most vacuum-tube and transistor oscillators?

5·2 Draw the circuit of a simple vacuum-tube oscillator and then show its transistor counterpart.

5·3 Draw the circuit of a transistor multivibrator.

5·4 Explain in detail how the foregoing multivibrator circuit operates.

5·5 Explain the purpose of each component in Fig. 5·15.

5·6 Why is the collector of the transistor in Fig. 5·16 tapped down on coil L_1? Explain the purpose of C_6 and C_7 in the same circuit.

5·7 Draw the diagram of a simple crystal transistor oscillator.

5·8 How does the oscillator of Fig. 5·22 function?

5·9 What is the difference, in terms of feedback, between a common-emitter oscillator and a common-base oscillator?

5·10 Explain how the RC oscillator of Fig. 5·11 functions.

5·11 Could the number of RC sections in the oscillator of Fig. 5·11 be made greater? Explain.

5·12 What is f_{max}? Of what use is it?

5·13 How could you obtain a sine-wave signal from a multivibrator?

5·14 What advantages does a crystal offer in an oscillator? What disadvantages?

CHAPTER 6

Transistor Radio Receivers

AMONG THE FIRST commercial uses to which transistors were put was in small, portable radio-broadcast receivers. This is a natural application, since transistors lend themselves readily to compact, lightweight assemblies of the type required in such receivers. The portability feature is further enhanced by the fact that only small B-type batteries are required for power. Filament-heating batteries, which vacuum tubes require and which occupy considerable space, are completely dispensed with here.

The Regency Radio Receiver

A fairly standard portable radio receiver is the Regency model TR-1G, and it takes advantage of every space-saving feature afforded by the transistors and associated miniature components, Figs. 6·1 and 6·2. Overall dimensions of the unit are 5 by 3 by 1¼ in., enabling the entire set to fit easily into the pocket of a man's jacket. Weight of the set, with the batteries, is only 12 oz.

The schematic diagram of this receiver is shown in Fig. 6·3. There are four transistors and five stages. The extra stage is the second detector, and its function is performed by a germanium diode, here either a CK706A or a 1N132. The transistors are of the NPN variety, and three special designs are used for the converter, i-f, and audio stages. Manufacturer of these units is Texas Instruments, Incorporated.

The first stage, containing transistor X_1, is essentially a self-oscillating converter. The input signal is picked up by a tuned ferrite-core coil which possesses a high Q. A low-impedance winding on the antenna coil couples the signal to the base of X_1.

Local oscillations are generated by a parallel-resonant circuit in the emitter circuit which is inductively coupled to a coil in the collector

Fig. 6·1 The Regency model TR-1G transistor radio is small enough to fit in the pocket of a man's jacket. (Regency)

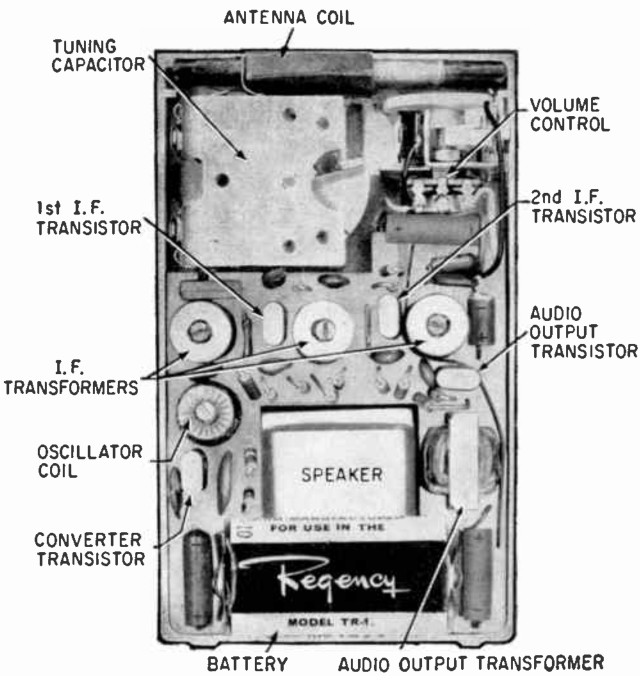


Fig. 6·2 Inside view of Regency model TR-1G transistor receiver showing layout of components. (Regency)

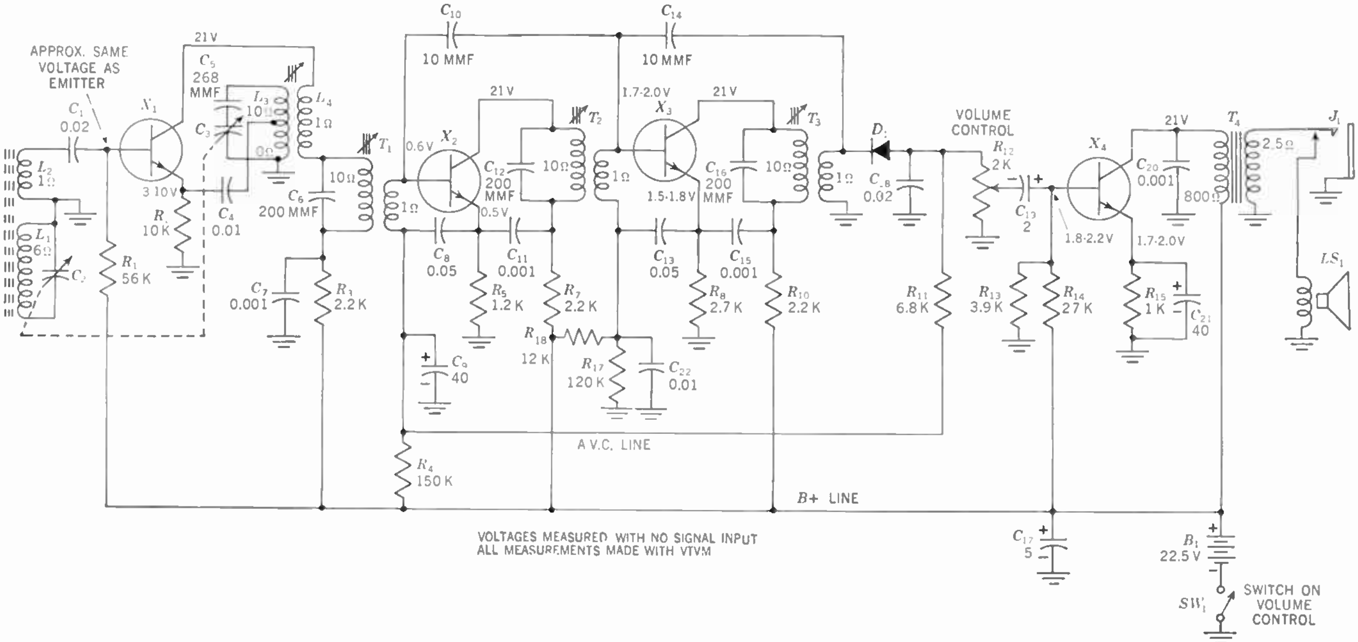


Fig. 6·3 The schematic diagram of the Regency model TR-1G receiver. (Howard W. Sams & Co., Inc.)

circuit. The low-impedance emitter is tapped down on the tuned circuit in order to provide the proper impedance match without lowering the Q of the circuit.

The foregoing oscillator arrangement is a fairly common one. Its equivalent vacuum-tube circuit is shown in Fig. 6·4. With the incoming signal and the local oscillator voltage both being applied to the converter transistor, the appropriate i-f signal is formed and then fed to transformer T_1 and the i-f stages beyond.

A 10,000-ohm resistor is placed in the emitter circuit to provide d-c stabilization against temperature changes and variations among different replacement transistors. The positive voltage which the emitter current develops across R_2 is counterbalanced by a positive voltage fed

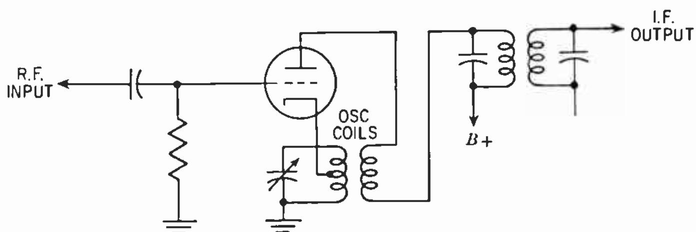


Fig. 6·4 **The equivalent vacuum-tube circuit of the converter stage shown in Fig. 6·3.**

to the base from the battery. The actual voltage difference between these two elements is on the order of approximately only 0.1 volt.

The proper biasing voltage for the collector of X_1 is obtained from a 2,200-ohm resistor which is tied to the 22½-volt B+ line. A 0.001-μf bypass capacitor C_7 keeps the signal currents out of the d-c distribution system.

There are two stages in the i-f system, and both operate at 262 kc. This frequency is considerably below the 455 kc common in vacuum-tube radio receivers, and it possesses the disadvantage of making this receiver more susceptible to image-frequency pickup. However, the lowered frequency of operation is advantageous in that it provides greater gain and more stability.

The primary of each i-f transformer is tuned with a fixed capacitor, while the secondary is untuned. This is done to match the high collector impedance of the preceding stage to the low input impedance of the following stage. Peaking of each i-f coil is achieved by varying the position of an iron-core slug.

Each i-f stage is neutralized by feeding back a voltage from the base of the following stage to the base of the preceding stage. The feedback occurs through a 10-$\mu\mu$f series capacitor.

Automatic gain control is applied to the first i-f stage only. A negative voltage is obtained from the second detector and applied to the base of X_2. Its purpose is to regulate the emitter and collector currents and, with this, the stage gain. When the incoming signal becomes stronger, the negative agc voltage rises, reducing the collector current of X_2 and, with it, the gain. The opposite condition prevails when the signal level decreases. This method is quite effective and provides a wide range of control. (A detailed discussion of automatic gain control in transistor receivers will be given after this analysis of the Regency model TR-1G.)

The base bias for the second i-f stage is obtained from the voltage divider network of R_{17} and R_{18}. This bias voltage is bypassed by C_{22} and then further bypassed by C_{13}, a 0.05-μf capacitor.

Both emitters have d-c stabilizing resistors. (If it were not for the presence of C_8, C_{11}, C_{13}, and C_{15}, signal degeneration would occur also. As it is, only the direct portion of the current passes through R_5 and R_8.) Note, however, that the emitter resistor of the first i-f stage is 1,200 ohms in value, whereas the emitter resistor of the second stage is 2,700 ohms. The reason for this difference stems from the compromise that must be reached in the first i-f stage between good agc action and the d-c stability of the amplifier. A value of R_5 greater than 1,200 ohms is desirable for stability purposes, but the degeneration that produces the stability would result in reduced gain-control action.

Each of the collectors of X_2 and X_3 receives its operating voltage through a 2,200-ohm dropping resistor. C_{11} and C_{15}, at the top end of the resistors, serve as decoupling and bypass capacitors.

Following the second i-f stage is the second detector D_1, and this function is performed by a germanium diode. The load resistor for the detector is the volume control. Note the impedance of the control, 2,000 ohms; this low value is needed to match the input impedance of the audio output stage X_4.

The final amplifier is operated with the emitter grounded through a 1,000-ohm resistor. Base bias is obtained from the voltage-divider network formed by R_{13} and R_{14}. The output transformer matches the 10,000-ohm collector impedance of X_4 to the low voice-coil impedance of the miniature speaker. Diameter of the speaker is only 2¾ in. Provision also exists for a small earphone plug which can be inserted into a small jack on the side of the receiver. When the earphone is in use, the speaker is disconnected.

The total power for the receiver is furnished by a hearing-aid type of 22½-volt battery. Total current drain is on the order of 4 ma.

The compactness of this receiver can be seen by an inspection of

Fig. 6·2. All components, including the two-gang tuning capacitor and the speaker, are miniaturized. Operating voltage on electrolytic capacitors C_9, C_{21}, and C_{19} is 3 volts; on C_{17}, it is 25 volts.

Automatic Gain Control of Transistor Amplifiers

The automatic control of the gain of a transistor r-f or i-f amplifier in radio receivers is almost universally achieved by varying the emitter or collector current in accordance with the amplitude of the incoming signal. Figure 6·5 illustrates the variation in gain that can be

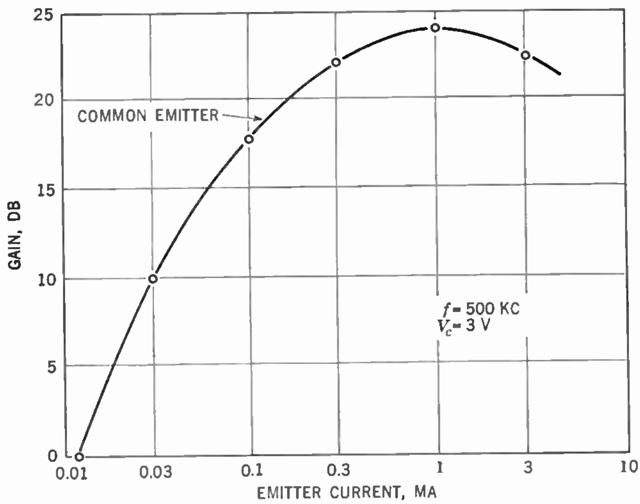


Fig. 6·5 Transistor amplifier gain as a function of the emitter current.

achieved. (No mention is made in the above discussion of the changes in transistor input and output impedances accompanying changes in emitter current flow. These do occur, and they also contribute to the changes in gain.) Now, to provide this control of gain, a certain amount of d-c power is necessary. This arises from the fact that current from the control source is required. In vacuum-tube amplifiers, little or no power is required because the control voltage is fed to the grid of a tube and this element, being negative with respect to the cathode, draws no current. A transistor, on the other hand, is a current-operated device, and to alter its current, we must have the control stage supply a suitable amount of its own current. This, in turn, means that power must be expended.

The control voltage is obtained almost invariably from the second detector; hence, this is the stage which must supply the control power. To assist the detector in this task, the controlled transistor i-f amplifier is usually made to function as a d-c amplifier for the control signal. For example, in the stage shown in Fig. 6·6 the emitter current is agc-controlled. Instead of varying the emitter current directly, however, the control voltage is applied to the base of the transistor, and the resulting changes in direct base current are amplified and appear as larger changes of emitter current. When the incoming signal is strong, then a voltage is fed back which serves to reduce the emitter current, and the stage gain is reduced. Conversely, when the incoming signal

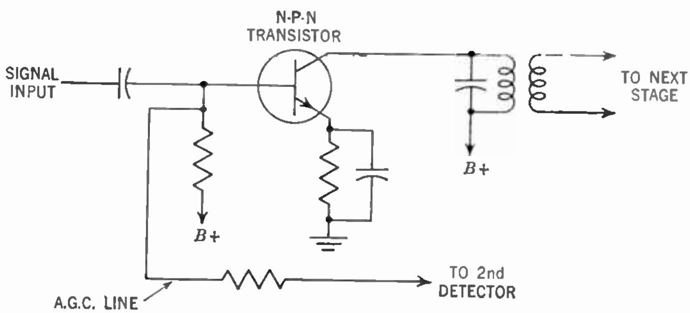


Fig. 6·6 An example of I_e control of an i-f amplifier. The agc voltage is applied to base.

is weak, the voltage fed back is reduced in value, permitting more emitter current to flow and raising the gain.

By reducing the emitter current, we also reduce the collector current and, hence, the strength of the signal developed across the output tuned circuit.

If sufficient power is available in the detector circuit, an attempt can be made to control the emitter current directly by introducing the control voltage in the emitter circuit of the i-f stage. However, in the absence of this power, the control voltage can be applied to the base of the i-f stage, as shown.

The advantage of the foregoing system of agc is its simplicity and economy. There is, however, the serious disadvantage that the transistor is unable to effectively cut off the signal, thereby giving rather incomplete agc action. To overcome this deficiency and provide a greater range of control, a delayed agc circuit is frequently added to supplement the primary agc. This addition consists of an auxiliary diode, as shown in Fig. 6·7. Here, the primary agc network extends from the detector through R_{108} to the base of the first i-f stage, an NPN transistor. The control voltage developed across C_{110} and R_{127} is negative, and it works to reduce the B+ voltage present at the top of R_{105}.

The auxiliary germanium diode is connected between the center tap of T_2 and the top end of the primary of T_1. On strong signals, the amount of agc voltage applied to X_2 is quite high and conduction through the transistor is reduced, thereby lowering the effect of the stronger signal. This causes the positive voltage at the collector to rise, in this instance toward 8 volts. Actually, 8 volts would occur only

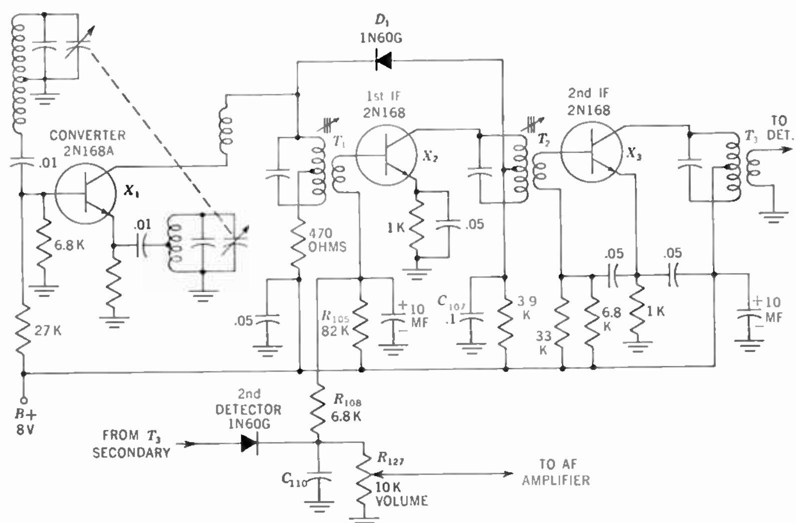


Fig. 6·7 An agc system that is supplemented by an auxiliary diode D_1.

if the transistor ceased conducting completely; but the voltage value does get slightly above 7.5 volts. When it does, D_1 will conduct because the anode will be more positive than the cathode. The potential on the latter element is normally about 7.5 volts, obtained from the connection at T_1.

When D_1 conducts, it acts as a low-valued resistor in series with the 0.1-μf capacitor C_{107}. This combination loads the primary of T_1 and reduces the signal fed to the first i-f stage and all subsequent stages. On moderate and weak signals the current passed by X_2 is enough to keep its collector potential below 7.5 volts. In such instances, D_1 cannot conduct and T_1 is permitted to pass a stronger signal on to the first i-f amplifier.

A Second Transistor Receiver

Another radio receiver, which is designed along somewhat similar lines as the prior set but which contains a greater number of stages, is shown in Fig. 6·8. The lineup of stages includes a converter, three i-f amplifiers, a germanium-diode detector, and three audio amplifier stages. The maximum power output is in excess of 200 mw, which is more than ample for this particular purpose. The B+ voltage is 6 volts, and the current drain, at low levels, is 10 ma. At high peak levels it may go as high as 20 ma. A set of four 1½-volt batteries could be expected to have a life in excess of 250 hr if used at the rate of 2 hr a day.

The transistor used in the converter stage is a drift transistor; all of the others have an alloy-junction structure. The transistor type numbers are unfamiliar because these units are of Japanese manufacture; however, the units are similar in construction to American-made transistors. All are PNP types.

The converter stage is closely similar to the same stage in the receiver previously discussed. The antenna is a ferrite rod on which is wound a tuned primary coil, for station selection, and a secondary coil to apply the signal to the base of X_1. Tuning range extends from 532 to 1,620 kc.

The oscillator transformer is somewhat more extensive; it has one winding which is connected into the emitter circuit, one winding which is connected to the collector, and one winding which serves to couple the other two windings to each other. This intermediate winding, in conjunction with the oscillator tuning capacitor, establishes the oscillator frequency. A small trimmer capacitor provides an alignment adjustment at the high end of the band, while a small adjustable slug permits alignment at the low end. The intermediate frequency produced in X_1 is 455 kc.

Only part of the primary winding of T_1 is utilized by the collector of the converter transistor in order to achieve an impedance match. As is customary in transistor i-f transformers, the primary is tuned to develop the necessary high impedance for X_1, while the secondary is untuned to better match the low input impedance of the following stage. Tuning of the transformer is accomplished by a movable slug in T_1.

A diode D_1 is connected between the primary of transformer T_1 and the center tap of transformer T_2. This is a delayed agc diode which supplements the normal agc network. Operation of this circuit has already been described.

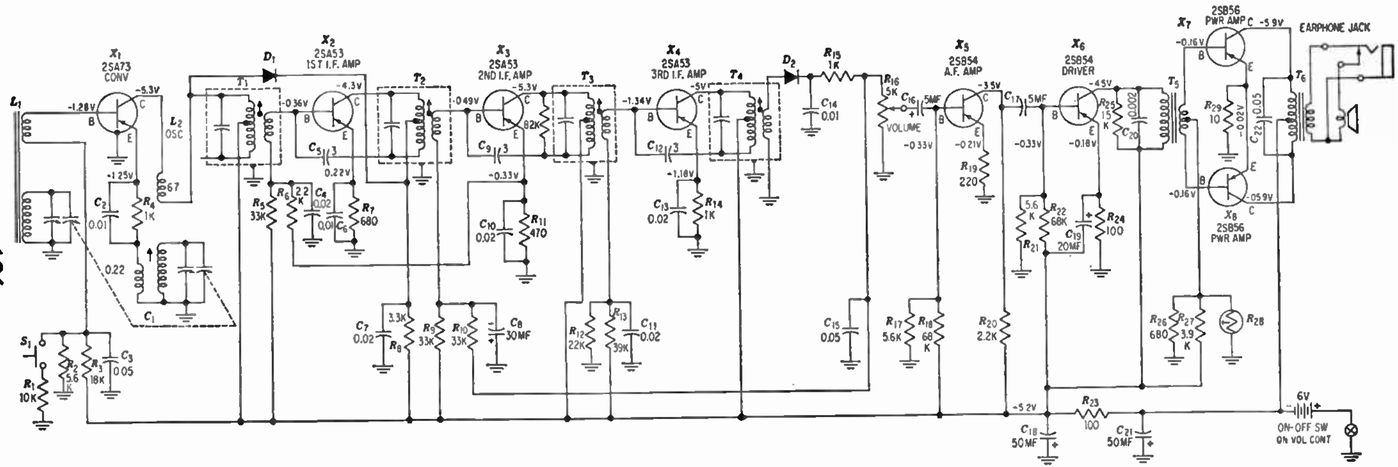


Fig. 6·8 A second, more extensive transistor receiver. (Motorola, Inc.)

Before we leave the converter stage, note should be made of the special switch S_1 in the base-bias network. The operating bias for the transistor is provided by R_2 and R_3. S_1 is normally open and does not take part in the operation of the stage. If it is desired to measure the battery drain for servicing purposes, however, then the receiver is turned off and a milliammeter is connected across the on-off switch and S_1 is depressed. The 10,000-ohm resistor which S_1 brings into the battery circuit establishes the same current drain as a normally operating receiver will. In this case, the milliammeter should read 9.5 to 11.5 ma. This, then, is a simple circuit arrangement to quickly check current drain.

There are three stages of i-f amplification, each providing more than 25 db gain. Each transistor is connected with the emitter common to input and output circuits. Input windings are untuned to present a low impedance to the base of the following stage, while the output windings are tuned to develop a high impedance.

In each instance, only part of the collector winding is utilized as an output load for the transistor to which it is connected. Part of the winding also provides a suitable neutralizing voltage which is fed back to the base (of that transistor) through a 3-$\mu\mu$f capacitor. Transformers T_1, T_2, and T_3 are basically the same in so far as primary and secondary windings are concerned. Transformer T_4 has a different set of impedances to match the loading imposed by the diode detector.

Automatic gain control is applied to the base of the second i-f amplifier. A positive voltage is obtained from the diode detector and fed back to the base of the second i-f transistor. Note that this agc voltage is applied in series with a negative d-c voltage (from the battery), and it is the latter which establishes the bias on the second i-f transistor base when no signal is being received. When a signal is received, the agc voltage reduces the negative biasing voltage by an amount dependent upon signal intensity.

The same i-f stage also has a small stabilizing resistor in the emitter lead. Its value is lower than the corresponding emitter resistors of the other i-f stages, and the reason for this was discussed previously. Base bias for the other i-f transistor amplifiers is provided in each instance by a voltage divider.

This primary agc circuit is supplemented by diode D_1 between the converter and first i-f stages. The combination of the two networks provides efficient control of the overall receiver gain.

The detector is a germanium diode (1N60 or equivalent) with the 5,000-ohm volume control as its d-c load. A 0.01-μf capacitor C_{14} acts as an i-f bypass. Capacitor C_{15} is a filter capacitor for the agc voltage.

Three stages of audio amplification follow the second detector. The first two stages employ similar transistors and are similarly biased. In each instance there is a voltage-divider network in the base circuit and an emitter resistor for d-c stabilization. In X_5 the emitter resistor is unbypassed, and so both signal and d-c degeneration or stabilization are provided. In X_6 the emitter resistor is bypassed by C_{19} and C_{18}; hence, only d-c degeneration takes place.

The output stage is a common-emitter push-pull class B amplifier. A thermistor is provided in the base-biasing circuit to maintain a steady base current with temperature changes. This is explained at greater length on page 209. Some additional d-c stabilization is also provided by the 10-ohm resistor in the common-emitter return path. No signal degeneration is introduced by leaving R_{29} unbypassed, because the signal currents of the transistors are 180° out of phase with each other. The a-f response is limited by C_{22}, the 0.05-μf capacitor across the output transformer.

A 15-ohm earphone can be connected to the output for personal reception if desired. When this earphone is in use, no sound is heard from the loudspeaker.

Two internal views of this receiver are shown in Fig. 6·9. The top photograph shows the location of the alignment adjustments. The other identifies the various components.

Receiver with Transistor Detector

The schematic diagram of another transistor portable radio is shown in Fig. 6·10. This circuit contains a separate mixer and oscillator, two i-f stages, a transistor second detector, an audio amplifier, and a class B push-pull output. Direct-current power is supplied by five 1½-volt flashlight batteries, and the audio power output is in excess of 250 mw.

The mixer-oscillator stage (with minor modifications) was discussed in Chap. 5. The feedback of energy from the collector of X_2 to the base is accomplished by means of the small tickler coil that is connected to the 1,500-$\mu\mu$f base capacitor. The collector itself is tapped down on L_2 so that an impedance match can be secured, while, at the same time, the Q of the coil is not loaded down to the point where frequency stability and tracking are affected.

The oscillator output is inductively coupled to the secondary of L_2 and then transferred via C_4 to the emitter of the mixer stage. The 2,700-ohm resistor paralleling C_4 serves to stabilize X_1. The mixer combines the oscillator voltage with the incoming signal, received from the base, and the resultant i-f signal appears in the collector circuit

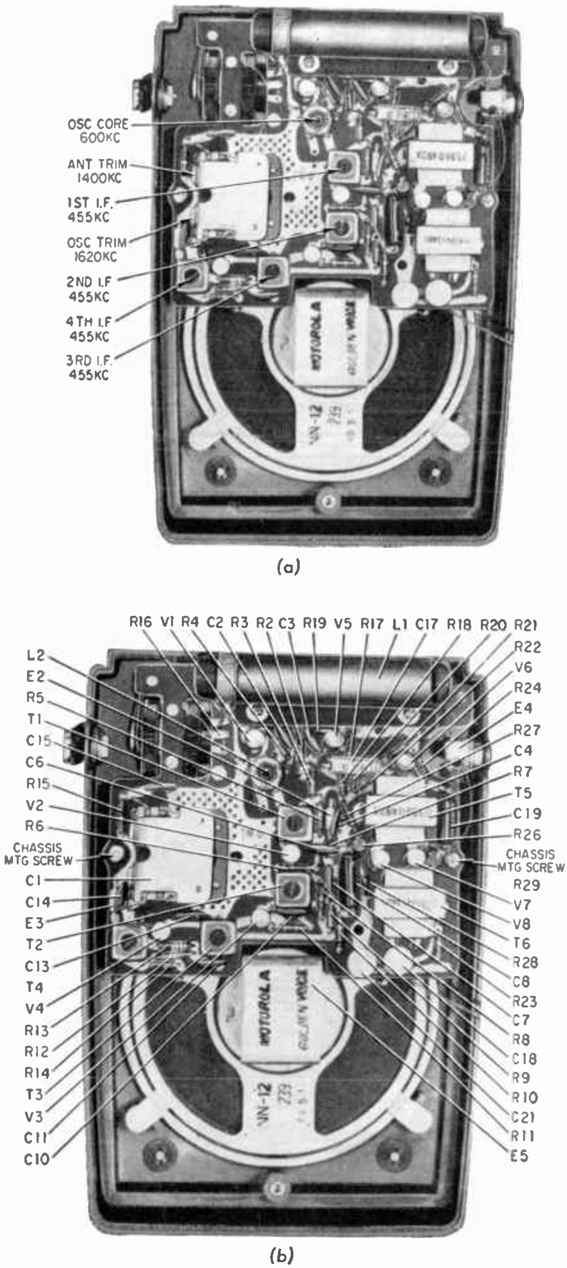
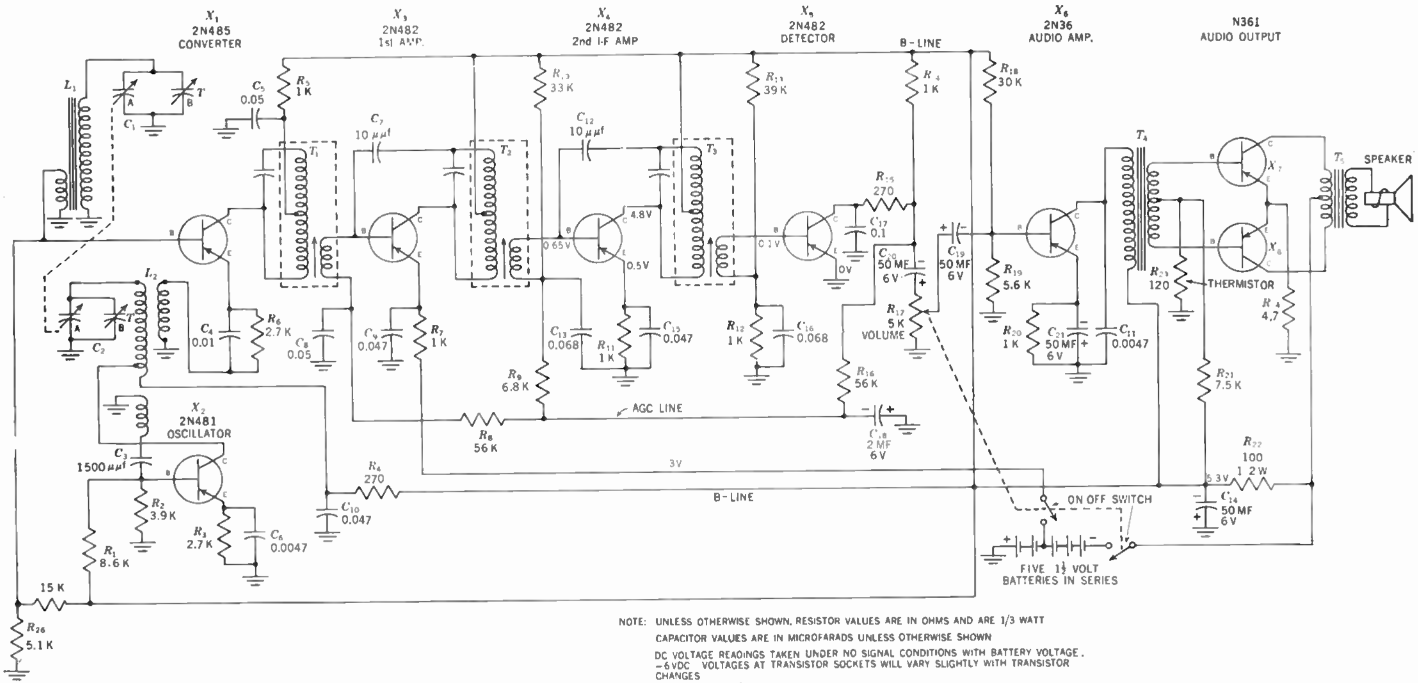


Fig. 6·9 (a) Alignment-point locations for receiver of Fig. 6·8. (b) Parts location in the receiver of Fig. 6·8 (Motorola, Inc.)



NOTE: UNLESS OTHERWISE SHOWN, RESISTOR VALUES ARE IN OHMS AND ARE 1/3 WATT
CAPACITOR VALUES ARE IN MICROFARADS UNLESS OTHERWISE SHOWN
DC VOLTAGE READINGS TAKEN UNDER NO SIGNAL CONDITIONS WITH BATTERY VOLTAGE. −6VDC VOLTAGES AT TRANSISTOR SOCKETS WILL VARY SLIGHTLY WITH TRANSISTOR CHANGES
USE ONLY VTVM

Fig. 6·10 Schematic diagram of a transistor receiver employing a transistor as a detector.

and is transferred by T_1 to the first i-f amplifier. R_5 and C_5 form a decoupling filter to prevent the signal from reaching the B— line, with R_5 chosen to limit the collector dissipation to a safe value should the oscillator fail to function properly.

The two i-f stages employ common-emitter arrangements, and both stages are agc-controlled. The agc voltage is obtained from a class B power detector and applied to each base. Since these are PNP transistors, the base should be negative with respect to the emitter. A negative voltage is supplied to each base from the d-c battery line. To vary the gain of each stage, the output voltage from the detector becomes less negative as the incoming signal level increases. This has the effect of reducing the bias between the base and emitter and, in essence, lowers the emitter and collector currents through the transistor. As the current decreases, the gain of the stage drops.

In the first i-f stage, the bottom end of the emitter resistor R_7 connects to the —3-volt terminal on the battery. An additional —0.75 volt develops across the 1,000-ohm emitter resistor, so that the total emitter voltage with respect to ground is —3.75 volts. The voltage of the base is —3.90 volts, and this is obtained through the agc line and the connection of this line to R_{14}. The latter resistor ties in, at its opposite end, to the battery. The net bias, then, between base and emitter is 0.15 volt with the base more negative than the emitter. When the agc bias is active, upon the arrival of a signal, it will act to reduce the bias difference between base and emitter.

The reason for returning the emitter of X_3 to a negative tap on the battery is that it permits the gain of this stage to be reduced sufficiently to prevent overload of the second i-f stage or the detector. In the second i-f stage, the emitter resistor is returned to ground, and because of this, the agc voltage cannot reduce the transistor current to as small a value as it can in the first stage.

The immediate frequency is 455 kc. C_7 (10 $\mu\mu$f) and C_{12} (10 $\mu\mu$f) are neutralizing capacitors.

The stage following the second i-f stage is the second detector, and a 2N482 transistor is employed here in a class B power-detector arrangement. This type of detector was once fairly popular in vacuum-tube circuits, and a good deal of this popularity stemmed from the fact that it will amplify the signal. It is the latter advantage which accounts for its use here, and in its present application 10 db of gain is obtained. It would be simpler and cheaper to use a germanium diode, as in the two preceding sets, but a diode introduces a loss, and gain here is important. In vacuum-tube receivers, the class B power detector is no longer used because sufficient prior amplification is available so that detector gain is not required, because diodes are cheaper, and because a vacuum-tube power detector introduces a considerable amount of distortion on relatively weak signals. In the transistor class B detector, less distortion is introduced because transistor characteristic curves are considerably more linear than vacuum-tube characteristic curves. As a matter of fact, transistors will give essentially linear detection at smaller power levels than even diodes.

Not to be overlooked also is the ability of a transistor detector to supply more agc power than a diode detector.

For those readers who may not be familiar with power detectors, the following explanation is given. The E_G-I_P characteristic curve for

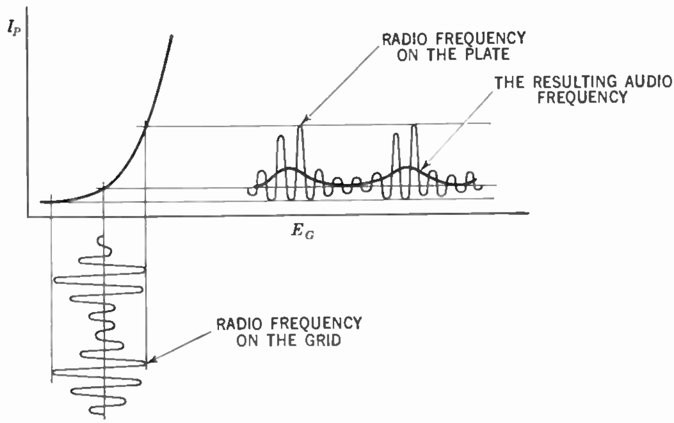


Fig. 6·11 The manner in which a class B power detector operates.

a triode is given in Fig. 6·11, and if we bias the tube close to the cutoff point, then the incoming signal applied to the grid will vary back and forth about this point. However, the negative half of the input signal will operate over the curved portion of the characteristic, producing considerably less plate current than the positive half cycles of signal. (Portions of the negative half of the signal will drive the tube to cutoff.) Essentially we have rectified (i.e., detected) the signal, and if we remove the i-f component, we shall obtain the desired audio intelligence.

In the transistor detector of Fig. 6·10, the emitter is connected directly to ground. The potential of the base is established by the divider network of R_{12} and R_{13}, and this voltage is so low that the stage is close to cutoff. Under no-signal conditions, the collector voltage is very close to the full B— voltage. The agc line connects also to the collector of X_5, and it is through this connection that the base elements of X_3 and X_4 receive their operating voltages.

When a signal is received, the collector current of the detector increases, and since this is a PNP transistor, electron flow will be from the battery to the collector. This will produce a voltage drop across R_{14} such that the collector will become *less* negative or *more* positive. This change will be transmitted to the bases of the controlled i-f amplifiers and result in a current decrease through these transistors. In this way the gain of the two i-f stages is controlled. R_{16}, R_9, R_8, C_{18}, C_{13}, and C_8 serve to filter out any audio components of the agc voltage and to establish the time constant of this network.

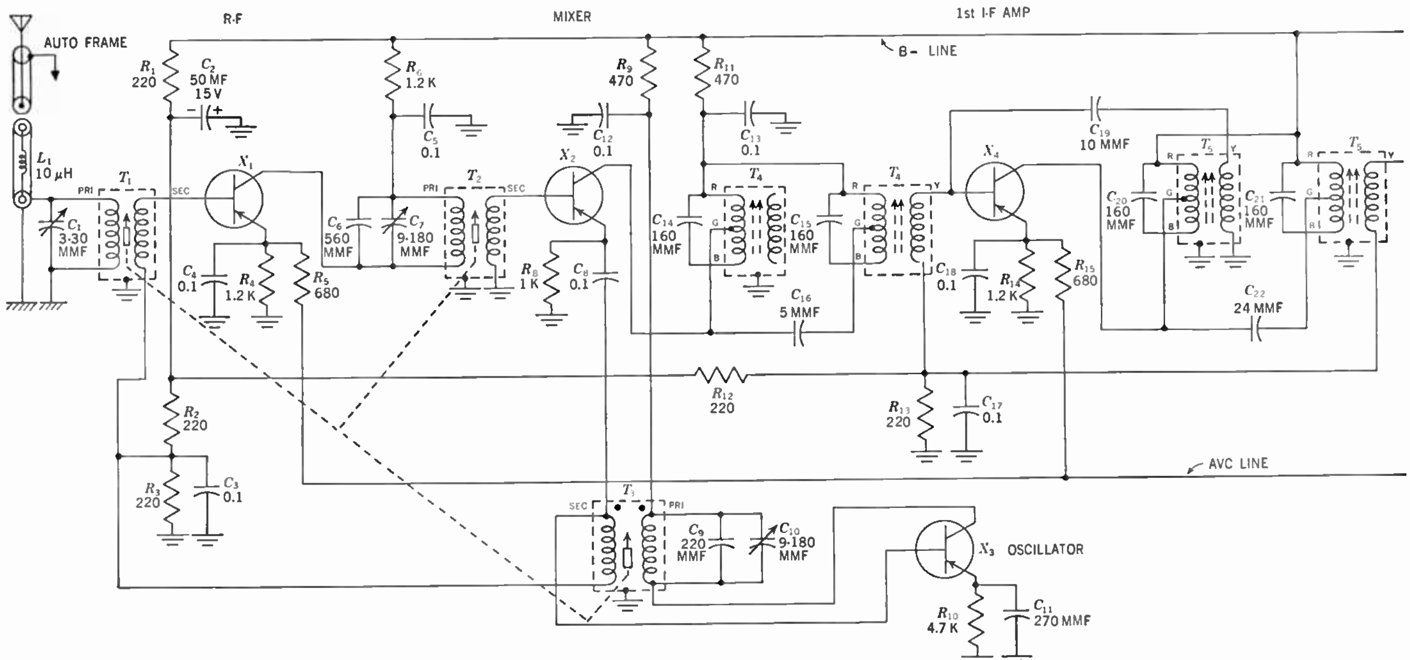
The remainder of this receiver circuit is fairly straightforward and will present little difficulty to the reader. The signal from the detector appears across the volume control R_{17} and is taken from there and applied to the base of X_6, a 2N362 audio amplifier. R_{18} and R_{19} form a voltage-divider network to provide the desired bias for the stage. A stabilizing resistor is present in the emitter circuit, and this is suitably bypassed by a 50-μf capacitor. The 0.0047-μf capacitor from collector to ground serves to remove any stray i-f voltage that may have reached this point.

The final stage is a class B push-pull output amplifier. The full 7.5 volts is applied to the collector elements to obtain the desired power output. A small base-to-emitter bias is used to minimize crossover distortion and to make it easier to substitute other 2N361 transistors should replacement become necessary.

Transistor Automobile Radio

The low-voltage requirements of transistors make their application in automobile receivers particularly desirable, since they can operate directly from the 6- or 12-volt battery and thereby eliminate the need for a vibrator, a power transformer, and a rectifier. This will not only result in a lowering of costs and a substantial saving in space requirements but also help to reduce servicing expense. It has been found that 85 per cent of the defects which occur in automobile radios stem from a breakdown of the vibrator, transformer, or rectifier. Eliminated, too, is vibrator hum. Finally, the drain on the car battery by the radio is lowered by a factor of 10.

The schematic diagram of a transistor automobile radio is shown in Fig. 6·12. In common with many such receivers, there is an r-f stage, a converter, two i-f amplifiers, a transistor detector, an audio amplifier, and a push-pull output stage. Power output is 2 watts, and the overall receiver sensitivity is 2 μv. After we have investigated the circuit, the


R-F
MIXER
1st I-F AMP
AUTO FRAME
B- LINE
AVC LINE
OSCILLATOR
R_1 220
C_2 50 MF 15 V
R_6 1.2 K
C_5 0.1
R_9 470
C_{12} 0.1
R_{11} 470
C_{13} 0.1
C_{19} 10 MMF
L_1 10 μH
C_1 3-30 MMF
PRI
T_1
SEC
X_1
C_4 0.1
R_4 1.2 K
R_5 680
C_6 560 MMF
C_7 9-180 MMF
T_2
X_2
R_8 1 K
C_8 0.1
C_{14} 160 MMF
T_4
C_{15} 160 MMF
C_{16} 5 MMF
X_4
C_{18} 0.1
R_{14} 1.2 K
R_{15} 680
T_5
C_{20} 160 MMF
C_{21} 160 MMF
C_{22} 24 MMF
R
G
B
Y
R_{12} 220
R_{13} 220
C_{17} 0.1
R_2 220
R_3 220
C_3 0.1
T_3
C_9 220 MMF
C_{10} 9-180 MMF
X_3
R_{10} 4.7 K
C_{11} 270 MMF

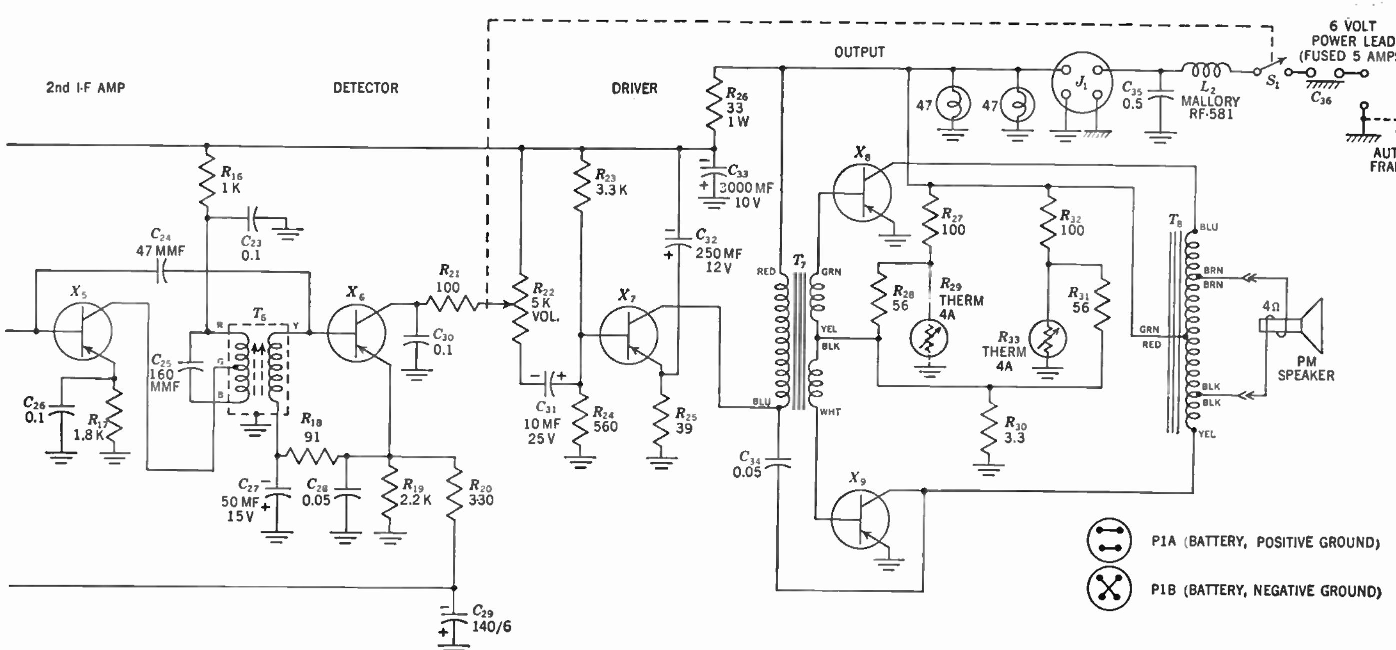


Fig. 6·12 Schematic diagram of a transistor automobile receiver. (RCA) performance characteristics will be considered in somewhat greater detail.

The antenna for this receiver is a conventional automobile rod antenna, and this feeds the tuned primary of the antenna transformer T_1. The secondary of T_1 is untuned in order to match the low input impedance of X_1, the r-f amplifier. Tuning of the transformer is carried out by means of an adjustable powdered-iron tuning slug which is ganged to similar slugs in the mixer and oscillator coils. Each slug is 1.2 in. long by 0.18 in. in diameter. Slug travel is about 1 in.

Bias for the base of X_1 is obtained at the junction of R_2 and R_3 and is on the order of —1.5 volts. The emitter contains a stabilizing resistor R_4 and a suitable r-f bypass capacitor C_4. Automatic-gain-control voltage from the second detector is applied to R_4 by resistor R_5. More on this in a moment.

The output signal of X_1 reaches the base of the following mixer by means of transformer T_2. The turns ratio of this transformer is such that the 10,000-ohm output impedance of X_1 is matched to the 500-ohm input impedance of the mixer. Transformer operating Q is on the order of 15 to 20. Owing to the step-down turns ratio necessitated by the impedance match, the transformer reduces the signal by 3.7 db. However, this is greatly overshadowed by the gain of the r-f stage, which is 20 db, and the conversion gain of the mixer, which is also 20 db.

The oscillator X_3 possesses a tuned primary and an untuned secondary. A relatively high capacitance is shunted across the primary in order to improve the stability of the oscillator. Feedback of energy is between collector and base. Bias voltage for the base is obtained from the junction of R_2 and R_3 and is about —1.5 volts. The collector receives its operating voltage through R_9. The $R_{10}C_{11}$ network in the emitter circuit introduces degeneration in the oscillator circuit. However, this also reduces the loading on the oscillator-tuned circuit by the transistor input circuit and provides greater oscillator stability.

The oscillator voltage is injected into the mixer circuit by way of C_8 and R_8. Approximately 0.4 volt rms is used, this having been found to provide the highest conversion gain in the mixer. Too strong or too weak an oscillator injection voltage will result in less i-f signal being made available to the i-f system, and the audio output will be correspondingly affected.

Both coupling capacitor C_8 and the secondary of T_3 present a low impedance at both radio and intermediate frequencies, so that R_8 is effectively bypassed to ground for both input and output signals.

Three i-f interstage coupling networks are employed for the two i-f amplifiers, with each network consisting of two capacitively coupled double-tuned transformers T_4–T_4 and T_5–T_5 and a single-tuned transformer T_6. These provide the i-f response curve shown in Fig. 6·13. The overall response curve (radio and intermediate frequency) is also shown, and it can be seen that the overall selectivity is determined almost entirely by the i-f coupling networks.

Biasing of the first i-f stage is similar to the r-f stage, with approximately −1.5 volts being applied to the base from the junction of R_{12} and R_{13}. The emitter possesses a 1,200-ohm stabilizing resistor, and a 680-ohm resistor brings the agc voltage into the stage. The collector

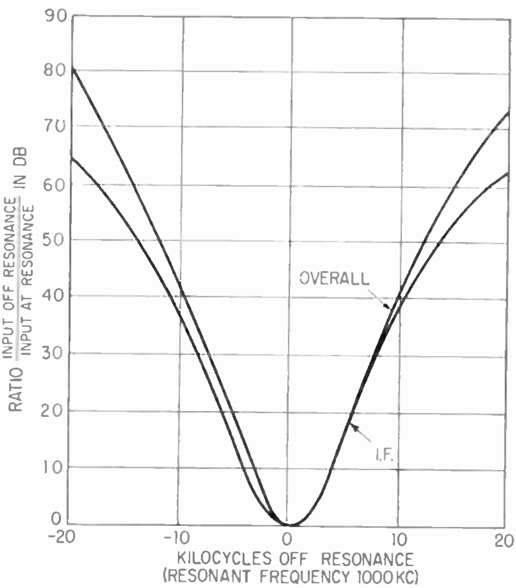


Fig. 6·13 The i-f and overall selectivity curves of the receiver shown in Fig. 6·12. (RCA)

of X_4 ties into the B− line without a decoupling network. The second i-f stage also receives its base bias from the junction of R_{12} and R_{13}, but its emitter does not tie into the agc line. However, the emitter does possess an 1,800-ohm stabilizing resistor and a 0.1-μf bypass capacitor to prevent signal degeneration. R_{16} and C_{23} form a decoupling filter for the collector of X_5.

Neutralization of the i-f stages is provided by C_{19} and C_{24}. An overall gain of 50 db is obtained from the base of the first i-f amplifier X_4 to the base of the second detector X_6.

The second detector is operated as a class B power detector. Zero bias is employed between base and emitter, and very little direct emitter current flows in the detector under no-signal conditions. When a signal is received, current flows in the collector circuit of the detector, and this current, passing through the volume control, provides the audio signal for the following stages. C_{30} shunts the i-f components of the detector signal around the volume control and back to the emitter.

Whatever current flows in the collector circuit passes through R_{19}, with the emitter end of R_{19} becoming negative with respect to the ground end. This negative voltage, properly filtered to remove audio variations, is fed back to the emitters of the r-f and first i-f stages, thereby providing automatic gain control. Since the voltage across R_{19} is negative (with respect to ground), an increase in signal strength will cause the value of this voltage to rise. This will drive the controlled emitters more negative than they were, in effect reducing the bias difference between each emitter and its base. This is equivalent to bringing each transistor closer to cutoff, thereby reducing its current flow and, in consequence, its gain.

The reader will recognize that in this system we have direct I_e control of gain. In order to obtain sufficient power to effect this control, a transistor detector was needed.

The RC networks in the emitter circuits of the detector, the r-f amplifier, and the first i-f amplifier provide suitable filtering of the r-f and audio components of the agc voltage. R_{18}, between the base and emitter of the detector, improves the linearity of the rectifying action.

The audio signal which is developed across the volume control is applied to the base of X_7, the audio amplifier. The design of this stage is straightforward, with R_{23} and R_{24} forming a voltage divider to provide X_7 with the proper base bias voltage. Control of the operating current of this amplifier is also provided by emitter resistor R_{25}. The collector current of X_7 is approximately 15 ma at moderate temperatures; it increases to 30 ma at 80°C and drops to 10 ma at −40°C. C_{32} is an emitter bypass capacitor, being essentially in parallel with R_{25}.

The signal from X_7 is transformer-coupled to a class B push-pull output stage consisting of X_8 and X_9. For best results (i.e., minimum distortion), X_8 and X_9 should be selected with characteristics as nearly similar as possible. Both output transistors are connected with emitters grounded. Furthermore, a small base-to-emitter bias is employed to minimize nonlinearity in the crossover segment of the characteristic curves. There is an optimum value of threshold emitter current which results in the least amount of nonlinearity, and the bias is chosen accordingly. While this value of current is independent of temperature, the corresponding base-to-emitter voltage needed to develop this current does vary with temperature, in this case approximately −0.0025 volt per °C. To develop the desired voltage variation with temperature automatically, two thermistors, R_{29} and R_{33}, are inserted in the base-to-emitter bias path. The resistance of these thermistors will vary with temperature in such a way that the desired voltage variation is obtained and the emitter current is kept steady. (A thermistor, for those readers not acquainted with it, is a thermally sensitive resistor which has a high negative temperature coefficient of resistance. That is, its resistance decreases as the temperature rises and increases as the temperature falls. In addition to the present application, thermistors are also employed for automatic volume compression and expansion, temperature control, and temperature and power measurements.)

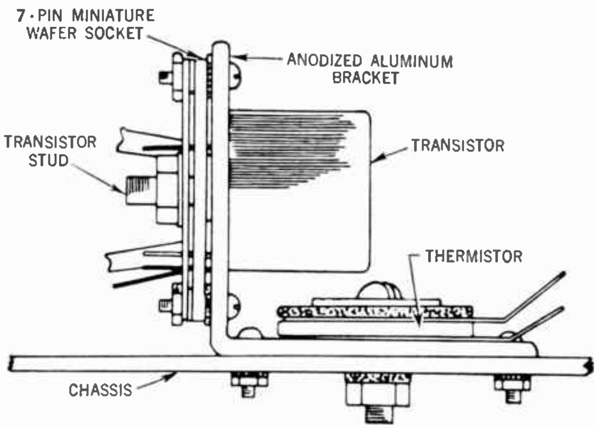


Fig. 6·14 Power transistor and thermistor mounting arrangement in the receiver in Fig. 6·12. (RCA)

The resistance network consisting of R_{27} to R_{33} may appear to be somewhat complex, but its only purpose is that just indicated. Also, the transistor-thermistor mounting arrangement shown in Fig. 6·14 is employed to maintain close thermal contact between these components so that they will be at the same temperature. This is particularly significant; without it the effectiveness of this control circuit is reduced.

The speaker is coupled to the collectors of X_8 and X_9 through autotransformer T_8. This method was chosen because of the close coupling it provides, with resultant high transformer efficiency. Power gain of the output stage is on the order of 24 db.

Capacitor C_{34} introduces negative feedback in the output stage for frequencies above 2 kc. This controls the extent of the high frequencies passed on to the speaker. C_{34} serves the same purpose as the common practice of shunting a capacitor across the primary winding of the output transformer.

Since this is an automobile radio, a number of precautions must be taken to eliminate interference from the car ignition system. A spark plate C_{36} and an r-f choke and capacitor L_2 and C_{35} prevent high-frequency interference signals from entering the radio via the battery line. Rejection of high-frequency impulse type of ignition interference appearing on the antenna is accomplished by the choke L_1 in series with the antenna lead, which, together with the shunt capacitance

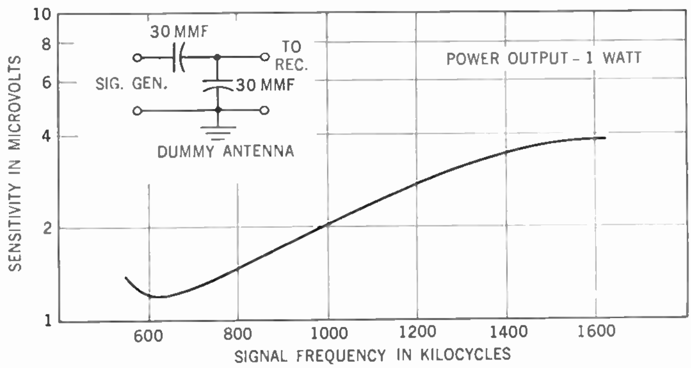


Fig. 6·15 Automobile radio-receiver sensitivity as a general function of signal frequency.

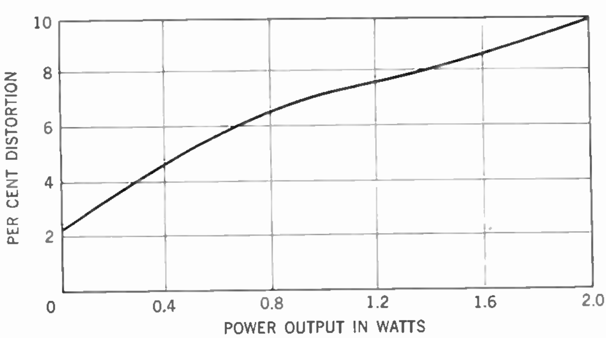
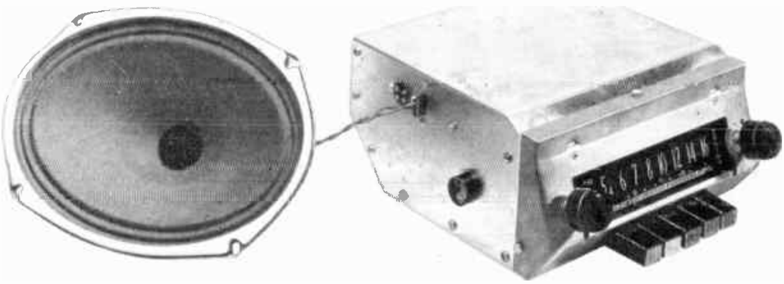


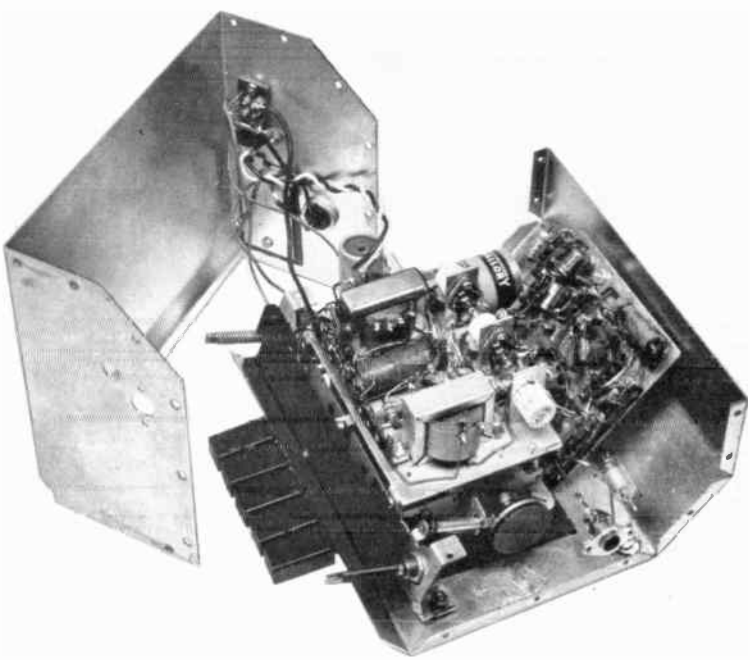
Fig. 6·16 Distortion vs. power output for automobile receiver of Fig. 6·12. (RCA)

across the antenna primary, forms a low-pass filter. Also, the receiver chassis is insulated from the receiver case.

Some of the performance characteristics of this receiver are shown in Figs. 6·15 and 6·16. In Fig. 6·15 we have the receiver sensitivity as a function of signal frequency. The increased sensitivity of the receiver at the lower signal frequencies is due almost entirely to the higher gain of the r-f and mixer stages at these frequencies. The overall sensitivity, however, is still quite good. In the second performance



(a)



(b)

Fig. 6·17 (a) Outward appearance of the RCA developmental transistor automobile receiver. The reduced size of the receiver can be seen by comparison with the nearby speaker. (b) An internal view of the RCA developmental transistor automobile receiver. The power transistors are seen in the foreground. (RCA) curve, Fig. 6·16, we have the variation in distortion with power output. At 2.0 watts, the distortion is 10 per cent, the latter being the recognized standard. Distortion rises fairly linearly with power output, starting from a low of 2 per cent.

Two views of this automobile receiver are shown in Fig. 6·17.

Reflex Amplifiers

In an effort to economize and save space, a number of designers have resorted to the use of reflex amplifiers in portable transistor receivers. The reflex action usually occurs in the final i-f amplifier which

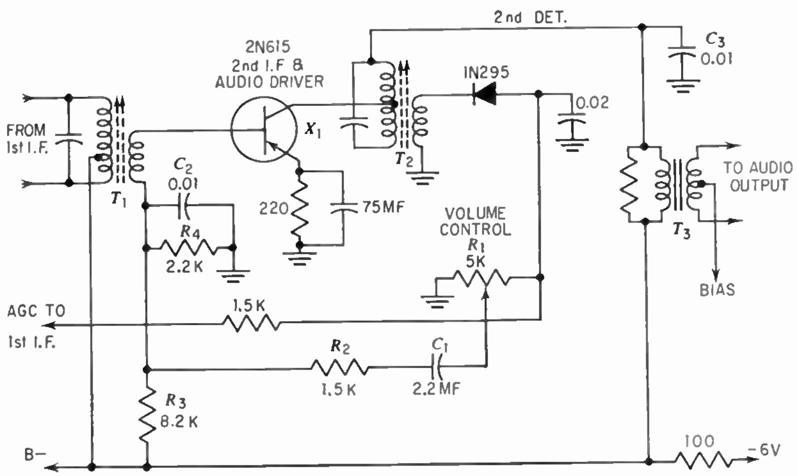


Fig. 6·18 A reflex-amplifier circuit.

functions not only as a normal i-f amplifier but as the first audio amplifier as well. The two functions are carried on simultaneously without interference to each other. This can be done because the i-f and audio frequencies are so widely separated.

A typical reflex circuit is shown in Fig. 6·18. As a normal i-f amplifier, X_1 receives the i-f signal from the preceding stage through T_1, amplifies the signal, and then transfers it by way of T_2 to the diode second detector that follows. At the detector, the signal is demodulated, with the audio and d-c components appearing across the volume control R_1. From this point, the d-c voltage, properly filtered, is fed back to the base of the first i-f amplifier for control of the gain of that stage.

To this point, circuit operation, as described, is normal. However, the audio signal at the volume control is now coupled back to the base of X_1, the second i-f amplifier, by C_1 and R_2. The audio signal sees the secondary of T_1 as just another length of wire, with negligible impedance. Likewise, after it has passed through X_1, the audio signal sees T_3 as its load, and not T_2. Transformer T_3 is an a-f transformer, with its primary in the collector circuit of X_1 and its secondary connected to the bases of the output class B amplifiers.

Just as the i-f transformers do not impede the audio signal, so the audio circuits must be suitably bypassed for the i-f signals. In the input circuit of X_1, capacitor C_2 keeps the i-f signal away from the

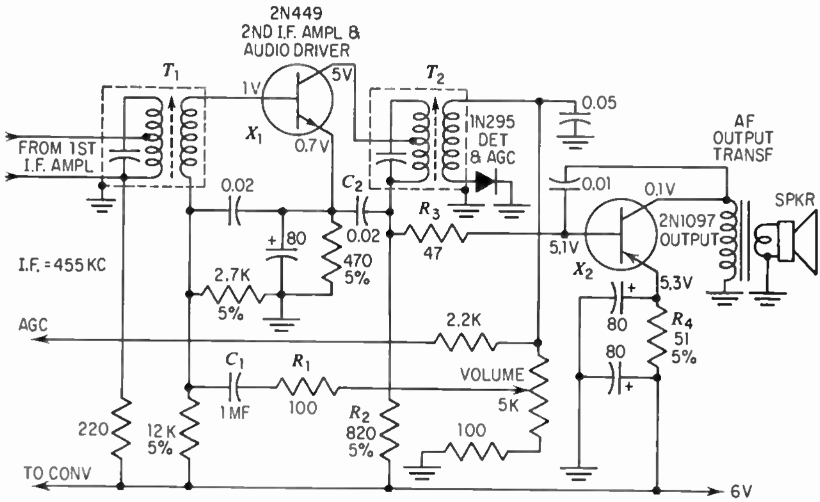


Fig. 6·19 A second reflex circuit.

volume-control network. In the output of X_1, C_3 permits the i-f signal to travel around T_3 on its way back to the emitter. Thus, the two signals are kept in their respective paths while X_1 amplifies them both (although generally not equally).

Capacitors C_2 and C_3 have some effect on the audio signal, but this is small in comparison to the signal that does not pass through them. Resistors R_3 and R_4 establish the d-c base bias for X_1.

The reflex circuit in Fig. 6·19 is similar to the preceding circuit with several variations. The audio signal is fed from the volume control through R_1 and C_1 to the base of X_1. After the audio signal has passed through the transistor, it is developed across the 820-ohm resistor R_2, which is connected in series with the primary of T_2. From R_2 the audio signal is transferred directly to the base of the audio output transistor

X_2 through R_3. Intermediate-frequency signals are kept away from R_2 by capacitor C_2.

In the audio output stage, the collector d-c potential is close to zero because its output transformer winding is grounded at its opposite end. Note, however, that the emitter resistor R_4 is returned to +6 volts, so that the potential difference between emitter and collector is 6 volts. The base element is also provided with a fairly high d-c voltage, here 5.1 volts, to bring its d-c level up to that of the emitter. The actual potential difference between base and emitter is on the order of 0.2 volt.

Transistorized FM Circuits

One of the difficulties that had to be surmounted in order to commercially produce a transistorized FM receiver was a low-cost high-frequency transistor. The FM r-f range extends from 88 to 108 Mc, far beyond the radio-broadcasting frequencies. Even in the i-f section, FM receivers use a center frequency of 10.7 Mc, a value still considerably above the radio-broadcast frequencies. Fortunately, it is possible to produce MADT, drift, and mesa transistors in quantity and with fairly uniform characteristics and at relatively low cost.

FM broadcast receivers are usually designed to incorporate AM as well, since the additional expense involved is quite small. This is because one set of transistors can handle both signals, and only a few additional i-f transformers and a detector diode constitute the major portion of the added cost. In the discussion to follow, a combination AM/FM receiver will be considered.

Block diagram. A block diagram of an AM/FM receiver is shown in Fig. 6·20. The complete front end of the set, from the r-f amplifier to the second i-f stage, is utilized in common by both AM and FM signals. It is only necessary to change the tuning circuits in order to switch from AM to FM or vice versa. As a matter of fact, the i-f tuned transformers remain series-connected without any change. Switching in the front-end section occurs principally in the oscillator, r-f amplifier, and mixer.

Beyond the second i-f stage, the signals separate, the AM signal going to a diode detector where it is demodulated. At the same time, a d-c control voltage for agc is developed and fed back to one or both i-f transistors. Both actions are completely conventional. Thereafter, the audio signal is amplified first by an a-f driver and then by a class B power-output amplifier. At this point, it is sufficiently powerful to drive one or more loudspeakers.

For the FM signal, a limiter stage follows the second i-f amplifier.

Then the signal goes to a discriminator where the sound intelligence is abstracted. The signal, now in its audio form, is fed to the driver where sufficient power is developed to adequately drive the class B output stage.

In the entire receiver, then, there are only three stages which the AM and FM signals do not use in common. This makes for a highly efficient arrangement, not only from the standpoint of design and cost but also from a servicing viewpoint.

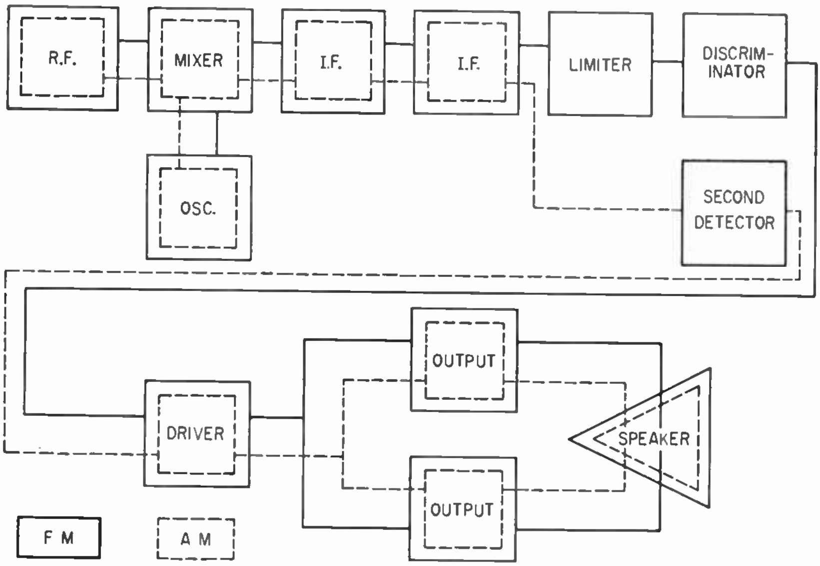


Fig. 6·20 Block diagram of a transistorized AM/FM receiver. (Philco Corp.)

Circuit description. *The Tuner.* The front-end section, the tuner, of this AM/FM receiver consists of an r-f amplifier, a local oscillator, and a mixer, Fig. 6·21. The r-f stage is a common-emitter neutralized tuned amplifier utilizing a MADT transistor. The AM and FM input circuits are switched at the base of X_1. The FM antenna is a 75-ohm whip type which simply extends straight up in the air. The AM antenna is a 6-in. ferrite loop. Both input circuits receive agc voltages from separate agc systems. When X_1 is AM-operated, reverse agc is employed just as it is in the broadcast receivers previously described. When X_1 is FM-operated, forward agc is employed. This differs markedly in operation from reverse agc, and it will be described presently. Both methods serve the same purpose, however: to control the signal level reaching the FM and AM detectors with varying input signal levels.

In the collector circuit of X_1, L_1 is the tuning coil for FM signals. The coil is resonated by C_6, a small variable capacitor which is part of a six-gang capacitor which tunes all of the tuning circuits in the front-end section. The other members of this assembly are C_1, C_3, C_8, C_9, and C_{18}. Each of these capacitors is shunted by small trimmer capacitors which permit tracking of each set of capacitors over its respective band, AM or FM. A small tap at the bottom of L_1 provides a neutralizing voltage which is fed back to the base of X_1 by capacitor C_N.

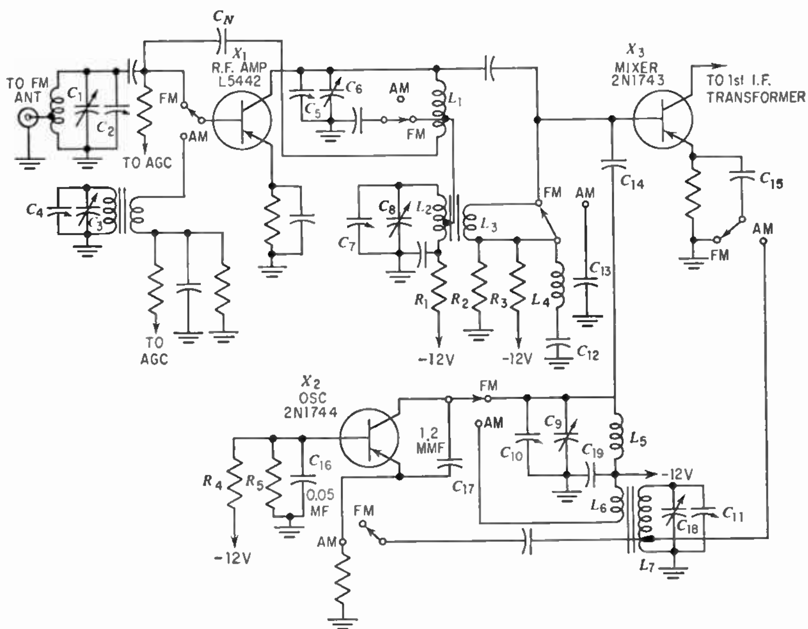


Fig. 6·21 The front-end section of the AM/FM receiver.

AM operation, occurring at much lower frequencies, requires no neutralization, and hence none is provided.

During FM reception, coils L_2 and L_3 are inactivated by the short circuit placed across L_3 by the AM/FM selection switch. Resistor R_1 is the collector dropping resistor for X_1, while R_2, R_3 form a voltage-divider network for the base of the mixer X_3. L_4 and C_{12} form a 10.7-Mc trap to prevent any 10.7-Mc voltage from developing in the mixer-base circuit and leading to possible oscillation in the mixer. In essence, these two components serve to neutralize the stage. This i-f trap also prevents any spurious signals from the r-f stage at the i-f frequency from penetrating further into the receiver. When AM signals are received, the trap is shunted by C_{13}, a relatively high value capacitor which effectively shunts the network out of the circuit.

The oscillator transistor X_2 appears at first glance to be connected in the common-emitter configuration. A closer look, however, reveals that the base is placed at ground potential by capacitor C_{16}. Hence, this is a grounded-base oscillator. Capacitor C_{17} is externally added to supplement the internal emitter-collector capacitance of X_2 to encourage oscillations. L_5, C_9, and C_{10} form the tuning circuit for FM, developing the required oscillator signal which is injected at the base of X_3 by capacitor C_{14}. The AM tuning network, composed of L_6, L_7, C_{18}, and C_{11}, is effectively kept out of the active circuit by C_{19}. R_4 and R_5 provide the proper biasing voltage for the base of X_2.

On AM, L_5, C_9, and C_{10} are switched out of the oscillator circuit, while L_6, L_7, C_{18}, and C_{11} are brought in. Injection of oscillator signal at the mixer is now made at the emitter of X_3 with C_{15}. (During FM operation, C_{15} serves as an emitter bypass capacitor.)

The mixer is completely conventional in design, employing X_3 as in the common-emitter configuration. In the collector circuit of X_3, the FM and AM transformers are series-connected, Fig. 6·22. This arrangement is possible because the impedance of the FM transformer windings is low at AM frequencies while the capacitor shunting the AM transformer winding provides a low-impedance path for FM signals when these are active.

I-F Amplifiers and Limiter. The i-f section consists of three stages of FM i-f amplification and two stages of AM, Fig. 6·22. All stages employ the common-emitter arrangement for both FM and AM signals. If we examine these stages from the standpoint of the AM signals, we find that interstage coupling is achieved by T_1, T_2, and T_3. An agc voltage is brought in at the base of X_4, where it combines with the normal d-c bias. This agc voltage is a reverse bias in that it tends to reduce the current flowing through X_4 when the signal is too strong and increase the collector current when the signal is too weak. An amplified agc voltage is also obtained in the emitter circuit of X_4, and this is used to control the gain of the r-f stage. Since the emitter voltage follows the base voltage, the r-f amplifier is also subjected to a reverse agc action.

The second i-f stage operates without any agc voltage. The AM signal is further amplified here, then transformer-coupled to a diode detector. Small 220-ohm resistors are inserted in the collector circuits of each i-f transistor because it has been found that under strong signal conditions the collector-base diode becomes forward-biased, allowing feedback of the proper amplitude and phase to produce oscillation on

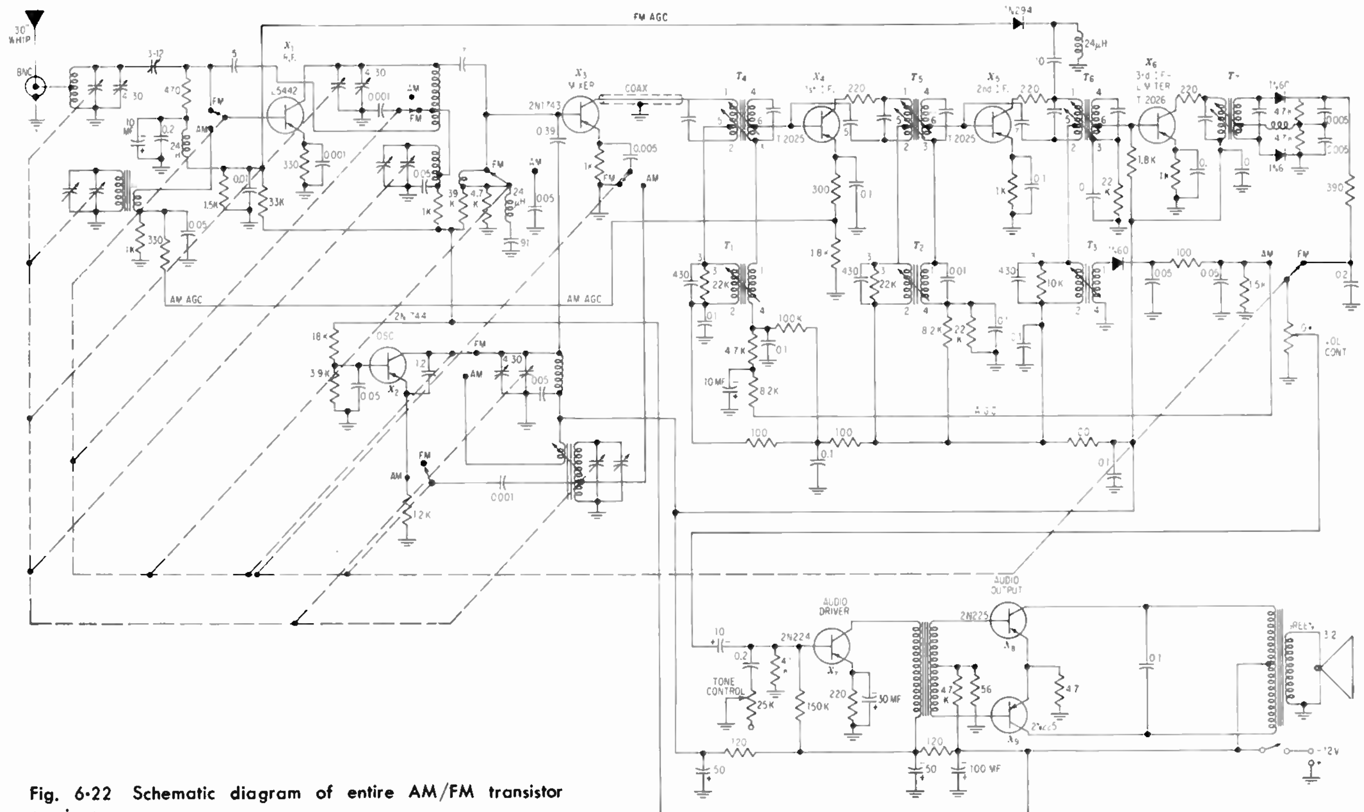


Fig. 6-22 Schematic diagram of entire AM/FM transistor receiver.

the signal peaks. These series collector resistors limit the signal fed back and maintain good stability.

For FM operation, transformers T_1, T_5, T_6, and T_7 couple the signal from stage to stage. The first two stages are neutralized, while the third stage, a limiter, is not. These neutralization capacitors on FM add to the internal capacity of the transistors on AM. To prevent oscillation on AM, a certain amount of mismatching is purposely incorporated into the AM transformers. This will reduce the AM gain, but so much is still available that no particular difficulty is presented.

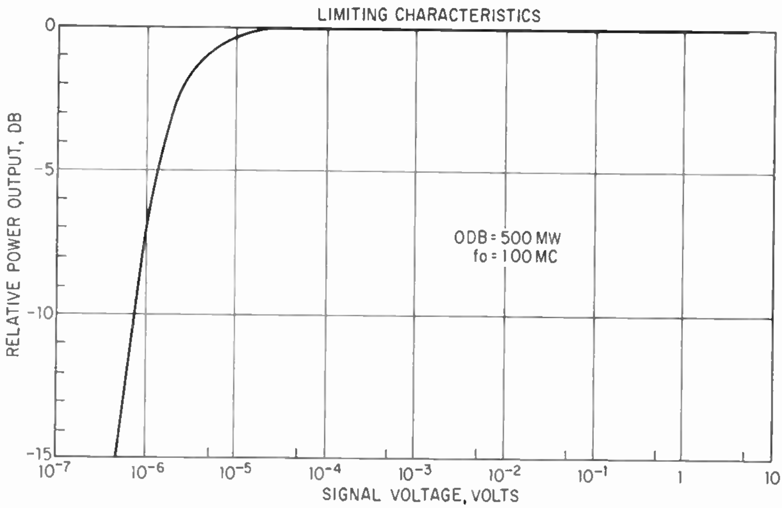


Fig. 6·23 The limiting characteristics of the i-f system.

While the third i-f stage is labeled as the limiter, practically, the first and second i-f stages also provide some limiting. The major portion of the limiting, however, occurs in the final stage. This additional limiting is required because when the third stage is strongly overdriven, it becomes inoperable.

Figure 6·23 shows the limiting characteristics of this circuit. Notice that full limiting occurs with an input signal of about 13 μv.

Detector and Audio Circuit. The AM detector is a conventional circuit providing a demodulated audio signal for the audio amplifiers and a reverse agc voltage for the first i-f amplifier, Fig. 6·24. The low-valued volume control was selected primarily because of the low input impedance of the following audio amplifier stage.

The FM detector is a Foster-Seeley discriminator utilizing two 1N60 diodes. When this circuit is employed with tubes, the load resistors R_1

and R_2 are ordinarily 100,000 ohms each. Because of the low input impedance of the following audio amplifier, however, these resistors have been each reduced to 4,700 ohms. The FM detector output is applied to the same 10,000-ohm volume control; the latter is switched to the system in operation. (For additional information concerning the operation of this and other FM detectors see Milton S. Kiver, "F-M Simplified," 3d ed., D. Van Nostrand Company, Inc., Princeton, N.J., 1960.)

The audio circuit consists of a 2N224 driver, transformer-coupled to a pair of 2N225's which are operated in class B push-pull. The output

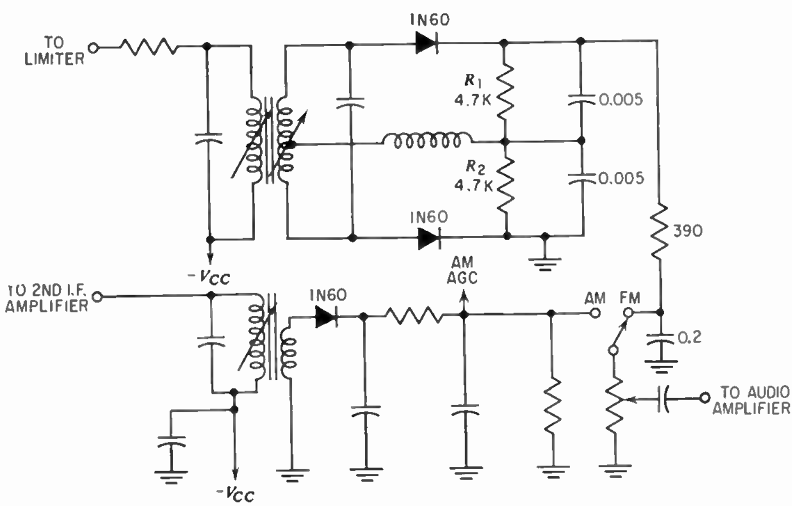


Fig. 6·24 FM discriminotor ond AM detector circuit.

stage is transformer-coupled to a 3.2-ohm speaker providing a maximum undistorted output in excess of 500 mw.

FM AGC Circuit. The FM agc circuit obtains its drive from the second i-f amplifier through a 10-$\mu\mu$f capacitor, Fig. 6·22. The 10.7-Mc signal is rectified and filtered to provide a negative voltage which is then applied to the base of the r-f amplifier. When the signal level increases, the negative agc voltage increases. Since the r-f transistor is a PNP type, a greater negative base voltage means more collector current. Thus, what we have here is a forward agc voltage in distinction to the reverse agc voltage provided during AM reception.

Forward agc operates not by current variations directly affecting transistor gain, but by changing the collector-to-emitter voltage to alter gain. A resistor is inserted in the collector lead. When the transistor current is increased, more voltage is dropped across this transistor. Less voltage is left between the collector and emitter and, as a result, the gain drops. The stage gain is increased by reducing the collector current. This reduces the voltage drop across the resistor and thus permits the d-c collector voltage to rise.

For high-frequency operation using MADT transistors, forward agc provides better overload characteristics than reverse agc. That is, there is a wider range of control and the system is better able to cope with strong signals than if the other form of agc network had been employed. At the lower AM i-f frequencies, it is more desirable to employ reverse agc.

As a broad distinction, then, forward agc achieves its control through variation of the collector-to-emitter voltage, whereas reverse agc achieves its control through variation of the collector current. To effectively employ forward agc, the transistor must be internally fabricated to provide a wide, linear reduction in gain at reasonable collector-to-emitter voltages. Without this characteristic, useful control by this method is not feasible.

Note: Nearly every transistor can have its gain reduced if its collector-to-emitter voltage is reduced sufficiently. However, in many transistors this reduction occurs when the collector voltage is so low that only very small amplitude of signal can be passed through the stage. With MADT transistors, the control can be achieved while the collector voltage is still relatively high. Hence, the amount of signal passing through the stage is still fairly large.

QUESTIONS

Answer the first four questions using the transistor receiver circuit of Fig. 6·3.

6·1 a. How many transistors does the receiver employ?

b. How does the first stage (X_1) function? Draw its equivalent vacuum-tube circuit.

6·2 a. What purpose does R_2 serve? R_3? C_7?

b. Why is the emitter of X_1 tapped down on L_3?

6·3 a. Why is the secondary of transformer T_1 untuned?

b. What is the purpose of C_{10} and C_{14}?

c. Why is $R_{\text{[illegible]}}$ much larger in value than R_5?

6·4 How does the agc system operate in the receiver of Fig. 6·3?

6·5 What points of similarity exist between the agc systems of vacuum-tube and transistor receivers? What are the differences?

6·6 Answer the following questions about the audio stages of the receiver in Fig. 6·8.

a. Identify the agc circuit. Include all components.

b. How is the crossover distortion minimized in the output stage?

c. How do the output transistors receive their biasing voltages?

Answer the following four questions about the transistor receiver circuit of Fig. 6·10.

6·7 a. How does the oscillator signal reach the mixer stage?

b. Where is the incoming signal applied to the converter?

6·8 Explain how the agc system operates.

6·9 Explain how the detector stage functions.

6·10 What advantages does this form of detection offer? What disadvantages in comparison to diode detectors?

Answer the following three questions about the circuit in Fig. 6·12.

6·11 a. Which stages are agc-controlled?

b. What form of control system (I_e or V_c) is utilized? Prove your answer.

6·12 What is the purpose of each of the following components: R_6, R_{18}, L_2 and C_{35}, R_5, and C_{19}?

6·13 What type of audio output stage is employed? What purpose do R_{29} and R_{33} serve?

6·14 Why is it important to inject the proper amount of oscillator voltage into the mixer? What happens if this voltage is too large or too small?

6·15 What is a reflex amplifier? Where is it most likely to be used in a transistor receiver? What are its advantages?

6·16 Explain how the reflex amplifier of Fig. 6·19 operates. Trace the paths of both the i-f and audio signals.

6·17 Which stages in the circuit of Fig. 6·22 are in operation during reception of an FM signal? During reception of an AM signal?

6·18 What is the difference between reverse agc and forward agc? Explain how each operates to control the level of the output signal. When is each used in the circuit of Fig. 6·22?

6·19 Trace the FM agc and AM agc circuits in the receiver of Fig. 6·22.

6·20 How is it possible to employ the same i-f transistors for AM and FM signals? How are these two signals kept in their respective paths?

CHAPTER 7

Transistors in Television Receivers

THE TASK of producing an all-transistor television receiver is considerably more difficult than that of bringing forth a commercial all-transistor radio receiver for two reasons. First, many of the circuits in television receivers operate at fairly high frequencies; second, the power requirements of a number of stages, particularly those in the vertical- and horizontal-deflection systems, are quite high. Both of these requirements must be met by transistors which are economical, stable, and (within each type) uniform in characteristics from unit to unit if transistorized television receivers are to become acceptable to the public.

Considerable research work is being done by the industry in this direction, and transistors which meet the foregoing criteria are gradually being developed. Already, several fully transistorized television receivers have been designed and marketed. The discussion to follow will analyze the various sections of a television receiver and indicate typical circuits that can be employed in them.

The block diagram of a black-and-white television receiver is shown in Fig. 7·1. The number of stages contained in each box is not shown and, to a certain extent, is not important. What is important is that the function represented by that box be achieved.

The R-F Stages

The front-end section of the receiver contains the tuner and the r-f amplifier, mixer, and local oscillator. In the vhf band, signal frequencies extend from 54 to 88 Mc and from 174 to 216 Mc. For uhf reception, a frequency coverage from 470 to 890 Mc is required. The local oscillator often generates frequencies which are 25 to 45 Mc above the incoming signals, but this, in itself, is of minor significance. If transistors can be made to operate in either the vhf or vhf-uhf bands, then they will certainly function up to frequencies which are 25 to 45 Mc higher.

For a transistor to be usable in these stages, it not only must be capable of operating at these frequencies but must also do so with a fair amount of gain. Also, the noise factor of a transistor is important here because of the very low level of the incoming signal.

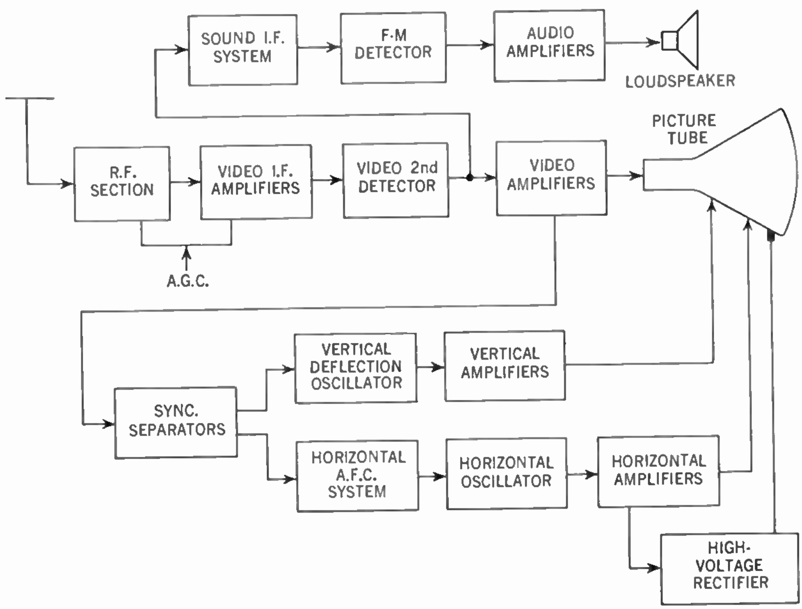


Fig. 7·1 A simplified block diagram of a black-and-white television receiver.

Of the three front-end stages, the requirements of the local oscillator are probably the easiest to satisfy. This is because noise is not a significant operating characteristic and because transistors will almost always oscillate at considerably higher frequencies than their cutoff value. On the other hand, for purposes of amplification, it is desirable to keep well below the transistor cutoff frequency, and this will restrict transistor application.

The r-f amplifier. The r-f amplifier is perhaps one of the most important stages in the television receiver. The signal which it receives from the antenna is exceedingly weak, perhaps no more than 30 or 40 μv. It is the function of the r-f amplifier to amplify this signal to whatever extent it can, at the same time keeping the amount of noise voltage which it (the amplifier) introduces at the lowest possible value. Thus, selection of a suitable r-f transistor is primarily based on two factors: (1) ability to amplify at the television frequencies and (2) as low a noise figure as possible.

The most widely employed configuration for the r-f amplifier is the common-emitter one, Fig. 7·2. The input signal from the antenna is brought to the base of the transistor through a filter network which, in Fig. 7·2, consists of a balun, an FM trap, and a 45-Mc trap. The balun serves primarily to match the balanced 300 ohms impedance of

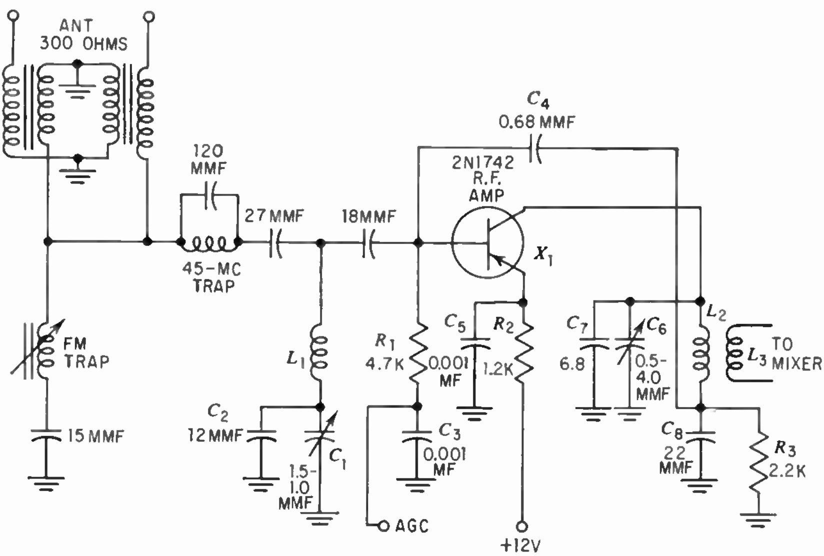


Fig. 7·2 The r-f amplifier of a television tuner.

the lead-in line to the unbalanced input impedance of the transistor. With a collector current of 2.5 ma and a collector-to-emitter voltage of 10 volts, the input impedance will generally be under 200 ohms over the range of frequencies to be received. The purpose of the FM trap is to prevent any 88- to 108-Mc FM signals from entering the tuner in strength. And the 45-Mc series trap is designed to keep signals in this range from also affecting the receiver. Such signals can be particularly destructive to the final picture, because the video i-f section operates in the 45-Mc range and the selectivity of the tuner circuits is not great enough to effectively kill a strong 45-Mc incoming signal.

The input tuning coil for X_1 is L_1. This, in conjunction with C_1 and C_2, provides a certain amount of selectivity. L_1, as well as L_2, would be changed with each channel. C_1 is an adjustment capacitor which is adjusted on channel 10 and then left alone for the remaining channels.

In the r-f amplifier shown in Fig. 7·2, the end of the emitter resistor R_2 is connected to +12 volts. The agc line, which is connected to the base of X_1, also ties into the 12-volt d-c line so that both emitter and base are within 0.1 or 0.2 volt of each other, with the base more negative than the emitter. Control of the gain of the stage is then achieved by varying the agc voltage.

The output circuit consists of L_2, C_6, C_7, and C_8. To see this circuit more clearly, it is redrawn in Fig. 7·3. Note that the connection point of C_6 and C_7 with C_8 is grounded, and this has the same effect as placing a ground connection on L_2. As a result, the r-f voltage at the bottom end of L_2 is 180° out of phase with the r-f voltage at the top

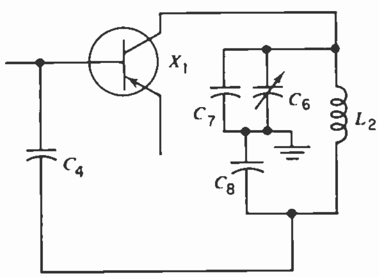


Fig. 7·3

(or collector) end. Capacitor C_4, by tying into the bottom end of L_2, is able to bring to the base of X_1 a neutralizing voltage to counteract any energy feedback that may take place between the collector and base internally. Capacitor C_4 is thus the neutralizing capacitor.

Note that the collector has a d-c path to ground through L_2 and resistor R_3. Since the emitter (and base) tie into the +12 volt line, the collector can be considered to be 12 volts negative with respect to the two latter elements. This provides the proper voltage polarity for the collector (of a PNP transistor) and enables the transistor to function properly.

The 2,200-ohm resistor R_3 is inserted in the collector circuit because forward agc is applied to the base. As indicated in the preceding chapter, with forward agc an increase in signal produces an increase in collector current through the agc voltage change. This increase in collector current produces an increased voltage drop across R_3, thereby reducing the collector potential (with respect to the emitter), and this, in turn, reduces the stage gain. A signal reduction produces the reverse action, leading to a rise in gain.

The r-f mixer. Coils L_2, L_3, and L_4, Fig. 7·4, are wound on the same form and all are inductively coupled together. These coils would be changed, with L_1, for each different television channel. The signal from L_2 is brought into the base of X_2 by L_3 at the same time the oscillator signal is received by L_3 from L_4. The two signals mix in X_2, producing a difference-frequency signal which is then transferred to L_7. This difference signal is the i-f signal and has a frequency range from 41.25 to 47.00 Mc. The sound carrier is at 41.25 Mc, while the video carrier is at 45.75 Mc.

Note that the first i-f amplifier taps down on L_7. This is done because the input impedance of the next stage (in the i-f system) is less than the output impedance of the mixer. The same effect could have been

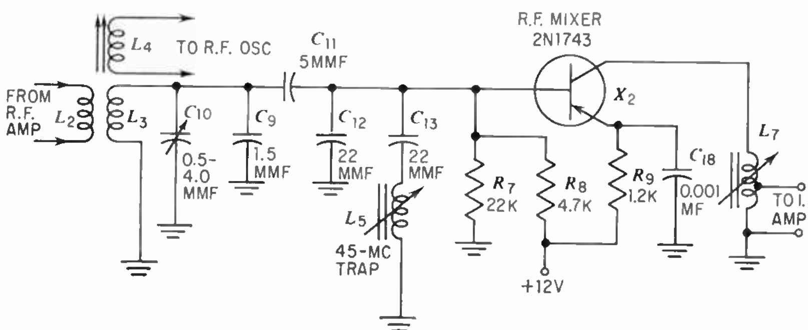


Fig. 7·4 The r-f mixer of a television tuner.

achieved by using a transformer in place of L_7 in which the secondary, connecting to the i-f amplifier, had fewer turns than the primary.

Capacitors C_9 and C_{10} parallel L_3 and help establish the resonant frequency range for this tuned circuit. C_{10}, being variable, is an alignment adjustment. Through its manipulation on a certain channel, generally 10, the proper response-curve shape is achieved. This then generally holds for all of the other channels as new coils are brought in for L_2, L_3, and L_4.

Capacitors C_{11} and C_{12} serve to match impedances of the tuned circuit and the input of X_2. C_{13} and L_5 form a low-impedance path for i-f signals (in the 41- to 47-Mc range). By placing such a low-impedance network across the input, any i-f signal fed from the collector to the base is, in essence, grounded and regeneration of the mixer at 45 Mc cannot occur. In this way, the mixer is neutralized.

Resistors R_7 and R_8 provide the proper bias current for X_2. R_9 is employed for thermal stability. At the same time, it is bypassed by C_{18} to prevent signal degeneration. The operating voltages are selected to bring the noise level of the mixer to as low a value as possible. Also instrumental in this action is the amount of oscillator signal voltage injected into the mixer base circuit. This is shown in Fig. 7·5. Note how the gain of the stage rises and the noise generated drops as the injected oscillator voltage rises. Beyond a certain point, however, the gain starts to drop and the noise begins to rise. From these curves, it is apparent that too little oscillator voltage at the mixer base is just as bad as too much.

The collector of X_2 connects to d-c ground through L_7, making the d-c potential at the collector close to zero volts. As with the r-f amplifier, the emitter and base tie into the +12-volt line.

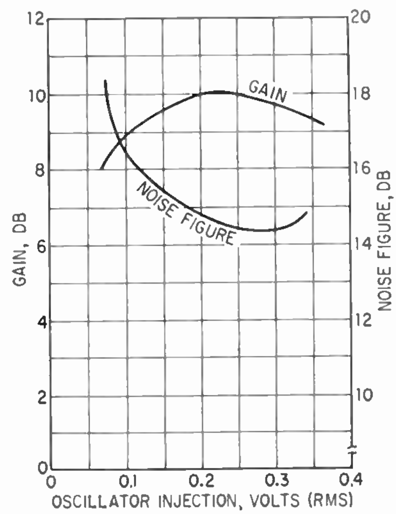


Fig. 7·5 Mixer goin ond the noise figure voriotion with oscillotor injection level.

The r-f oscillator. The final stage in the front-end tuner is the r-f oscillator. This is most frequently a common-base arrangement, as shown in Fig. 7·6. The base is brought to a-c ground potential by capacitor C_{14}. At the same time, to encourage oscillation, a small capacitor C_{15} is shunted between emitter and collector. In the common-base arrangement, the a-c voltages at emitter and collector are in phase, and through the internal capacitance in X_3 between emitter and collector, plus the assisting action of C_{15}, the stage oscillates readily over a wide range of frequencies.

L_4 and C_{17} help establish the operating frequency for the oscillator. L_6, an adjustable inductance, shunts L_4, and hence variation of the inductance of L_6 will alter the operating frequency. In a television receiver, the adjustment of L_6 can be made from the front panel to tune in a station properly. L_4 also has an adjustable core to permit major adjustments in the event that rotation of the core of L_6 does not bring in the station properly. Thus, the adjustment in L_4 can be considered as a coarse adjustment, while the adjustment of L_6 is a fine adjustment.

Resistors R_4 and R_5 establish the proper operating point for X_3. R_6 also ties into the +12-volt line to bring the emitter potential close to that of the base. The d-c path for the collector is completed to ground through L_4.

The circuit diagram of a complete tuner developed with the three stages just discussed is shown in Fig. 7·7. The overall gain will range

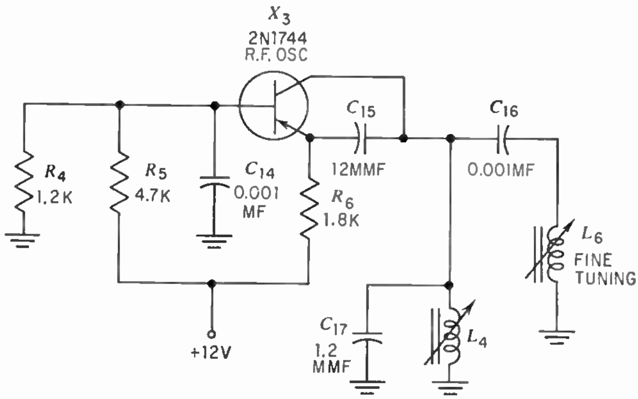


Fig. 7·6 The r-f oscillator of a television tuner.

from about 25 db on the high channels (7 to 13) to about 40 db on the low channels (2 to 6). The noise figure will be fairly comparable to that of a tube tuner.

Video I-F System

The video i-f system of a television receiver supplies most of the signal gain as well as most of the selectivity. To give some indication of the quantities involved, assume that a 30-μv signal is developed across a 300-ohm antenna feedline. This must become 50 volts of composite video signal at the cathode of the picture tube. To achieve this, approximately 110 db overall gain is required. Of this total value, 20 db may be assigned to the tuner and 30 db to the video amplifiers following the video second detector. This then leaves 60 db of gain to be developed in the i-f amplifiers at the picture carrier frequency.

Both common-base and common-emitter configurations have been employed for the video i-f system. The first arrangement permits using

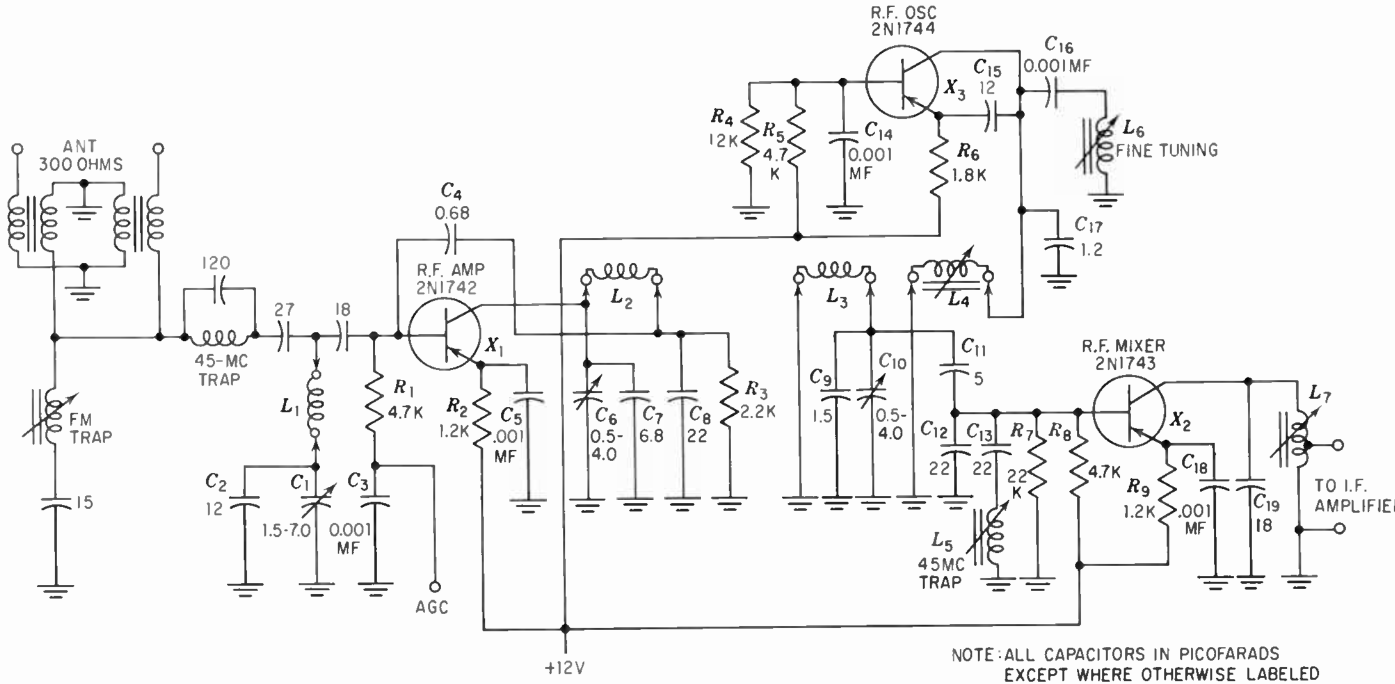


Fig. 7·7 A complete transistor tuner for a television receiver.

a transistor with a lower cutoff frequency, while the second provides much higher gain per stage. (The reason for this behavior will be explained mathematically in Chap. 12.) Because of the higher gain and because high-frequency transistors are readily available (at least in this range), the common-emitter arrangement is most frequently employed. For either approach, essentially similar circuits are used; they involve either single-tuned or bifilar coupling coils and trap circuits for the sound and adjacent channel carriers. These combinations possess the same form they do in vacuum-tube video i-f stages, and any knowledge of these stages in vacuum-tube systems will apply here in large measure.

A typical video i-f amplifier capable of operating in the 45-Mc range is shown in Fig. 7·8. The i-f signal is brought to the transistor

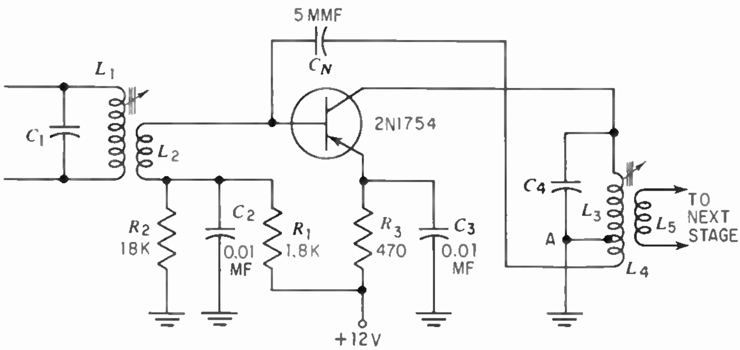


Fig. 7·8 A typical video i-f amplifier.

by the bifilar coil arrangement L_1 and L_2. The signal is applied to the base of the transistor 2N1754, and it appears in the collector circuit in amplified form. Here it is coupled by L_3 to L_5 and on to the next stage. In the arrangement shown, and with a 2N1754 transistor, 15 to 20 db of gain is obtainable. Thus, for a total of 60 db overall i-f gain, three to four i-f stages would be required.

A 12-volt supply is required to power the stage. R_1 and R_2 provide the base with the proper current (or voltage) to establish the operating point. C_2 places the bottom end of L_2 at signal ground in order that the full incoming video signal will be applied between the base (connected to the top end of L_2) and the emitter (connected by C_3 and C_2 to the bottom end of L_2). R_3 provides d-c stabilization for the 2N1754, while C_3 prevents signal degeneration across R_3.

In the output circuit, point A is at ground so that L_3 and L_4 have reverse voltages appearing across them. By connecting the end of L_4

to the base through a neutralizing capacitor C_N, any feedback voltage fed internally from collector to base can be neutralized from the top end of L_3. The value of C_N will depend on the internal feedback capacity of the transistor and on the number of turns in L_4.

Both the input and output sets of bifilar coils have movable tuning slugs to permit frequency adjustment for alignment purposes. As in vacuum-tube video i-f circuits, stagger-tuning of the interstage coupling circuits is done, although in transistor circuits the amount of stagger-tuning is much less than in vacuum-tube circuits. The reason for this difference is the much lower impedance presented by transistors. With this lower impedance shunting the tuning circuits, a wider bandspread is obtained. Hence, it is not necessary to offset the

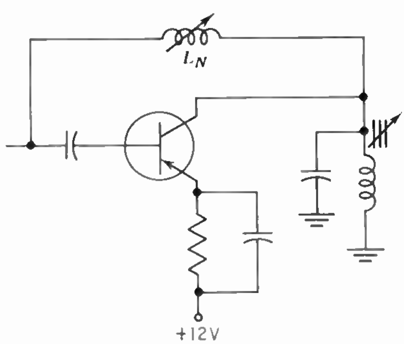


Fig. 7·9 Neutralization of an i-f amplifier by using a neutralizing inductance L_N.

resonant frequencies of the various tuning coils as much to obtain the desired bandspread required by the video signal.

Neutralization may also be achieved by using an inductance in place of C_N of Fig. 7·8, as in Fig. 7·9. If this method is employed, care must be exercised in the placing of parts to minimize any mutual coupling between the tuned circuits and the neutralizing inductance.

A three-stage video i-f amplifier system operating at 45 Mc is shown in Fig. 7·10. The input to the system from the preceding r-f mixer occurs through a short length of coaxial cable which links transformer T_1 to transformer T_2. A 41.25-Mc shunt trap is inserted in this path to reduce the amplitude of the sound portion of the received signal to approximately 5 to 10 per cent of the peak value of the video signal, Fig. 7·11. The 47.25-Mc trap reduces interference that might arise from the sound carrier of the next lower adjacent channel. At the same time, this trap also provides the proper slope for the high-frequency end of the video i-f response curve, Fig. 7·11. (By the same token, the 41.25-Mc trap helps to fashion the low-frequency end

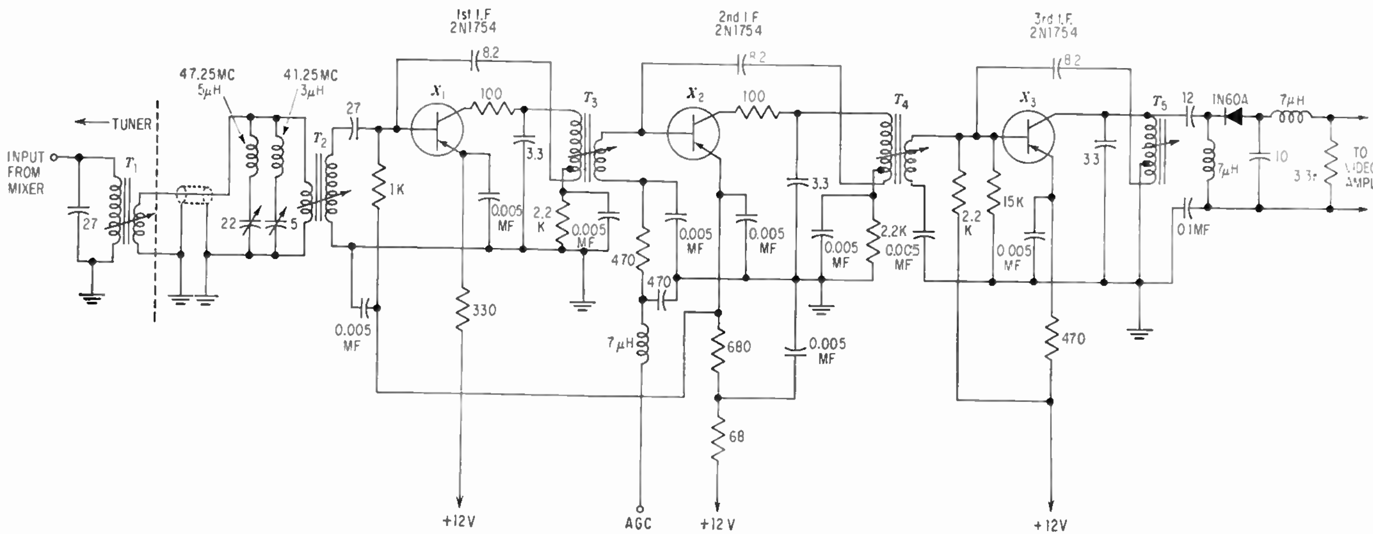


Fig. 7·10 A three-stage videa i-f amplifier system operating at 45 Mc.

of this same video i-f response characteristic.) Overall bandwidth is on the order of 3.5 to 4.0 Mc.

The stages are coupled together by bifilar transformers, T_2, T_3, T_4, and T_5. These transformers are peaked at slightly different frequencies to provide the necessary bandpass. The amount of detuning, however, is quite small. The primary of each unit has a high impedance (10,000 ohms) to match the relatively high collector impedance of the transistor feeding this transformer. The secondary of each transformer has a low impedance to match the low impedance of the following base input circuit.

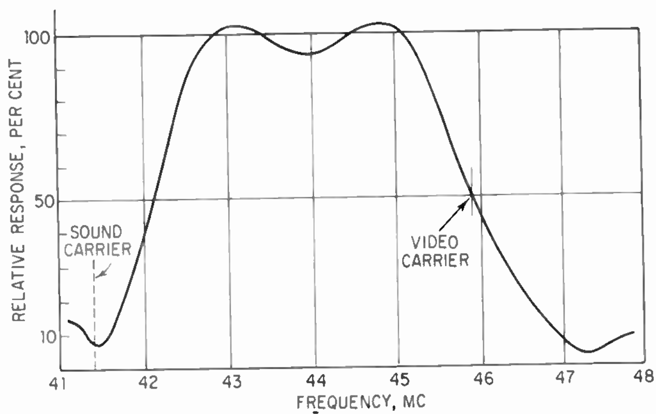


Fig. 7·11 Overall response curve of the i-f system of Fig. 7·10.

Each secondary winding has an extension containing a few turns to provide the necessary out-of-phase neutralizing voltage. An 8.2-$\mu\mu$f capacitor returns this voltage to the base of the stage.

Each of the transistors possesses a bypassed emitter resistor to provide temperature stability. In the first stage, the emitter resistor is 330 ohms; in the second stage, it is 680 ohms; and in the third stage, it is 470 ohms. In each instance a 0.005-μf capacitor provides the bypass action. In the second stage, the additional 68-ohm resistor is designed to decouple this stage from the third stage. This might occur through the common 12-volt power supply.

Forward agc is applied to the base of X_2. With an increase in signal, an increasing negative voltage is brought by the agc network to the base of X_2. This increases the current through this transistor, resulting in a lowered collector voltage because of the increased voltage drop across the 2,200-ohm resistor in the collector circuit. This decrease in collector voltage results in less stage gain, thereby tending to counteract the signal rise.

The first stage is also controlled by the agc voltage because the base of X_1 connects to the emitter of X_2 through a 1,000-ohm resistor. An increase in current through X_2 produces a rising negative voltage at the emitter, and this, fed to the base of X_1, likewise causes the current X_1 to rise. (Both X_1 and X_2 are PNP transistors.) Since the collector circuit of X_1 also has a 2,200-ohm resistor, it will also suffer a reduction in collector voltage and a subsequent drop in gain. Thus, both stages are directly controlled by the agc line. Stage 3 has fixed bias only and is not involved in the agc action.

The collector circuits of the first two stages also have 100-ohm series resistors in them. These are inserted to prevent the appearance of spurious oscillations that are sometimes found to develop under strong-signal conditions.

The output of the final amplifier X_3 is coupled to the diode detector without any special impedance-matching network because the two impedances are approximately equal. The video detector, in turn, is d-c-coupled to the video amplifier that follows it. Because of this direct connection, it is not possible to directly ground the bottom ends of the detector network components. An a-c ground is provided by the 0.1-μf capacitor. The diode load is formed by the 3,300-ohm resistor and the 10-$\mu\mu$f capacitor. The 7-μh series choke resonates at 45 Mc and serves to keep the i-f signal out of the video amplifier system.

Overall gain of this three-stage system is about 60 to 70 db. With the agc circuit, the gain can be reduced all the way down to 0 db.

One final point: the emitter of each stage is tied into the 12-volt power line and the base must also be given a voltage which is somewhat comparable, since the two element voltages must be very close together. This is done in the first two stages by injecting 10.5 volts d-c onto the agc line (not shown in Fig. 7·10). This voltage is then made to vary by the agc network from 10.5 to 9.0 volts, at which point the overall system gain is 0 db. In the third stage, the base return line connects to the 12-volt supply.

Since the emitters connect to the positive 12-volt line, each of the collectors can have its circuit returned to ground. This makes each collector properly negative with respect to its emitter (and base), which is the desired conditon for a PNP transistor.

The foregoing is a typical video i-f system. Other variations may be developed, but the basic overall operation will remain the same.

Video Detector and Amplifiers

The video second detector in most vacuum-tube television receivers is currently formed by a germanium diode; hence, transistorization is not required here. A typical circuit arrangement is shown in the i-f system of Fig. 7·10, and this is seen to be similar to video detectors in vacuum-tube television receivers.

The video-frequency amplifiers following the second detector must be capable of amplifying a band of frequencies extending from about 30 cycles to 4 Mc. Actually, with present monochrome receivers, the

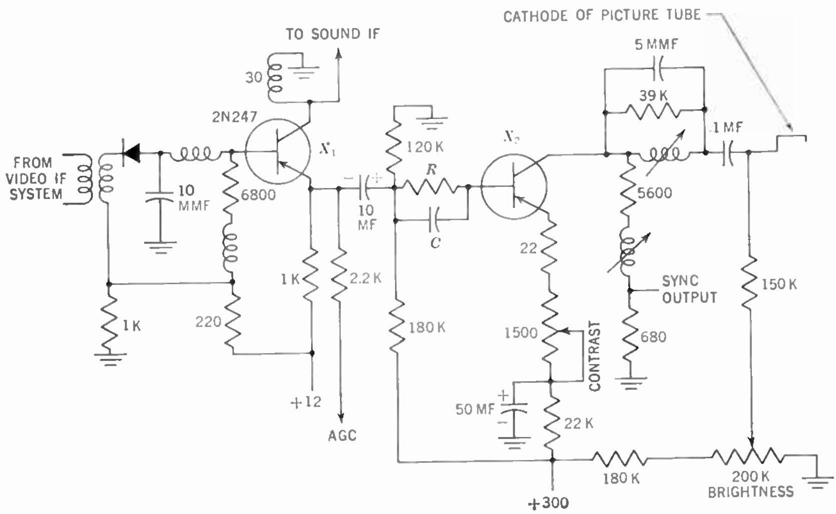


Fig. 7·12 A two-transistor video-amplifier system. (After M. C. Kidd, Transistor Receiver Video Amplifiers, *RCA Review*, September, 1957)

gain begins to fall off at about 3.2 Mc. A suitable two-stage video amplifier using high-frequency transistors is shown in Fig. 7·12.

The circuit uses a common-collector to common-emitter combination with two supply voltages, +12 and +300 volts. A common collector was selected for the first stage so that it would present a high enough input impedance to permit a fairly high video-detector load resistor. This assures a high detector efficiency and a high detector output voltage. A common-emitter or a common-base amplifier, in the same position, would present a low input impedance, and a 6,800-ohm load resistor, such as that being used in Fig. 7·12, could not be employed effectively. What the video detector would see would be the lower input impedance of the common-emitter or common-base stage and not the 6,800-ohm load resistor.

The bias for the first amplifier stage is provided by the 220- and 1,000-ohm network. These two resistors are connected in such a way that the biasing voltage *does not* bias the diode detector. This precaution is required here because of the d-c coupling between the diode and the first video amplifier.

A 4.5-Mc sound takeoff coil is inserted in the collector circuit of X_1. Whatever signal develops here is transferred to the sound i-f system. Also present in the first stage is the takeoff for the agc circuit. This is a good place to make the tap because the full video signal, including the d-c component, is available. The latter is needed for some agc systems, such as keyed automatic gain control.

The second video amplifier, X_2, operates from a +300-volt supply, and for this reason a 10-μf capacitor is inserted between the emitter of X_1 and the base of X_2. Without this blocking capacitor, the bias conditions in the first stage would be completely upset. The 180,000- and the 120,000-ohm resistors establish the bias for X_2. Resistor R and capacitor C form a peaking circuit to improve the frequency response of the system. The same purpose is served also by the peaking coils in the output of the video detector and in the collector circuit of X_2.

Because a high bias voltage is applied to X_2, it is possible to employ a 22,000-ohm resistor in the emitter circuit. This high a resistor, in this section of the transistor circuit, provides the stage with a high degree of stability and makes it relatively insensitive to temperature variations. The resistor is adequately bypassed by a 50-μf capacitor, so that no signal degeneration is introduced. The contrast control, on the other hand, which is also in the emitter leg of X_2, does operate by varying the amount of signal degeneration it introduces. The video output will be lowest when the contrast control is fully in the circuit.

Sync-pulse output is obtained from a 680-ohm resistor in series with the load circuit for X_2. Also present between the collector of X_2 and the cathode of the picture tube is a 4.5-Mc trap circuit.

A peak-to-peak signal of the order of 10 volts is supplied to the cathode of the picture tube. The cathode is chosen because it requires 15 to 20 per cent less drive than the control grid of the picture tube. When the cathode is driven, the video signal is in the direction to cancel the screen voltage and thus increase the effective output, since the drive required is proportional to the screen voltage.

Curves showing the frequency response of this two-stage amplifier are given in Fig. 7·13. Overall voltage gain is 30 db for a bandwidth of 3.5 Mc. Note that the bandwidth changes slightly with contrast control setting, although the change is negligible so far as the actual viewing image is concerned.

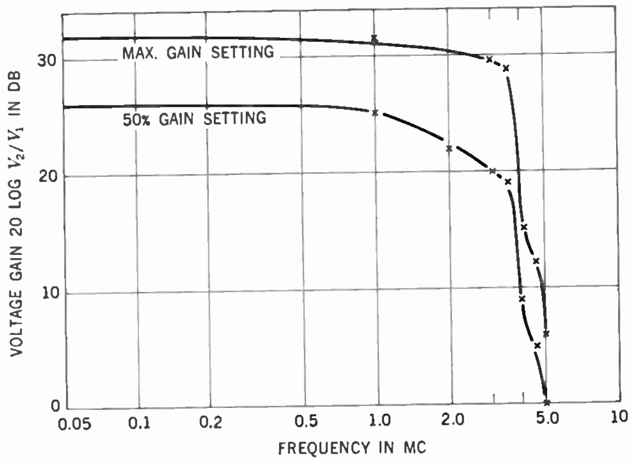


Fig. 7·13 Frequency response of the two-transistor video amplifier of Fig. 7·12. (RCA Review)

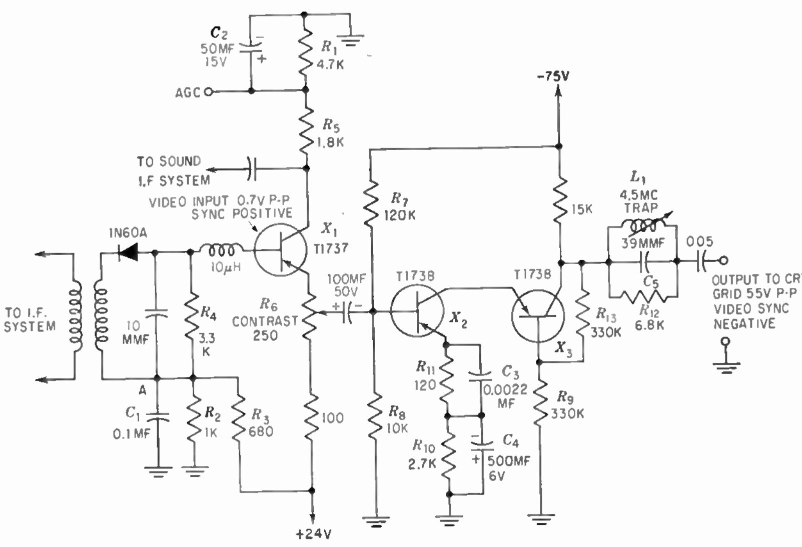


Fig. 7·14 A typical voltage-doubler video output circuit and driver.

Voltage-doubler video system. Still another approach to a video system which can produce a large output swing is shown in Fig. 7·14. The incoming video signal is fed to the 1N60A germanium diode where it is rectified and then transferred to the base of the first video amplifier, X_1. Since a d-c path is employed between the detector and X_1, the video stage receives the full video signal, including the d-c component. This is useful in that X_1 can also act as an agc amplifier.

The variations in this d-c component, due to fluctuations in signal strength, will vary the voltage across the 4,700-ohm resistor in the collector circuit of X_1 between 2 and 3 volts, and this is sufficient to adequately control the gain of the video i-f system. X_1 is operated from a +24-volt supply, producing a +10 volts at the agc takeoff point at R_1. This is the reference voltage required by the i-f system of Fig. 7·10. Voltage variations across R_1 due to changes in the d-c component of the demodulated video signal then produce the necessary agc action.

Note that R_1 is shunted by a 50-μf capacitor which prevents video signal variations from affecting the voltage developed here. This, of course, is necessary since the agc voltage should not be affected by the video content of the signal.

X_1 is connected as an emitter follower in order not to load down the video detector. Base bias for X_1 is established principally by R_2 and R_3, although video load resistor R_4 is also in this bias path. C_1 is needed to provide an a-c ground for the detector circuit. A d-c ground is not feasible at point A because of the direct connection between the detector and X_1.

A signal for the sound i-f system is taken from R_5. The signal here is the full video signal as well as the sound signal. However, a subsequent 4.5-Mc resonant circuit, not shown, removes all but the sound intermediate frequency and this is then passed on to the sound i-f amplifiers.

Video-signal drive voltage for X_2 is obtained from the contrast control R_6. X_2 is operated as a conventional common-emitter amplifier, except that its collector is directly attached to the emitter of X_3, a common-base amplifier. This latter stage functions as a voltage doubler, multiplying by 2 whatever voltage it receives from X_2. Using a common-base amplifier at this point provides several advantages. First, it will provide a wide bandwidth, reducing the need for peaking circuits. Second, the high output impedance of the common-base amplifier enables it to better provide a sizable voltage swing to the high input impedance of the cathode-ray picture tube. And finally, this arrangement can better withstand a greater output-voltage swing than a common-emitter amplifier can.

Proper bias for X_2 is provided by R_7 and R_8. The bias for X_3 is set by X_2 and R_{13} and R_9. Resistor R_{10}, heavily bypassed, provides thermal stability. Resistor R_{11}, however, has a very low valued capacitor C_3 across it, and this arrangement provides peaking to help maintain the frequency response of the system. At the low frequencies, C_3 offers a high resistance, increasing the effective signal impedance in the emitter circuit and thereby lowering the current through X_2 and its gain. As the frequency rises, the effect of C_3 is to lower the emitter impedance, permitting the current through X_2, and with it the gain of the stage, to rise. The net result is an improvement of the high-frequency gain of the system.

L_1, C_5, and R_{12} form a 4.5-Mc trap to prevent any sound signal from reaching the picture tube and producing a visible sound beat. The overall frequency response of the amplifier in Fig. 7·14, from the second detector to the output, is shown in Fig. 7·15.

The Sound Section

The sound system of a television receiver functions initially at 4.5 Mc and, beyond the FM detector, at ordinary audio frequencies.

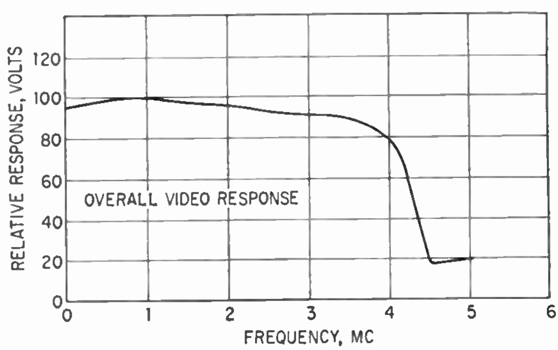


Fig. 7·15 Overall response of the amplifier in Fig. 7·14.

Neither the i-f nor the audio stage offers any particularly difficult problems other than, in the case of the i-f system, that of obtaining transistors with a suitably high α cutoff frequency. Typical stages for both sections were discussed in preceding chapters.

Either the FM detector can be transistorized, as shown in Fig. 7·16, or a pair of matched germanium diodes can be employed in one of the arrangements shown in Fig. 7·17. The latter two circuits are quite familiar by now, being direct germanium-diode equivalents of vacuum-tube Foster-Seeley and ratio detectors. One of the important features here is the use of closely matched diodes; the greater their differences, the less effective the circuit in minimizing distortion and combating amplitude modulation.

A particular point of interest is the use of low-valued load resistors in both circuits in place of the fairly high values employed in the vacuum-tube versions. This change is necessary in order for the de-

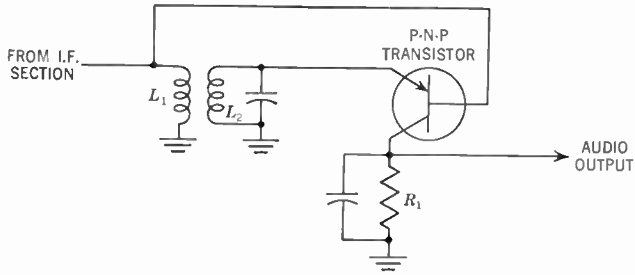


Fig. 7·16 A transistor FM detector.

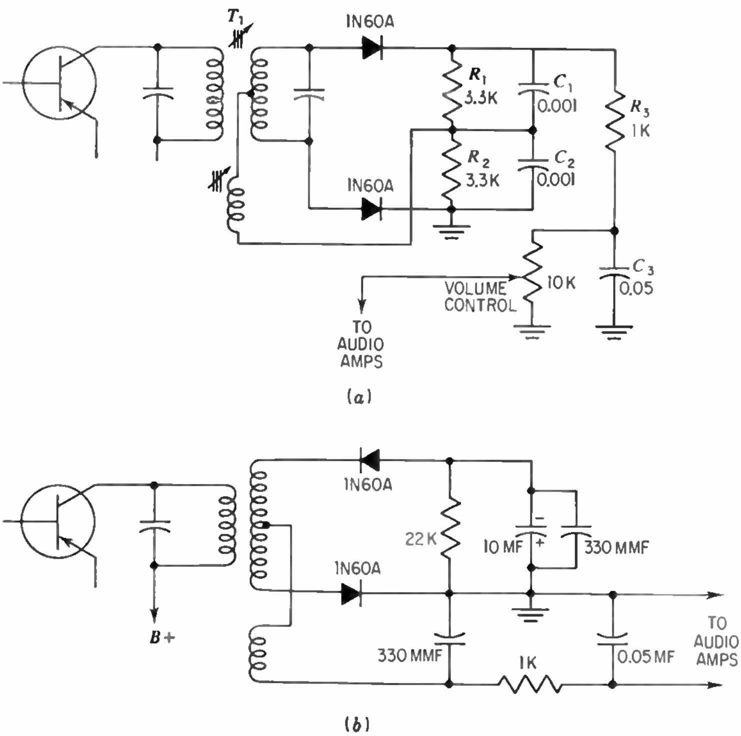


Fig. 7·17 Two types of FM detector circuits. (a) Discriminator. (b) Ratio detector.

tector output impedance to match the low input impedance of the audio amplifier that follows.

It is possible to design a transistor FM detector using a symmetrical transistor. Briefly, this is a unit in which the emitter and collector sections are made identical so that, with the proper biasing voltage, either section could operate as the emitter or collector. (We shall refer to this type of transistor again in connection with a horizontal phase detector.)

The circuit of this FM detector is shown in Fig. 7·16. The FM signal appearing across L_1 is applied to the base, while the voltage developed across L_2 is applied to the emitter. During the positive portion of the signal applied to the base, the emitter-collector path is open and there is no current flow through the load resistor R_1. During each negative swing, current does flow. We are discussing here a PNP transistor; for an NPN unit, the periods of conduction would be reversed.

Now, the amplitude and direction of the current flow depend upon the phase relationship of the signal developed across the secondary with respect to the primary signal. At the resonant frequency of L_2, the voltage it develops is 90° out of phase with the voltage across the primary. During this condition, the average voltage drop across R_1 will be zero. As the applied frequency is changed, the secondary voltage lags the primary voltage by an angle less than 90° if the frequency rises, or it will lag by more than 90° if the frequency drops below the resonant (or mid) frequency of L_2. As the phase relationship changes, the voltage developed across the load resistor will vary in step with the frequency modulation. (A full discussion relating phase changes to FM detection will be found in Milton S. Kiver, "F-M Simplified," 3d ed., D. Van Nostrand Company, Inc., Princeton, New Jersey, 1960.)

Changes in amplitude of the incoming signal will not affect the output as long as the signal amplitude is strong enough to operate the transistor beyond the knee of its charactertistic curve, Fig. 7·18. Once past this region, the collector current remains fairly constant with changes in collector voltage. Thus, because of its characteristics, the transistor will function as a limiter, too.

Sync Separators

Returning to the video system, a portion of the signal is taken from one of the video amplifiers and applied to the sync section. Here the vertical and horizontal sync pulses must be separated from the rest of the video signal. The latter voltages are then suppressed, while the sync pulses are passed on to their respective deflection systems. It is also desirable during this separation process to suppress or at least reduce the effect of any noise pulses that may be present.

In adapting transistors for sync separation, advantage can be taken of the fact (just noted) that beyond the knee of the characteristic curves, the collector current changes very little with change in collector voltage. Thus, if we drive a sync separator amplifier from cutoff to saturation, a double-clipped output voltage can be obtained possessing an amplitude that is only a few tenths of a volt less than the collector supply voltage. This is useful not only in securing a flat-topped output pulse but also in clipping any noise spikes that may be present at the sync-pulse level.

A two-stage sync separator is shown in Fig. 7·19. The first transistor is of the PNP variety, while the second is an NPN unit. Both transis-

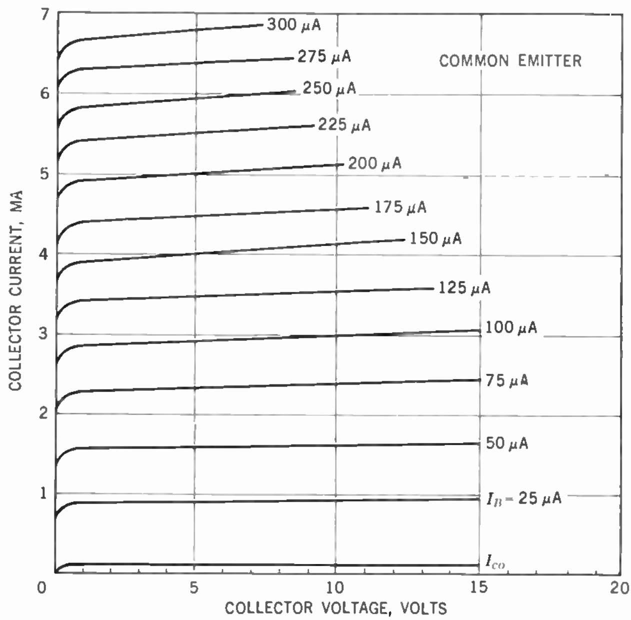


Fig. 7·18 Characteristic curves of a typical transistor. Note that collector current remains fairly constant with changes in collector voltage beyond the knee of each curve. This behavior is utilized in sync separators and limiters.

tors should preferably have α cutoff frequencies in excess of 3 Mc so that the steep sides of the vertical and horizontal sync pulses will be reproduced. (A low-frequency transistor would tend to slow the rate of voltage rise and change the steep sides to sloping sides.) The video input signal to the sync separator should be in the sync-pulse negative phase and should come from a low-impedance source. It is desirable to have the first sync separator stage conduct only while the sync pulses are active and to cut off or become nonconductive in the interval between sync pulses. This is achieved in X_1 of Fig. 7·19 through the combination of R_1 and C_1. When the sync pulse arrives, it causes the transistor to conduct, with emitter current flowing through R_1 in direction indicated by the arrow. This surge of current develops sufficient bias across the R_1C_1 combination so that at the end of the pulse interval, the base-emitter junction is reverse-biased and all current flow through the transistor is halted until the arrival of the next sync pulse.

Note that this is a self-biasing arrangement in which the emitter current will vary with signal amplitude, producing corresponding voltage variations across R_1 and C_1. In this respect it is similar to the cathode-biased vacuum-tube separator frequently used. Self-biasing is desirable here because it enables the operating condition of the stage to change in step with the level of the incoming signal. The reader will appreciate that such variations exist even in receivers employing automatic gain control.

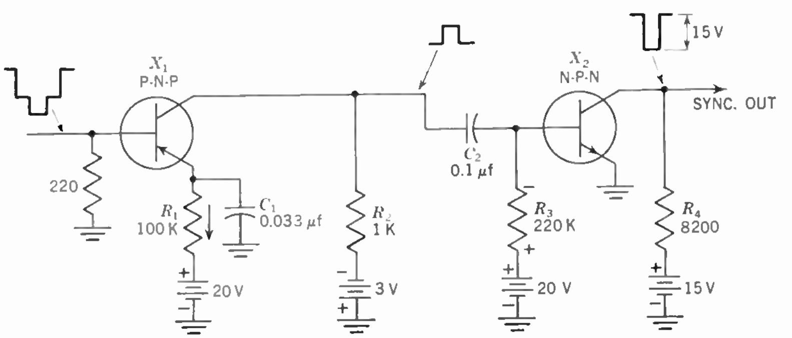


Fig. 7·19 A two-stage sync separator. (After H. C. Goodrich, Transistorized Sync Separator Circuits for Television Circuits, *RCA Review*, December, 1955)

The amplified sync signal appears at the collector of X_1 and is passed on to X_2. This second stage is so operated that it is driven into saturation during sync by any usable signal. This provides sync pulses of uniform amplitude and cuts off any noise pulses at sync level.

Strong overdriving of a sync amplifier must be avoided because it leads to an output pulse which is wider than the input pulse. (This effect is explained in Chap. 8.) This broadening will cause a phase shift in many types of horizontal phase detectors and is therefore undesirable. To prevent this overdriving, self-bias is employed on transistor X_2. This is achieved through the combination of C_2 and R_3. When the incoming signal tends to drive X_2 far beyond saturation, the base-emitter circuit develops a voltage across R_3 (with the polarity indicated) which reduces the extent of the overdriving.

The sync pulse developed at the collector of X_2 has a peak-to-peak amplitude of 15 volts and is negatively phased. If the opposite polarity is desired, a phase inverter may be employed.

Advantage can be taken of the sync clipping capabilities of a transistor driven from cutoff to saturation to develop a useful one-stage sync separator, Fig. 7·20. R_1 and C_1 again form a self-biasing arrangement to prevent pulse broadening while at the same time providing clipping action at all levels of input signal, from the very weak to the very strong. It will be appreciated, however, that because a single

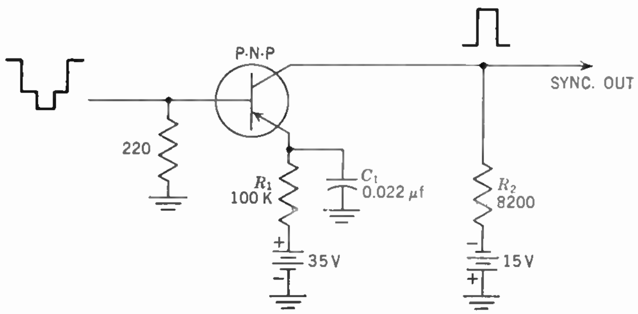


Fig. 7·20 A one-stage sync separator. (*RCA Review*)

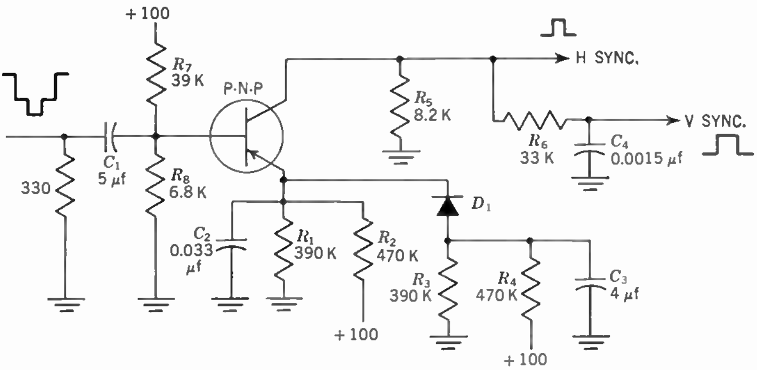


Fig. 7·21 A one-stage sync separator with a dual-time-constant network. (*RCA Review*)

stage is being employed, the preceding amplifier must be capable of supplying a greater video input current for proper separation.

A more elaborate version of the foregoing sync separator is shown in Fig. 7·21. This circuit not only produces flat-topped sync pulses but also provides a measure of immunity against blocking caused by strong noise pulses. It accomplishes the latter goal by means of a dual time-constant network in the emitter circuit, coupled with a special germanium diode D_1.

The manner in which this dual network functions is as follows: Upon the arrival of a sync pulse, the emitter current of the transistor will bias D_1 in the forward direction, causing it, in effect, to become a closed switch. This will bring the long-time-constant network of R_3, R_4, and C_3 into the circuit and permit the effective separation of the vertical sync pulses. (A long separation time constant is needed to accomplish this separation.) Now, if a strong noise pulse should come along, capacitor C_3 will charge to its peak value. When the noise passes, the weaker normal sync pulses will be unable to provide D_1 with enough countervoltage to cause it to conduct, forcing D_1 to remain open and removing C_3 with its excess charge from the circuit. Horizontal sync-pulse separation will now be achieved by using the bias developed across the short-time-constant circuit of R_1, R_2, and C_2. When the excess charge on C_3 has drained off, the long-time-constant network will reestablish itself actively in the circuit. By means of this arrangement, the horizontal noise immunity of the sync separation is improved by an average factor of 8:1 over systems not employing a double-time-constant circuit.

Note: The excess charge on C_3 drains off long before the next vertical sync pulse arrives. It is permissible to use a long-time-constant network to assist in the separation of the horizontal and vertical sync pulses *as long as there are no strong noise pulses present.* If noise pulses are present, they could easily block or inactivate the sync separator for many horizontal sync pulses, permitting the horizontal sweep oscillator to slip out of synchronization. In the above arrangement, this is avoided by the use of D_1 and the alternate short-time-constant circuit.

Vertical-deflection System

Vertical oscillator. Beyond the sync separators, the pulses are fed to the vertical- and horizontal-deflection sections. Because transistorized vertical-sweep circuitry is simpler and has progressed farther, let us examine it first, starting with the oscillator. In vacuum-tube circuits, either multivibrators or blocking oscillators have been utilized. In transistorized systems, at least to the present time, only blocking oscillators have been used because transistorized multivibrators have thus far exhibited frequency changes due to temperature variations, whereas the blocking oscillator is considerably more stable in this respect.

In many ways the blocking oscillator in Fig. 7·22 is similar to its vacuum-tube counterpart. For example, the blocking transformer T_1 serves to provide feedback from the collector output circuit to the base input circuit. Also, sync pulses are fed to the base while a sawtooth deflection wave is developed across the 8-μf capacitor from collector to ground. Furthermore, charge capacitor C_1 receives its voltage buildup while the transistor is cut off and then discharges when the transistor is pulsed into conduction. There are, however, some significant differences between the two circuits, and these stem from the dissimilarity in operational characteristics of tubes and transistors. The differences will be evident from the following analysis of circuit operation.

The transistor is biased beyond cutoff by the negative voltage present across R_1. The value of this cutoff voltage is determined by the resistance setting of R_2; hence, R_2 is equivalent to the conventional

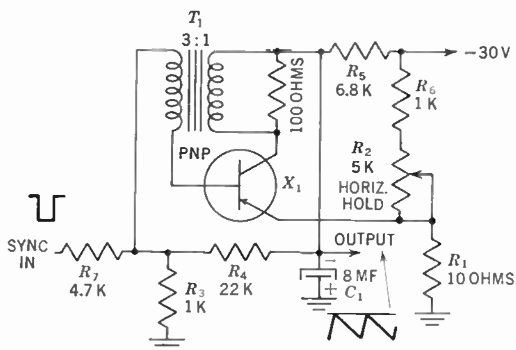


Fig. 7·22 A transistor blocking oscillator. (After M. B. Finkelstein, "Transistor Vertical Deflection for Television Receivers," Transistors 1, published by RCA)

hold control. At the same time, capacitor C_1 charges through R_5 until the voltage across R_3 becomes more negative than the voltage at the emitter. When this occurs, current will flow through the transistor, and the conductive part of the cycle will begin. The current flowing through the collector circuit causes a voltage to be induced in the base side of T_1, increasing the forward bias and the transistor current. This, in turn, couples a greater voltage into the base winding and increases the collector current even more. As a result, the transistor very rapidly becomes a virtual short circuit, permitting capacitor C_1 to discharge quickly through T_1 and R_1. The discharge is practically complete, and the voltage across C_1 essentially drops to zero.

At this point the voltage across R_3 and R_4 is also zero, and the transistor is driven into cutoff by the difference between the emitter voltage at the arm of R_2 and the voltage induced at the base by the collapsing field around T_1. The sequence of events now repeats itself at a frequency determined by the various resistance and capacitance values in the circuit.

The major frequency-determining components are C_1, R_5, and R_2; C_1 and R_5 also develop the output wave. Here is a major difference between transistor and vacuum-tube blocking oscillators. In the latter circuits, the frequency-determining parts are separate and distinct from the wave-forming circuit; in the transistor circuit, one set of components performs both functions.

The sawtooth output waveform produced by the circuit of Fig. 7·22 is negative; if a positive-going signal is desired, an NPN transistor would be employed in conjunction with a positive voltage source. Obtaining reversal in this way is another unique feature of transistors and cannot be duplicated by tubes.

Another sawtooth blocking oscillator, in which the output wave is developed in the emitter leg, is shown in Fig. 7·23. This circuit is sim-

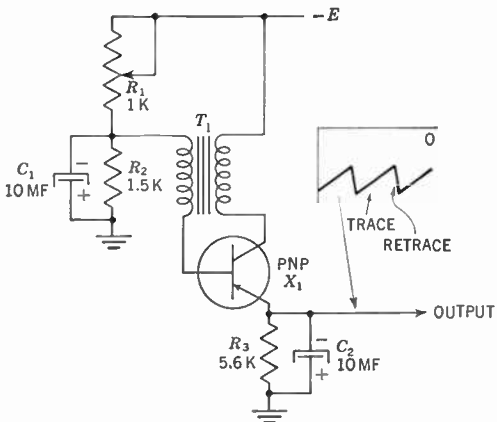


Fig. 7·23 Development of the sawtooth wave in this circuit differs from that in Fig. 7·22.

ilar to the preceding arrangement in that one *RC* combination (here, R_3 and C_2) provides both the timing and the output deflection waveform. It differs, however, in the way in which the sawtooth wave is produced. It has become customary in vacuum-tube circuits to have the charge time of the deflection capacitor correspond with beam trace and the discharge period with beam retrace. In the arrangement of Fig. 7·23, C_2 charges through the transistor during the retrace period and discharges through R_3 during trace time. How this is accomplished can be seen from the following discussion.

The base-emitter circuit is biased in the forward direction by R_1 and R_2. This produces a current flow through the transistor, which starts charging capacitor C_2. The collector current flowing through the primary of T_1 induces a voltage in the base winding which acts to increase the base current. This serves to further raise the collector current and the induced base voltage until the transistor is conducting to its fullest extent. This current buildup through the transistor is exceedingly rapid, and during this interval capacitor C_2 is charging through the resistance of the transformer primary and the transistor. The ultimate value of the charge is determined by the fixed base bias and the induced voltage in the transformer secondary winding.

When this point is reached, the base current ceases and the transistor is driven into cutoff because the voltage across C_2 is sufficient to bias the base-emitter circuit in the reverse direction. While the transistor is cut off, C_2 discharges slowly through R_3, thereby developing the trace portion of the deflection wave. Note that this wave gradually rises toward zero from its initial high negative value.

When C_2 has discharged to the point where its voltage is equal to the fixed bias voltage, current again starts flowing in the base-emitter circuit, and the same sequence of events recurs. Since R_1 determines the value of fixed base bias, it determines the frequency of operation (together with the time constant of R_3 and C_2) and hence would function as a hold control.

A negative sync pulse could be introduced to the base circuit through either the use of capacitive coupling or a third winding on the blocking transformer. Both methods have been employed and both are satisfactory.

A positive-going sawtooth wave is produced by the above circuit. If a negative-going sawtooth is desired, an NPN transistor could be used with a positive power supply.

All of the circuits discussed above are practical designs which have been employed for the purposes indicated, and each has performed in an entirely satisfactory manner.

Complete vertical-deflection systems. In a vacuum-tube deflection circuit, the vertical oscillator is followed by the output amplifier. In order to develop the necessary current swing, this stage requires a certain amount of driving voltage. In transistor circuitry, a similar two-stage arrangement requires that a power transistor be used in the oscillator stage because of the power needed by the output amplifier in delivering sufficient sweep current to the vertical-deflection yoke.

A suitable two-stage vertical-deflection system for a 21-in., 90° tube is shown in Fig. 7·24. The oscillator, using a power transistor, develops a negative-going sawtooth wave (across C_1) which is applied to the base of the output amplifier. (Note that this is a forward-biasing wave because it is being fed to a PNP transistor.) Frequency control of the oscillator is achieved by varying its base-bias resistor R_1, and synchronization is accomplished by coupling a negative triggering pulse into the base via a tertiary winding on the pulse transformer.

The waveform of the voltage fed to X_2 is sawtooth, as shown in Fig. 7·25*a*. This differs from conventional vacuum-tube practice, in which the input wave to the output amplifier is usually peaked. The

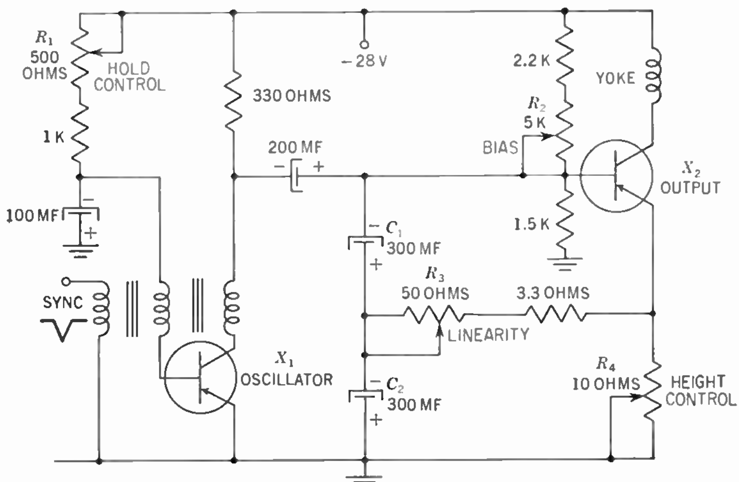


Fig. 7·24 A two-stage transformerless vertical-deflection system. (After W. Palmer and G. Schiess, "Transistorized Television Vertical Deflection System," Sylvania Electric Products, Inc., 1958)

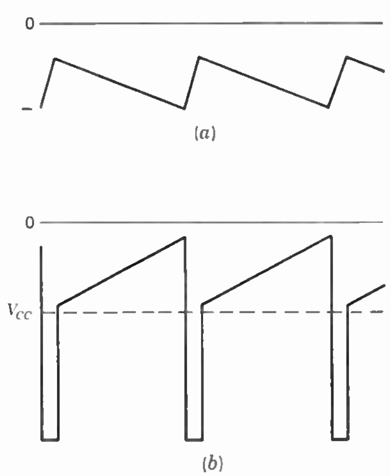


Fig. 7·25 (a) Sawtooth voltage applied to base of X_2, Fig. 7·24. (b) Yoke-voltage wave. V_{cc} is battery voltage.

reason for the difference is that, comparatively, collector resistance of a common-emitter transistor stage is much higher than the plate resistance of a triode tube. In a triode vertical output circuit, both plate resistance and plate-load inductance must be considered; i.e., to obtain sawtooth deflection current, a sawtooth voltage must be applied across the resistive portion and a square-wave voltage must be applied across the inductive portion. If, however, the ohmic value of the resistance is at least 10 times that of the inductance, a sawtooth driving voltage must be used.

Returning to Fig. 7·24, we find that C_2 and R_3 form a linearity-correction network. The sawtooth current flowing through emitter resistor R_4 is integrated into a parabola across C_2 and fed back to the base through capacitor C_1. The special correction is needed to compensate for the nonlinear characteristics of transistor X_2. Adjustment of linearity control R_3 affects the amplitude of the parabola and hence the degree of linearity compensation. It will be found that the linearity control also affects the amplitude of the yoke current; hence, readjustment of the amplitude (height) control R_4 is necessary with each linearity adjustment. Finally, a bias control R_2 is included to permit adjustment of the bias applied to X_2 for replacement transistors. This is needed because the current amplification (β) may vary considerably among different transistors of the same type.

The yoke itself is located in the collector leg of X_2. If we apply a sawtooth wave of the form shown in Fig. 7·25*a* to the base of X_2, the voltage developed across the yoke will be as shown in Fig. 7·25*b*. The sharp voltage spike developed when the transistor is brought to cutoff is due to the inductive reactance of the yoke, and its amplitude is equal to di/dt. For a yoke inductance of 40 mh, a peak-to-peak yoke current of 450 ma, and a retrace interval of 350 μsec, the pulse generated will have an amplitude of about 52 volts. To this would be added the negative voltage of the collector battery (about 25 volts), for a total of 77 volts. The collector must be capable of withstanding this surge; otherwise, it will periodically break down, damping the pulse and increasing the retrace time of the beam.

Since the yoke is directly positioned in the collector arm of the output amplifier, it will have the d-c component of the collector current flowing through it (yoke current does not reverse in this circuit). This will cause vertical picture decentering to a considerable extent; so much so, in fact, that conventional centering methods are incapable of bringing the picture back to its proper place on the screen. There are several ways to solve this problem, but the most attractive one employs permanent magnets. Small ceramic bar magnets are placed longitudinally in the opening or window formed by the vertical yoke windings, Fig. 7·26*a*. An alternative method would be the placement of ring segments of the ceramic magnetic material between the yoke and the picture-tube neck, Fig. 7·26*b*. Both methods provide a strong enough magnetic field to recenter the picture on the screen.

Another vertical-deflection system, utilizing a driver stage between the oscillator and the output amplifier, is shown in Fig. 7·27. The oscillator circuit, previously discussed, develops the sawtooth wave in

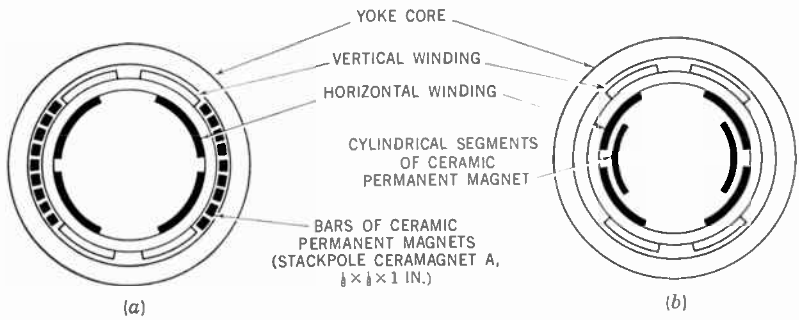


Fig. 7·26 Magnetic centering used with transformerless deflection amplifier.

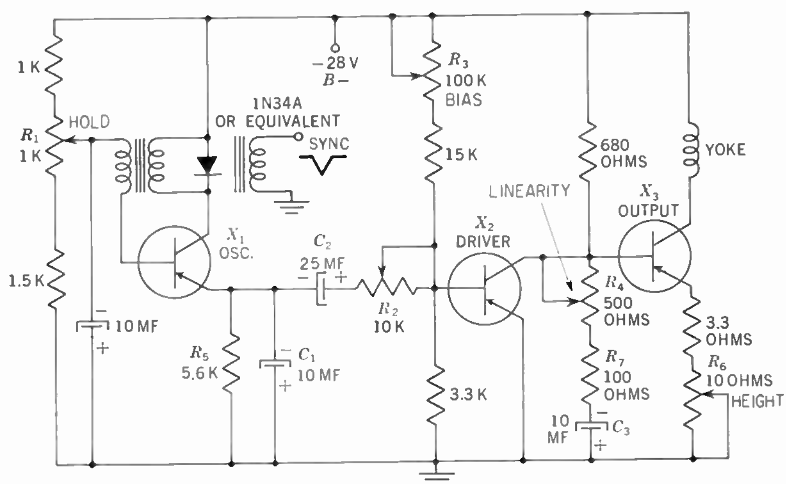


Fig. 7·27 A three-stage vertical-deflection system. (Sylvania Electric Products, Inc.)

the emitter leg of the transistor. R_5 and C_1 form the basic timing network, although R_1 can exert enough influence to function as the hold control.

Pulses from the sync-separator section are brought in via a third winding on the blocking oscillator transformer. Another winding in the collector circuit is shunted by a diode to remove the long narrow voltage spike that is developed by the inductance of the blocking transformer when the oscillator is cut off sharply. By removing the pulse, which serves no useful purpose, the chances for collector breakdown are minimized.

The sawtooth wave developed by the blocking oscillator is transferred via C_2 and R_2 to a common-emitter driver. R_2 is made variable to permit selection of the proper current drive for the low-impedance driver stage (a-c input impedance is on the order of 300 ohms). R_3 is a bias control for X_2 and is incorporated to establish the correct operating conditions for the driver transistor. Adjustment would be essential if X_2 were replaced.

The network comprised of R_4, R_7, and C_3 between the base of X_3 and ground is designed to improve the linearity of the signal. The three components function as an integrating network, adding a parabolic waveform of variable amplitude to the base signal of the output stage. The variable feature is provided by R_4, the linearity control.

The final stage, also a common-emitter arrangement, possesses still another amplitude control, R_6. Thus, this system has two such controls, providing a greater flexibility in the selection of replacement transistors. When the β values of the transistor type used can be kept within a fairly narrow range, one of the amplitude controls could probably be dispensed with.

In comparing the two- and the three-transistor deflection systems, the following comments by the designers of these systems may be enlightening:

Circuitry. The two-transistor circuit is simpler in construction, requiring fewer components. This advantage is offset somewhat by the fact that high capacitor values have to be used throughout because of the low impedance level of the circuit. Furthermore, in the two-transistor system, a higher-rated power transistor must be employed in the blocking oscillator stage.

Linearity. In the three-stage circuit, there is greater leeway in providing linearity compensation such as we found in Fig. 7·27. Also, with the driver stage and the amplification it provides, degeneration can be employed in the output amplifier.

Synchronization. The two-transistor circuit employing a power transistor in the oscillator was found to require more power for synchronization. This might necessitate the use of higher-power transistors in the sync-separator stage.

Frequency stability. Both circuits are comparable in performance.

Operational stability. It will be noted that the two-stage vertical deflection system utilizes a-c coupling throughout, while the three-stage circuit has d-c coupling between the driver and the vertical output amplifier. In general, a-c-coupled systems tend to be more stable than d-c systems, particularly in transistor circuits where temperature changes have such a marked effect on operation. Hence, the circuit of Fig. 7·24 would be more stable than the circuit of Fig. 7·27.

Several things can be done to increase the stability of Fig. 7·27. Direct-current feedback, from the emitter of the output stage to the base of the driver stage, would provide some improvement. Another approach would be the use of a reverse-biased diode between driver base and ground. The back resistance of the diode would decrease with temperature and thus reduce the base bias. Still a third solution

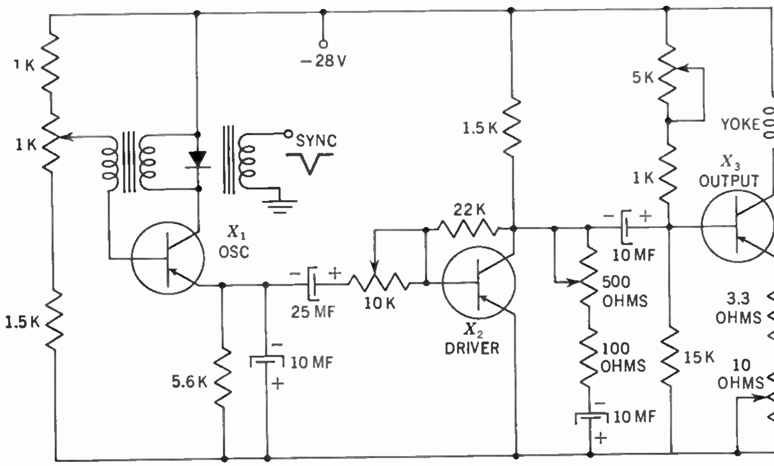


Fig. 7·28 A variation of Fig. 7·27 employing a-c coupling between all stages. (Sylvania Electric Products, Inc.)

would be the use of a-c coupling between the driver stage and the output stage. This is shown in Fig. 7·28.

Horizontal-deflection System

A second sync-pulse output from the sync separator stage is directed to the horizontal-deflection system. Let us now turn our attention to this section of the receiver.

Horizontal phase detectors. The susceptibility of the horizontal-sweep oscillator to noise pulses and other forms of interference has led to the universal use of automatic-frequency-control networks ahead of the horizontal oscillator. Whatever the form of the control system, its method of achieving control is by comparing the frequency of the generated sweep voltage with the frequency of the arriving horizontal sync pulses. If a frequency difference exists, there is developed a corrective voltage which, when fed back to the horizontal oscillator either directly or indirectly, causes the generated frequency to change until it is equal to that of the incoming pulses.

A widely employed afc circuit in vacuum-tube television receivers is shown in Fig. 7·29. D_1 and D_2 are shown as germanium diodes, although they can be vacuum-tube diodes as well. In a transistor television receiver, they would be germanium diodes, and that is the way the circuit will be shown here.

The two diodes are seen to be connected in series with each other at point A, and from this point, a resistor connects to ground. Coming into this network are two horizontal-sync pulses of opposite polarity (representing the received signal) and a sawtooth wave (representing

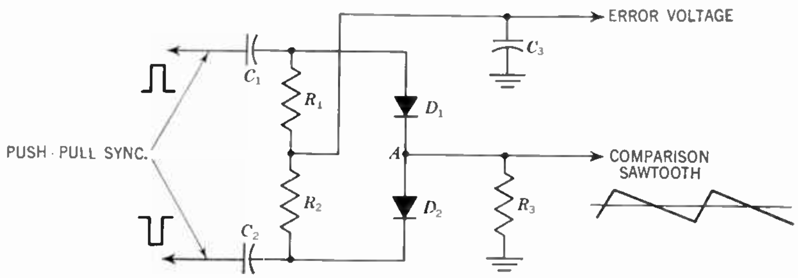


Fig. 7·29 A double-diode phase detector.

the generated deflection voltage). The latter signal is obtained from a point beyond the horizontal sweep oscillator so that it will reflect the frequency being generated by that stage.

As a first step, let us disregard the sawtooth voltage and observe the effect of the two sync pulses. The positive sync pulse is applied to D_1, and if we were concerned only with this sync pulse and the circuit of D_1, then the simplified circuit would appear as shown in Fig. 7·30. Application of the positive pulse causes current to flow from D_1 to the right-hand plate of C_1 and from the left-hand plate of C_1 through the signal source (i.e., a prior stage) to ground and then up through R_3 to D_1. The time constant of this circuit is low enough that C_1 charges to the peak value of the applied pulse. During the interval between pulses, C_1 discharges through R_1 and C_3, developing voltage drops across these two components with the polarity as indicated in Fig. 7·30. When the next pulse arrives, C_1 is recharged to the full peak value. The current flow through D_1 is thus in spurts which are generally shorter than the applied pulses themselves.

At the same time that this is happening, negative sync pulses are being applied to D_2 and causing current to flow through this diode. A simplified arrangement of this portion of the network is shown in Fig. 7·31. The current travels from D_2 down through R_3 to ground and from there to the signal source and C_2 and then to D_2. The polarity of the voltage drop across R_3 and C_2 caused by this current is indicated in Fig. 7·31. Note that the voltage drop across R_3 produced by the current from D_2 is opposite in polarity to the voltage drop developed across this same resistor by D_1. If, as is usual, both incoming sync pulses possess the same amplitude and both diodes conduct equally well, then the net resultant voltage across R_3 is zero.

During the interval between pulses, capacitor C_2 discharges through R_2 and C_3, developing voltage drops across these two components with the polarity indicated in Fig. 7·31. Again the net resultant voltage

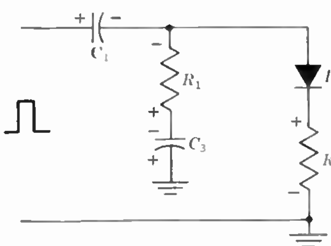


Fig. 7·30 The circuit of D_1 in Fig. 7·29.

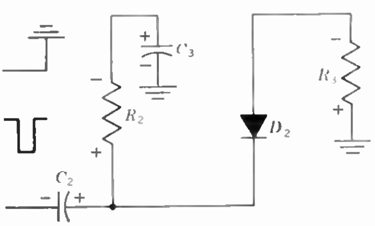


Fig. 7·31 The circuit of D_2 in Fig. 7·29.

across C_3, owing to the two discharge currents that flow through it, is zero. And since it is the voltage present across C_3 that represents the corrective or error voltage to the horizontal oscillator, then with the sync pulses acting by themselves, no net voltage is produced. This is as it should be.

By the same line of reasoning, if we ignore the sync pulses and concern ourselves solely with the sawtooth wave applied to R_3, then we see that since D_1 and D_2 will be driven alternately into conduction for equal periods of time and with equal-amplitude voltages, the net output voltage across C_3 will again be zero.

With both types of voltages applied to this circuit simultaneously, comparison of the two signals will take place only at the instant that the sync pulses arrive, for it is only at this moment that D_1 and D_2 conduct and are therefore in a position to respond to the sawtooth voltage applied across R_3. Three situations are possible.

First, if the sync pulses arrive at a time when the sawtooth wave is passing through zero, then we have a situation which is similar to that discussed above when the sawtooth voltage was ignored. The net voltage developed across C_3 is zero. This indicates that the frequencies of the sweep oscillator and the sync pulses are in step with each other.

The second situation occurs when the sync pulses arrive and the sawtooth voltage is positive at this instant. Under this condition, D_2 will conduct more strongly than it will if the sawtooth wave is zero and C_3 will charge to a higher peak value (because now two series-aiding voltages are driving current through D_2). At the same time, the positive sawtooth voltage is also being applied to D_1 and for this diode it is working against the applied sync pulse. Hence, the total current through D_1 will decrease and produce a smaller voltage drop

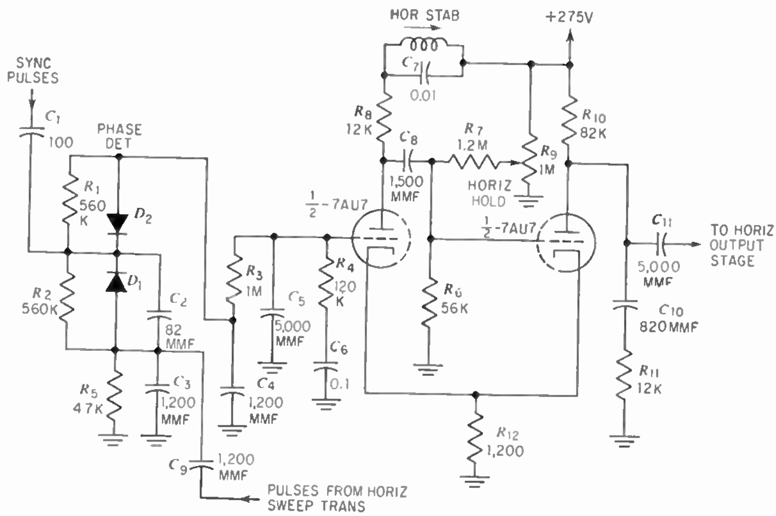


Fig. 7·32 A phase detector which requires only one set of input sync pulses.

across C_3. The net voltage across C_3 will be governed by the current from D_2 and will be positive with respect to ground. The horizontal sweep oscillator will thus receive a corrective voltage which, if the circuit is designed properly, will serve to alter its frequency so that the sawtooth voltage at R_3 will be passing through zero when the sync pulses arrive.

The third situation occurs when the sawtooth voltage is negative when the pulses arrive. Now D_1 conducts more strongly than D_2, and a net negative voltage will develop across C_3. This opposite-polarity voltage will have an opposite effect on the frequency of the horizontal sweep oscillator.

A two-diode circuit in which only one set of sync pulses is required is shown in Fig. 7·32. The two cathodes of the diodes (here, germanium diodes) are connected together, and a negative-going sync pulse is applied to their junction. This applies the sync pulse equally across D_1 and D_2, because C_3 and C_4 are so much greater than C_1 that D_1 and D_2 are effectively connected in parallel. This being the case, current will flow in each diode, causing equal currents to flow in the load resistors R_1 and R_2. The currents, of course, flow in opposite directions and the voltage drops across R_1 and R_2 will have opposing polarities and, therefore, will cancel out, producing zero volts output.

The sawtooth voltage, formed from pulses from the horizontal output stage, is a sample of the horizontal oscillator frequency. This voltage

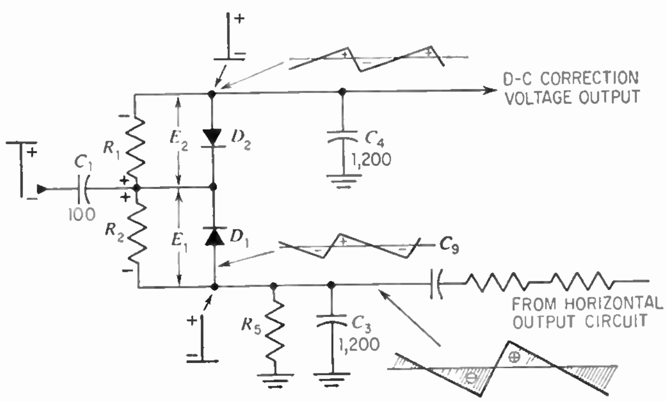


Fig. 7·33 A simplified diagram of the phase detector of Fig. 7·32 showing the waveforms in the circuit. (General Electric Co.)

is applied across D_1 and D_2, effectively bringing one-half of the original sawtooth wave across each diode. It can be shown that the sawtooth wave across D_1 will be going positive when the voltage across D_2 is going negative and vice versa, Fig. 7·33. The currents of the two diodes will be equal but opposite in polarity, so equal and opposite voltages across R_1 and R_2 will produce zero volts output.

From this it can be seen that the incoming sync pulses alone will not cause the phase detector to produce any voltage output. In like manner, the sawtooth wave alone will not cause the phase detector to produce any voltage output.

The sync pulse, being of much greater amplitude than the sawtooth wave, keeps the diodes biased so that they operate only when the sync pulse is applied to them. Therefore, only that portion of the sawtooth wave that occurs at the instant of the sync pulse has any effect on the output of the phase detector.

Now, if the sync pulse occurs in the exact center of the sawtooth retrace (i.e., retrace passing through its a-c axis), Fig. 7·34*a*, equal but opposite currents will flow and no output voltage will be developed.

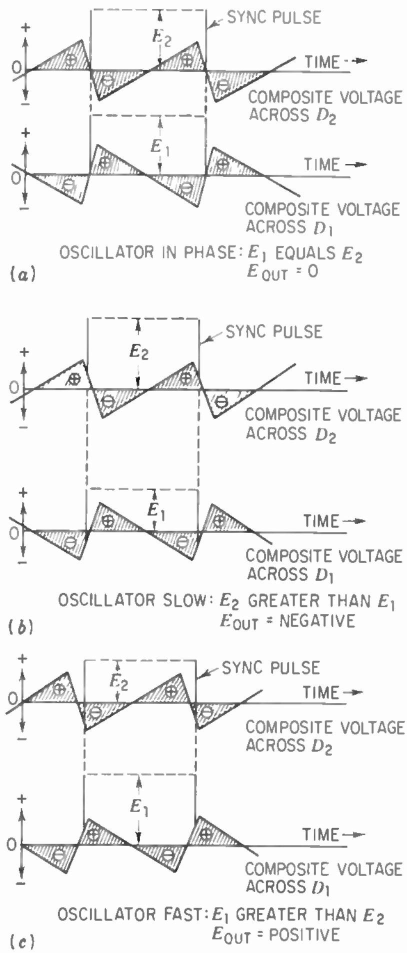


Fig. 7·34 Operation of the phase detector of Fig. 7·32 when the sync-pulse frequency and sawtooth frequency are (a) equal and (b) and (c) unequal. (General Electric Co.)

If the oscillator is slow, the sync pulse will occur before the sawtooth retrace passes through its a-c axis, Fig. 7·34*b*. On D_2, therefore, some of the sawtooth voltage will be added to the sync-pulse voltage because the sawtooth voltage is on the positive half of its cycle when the sync pulse occurs. Some of the sawtooth voltage on D_1 will be subtracted from the sync-pulse voltage because the sawtooth retrace there is still in the negative half of its cycle. The output voltage of the phase detector in this case will be negative because the voltage drop across R_1 is greater than the drop across R_2.

If the oscillator is fast, the sawtooth retrace will pass through its a-c axis before the sync pulse occurs, Fig. 7·34*c*. On D_2, therefore, some of the sawtooth voltage will be subtracted from the sync pulse. On D_1, some of the sawtooth voltage will be added to the sync pulse, producing a higher voltage drop across R_2 than across R_1. This will produce a positive output voltage in order to slow down the horizontal oscillator.

Application of the foregoing phase detector to a commercial transistor television receiver is shown in Fig. 7·35. Incoming horizontal

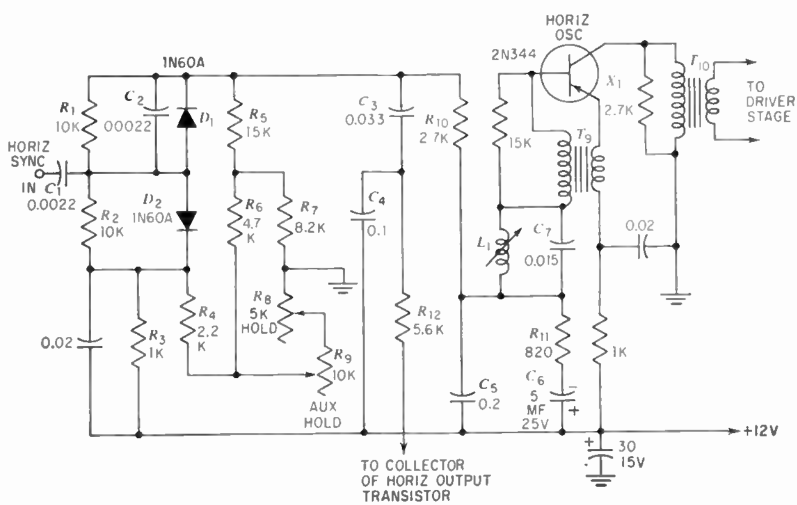


Fig. 7·35 Commercial application of the phase detector of Fig. 7·32. (Philco Corp.)

sync pulses are brought to diodes D_1 and D_2 by capacitor C_1. At the same time, a sawtooth wave is brought into the circuit by R_{12} and applied to capacitors C_3 and C_4. The two waveforms are compared in frequency by D_1 and D_2, and if any difference exists, a voltage is produced across R_1 and R_2. This voltage is then brought to the base of X_1 by R_{10}, R_{11}, C_5, C_6, L_1, and C_7. The latter six components serve as filter elements to smooth out instantaneous variations in the control voltage so that smoother control of the oscillator is achieved. The network also tends to prevent the oscillator from shifting back and forth in frequency as it attempts to find the correct operating value.

Incorporated into the phase-detector circuit is the network which permits the set viewer to adjust the operating frequency of the horizontal oscillator. Twelve volts is brought to the emitter of X_1 from the power line. The same +12 volts is also brought to the base of X_1 through R_3, R_4, R_5, R_6, R_7, R_8, and R_9. Included in this network are two hold controls, R_8 and R_9. By varying either or both of these potentiometers, the base voltage can be altered and thereby the operating frequency of the oscillator. R_8, the hold control, is the front-panel adjustment most frequently adjusted by the set viewer. R_9, the aux-

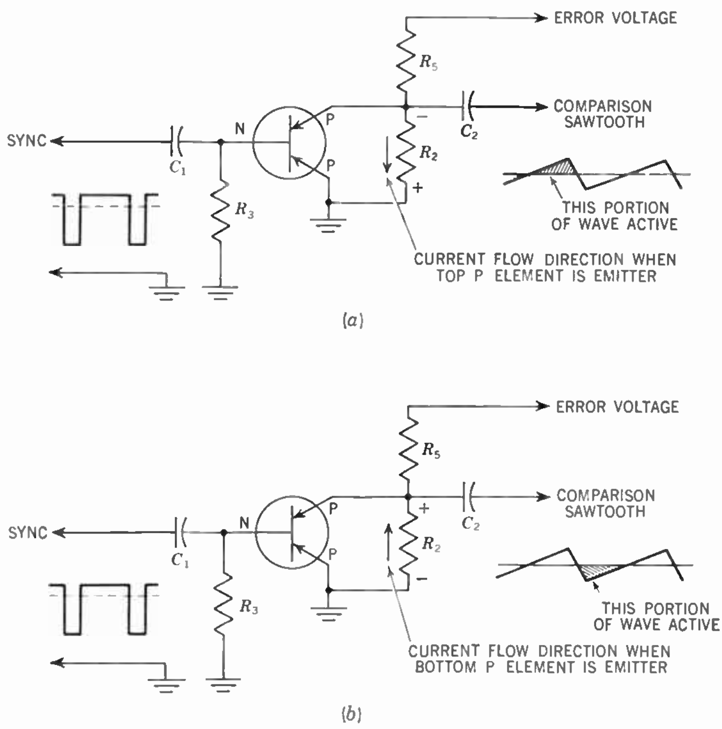


Fig. 7·36 The changing direction of current flow through R_2 when the polarity of sawtooth voltage reverses during sync interval.

iliary hold control, would be mounted on the back panel for occasional adjustment.

Transistor phase detector. In a transistor phase-detector circuit, similar operation can be achieved by means of a single transistor in a circuit such as that shown in Fig. 7·36. The transistor, however, is specially constructed so that the collector and emitter junctions are equal in area. (In the transistor designed for general usage, the collector possesses a greater area than the emitter.) This forms a symmetrical transistor. Advantage is taken of the fact that transistors will conduct in either direction to form this phase detector. That is, the biasing voltages to emitter and collector can be reversed so that the element which serves as an emitter under one set of conditions becomes a collector under another set of voltages. In short, either element can serve as the emitter or collector, depending upon the applied potentials. It is common, therefore, to refer to either element as a "collector-emitter." This behavior, too, is the reason both emitter and collector in Fig. 7·36 are shown with arrowheads.

Operation of this transistor phase-detector circuit depends on the transistor conducting only when the sync pulses are present. Thus, if the instantaneous sawtooth voltage across R_2 is positive at the instant the negative sync pulse triggers the base on, then the uppermost element with the arrowhead is serving as the emitter and the lower arrowed element is the collector. The reason is that this is a PNP transistor and, for conduction to occur, the emitter must be positive with respect to the base. Electrons then travel in the direction indicated in Fig. 7·36*a*, producing a negative voltage drop across R_2. The potential represents the error voltage which is transferred, via R_5, to the horizontal-control tube.

Conversely, if the sawtooth voltage is negative when the sync pulses arrive, then the element previously acting as the emitter now becomes the collector and the other P section becomes the emitter. Current flow is now reversed through R_2, Fig. 7·36*b*, and a positive error voltage is produced. In this way the transistor acts as a bidirectional switch producing an error (or correction) voltage whose polarity depends on the part of the sawtooth cycle active at the instant the sync pulses arrive.

If the sawtooth wave is passing through zero when the pulses arrive, no voltage appears across R_2.

In order for the circuit to function properly, the transistor must be completely cut off between sync pulses. This is achieved by having the peak-to-peak sync voltage at the base exceed the peak-to-peak sawtooth voltage. When the sync pulses are active, the current that flows in the base circuit causes C_1 to charge to their peak value. In the interval between pulses, this charge decreases very slowly, keeping the base at all times positive enough with respect to the collector-emitter voltage to prevent conduction.

A commercial circuit employing the transistor phase detector just discussed is shown in Fig. 7·37. Positive horizontal-sync pulses are brought to the base of X_1 by capacitor C_1. At the same time, a portion of the horizontal-deflection voltage is taken from the horizontal-output transformer (not shown), converted into a sawtooth wave by C_6 and R_7, and brought to the collector of X_1 by C_5. Direction of current flow through X_1 will depend on whether the sawtooth wave is positive or negative at the moment a horizontal pulse appears at the base of X_1. This, in turn, will determine the polarity of the control voltage developed across R_2. This voltage will be applied through R_6, R_8, C_3, and C_4 to the base of X_2, the horizontal blocking oscillator. Note that R_2 is in series with R_5, a potentiometer that supplies a negative d-c voltage to the base of X_2. Rotation of the center arm on R_5 will alter

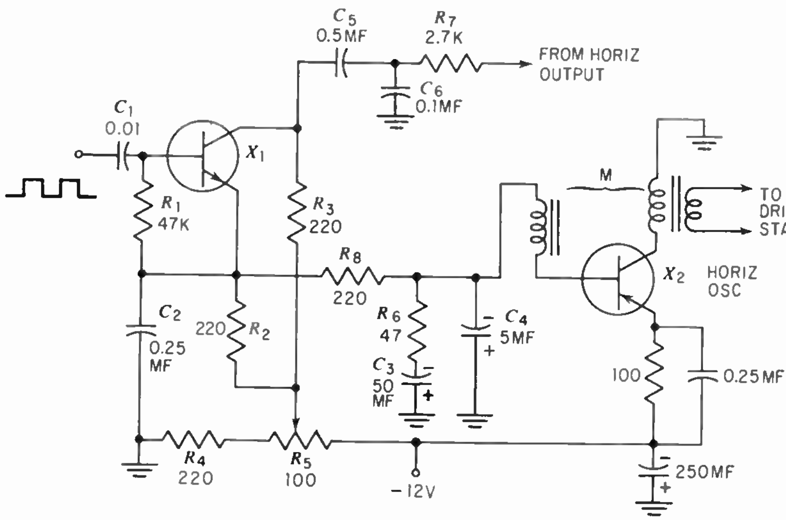


Fig. 7·37 A transistor phase detector, X_1.

the base-collector voltage of X_2 and, from this, the operating frequency of the oscillator. Thus, R_5 is rightfully the horizontal-hold control.

The positive or negative voltage variations introduced by R_2 will shift the oscillator operating frequency above or below a center frequency established by the setting of R_5. In this way, automatic control of the oscillator frequency is maintained.

Horizontal oscillator. The horizontal oscillator follows the phase detector, receiving the correction voltage developed by the latter circuit. A frequently employed oscillator is the same blocking oscillator just described for the vertical system. Essentially the same considerations apply, modified only by the higher operating frequency and the need to respond to the afc voltage.

A typical oscillator circuit, integrated into a complete horizontal-deflection system, will be considered presently.

Horizontal-output stages. The driving signal from the oscillator, if it is powerful enough, may be applied directly to the horizontal-output stage. At the present level of transistor development, however, there is more likely to be a driver stage between the oscillator and the output amplifier. This driver simply takes the output voltage developed by the oscillator, strengthens it, and then applies it to the horizontal-output amplifier.

Whether there is a driver and output amplifier or simply an output stage alone, the best place to start is with the output stage, because it is the requirements of this amplifier which will determine, to a large

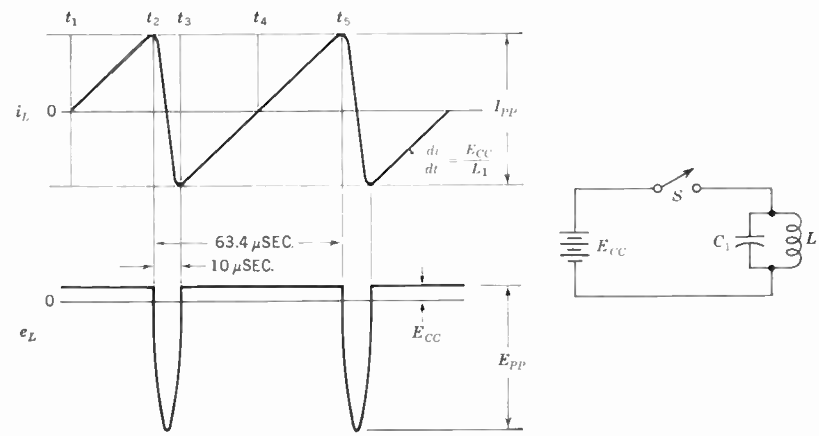


Fig. 7·38 The basic output circuit of a horizontal-deflection system and its waveforms. (After H. C. Goodrich, A Transistorized Horizontal Deflection System, *RCA Review*, September, 1957)

extent, the form of the stages which precede it. The chief purpose of the output transistor is to develop a sawtooth current through a deflection yoke. A secondary goal is to produce a high-voltage pulse which can be rectified and employed as the accelerating d-c voltage for the picture tube.

An output stage and driver circuit which has been used successfully is shown in Fig. 7·41. In order to understand fully the operation of this circuit, let us consider the basic circuit from which it was derived. (The method of explaining the operation of this system is essentially the same as that suggested by Goodrich.) The circuit, shown in Fig. 7·38, consists of a battery E_{CC}, a switch S, a coil L_1, and a capacitor C_1. If switch S is closed at time t_1, the current flowing from the battery through coil L_1 will increase at a linear rate. At time t_2, the switch is opened, interrupting this flow. The interruption shock-excites the capacitor-coil combination and the current starts to oscillate between L_1 and C_1. Since the current had been passing through L_1 when the switch was opened, the inductance will attempt to keep the current flowing, forcing it to flow into C_1, and it is this particular action that starts the oscillatory motion.

For one half cycle, or from time t_2 to time t_3, the circuit is permitted to oscillate. At t_3, switch S closes again, and now the current will flow back into the battery because the half cycle of oscillation has caused it to reverse itself. The flow of current into the battery will continue until t_4, at which point it will have decreased to zero. From t_4 to t_5, the current travels from the battery to L_1, repeating the sequence of events just described.

In terms of horizontal-output circuit operation, the steady rise of current from t_3 to t_5 represents the interval when the electron beam is steadily scanning across the picture screen from extreme left to extreme right. At t_5 (or t_2), the beam is blanked out and swung quickly back to the left-hand side of the tube. This is the retrace interval, and it must be completed in about 10 μsec. It is during this interval that a sharp pulse of voltage is developed across L_1 by the opening of switch S. This pulse is generally stepped up, rectified, and then applied to the second anode of the picture tube.

Note that if we were to represent the waveform of the voltage obtained from the battery, it would be a square wave. When the switch is closed, the full value of E_{cc} is instantly applied to L_1C_1; when the switch is open, the applied voltage drops instantaneously to zero.

In a practical circuit, the switch can be replaced by a junction transistor and the driving voltage by square waves from either an oscillator or driver. The square waves are required to turn the transistor on or off automatically. The basic transistor circuit and the waveforms in that circuit are shown in Fig. 7·39. From t_1 to t_2, the input signal biases the base-emitter circuit in the forward direction and the transistor conducts, completing the circuit. The input signal is strong enough to place the transistor in saturation, and in this condition the internal resistance (emitter-to-collector) is reduced to a fraction of an ohm.

At t_2, the incoming pulse goes sharply positive which, for a PNP transistor, reverse-biases the base-emitter circuit and cuts off the transistor. The cutoff does not occur as sharply as the pulse change because when a transistor becomes saturated, carriers are accumulated in the base region and a short period is needed to clear this region after the forward bias has been removed.

From t_2 to t_3, the transistor switch is open and L_1 and C_1 are shock-excited into oscillation. For one half cycle, the circuit oscillates; then the reverse bias is removed and the transistor is once again driven into strong conduction. The current now flows in the reverse direction through the transistor, just as it did in the preceding circuit. At t_4, this reverse current has decreased to zero and the forward flow begins again, repeating the sequence of events.

It is interesting to observe that the same step-by-step process takes place in horizontal-output stages utilizing vacuum tubes. However, since the horizontal-output tube cannot conduct in the reverse direction, a diode (i.e., the damper tube) must be connected across the

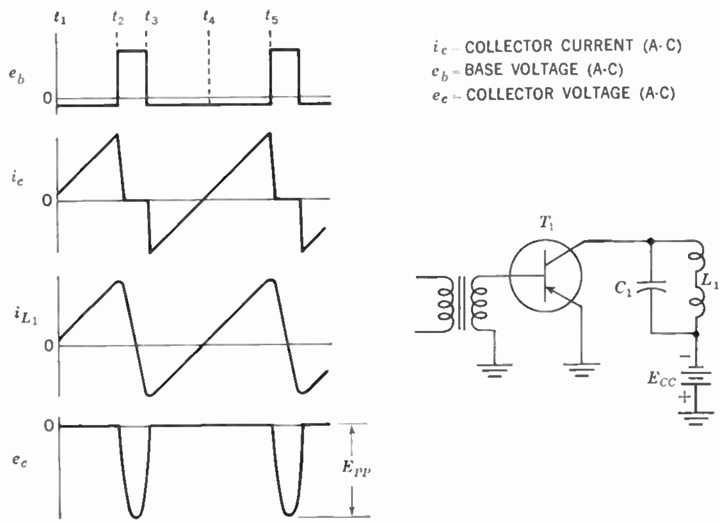


Fig. 7·39 The basic transistor output circuit and its waveforms. (RCA Review)

horizontal-deflection winding. With a transistor, this added item need not be used because the transistor itself will serve the purpose. For best results, the unit should be capable of carrying current equally well in both directions although some dissymmetry is tolerable.

During the interval from t_2 to t_3, when L_1 and C_1 are oscillating freely, a large, sharp pulse is developed across L_1, Fig. 7·39. Care must be taken to see that the peak voltage of the pulse does not exceed the collector breakdown voltage. The designer of this circuit devised an interesting method of blunting this peak. Another resonant circuit, L_2C_2, tuned to approximately the third harmonic of the resonant frequency of L_1C_1, is connected across L_1, Fig. 7·40. The voltage developed across L_2C_2 (labeled e_{L2} in Fig. 7·40) during the flyback interval possesses the form shown in Fig. 7·40. This voltage is 180° out of phase with the voltage across C_1 (e_c) at the center of the flyback pulse, and thus reduces the peak collector voltage by about 30 per cent.

With the foregoing explanation of circuit operation understood, we are ready to consider the actual horizontal-output circuit developed by this designer. The output stage and the driver that precedes it are shown in Fig. 7·41. If we examine the driver stage first, we see that it receives pulses from a preceding oscillator. During the scanning period,

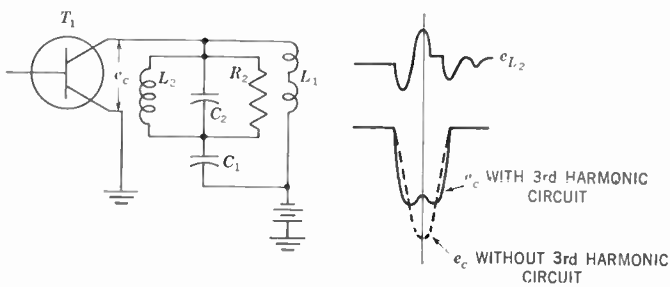


Fig. 7·40 The special resonant circuit L_2, C_2, and R_2 helps to reduce the voltage peak at L_1 during the flyback period. (*RCA Review*)

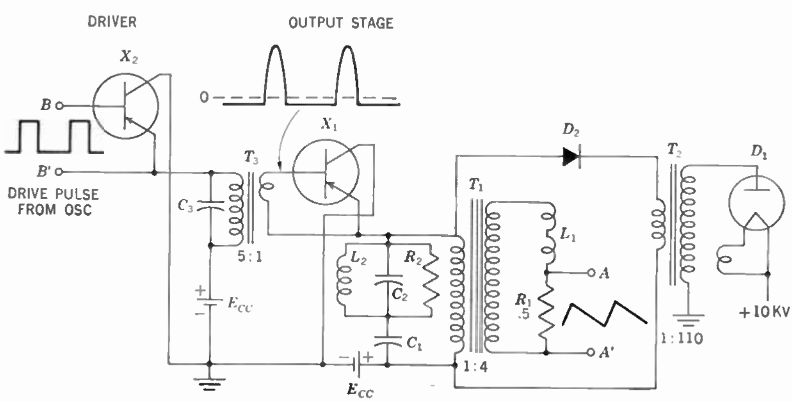


Fig. 7·41 Transistorized output stage and driver. (*RCA Review*)

the base emitter of the driver is strongly forward-biased and the transistor is operating in a saturated condition. During retrace, the incoming pulse goes sharply positive, cutting off X_2. Since the transistor is working into a transformer T_3, the sharp cutoff causes a pulse to appear across the primary winding. This pulse, after being stepped down in voltage to reduce the circuit impedance, is employed to cut off the output stage.

Alternating-current coupling between the driver and the output stage results in a d-c (i.e., zero) axis about 10 per cent above the signal-level scan, as indicated by the base waveform for X_1. (Because of the a-c coupling, the waveform distributes itself about the axis so that there is just as much area above the zero line as below.) The wave below the axis serves as a forward bias, which eliminates the need for an additional bias source. Furthermore, it tends to protect the output stage in case of drive failure, since with no drive both stages are cut off.

In both the driver and the output stage, a bootstrap circuit is used to permit the collector and case to be grounded for best heat conduction. Operation of the two stages is exactly as indicated above. High voltage is obtained from transformer T_2. Whatever flyback pulse appears across the primary winding of T_1 is stepped up by T_2 and then rectified by D_1. The latter is shown here as a tube, which for the present may be a more economical arrangement. However, the same function can be served by several semiconductor diodes connected in series. In time, a suitable single rectifying device will undoubtedly be developed. Diode D_2 is so connected that it is closed during the retrace interval and opened during the scan period. Its purpose is to prevent damped oscillatory voltages in T_2 from causing ripples in the forward scanning deflection current.

L_1 is the horizontal-deflection winding of the yoke. R_1 is a small resistance, inserted in series with L_1, to obtain a sample of the deflection current. This voltage is applied to a preceding phase detector where it is compared in frequency with the incoming horizontal-sync pulses. Any difference between the two signals produces a correction voltage for the horizontal oscillator.

It may be of interest to see the oscillator and phase detector that precede the driver and output stages of Fig. 7·41. These are shown in Fig. 7·42. The phase detector is similar to the phase detectors previously discussed. The signal voltage developed across R_1 in Fig. 7·41 is applied between emitter and collector of X_4. At the same time, negative sync pulses (from a sync separator) are applied to the base of X_4. The two voltages are compared in frequency, and any difference produces a voltage across R_{12}. This error signal is applied through the antihunt integrating circuit $R_6C_4C_5R_5$ to the oscillator base winding.

A blocking oscillator is employed for the horizontal oscillator. Oscillator repetition frequency and pulse width depend primarily on R_2, C_3, and the ratio of X_3 base bias to collector supply voltage. In this circuit, the base bias is varied by the horizontal-hold control. This bias is placed in series with the d-c correction voltage from the phase detector.

Diode D_3 serves to limit the transformer inductive overshoot during the time X_3 is cut off in order that the collector breakdown voltage is not exceeded. The output signal for the driver is obtained from a third winding on the blocking transformer. This method of a-c coupling provides forward bias for the output stage. Since the forward bias tends to be higher than necessary, the operating point is shifted in the reverse direction by bias resistor R_4.

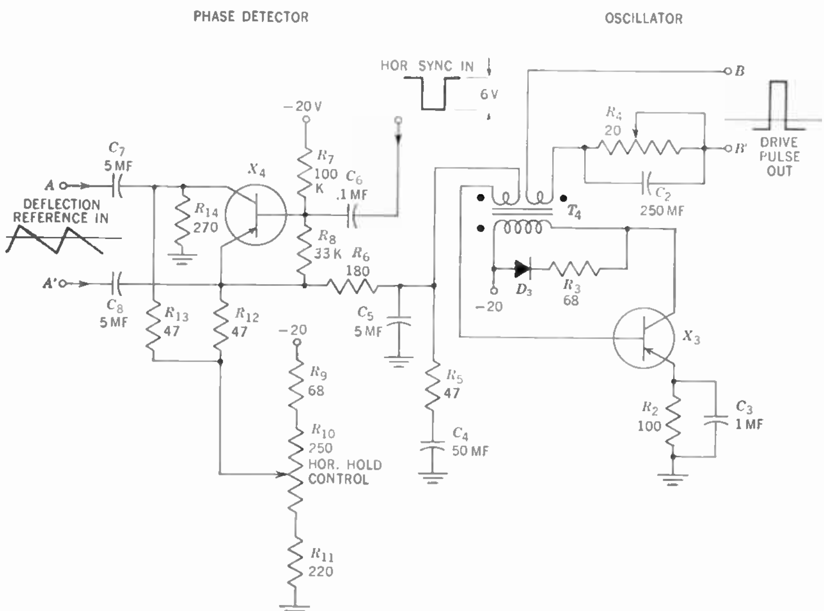


Fig. 7·42 The transistorized horizontal-sweep oscillator and phase detector which precede the driver and output stages in Fig. 7·41. (*RCA Review*)

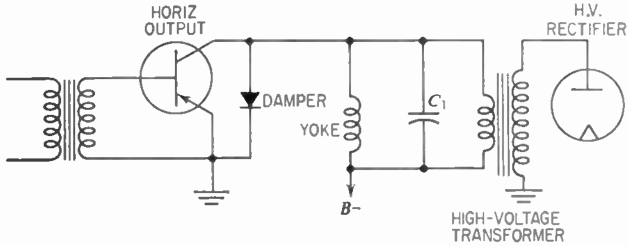


Fig. 7·43 A horizontal output circuit in which the transistor is directly coupled to the deflection yoke.

A transistorized horizontal-output circuit in which the transistor is directly coupled to the yoke is shown in Fig. 7·43. When this approach is employed, a low-impedance yoke is used, one with an inductance on the order of 100 μh or less. By way of contrast, yoke inductances of

8 to 30 mh are common when a transformer is interposed between the yoke and the output transistor, as in Fig. 7·41.

The transistor in Fig. 7·43 has a damper diode connected in parallel with it. This diode is useful in providing a linear sawtooth current flow through the transistor. It achieves this by conducting current (together with the transistor) during the interval when the output transistor is also carrying the reverse current (during the interval t_3 to t_4 in Fig. 7·38). The sawtooth current wave is made more linear, Fig. 7·44, because a practical transistor will seldom conduct equally in both directions. By paralleling a diode across the transistor, increased current can be passed during the reverse interval to help keep the flow steady.

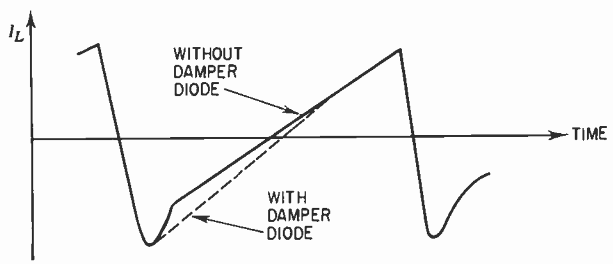


Fig. 7·44 **The shunt damper diode in Fig. 7·43 helps to achieve a linear sawtooth current through the yoke.**

In conclusion, it is interesting to note that the total power dissipated in the output transistor developing 90° deflection at 10,000 volts is of the order of 2 watts. In the output circuit of a vacuum-tube deflection circuit performing the same function, 20 watts would be dissipated.

QUESTIONS

7·1 Why is it more difficult to transistorize a television receiver completely than a radio receiver? Give several specific examples.

7·2 Draw the circuit of an r-f amplifier suitable for use with a television tuner. Indicate what characteristics a transistor for this stage should possess.

7·3 Explain how the circuit drawn for Question 7·2 operates.

7·4 Explain the function of each of the components in the circuit of Fig. 7·4.

7·5 Could the neutralization method employed in Fig. 7·4 be applied to the circuit of Fig. 7·2? Explain your answer.

7·6 How does the circuit of Fig. 7·8 function? Indicate how the transistor can operate even though the collector has zero volts on it.

7·7 Explain the operation of the agc network of Fig. 7·10.

7·8 Draw the diagram of a diode video detector.

7·9 Draw the diagram of a two-stage video amplifier system. Explain the purpose of each component.

7·10 Indicate what precautions would have to be observed in designing a transistorized video i-f system for a television receiver. (*Note:* Cover such items as α cutoff frequency, temperature stability, and impedance matching.)

7·11 What characteristics of a transistor enable it to be used successfully as a limiter? As a sync separator?

7·12 Explain how the circuit of Fig. 7·19 functions.

7·13 Why is a double-time-constant arrangement, such as used in Fig. 7·21, more desirable than a single-time-constant network?

7·14 Compare the circuits of Figs. 7·19 and 7·21 as to advantages and disadvantages.

7·15 Describe the operation of the sync separator of Fig. 7·21.

7·16 Draw the circuit of a transistor blocking oscillator that can be synchronized by pulses and that develops a sawtooth output voltage.

7·17 Explain the operation of the circuit drawn in response to Question 7·16.

7·18 What is the purpose of each of the controls in Fig. 7·24?

7·19 What is a symmetrical transistor? What counterpart does it have among vacuum tubes?

7·20 How does a double-diode phase detector operate?

7·21 Draw the basic circuit of a transistor phase detector. Explain briefly how it operates.

7·22 Why must the incoming sync pulses in the circuit of Fig. 7·32 possess much larger amplitudes than the sawtooth wave applied to the same circuit?

7·23 Draw the complete circuit of an actual transistor phase detector. Explain the purpose of each component.

7·24 How is bias provided for the two transistors in Fig. 7·41?

7·25 What is the purpose of L_2, R_2, C_2, and C_1 in Fig. 7·41? Explain in detail.

7·26 In what ways does X_1 in Fig. 7·41 differ in operation from a horizontal-output stage using a vacuum tube?

CHAPTER 8

Industrial Applications of Transistors

IN THE TWO preceding chapters, we have investigated the utilization of transistors in radio- and television-receiver circuits. We have seen how transistors can be employed as low- and high-frequency amplifiers, as mixers, converters, detectors, sync separators, pulse amplifiers, phase discriminators, and sawtooth generators. These form a wide range of applications, but they do not exhaust the possibilities by any means. In the general realm of electronics, beyond radio and television, lie even more diverse uses of the transistor, and some of them will be touched on in this chapter.

Transistors in Switching Circuits

There are many instances when exceedingly fast operating switches, on the order of microseconds, are required. That is, it is desired to go instantaneously from one state, say an "off" condition, to the opposite "on" condition. This occurs in high-speed counters, in computers, and in circuits where something must happen before or must not happen until one or more triggering pulses arrive to signal the start of a certain sequence. For this function, junction transistors are admirably suited, not only because they can be built to provide switching times of a few millimicroseconds, but also because the additional circuitry required, such as resistors and capacitors, is frequently no more, in total number, than the number of transistors and sometimes even less than this. In place of millimicrosecond, the term *nanosecond*, abbreviated nsec, has been employed. Both mean the same thing, namely 10^{-9} sec.

To see what happens when a transistor is employed as a switch, consider the simple circuit shown in Fig. 8·1. Here we have a transistor, a load resistor, and a source of power. The transistor is connected in the grounded-emitter configuration. Since a PNP transistor is used, the collector is biased negatively. To produce current through the unit, the base-emitter circuit must be biased in the forward direction, which here requires that the base be made negative with respect to the emitter. If the base and emitter are kept at the same voltage (which also includes zero voltage), the transistor will be essentially cut off and no current other than I_{co} will flow through the system.

If a negative pulse is applied to the base, the base-emitter circuit will become forward-biased and current will flow in the collector circuit. The electrons will travel from the negative terminal of the battery to the collector, producing a voltage drop across R_L, which will make the collector (point A) less negative than before. Also, by virtue

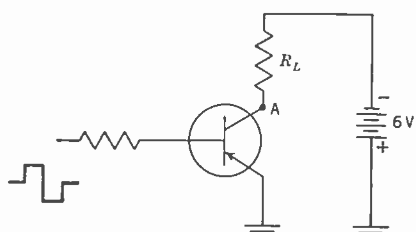


Fig. 8·1 Single-transistor switching circuit.

of the increased current, the impedance of the transistor, from point A to ground, will be less than it was prior to the application of the pulse. This can be seen from the fact that before the arrival of the pulse, the voltage between point A and ground was equal to the full battery voltage of 6 volts. Now, because of the current flow, it is less.

As we increase the amplitude of the pulse at the base, the collector current will increase until eventually practically all of the battery voltage will be dropped across R_L, leaving the collector voltage at point A essentially zero. A further increase in input pulse amplitude will act to increase the base current, but it will have very little effect on the collector current. When this condition is reached, the base-to-emitter voltage is larger than the base-to-collector voltage and the base becomes more negative than the collector. Actually, it will be found that the collector-to-base diode becomes forward-biased and begins to inject carriers into the base. At this point, the transistor is said to be saturated, and its collector impedance is nominally that of a forward-biased diode, which is very low. Since the base-emitter circuit is also forward-biased, the entire transistor impedance between point A and ground is no more than a few ohms.

In this saturated condition, more current is flowing in the base than is required to maintain the collector current at the value determined by the transistor. At the same time, the collector is also injecting a number of carriers into the base, with the result that there is a large minority carrier density in the base region. (In a PNP transistor, the minority carriers in the base region are holes. In an NPN transistor, they would be electrons.) This is in contrast to the normal operation of a transistor, where there is a high concentration of minority carriers at the emitter end of the base region and a much lower minority charge density at the collector end of the base region. In the saturated condition, minority carriers are being injected into the base region from both collector and emitter, and a fairly high charge is thus stored in the base.

Now let us see what happens when the input driving pulse ends and the base-emitter voltage drops to zero or even extends in the opposite direction, causing a reverse-biased condition. It may at first be supposed that the collector current similarly cuts off, and this would normally happen in a transistor that was functioning in other than the saturated state. With saturation, however, excessive minority carriers are stored in the base. Hence, when the input signal terminates, the excess-charge-density condition will prevail until, by recombination and by removal of carriers at the emitter and collector junctions, the charge density at the collector end of the base region becomes zero. When this happens, the collector impedance rises very rapidly and reassumes its normal high value.

Thus, there will be a short but finite time after the pulse at the base is terminated that the pulse across the collector resistor is similarly ended.

To see the foregoing action in terms of the input and output pulses, consider Fig. 8·2. The input pulse of current applied to the base is shown in Fig. 8·2*a*. The resultant output pulse is indicated in Fig. 8·2*b*.

The interval from t_0 to t_1 is the time it takes from the application of the input voltage until the output voltage has reached 10 per cent of its final value. This is then followed by the time interval, t_1 to t_2, required for the output to go from 10 to 90 per cent of its saturation value. This is the rise-time period. The transistor is now fully turned on and will stay on so long as the input voltage is maintained. At time t_3 the input pulse ends; this starts the turn-off. From t_3 to t_4 is the storage time, and during this interval the output voltage goes from its saturation value to 90 per cent of that value. Finally, there is the fall time, t_4 to t_5, when the output drops to 10 per cent of its saturation value.

The initial delay time is due to two causes. If the emitter-base circuit was initially reverse-biased to keep the transistor off, then it will take a short time to discharge the emitter-base internal capacitance from this reverse-biased condition through the base resistance and charge it to a forward-bias condition. Secondly, time must be allowed for the emitter current to diffuse through the base region.

The rise time refers to the turn-on of the collector current. Thus, the total turn-on time of a transistor is equal to the sum of the delay and rise times.

The storage time is due to the length of time required to sweep out the stored charge carriers in the base region which resulted from the

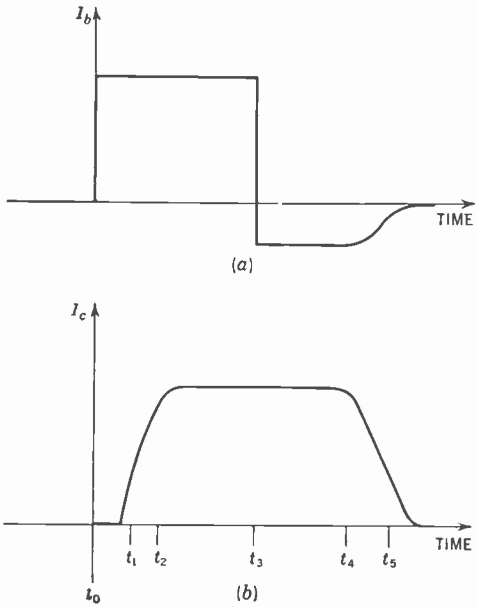


Fig. 8·2 The action in a transistor operated in the saturated condition. (a) Current in base circuit. (b) Current in collector circuit. (After J. L. Moll, Large Signal Transient Response of Junction Transistors, *Proc. IRE*, December, 1954)

collector-base junction being forward-biased during saturation (t_2 to t_4). This is true for alloy transistors. For grown-diffused and mesa transistors, the primary storage takes place in the collector region rather than the base region. It will be noted from Fig. 8·2*a* that a reverse current flows in the base circuit from t_3 to t_5. When the input driving pulse is cut off, and in those cases when the base is actually driven positive at time t_3, the emitter will, in effect, be made to function as a collector (remember that the collector is still forward-biased). Holes present in the base (for a PNP transistor) will be collected by the emitter, and instead of the emitter current falling to zero, a considerable reverse current will flow in the base-emitter circuit. This reverse bias is beneficial in reducing the delay in the fall of the collector current and is frequently employed. As soon as the carriers have been swept out of the base region or the collector region, the transistor begins to turn off.

Total turn-off time is equal to the sum of the storage and fall times.

The time required to turn off a transistor depends greatly on the construction of the unit. In low-frequency transistors, it may extend over an appreciable period of time. In high-frequency transistors, both the turn-on and turn-off times are made very small. Thus, in a high-frequency surface-barrier transistor, the turn-on time is from 22 to 27 nsec. The time required to turn the transistor off is on the order of 45 to 85 nsec.

Note that if a transistor is not driven to saturation, the carrier storage time, t_3 to t_4 in Fig. 8·2*b*, does not occur. However, the fall time, t_4 to t_5, still takes place, because a small time is required for the final carriers, injected by the emitter into the base region, to reach the edge of the collector. Hence, if a shorter total turn-off time is desired, the transistor is not driven to saturation.

Gating Circuits

Now that we have seen how a transistor operates in the switching mode, let us examine a number of typical circuits in which the transistor is used as a switch. Figure 8·3 shows a gating circuit in which

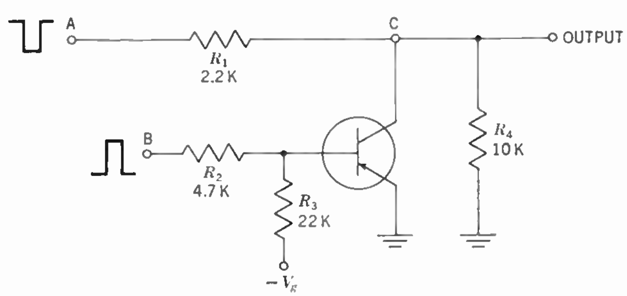


Fig. 8·3 A transistor gating circuit.

the transistor is effectively shunted across the signal path. With no signal applied to the base, the negative d-c voltage from $-V_g$ will bias the transistor on. This voltage is so chosen that the transistor is in the saturation state, with the result that it will shunt a very low impedance from point *C* to ground. While the transistor is in this state, any signal voltage at point *A*, such as a negative pulse, will be dropped across R_1 and none (or very little) will appear across the output terminals.

However, if a positive pulse appears at point *B* at the same time that a negative pulse is applied to point *A*, and if the positive pulse is strong enough to overcome the negative base bias and cause the emitter-base circuit to become reverse-biased, then the transistor will be turned off and the output voltage will be determined by the ratio of R_1 to R_4.

If, for circuit reasons, it is desired to prevent certain pulses appearing at point *A* from reaching the output, then the circuit may be modified in the following manner. A small positive voltage is applied to the base through resistor R_3. This will bias the transistor off, and any negative pulses appearing at point *A* will reach the output. If, now, a negative pulse is brought to point *B*, it will drive the transistor

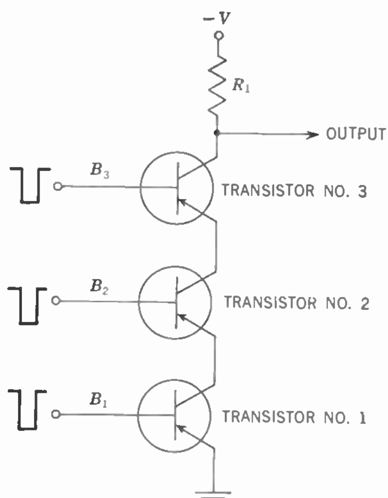


Fig. 8·4 A series type of AND gate.

on, in effect throwing a virtual short circuit across the upper signal line and, during this interval, no pulses at point *A* will reach the output.

This type of circuit arrangement is known as an AND gate because it requires pulses at points *A* and *B* to achieve the desired action.

Another AND gate is shown in Fig. 8·4. Three transistors are connected in series and the combination is in series with a load resistor. To energize the entire circuit, negative pulses must be applied to each base at the same time. Current, when it does flow, travels from emitter to collector of each unit, then on to the emitter of the next transistor, through this transistor to its collector, on to the next emitter, etc., until the load resistor R_1 is reached. The current passes through the resistor and the battery and then back to the bottom of the transistor chain, completing the circuit.

The voltage and current relations in a typical three-input AND gate with a supply voltage of −1.5 volts and a load resistance of 1,000 ohms will be as in Table 8·1.

Table 8·1 Voltage and Current Relations in a Three-input AND Gate

Transistor 3:	
Collector-to-ground voltage	−0.09 volt
Collector current	1.41 ma
Base-to-ground voltage	−0.36 volt
Base current	0.28 ma
Transistor 2:	
Collector-to-ground voltage	−0.06 volt
Collector current	1.69 ma
Base-to-ground voltage	−0.36 volt
Base current	0.34 ma
Transistor 1:	
Collector-to-ground voltage	−0.03 volt
Collector current	2.03 ma
Base-to-ground voltage	−0.36 volt
Base current	0.40 ma

In the foregoing arrangement, a circuit-current gain of 5 is assumed. This is generally less than is actually obtained in such gating circuits.

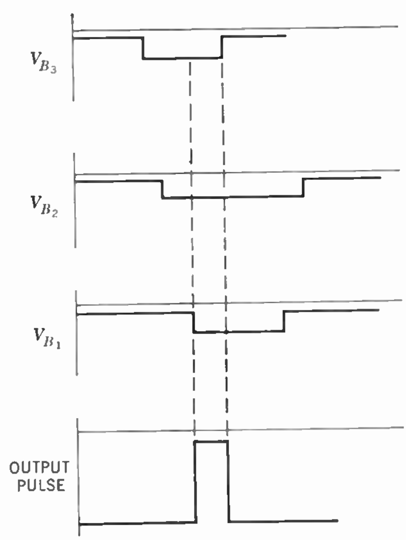
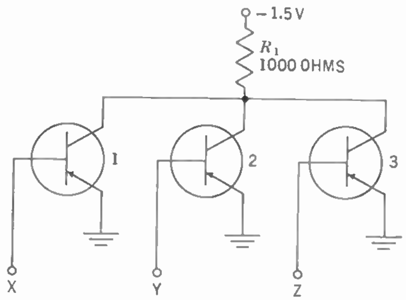


Fig. 8·5 In the AND gate of Fig. 8·4 the start of the output pulse is governed by the start of the last pulse at one of the transistor bases. Similarly, the end of the output pulse is set by the base pulse which ends first.

If any transistor in the chain is not turned on, while the others are, nothing happens because the path is kept open. Also, if the negative pulses applied to the various bases differ in length and starting and ending times, then the output pulse width will be determined as shown in Fig. 8·5.

OR gates. Another type of gate circuit, known as an OR gate, is shown in Fig. 8·6. The collectors of all transistors are connected to a common load resistor. As shown, each of the transistors is cut off because there is no forward-biasing voltage between the various bases and their respective emitters. Consider, now, what will happen if we apply a negative pulse to any base. The pulse will drive the transistor

Fig. 8·6 A three-transistor OR circuit.



into conduction, possibly even saturation if the pulse amplitude is strong enough. The transistor so triggered will have its internal impedance drop to a very low value, and the full battery voltage will appear across load resistor R_1. When the base pulse ends, the transistor will lapse back into cutoff and the entire system again will become nonconductive. This arrangement is known as an OR circuit because a

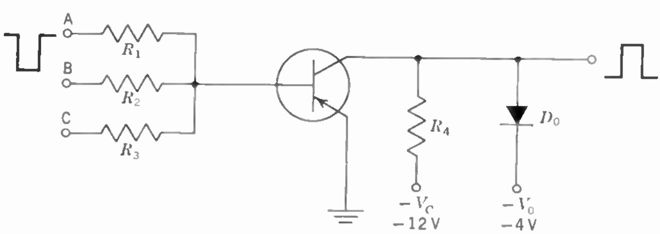


Fig. 8·7 Another OR-gate arrangement.

negative driving pulse at the base of transistor 1 *or* transistor 2 *or* transistor 3 will produce an output pulse.

Another form of OR circuit is shown in Fig. 8·7. The PNP transistor is connected in the common-emitter arrangement. A negative voltage, $-V_C$, is applied to the collector through resistor R_4. So long as no voltage is applied between emitter and base, the unit will be cut off. However, there are essentially three inputs, *A*, *B*, and

C, and a negative pulse applied to any one of these will drive the transistor into conduction and produce a positive output pulse at the collector.

Diode D_o serves two functions. It standardizes the output pulse amplitude and it prevents the transistor from going into saturation. One end of D_o connects to —4 volts, while the other end is attached to the collector of the transistor. As long as the collector is more negative than —4 volts, D_o will be nonconductive. However, the instant the collector voltage tends to become less negative than —4 volts, the

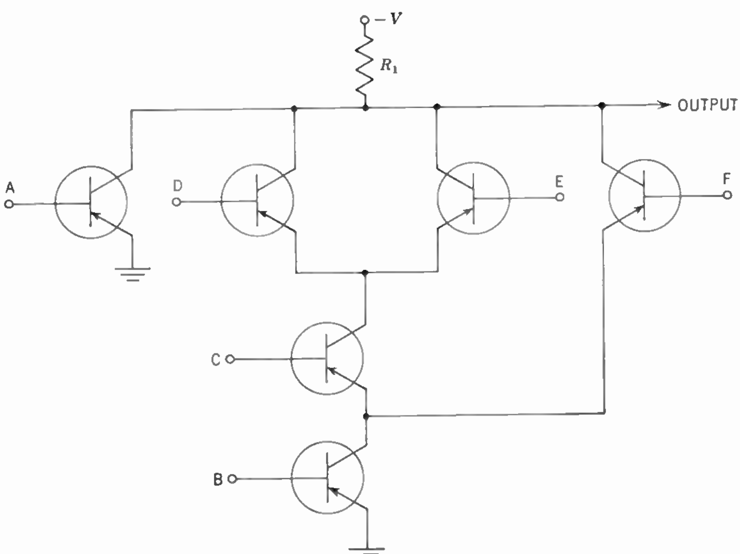


Fig. 8·8 A combination of AND and OR circuits.

diode conducts and clamps the collector at the —4-volt level. Thus, so long as sufficient drive is present at the input, the output pulse amplitude cannot exceed 8 volts (12 volts — 4 volts).

D_o also prevents the transistor from saturating, because it does not permit the collector voltage to drop so low that the collector becomes more positive than the base. And until that happens, saturation cannot occur.

It is readily possible to design combination AND–OR circuits to meet any given set of conditions. An example of a simple logic circuit combining AND and OR circuits is shown in Fig. 8·8. In order to produce a positive pulse across resistor R_1, one of the following conditions must occur:

1. A negative pulse at point *A*
2. Simultaneous negative pulses at points *B* and *F*
3. Simultaneous negative pulses at points *B*, *C*, and *D*
4. Simultaneous negative pulses at points *B*, *C*, and *E*

These combinations can be made as complex as possible, but they are all basically constructed from the simple AND and OR circuits discussed above.

Flip-Flop Circuits

In the foregoing gating circuits, the output conditions are maintained only as long as the input conditions are maintained. That is, they possess no memory, or the ability to remain "on" without being driven by the circuits that energized them originally.

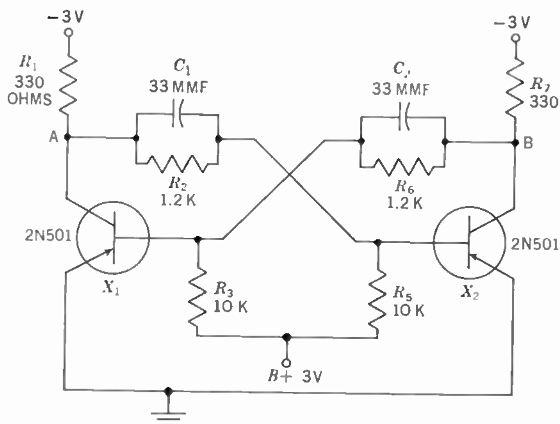


Fig. 8·9 **A bistable multivibrator or flip-flop circuit. Repetition or switching rates to 10 Mc are possible with this circuit and the 2N501 transistors.**

Bistable multivibrator. A circuit which can, in effect, remember or retain a certain condition indefinitely after the initiating force has been removed is a bistable multivibrator, commonly referred to as a flip-flop circuit. A typical arrangement of such a circuit is shown in Fig. 8·9. To understand the operation of this circuit, let us assume that transistor X_1 is conducting and transistor X_2 is cut off. For the "on" transistor, X_1 the collector voltage will be quite low because the collector current, passing through R_1, will drop enough voltage here to reduce the −3 volts from the power source to a value less than 1 volt.

The base of X_1 has applied to it a positive 3 volts. Since X_1 is a PNP transistor, this 3 volts would ordinarily be sufficient to cut X_1 off. However, the base receives enough negative voltage from the collector of

X_2 to overcome this positive 3 volts and keep X_1 in conduction. The voltage at the collector of X_2 is equal to the full -3 volts because X_2 is nonconductive.

Transistor X_2 is kept cut off because the positive voltage applied to its base is greater than the negative voltage received from point A.

The condition described above, with X_1 "on" and X_2 "off," will continue indefinitely because it is a stable one. To reverse this set of conditions, let us temporarily decrease the voltage at point B to zero. This will remove the negative offsetting voltage at the base of X_1 and permit the positive voltage to take over, cutting off this transistor and bringing the voltage at point A to -3 volts. This rise in negative voltage at point A will be transmitted through R_2 and R_5 to the base of X_2. Here it will offset the positive voltage from the power supply and enable X_2 to start conducting. The current flow that ensues will drop the voltage at point B to less than 1 volt. Thus, we now have reversed the operating conditions in the circuit, with X_1 cut off and X_2 conducting. Again, this is a completely stable state and will continue indefinitely unless something which will reduce the potential at point A, thereby initiating another switchover, comes along.

Note that this circuit does not possess any frequency of its own, but will flip from one state to the other at a rate determined by switching pulses applied to the circuit. (More on this in a moment.) There is a limit, however, to the number of switchovers per second that can be accommodated, since each switchover takes a period of time determined essentially by the turn-on and turn-off times of each transistor. Coupling capacitors C_1 and C_2 are inserted to bypass the coupling resistors R_2 and R_6 during the switching interval in order to bring the full force of the voltage change at point A or B to the base of the previously cut-off transistor as rapidly as possible.

It was seen above that the switchover from one conducting transistor to the other could be achieved by momentarily lowering the voltage at the collector of the nonconducting transistor. To achieve this by means of triggering pulses, two additional transistors, X_3 and X_4, are added to the circuit, Fig. 8·10. If, now, we assume that X_1 is conducting and X_2 is cut off, then a negative triggering pulse to the base of X_4 will drive this transistor sharply into conduction. The relatively large flow of collector current through R_7 will drop the voltage at point B and thereby initiate the switchover from X_1 to X_2.

By the same token, if X_1 is cut off and X_2 is conducting, a negative pulse at the base of X_3 will drop the potential at point A and cause a switchover.

In place of the two switching transistors in Fig. 8·10, simple steering diodes can be employed. A typical bistable circuit, this one using NPN

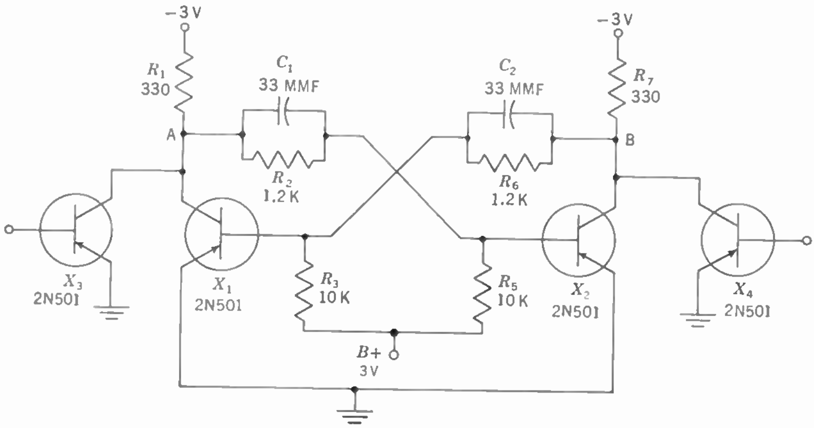


Fig. 8·10 A bistable multivibrator with triggering amplifiers X_3 and X_4.

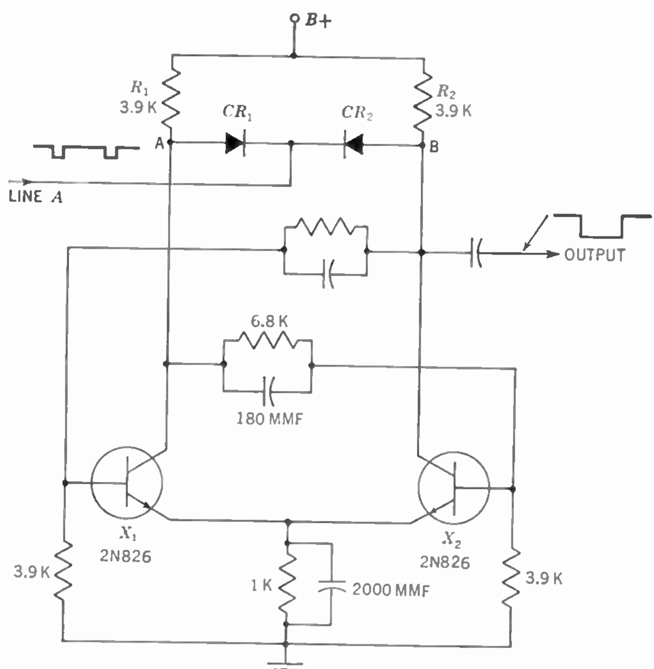


Fig. 8·11 A binary counter with switching diodes CR_1 and CR_2.

transistors, is shown in Fig. 8·11. The two diodes, CR_1 and CR_2, are connected in series opposition between the two collector elements of X_1 and X_2. If we assume that transistor X_1 is conducting and transistor X_2 is cut off, then the potential at point B will be more positive than the potential at point A. This particular voltage arrangement will forward-bias CR_2 and reverse-bias CR_1. This means that any pulses along line A will be able to pass through CR_2 and reach point B. They will, however, be unable to pass through CR_1.

The multivibrator will remain with X_1 conducting and X_2 cut off until a negative pulse arrives on line A. This pulse appears across R_2 and counteracts the B+ potential at point B. This, in effect, grounds point B and triggers the switchover from X_1 to X_2, as described above.

When the change has been completed, X_2 will be conducting and X_1 will be cut off. Now, point A will be more positive than point B, causing CR_1 to become forward-biased while CR_2 is reverse-biased. When the next negative pulse arrives, it will be directed to point A and R_1. Thus, by means of this shift in conduction from one diode to the other, incoming pulses are directed to the proper point in the circuit to initiate a switchover.

Direct-coupled Bistable Multivibrator

A modification of the foregoing bistable multivibrator which is especially interesting because of its extreme simplicity is the circuit shown in Fig. 8·12. If for a moment we disregard the switching tran-

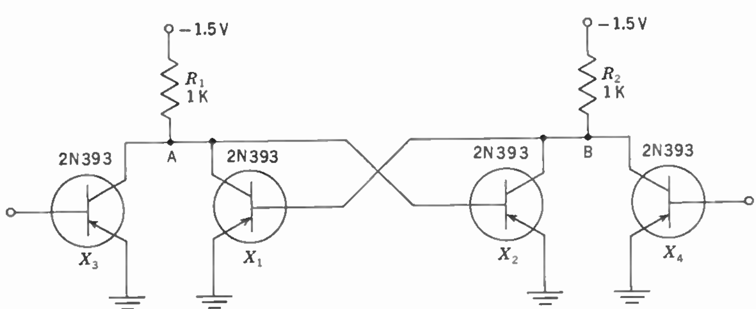


Fig. 8·12 Direct-coupled bistable circuit with switching transistors.

sistors X_3 and X_4 (which are the same here as in the previous arrangement), then the multivibrator consists essentially of two transistors X_1 and X_2 and two resistors R_1 and R_2.

Operation here is the same as before. If we assume that X_1 is conducting and X_2 is cut off, then the voltage at point A will be near zero and that at point B will be near 0.3 volt. The base voltage of X_1 will have a negative value determined by the resistance of R_2 and the input resistance of X_1. The second transistor, X_2, will be essentially cut off because its base is tied to point A, which is practically at zero voltage.

Switchover from one state to the other is accomplished by dropping the base voltage of the conducting transistor enough to cause this unit to stop conducting. This will bring the other transistor into conduction. The changeover can be readily accomplished by feeding a negative triggering pulse to the base of the proper switching transistor, X_3 or X_4. The action here is the same as before.

This particular circuit operates in the saturation mode so that the cutoff time will be lengthened by the storage effect. Furthermore, because of the direct coupling, the voltage swing across either load resistor is only on the order of 0.3 volt or so. The extreme simplicity of the circuit makes it very attractive, however.

Typical values of operating current and voltage for this transistor flip-flop with —1.5 volts supply potential and with load resistances of 1,000 ohms are given in Table 8·2. The voltage at the collector of

Table 8·2 Operating Current and Voltage for Transistor Flip-Flop

	"On" transistor	"Off" transistor
Collector-to-emitter voltage	−0.05 volt	−0.35 volt
Collector current	1.43 ma	20 μa
Base-to-emitter voltage	−0.35 volt	−0.05 volt
Base current	1.15 ma	5.0 μa

the "off" transistor is sufficient to energize the bases of several transistors. It does energize the base of the partner transistor in the flip-flop. Conversely, the "on" transistor collector voltage is low enough to turn off the base of any transistors connected to it.

Monostable multivibrators. Monostable, or one-shot, multivibrators have only one stable state, in contrast to the two stable states of the bistable circuit. To convert to one-shot operation, a capacitor C_1 and a resistor R_3 are added to the circuit of Fig. 8·12. The result is shown in Fig. 8·13. Transistors X_1 and X_2 form the multivibrator, while X_3 is added to trigger the circuit when this is required.

The operation of this monostable multivibrator may be explained by reference to the waveforms shown in Fig. 8·14. If R_1 and R_2 are chosen to be about 1,000 ohms, R_3 approximately 10,000 ohms, and capacitor C_1 about 50 μμf, the steady-state conditions are as follows:

Transistor X_2 will be conducting heavily, and its collector voltage will be approximately —0.03 volt with respect to ground. This is sufficient to ensure that X_1 is cut off. The collector voltage of X_1 will then be very near the supply voltage, in this case, —1.5 volts. Assume that, up to the time of triggering, the base of X_3 is *not* energized. This condition exists from time t_0 to time t_1. At time t_1, a negative pulse is applied to the base of X_3, causing its collector voltage, and hence the collector voltage of X_1, to be brought near ground potential. This positive pulse, appearing at the collector of X_1, is coupled through C_1 to the base of X_2 and drives X_2 to cutoff. During the time X_2 is cut off,

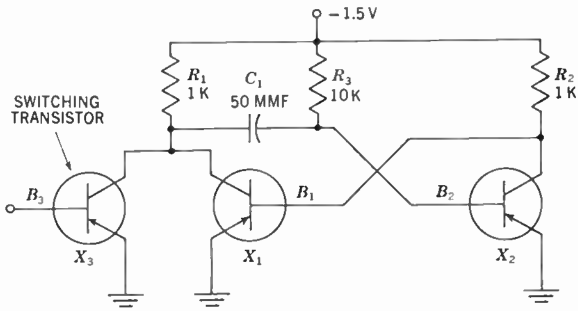


Fig. 8·13 A monostable multivibrator. (After R. H. Beter et al., Surface Barrier Transistor Switching Circuits, a paper of the Philco Corp., 1956)

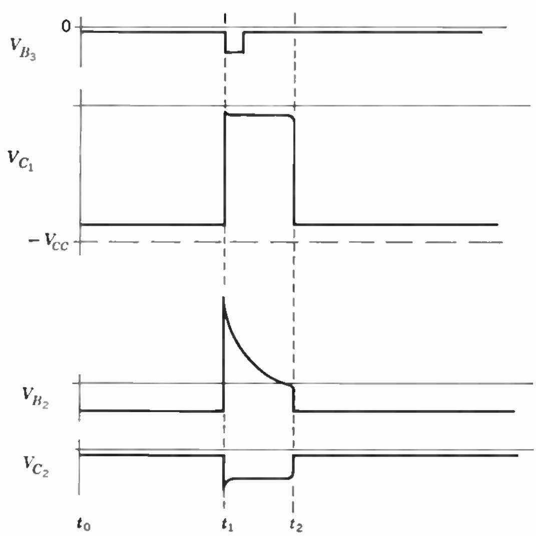


Fig. 8·14 Waveforms developed in the circuit of Fig. 8·13. V_{c_1} is the voltage at the collector of X_1; V_{c_2} is the voltage at the collector of X_2.

its collector voltage will be sufficiently negative to force X_1 into saturation. X_1 will remain in this condition until C_1 has discharged enough to permit X_2 to conduct again. This occurs at time t_2. Note that while the pulse obtained at the collector of X_2 is a function of the time constant of the circuit and is independent of the input or trigger pulse, above a minimum value, the pulse at the collectors of X_3 and X_1 may be a function of the input pulse if the input pulse has a duration longer than that determined by the time constant of the circuit. With fixed values of resistances, the output pulse duration varies linearly with the value of the capacitor in the range of 0.2 μsec to at least 130 μsec using good high-frequency transistors.

In this monostable multivibrator, the stable state occurs when X_2 is conducting and X_1 is cut off. The unstable, or transient, condition occurs when X_2 is cut off and X_1 is conducting. The output of such a multivibrator is an excellent square wave, and one of the most frequent applications of the circuit is to receive input pulses of different shapes and produce square waves at the same frequency.

The third remaining type of multivibrator is the astable, or free-running, type in which the circuit itself periodically switches one transistor from the conducting to the nonconducting state while its partner performs the reverse action. Such a multivibrator is described on page 174.

Transistor Choppers

Mention was made in a preceding chapter of the difficulties encountered in amplifying d-c or very low frequency signals. Special precautions must be taken to prevent drift of the operating point, and these almost always lead to more complex circuits with resultant higher costs, even if the desired stability can be achieved.

One alternate method which has proved successful in dealing with d-c and low-frequency signals is the transistor chopper. In this arrangement, the direct current or low-frequency alternating current is converted to a much higher frequency alternating current by opening and closing a switch, i.e., the transistor, in such a way that the signal is first switched across a load and then removed. The waveform of the a-c signal across the load will be a function of the waveform of the applied signal and the properties of the device used as the switch.

Certain types of transistor make excellent switches; i.e., they have two states: one in which they are conducting and one in which they are nonconducting. This switching action is entirely analogous to the operation of a conventional switch.

The basic circuit of a transistor chopper is shown in Fig. 8·15. A square wave is applied to the base, turning the transistor on when the voltage is negative (because this is a PNP transistor) and turning the transistor off when the voltage goes positive. At the output terminal, the collector, this has the following effect: When the transistor is cut off, the full signal appears at the collector; when the transistor is turned on, it is driven so strongly that the collector-emitter resistance is zero (ideally). This effectively places the collector at ground potential and reduces the output to zero.

Note that the amplitude of the resulting output signal, during transistor cutoff, is governed entirely by the signal voltage applied to the other end of R_1. The square-wave switching voltage does not have any effect on this output. By the same token, when the transistor is turned on, the output drops to zero and again the switching voltage does not have any effect on the output.

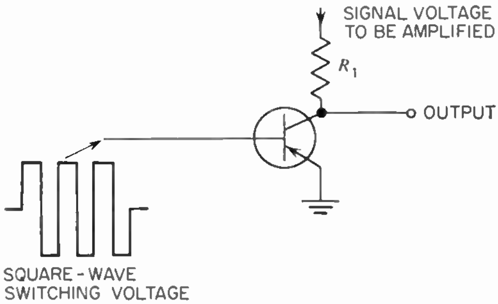


Fig. 8·15 Basic circuit of a transistor chopper. (After P. L. Schmidt, Voltage Conversion with Transistor Switches, *Bell Laboratories Record*, February, 1958)

If the applied signal is a d-c voltage, the output will be a series of square waves with an amplitude equal to this d-c voltage. If the applied signal is a low-frequency a-c voltage, the output will be a series of squares whose amplitude varies sinusoidally, Fig. 8·16.

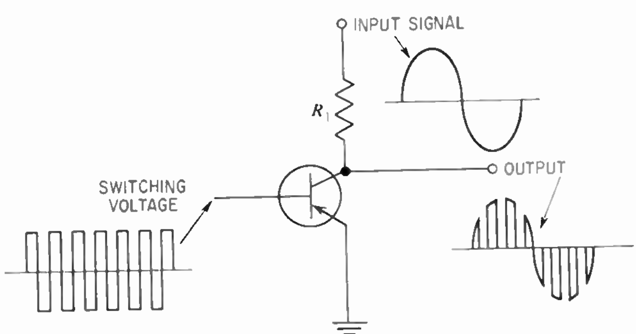


Fig. 8·16 The effect of a transistor chopper on a low-frequency a-c signal.

The switching-voltage rate is generally set at 300 to 400 cps. This is high enough to provide an output signal that is readily amplified by fairly conventional circuits and yet low enough that no difficulty is encountered from transistor junction capacitances or time constants in the amplifier.

The simple circuit of Fig. 8·15 will work satisfactorily when signals of 1 volt or greater are to be converted and amplified. However, it frequently happens that voltages of several hundred millivolts or less need to be amplified. At these levels, recognition must be given to the fact that when a transistor is cut off, its impedance is not infinite because of the leakage current that flows through the collector-base junction and causes some voltage to appear across R_1. This voltage bucks or opposes the signal voltage, producing an output voltage that is smaller than it should be. The leakage current is small, so the voltage drop it produces across R_1 is small, which is why this effect is important only with small signal voltages.

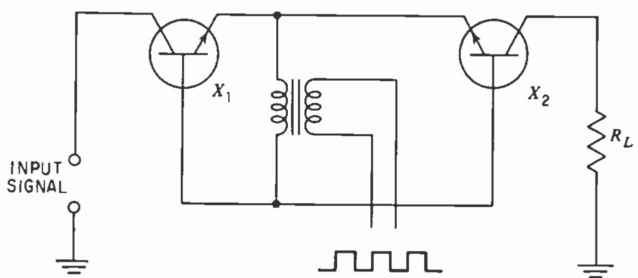


Fig. 8·17 A series-pair chopper circuit. (After R. B. Hurley, Transistorized Low-level Chopper Circuits, *Electronic Industries*, December, 1956)

The leakage current that flows when the transistor is off is referred to as the offset current.

By the same token, when the transistor is turned on, the collector-emitter resistance is not zero, but some finite value. The voltage drop across this resistance is referred to as the offset voltage. Again, this is disturbing for small signal voltages. Thus, the offset voltage and offset current limit the lowest magnitude of signal that can be accurately chopped.

Hence, when very small signals are to be amplified, the chopper circuit must be designed to minimize the magnitude of the offset voltage and current. Several ways in which this has been accomplished are described below.

The circuit in Fig. 8·17 is known as a series-pair chopper circuit. The input signal, i.e., the d-c or low-frequency signal to be converted to higher-frequency alternating current, is applied to the collector of X_1. The chopped signal is developed across R_L, in the collector circuit of X_2. The square-wave switching voltage is brought to both emitters simultaneously; thus, both X_1 and X_2 turn on and off together.

When the transistors are turned on, they simply act as a low-resistance path between input and output, permitting the input voltage to appear across R_L. The offset voltages that develop across X_1 and X_2 oppose each other because of the back-to-back connection of the two transistors. The collector currents of the two units flow in opposite directions; consequently, the voltage drops produced across the transistors by these currents will likewise be opposite. And if the transistors have been carefully selected, the two offset quantities will be equal or close to it.

Note that these collector currents flow in response to the square-wave switching voltage. Once X_1 and X_2 have been turned on, the

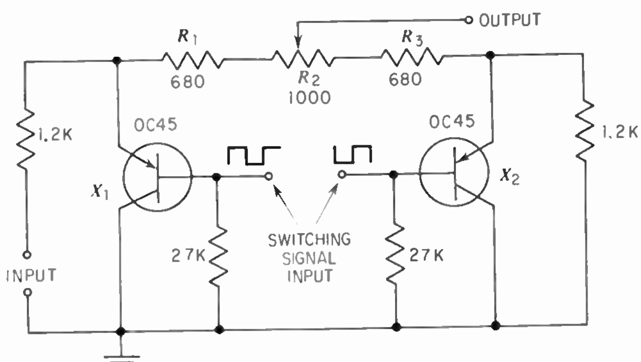


Fig. 8·18 A balanced transistor chopper.

input signal simply sees them as low-valued resistances and produces a current through them that develops an equal voltage across R_L. The transistors are here being employed as true switches.

When both transistors are turned off, the two offset currents likewise flow in opposing directions. Hence, the voltage they produce across the output is zero or very close to it.

A somewhat different approach is taken in the circuit of Fig. 8·18. The emitters of the two transistors X_1 and X_2 connect to opposite ends of a common network formed by R_1, R_2, and R_3. Both transistors are triggered on and off alternately; when X_1 is on, X_2 is off, and vice versa. With no input signal, no output voltage should be obtained. This will not be true if the leakage currents of the two transistors are unequal. To offset any difference that may exist, R_2 is made variable. Through its adjustment, a position where the output voltage is zero for zero input signal will be found. This arrangement permits a relatively inexpensive balancing of offset voltages and currents in the transistors used. When a signal voltage is applied, it will appear across the output for transfer to the a-c amplifiers that follow the chopper.

In both preceding circuits, a second transistor is required to help balance the offset voltage and current existing in the transistor to which the input signal is applied. If the effects of this voltage and current can be ignored because of the signal magnitude, then the simpler arrangement of Fig. 8·15 will serve satisfactorily. But where the signal is in the millivolt range, minimizing the effects of the offset voltage and current is quite important.

D-C to D-C Converters

The transistor, as a switching element, is being employed in a variety of other devices, one of the most interesting of which is the d-c to d-c converter. In these converters, low-voltage direct current is converted to alternating current having a frequency of several thousand cycles. The alternating current is stepped up and then rectified to provide a higher d-c voltage.

The foregoing sequence has been employed for many years in the power supplies of mobile equipment such as auto radios. Here the 6 (or 12) volts from the car battery is fed to a mechanical vibrator which changes the direct current to alternating current by simply interrupting the current mechanically several hundred times a second. The a-c voltage is then fed to a transformer, stepped up to the desired output level, and finally rectified back to direct current. Not only is this process inefficient, but the relative frailty of the mechanical vibrator results in frequent breakdowns. In the transistorized approach to this problem, efficiencies of 85 per cent can be realized, and there are no moving parts to fail. Structurally, too, the transistorized converter possesses considerable advantages.

The schematic diagram of a d-c to d-c converter is shown in Fig. 8·19. The heart of this circuit, where the conversion from direct to alternating current takes place, is shown separately in Fig. 8·20*a*. The operation of this circuit depends primarily on the properties of the transformer core material. Core materials which have a square *B-H* curve, Fig. 8·20*b*, are selected, so that very little magnetizing current change is needed until the saturation flux level is reached. Hypersil, Ferroxcube, Deltamax, and Supermalloy are typical core materials that have been successfully employed. The abrupt saturation of the core produces changes in the feedback voltage which force the transistors to switch instantly when the transformer voltages become negligible.

To understand the operation of the circuit in Fig. 8·20*a*, let us suppose that the d-c power has been applied. A slight unbalance in the circuit will cause more current to flow through one transistor than the other. Let us assume that more current flows through transistor X_1 and winding 2 than through transistor X_2 and winding 3. Assume also that

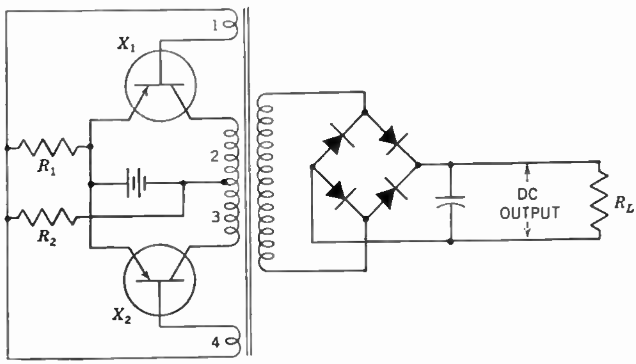


Fig. 8·19 **A d-c to d-c converter circuit. (*Bell Laboratories Record*)**

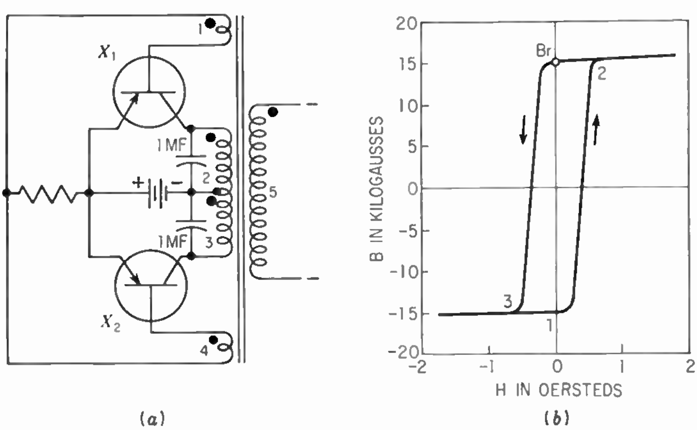


Fig. 8·20 **Basic circuit of the (a) transistor-core inverter and (b) dynamic hysteresis loop of the core. (*Bell Laboratories Record*)**

the core of the transformer is in the magnetic state represented by 1 in Fig. 8·20*b*. The flux in the transformer core now begins to change toward positive saturation, inducing a voltage in windings 1 and 4. This voltage makes the windings positive at the dots shown, while the other end of each winding becomes relatively negative.

The voltage induced in winding 1 drives transistor X_1 quickly into saturation, while the voltage across winding 4 forces X_2 to cut off. The total battery voltage, except for a small voltage drop across X_1, will appear across winding 2, and the flux in the core will change at a constant rate determined by the voltage, the number of turns, and the core area. This constant rate will continue until positive saturation of the core is reached at point 2, Fig. 8·20*b*. At this instant, the current through the winding will increase sharply, driving the core deep into saturation. The induced voltages in the windings drop sharply, forcing transistor X_1 to cut off. This quickly checks the flow of current in winding 2.

With this drop in winding-2 current, the core saturation swings back to point B_r, its residual value. This induces a voltage of opposite polarity in the windings, causing transistor X_2 to start conducting. With X_2 on and conducting quite heavily (actually, it is quickly driven into saturation), the full d-c battery voltage is essentially applied across winding 3. With the current now flowing in a direction opposite to its travel through winding 2, the flux increases steadily in the opposite direction. During the interval the flux is changing from positive to negative saturation, the induced voltages are high enough to keep X_1 completely cut off and to saturate X_2.

When point 3 is reached, the current in winding 3 increases sharply, driving the core far into saturation. The induced voltages again fall quickly to zero, cutting off X_2 and thereby checking the current in winding 3. It is at this point that the switchover occurs again, with the core flux now moving toward positive saturation. The action thus switches back and forth at a rate determined by the following factors: the saturation flux density of the core, the cross-sectional area and number of turns in the windings, and the value of the d-c voltage applied.

The output, which appears across winding 5, is close to a square wave. This is rectified by a full-wave rectifier to provide a d-c voltage of higher level than the energizing source. The necessary filtering can usually be provided by a single capacitor because of the fairly high frequency of the a-c square wave.

In the complete circuit, an additional resistor R_2 has been added to ensure circuit starting when the load on the system is high. This resistor biases the transistors into action when the circuit is energized initially and oscillation starts immediately because of the increased unbalance of the circuit. Also, 1-μf capacitors are shunted across windings 2 and 3 to suppress voltage spikes which would develop during the changeover periods due to the inductance in the circuit. These high-voltage spikes could damage the transistors.

An idea of the small size of some of these d-c to d-c converters can be seen in Fig. 8·21. The unit shown is designed for 30-watt output and 24-volt input. The casing has outside dimensions of 2 by 3 by 4 in. and is frequently hermetically sealed.

The high efficiency of the converter stems chiefly from the fact that when a transistor (X_1 or X_2) is driven into saturation, the voltage drop from emitter to collector is less than 0.5 volt. An interesting feature of this device is that the transistor can safely handle power which is perhaps as much as 10 times greater than its dissipation rating. This is because the switching time can be made small enough that the average dissipation is low. Maximum power obtainable from the con-

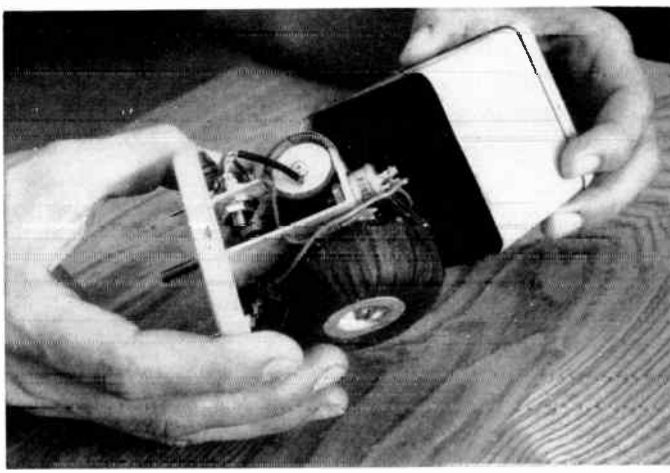


Fig. 8·21 The physical size of the d-c to d-c converter shown in Fig. 8·19. (*Bell Laboratories Record*)

verter is equal to the product of the battery voltage, the maximum current that flows through the transistors, and the circuit efficiency. Typical efficiency values range beween 80 and 90 per cent.

Before we leave the subject of d-c to d-c converters, it should be noted that the circuit of Fig. 8·19 can also be employed without a rectifier in the output. In this case, the direct current of the battery is converted to a higher-value alternating current and then this alternating current is employed directly. In this form, the unit is known as a d-c to a-c inverter.

One application of inverters is with magnetic amplifiers, where the direct current of a battery is transformed to alternating current and then utilized as a high-frequency carrier supply for such magnetic amplifiers. One of the disadvantages of a magnetic amplifier is that it possesses a slow response when operated from the usual 60- or 400-

cycle power source. This can be overcome by raising the frequency of the power source feeding the magnetic amplifier, however. With conventional means, such as generators, this presents a formidable obstacle; with a d-c to a-c inverter, the job can be accomplished simply and with considerable compactness.

Another use for inverters is in telemetering equipment, where information concerning the conditions in a system must be sent over considerable distances. A d-c to a-c inverter is extremely useful here because its output frequency is proportional to the applied d-c voltage. The latter quantity, in turn, can represent temperature, current flow, pressure, or any other quantity about which information is desired. Guided missiles and other unmanned aircraft make extensive use of telemetry for this purpose.

Transistors in Power Supplies

One extensive application which transistors have in common with tubes is in the regulation of the voltage or current output of a power supply. There are a variety of circuit arrangements possible, and some of the more basic ones are examined below.

Transistor shunt regulator. A shunt-regulator arrangement, where the transistor is actually in shunt with the output, is shown in Fig. 8·22. The collector of the regulating transistor is connected to the

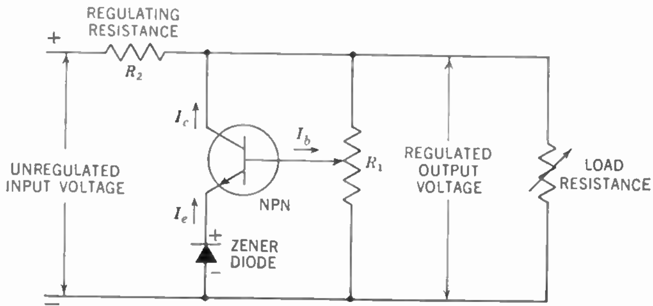


Fig. 8·22 A transistor shunt-regulator circuit.

positive side of the output voltage line. Since an NPN transistor is being used, the positive voltage will reverse-bias the collector. The emitter is connected to the negative side of the output voltage line through a reference voltage diode. Another name for this reference diode is Zener diode, and if the potential applied to the diode is examined carefully, it will be seen that it is reverse-biased.

The characteristic curve of the Zener diode is shown in Fig. 8·23. When this diode is biased in the forward direction, the current flow through the unit will rise quite sharply at fairly low biasing voltages. When it is reverse-biased, however, it will be found that the current is minute, on the order of microamperes, until a certain voltage, called the saturation voltage, is reached. At this point, the electrons or holes which form the leakage current are given sufficient energy to create other electron-hole pairs which add to the initial reverse current. This process builds up rapidly and leads to large increases in current for small further increases in voltage. The diode is now in the saturation region, and any attempt of the reverse voltage to rise is met by an increased current flow which tends to counteract the voltage

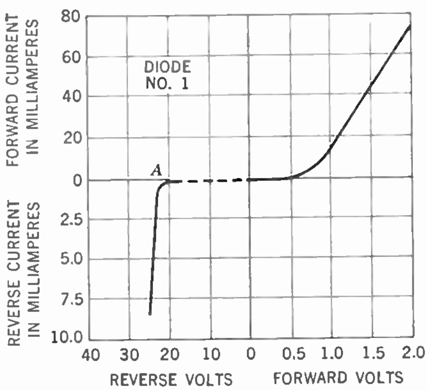


Fig. 8·23 Characteristic curve of a Zener diode.

increase and thus maintain a steady voltage drop across the diode. This is indicated in Fig. 8·23, where the characteristic curve at point *A* drops almost straight down. This signifies that any attempt to increase the reverse bias voltage is met by an increase in current, but practically no increase in voltage drop. (To maintain these conditions, it is apparent that the internal resistance of the diode must also change. This is implied in the foregoing sequence of events.)

In this state, the Zener reference diode establishes a fixed voltage. In the regulatory circuit of Fig. 8·22, this diode characteristic, in combination with the transistor, acts to maintain a fairly constant output voltage in the following manner. The base of the transistor is connected to the movable arm of potentiometer R_1 and the arm is so set that the base potential is slightly positive with respect to the emitter voltage. In this condition, current I_b flows in the base circuit while current I_c flows in the collector circuit. Both of these currents combine to form the emitter current I_e.

Now, if the regulated output voltage increases, the base will become more positive than it was before. Since the emitter voltage is fixed by the reference diode voltage, the net effect will be to increase the base-to-emitter potential. This will increase the current through the transistor, which, in turn, will increase the voltage drop across the regulating resistance R_2 and decrease the output voltage. The correcting process continues until the output voltage is brought back to its initial value.

The process is essentially the same for a decrease in output voltage, except that now the base-to-emitter voltage decreases, the collector current drops, less voltage is developed across R_2, and the output voltage is brought back to normal again. All these changes require less than a second to occur.

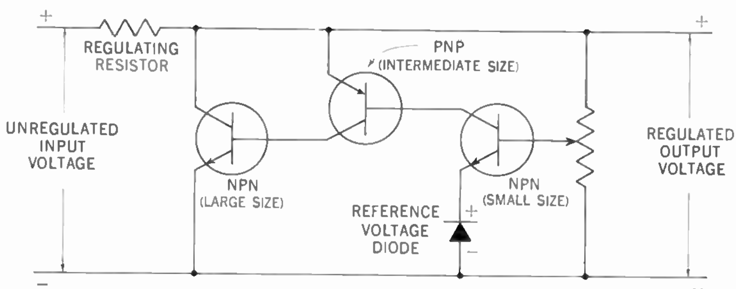


Fig. 8·24 A transistor shunt regulator using three transistors.

The value of the regulated output voltage is determined by the potentiometer setting. This, in turn, is limited by the operating range of the transistor as a function of base-to-emitter voltages. There can be no more voltage difference between the base and emitter than that which produces the safe maximum collector current.

Stability of this arrangement will depend on the constancy of the voltage drop across the reference diode with changing values of current. The smaller the current variation through the diode, the more constant the voltage drop. One way to achieve this, while still exercising effective control of the output voltage over a range of voltages, is through the use of several stages of transistor regulators. This is shown in Fig. 8·24.

In this circuit, two more transistors have been added to the preceding circuit. The first stage compares the output potential to the reference potential and drives the second stage, which, in turn, drives the third stage. The first stage functions as before, except that its collector current becomes the base current for the second transistor. The collector current of the second transistor is an amplified version of its base current (and, therefore, the collector current of the first stage). This second collector current becomes the base current for the third transistor and, at the output of the final transistor, the collector current is a still further amplified version of its base or input current.

No coupling networks are needed between stages because the second transistor is a PNP transistor while the first and third are NPN units. Here we are making use of complementary symmetry, and the result is a circuit which is reduced to bare essentials.

In Fig. 8·24 different sizes are specified for the three transistors. The first unit might be a 50-mw transistor operating at a collector potential of 10 volts. Then the maximum base current of the second-stage PNP transistor should not exceed 5 ma. If we assume an am-

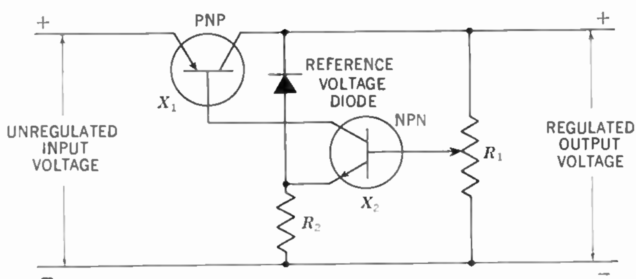


Fig. 8·25 A transistor series regulator circuit.

plification factor of 20 for the second transistor, its output collector current will be 100 ma. This, now, becomes the base current for the third transistor and, with the same amplification, a final collector current of 2 amp is obtained. Hence, the final transistor must be a power unit designed to operate at this level. Such transistors are available.

Series regulators. A regulation circuit which is more efficient than the preceding arrangement is the series circuit shown in Fig. 8·25. A PNP transistor is so connected that all of the load current must pass through it. The base of this transistor, X_1, is directly connected to the collector of another transistor, X_2, so that the base current of X_1 is determined by X_2. This base current, in turn, will determine how much collector current flows through X_1 and the output load.

In detail, the circuit operates as follows. The emitter of X_2 is held at a constant potential with respect to the *positive* output terminal. (This differs from the arrangement in Fig. 8·22, where the emitter is held at a constant potential with respect to the *negative* output terminal.) When the output voltage tends to increase, the base voltage rises, but not as much as the emitter voltage. Since the emitter voltage remains fixed with respect to the positive output voltage, the net effect of this action is to reduce the base-to-emitter voltage. This lowers the collector current of X_2 and also the base current of X_1. The collector current of X_1 is thereby reduced, lowering the current flowing through the output load resistance. Thus, the output voltage rise is counteracted.

When it is stated that the emitter of X_2 is held at a constant potential with respect to the positive output terminal, it means that when the output voltage rises, the emitter potential increases by exactly the same amount (in order to keep their difference constant). On the

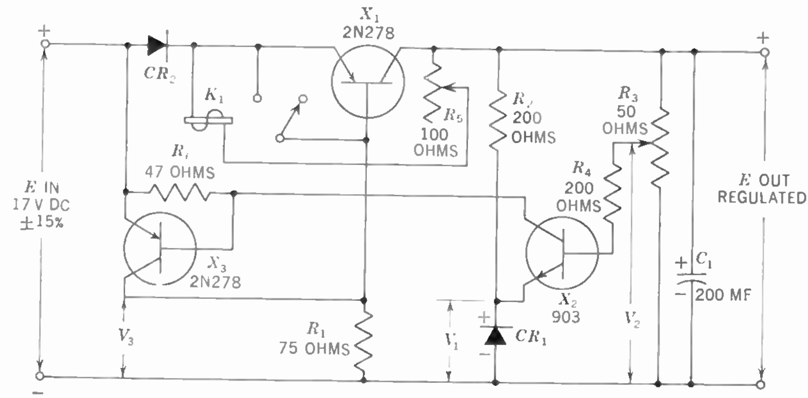


Fig. 8·26 A more elaborate transistor series regulator circuit. (Delco Radio Div.)

other hand, the base voltage rise, under the same conditions, is governed by the setting of the center arm on R_1. If this arm is at the midpoint of the resistor, the base receives only one-half of the rise. Thus, it is seen that the emitter voltage will rise more than the base voltage. With a decrease in positive output voltage, the emitter voltage will drop *more* than the base voltage. That is, the difference between the positive output terminal and the emitter of X_2 remains constant, no matter what the output voltage may do.

The setting of the arm on potentiometer R_1 will determine the value of the regulated output voltage. Resistor R_2 is selected to keep the reference diode in its saturated voltage region.

While the regulation circuit of Fig. 8·25 is more efficient than the circuit of Fig. 8·24, X_1 must be a high-current transistor because it passes all of the circuit current. This would probably make this arrangement more expensive than the prior one. A somewhat elaborate variation of this circuit is shown in Fig. 8·26. The basic purpose is still the same, i.e., to vary the collector current through X_1 as the output voltage varies. V_1 is the voltage drop across the reference diode CR_1. V_2 is a function of the output voltage. The voltage difference between V_1 and V_2 is used to bias the base of transistor X_2. To understand the operation of the circuit, assume that the output voltage has increased because of a change in the load. This calls for a decrease of the collector current in transistor X_1 so the output voltage will return to its former value. This decrease is accomplished in the following manner.

The rise in output voltage raises V_2 and increases the difference between V_1 and V_2. This causes the base and collector currents of X_2 to rise. In turn, the base and collector currents of X_3 rise. The increased collector current of X_3 produces a greater voltage drop across R_1, with the top end becoming more positive. This works counter to the bias requirement of X_1 and forces the collector current in this transistor to decrease, thereby bringing the output voltage back to its regulated value.

In this arrangement, the effect on X_2 for a change in output voltage is opposite to what it is for the same transistor in Fig. 8·25. This stems entirely from the difference in placement of CR_1. In one case, Fig. 8·26, it is connected directly to the negative voltage line, while in Fig. 8·25, it is tied to the positive line. Because of this difference, a second transistor is inserted between X_2 and X_1. This is X_3 in Fig. 8·26.

Should the voltage across X_1 rise above a preset value, it will cause the voltage-sensitive relay K_1 to operate. The excess voltage could be produced by a large increase in input voltage, by an overload, or by a short-circuited output. When relay K_1 operates, the emitter and base of X_1 are shorted together. This cuts off the collector current of X_1. In this circuit, a fuse or circuit breaker would not react in time to protect the circuit from damage. Resistor R_2 provides a bias current for the voltage reference diode CR_1. R_4 protects X_2 and CR_1 in the event X_1 should short from emitter to collector. Resistor R_5 provides an adjustment to control the operation of relay K_1. Capacitor C_1 prevents high-frequency oscillations by reducing the loop gain to a safe level for frequencies at which the internal phase shift of the transistors would be significant. The forward voltage drop across diode CR_2 provides additional collector bias for X_3. Thus, transistor X_3 is able to cut off the collector current of X_1. This diode is particularly useful at high temperatures. Finally, resistor R_6 reduces the collector leakage current of transistor X_3.

The efficacy of this arrangement is revealed by the performance curves shown in Fig. 8·27. Note that as the load (i.e., output wattage) varies from approximately 2 watts to 25 watts, the output voltage decreases only 0.6 per cent, or from 12.60 volts (normal) to 12.53 volts for an unregulated input voltage of 14 volts. When the input voltage is 20 volts, the percentage variation is even less.

Series current regulator. Thus far, the circuits have regulated for a constant output voltage. However, in many applications a constant current is desired and, to achieve this, the circuit shown in Fig. 8·28

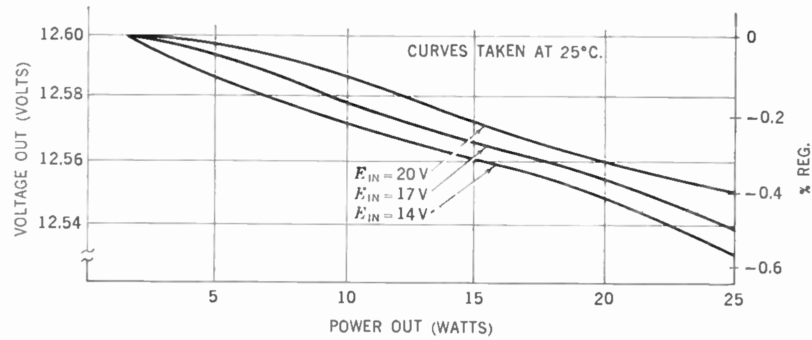


Fig. 8·27 Regulation characteristics of the circuit shown in Fig. 8·26. (Delco Radio Div.)

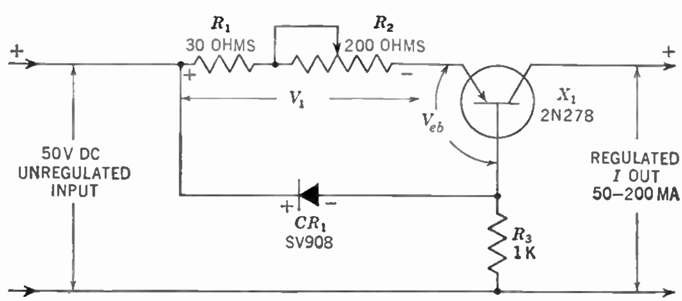


Fig. 8·28 A transistor current regulator. (Delco Radio Div.)

may be employed. All of the current must pass through transistor X_1; hence, the object of this circuit is to have changes in the output current alter the internal resistance of X_1 so that these changes are counteracted. This is achieved in Fig. 8·28 by causing all of the output current to pass through R_1 and R_2. A voltage drop is produced across these resistors with the polarity indicated. At the same time, reference diode CR_1 establishes a fixed (or reference) voltage with a similar polarity.

The voltage applied between base and emitter, V_{eb}, is then equal to

$$V_{eb} = V_d - V_1$$

where V_{eb} = base-to-emitter voltage
V_d = voltage established across reference diode
V_1 = voltage drop across resistors R_1 and R_2

A typical value of V_{eb} is 0.2 volt. In this specific circuit, V_d is approximately 8 volts.

Now, as the load current varies, it will vary the voltage developed across R_1 and R_2. This will cause the base-to-emitter voltage to change because V_d remains fixed. A variation in V_{eb} will cause the collector current to vary, and thus regulation is achieved.

R_2 is made variable to permit adjustment of the output current to a specified level. Thereafter the circuit tends to keep the current at this value. R_3, the 1,000-ohm resistor in the base load of the transistor, provides a keep-alive current for the reference diode. With loading, the transistor supplies a small portion of the current drawn by R_3; therefore, less current will flow through the reference diode. R_3 must draw enough current through the reference diode that the voltage drop across the diode will remain at 8 volts as the current regulator is loaded.

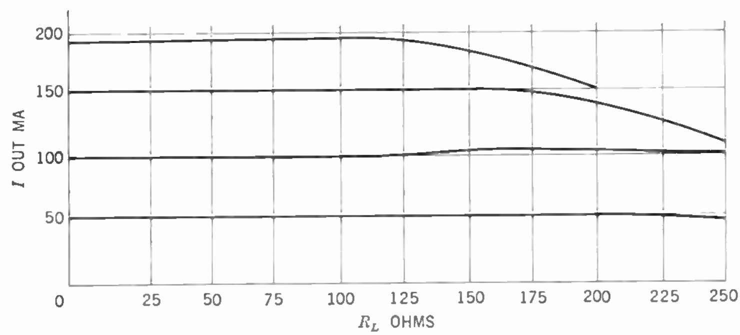


Fig. 8·29 Performance characteristics of the current regulator of Fig. 8·28. (Delco Radio Div.)

A graph showing the performance characteristics of this circuit is given in Fig. 8·29. In general, the current supplied to the load will remain essentially constant until R_L, the load resistance, is increased in value to the point where the voltage drop across R_L is as large as the voltage drop across R_3.

Null Detector

Null detectors are useful instruments in the laboratory, on the production line, and in the shop. One of their principal functions is to detect the low or zero point between two adjacent peaking points, such as we frequently find when aligning a resonant system. Or it may be that a certain control has to be adjusted to a specific setting and, by the proper test setup, a null detector will assist in this operation.

The circuit diagram of a null detector which has high sensitivity, moderately fast response to input-level changes, and the ability to withstand a large overload is shown in Fig. 8·30. Basically, the system is a high-gain audio amplifier whose output is rectified and then fed to a d-c microammeter. When an a-f signal is fed into this circuit, it is amplified by the four transistor stages and then applied to transformer T_2, where it is coupled to the secondary. Here, rectification takes place on both halves of the cycle. This can be seen by noting

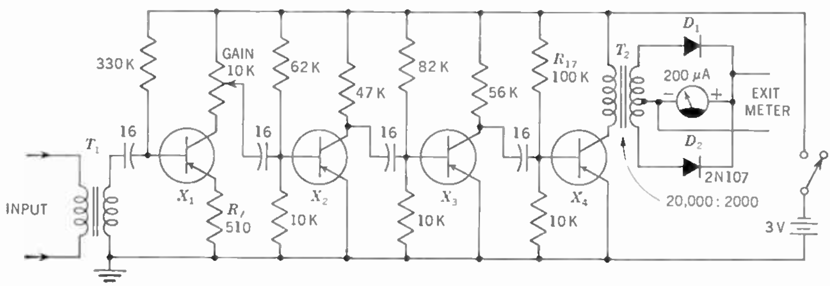


Fig. 8·30 A null detector. (After C. D. Todd, Transistor Null Detector Has High Sensitivity, *Electronics;* Feb. 1, 1957)

that when the induced voltage in the secondary of T_2 is such that the top end of the winding is positive (and the bottom end is negative), diode D_1 will conduct. Current will flow from the center tap, through the microammeter to D_1 and through D_1 to the upper end of the winding. During this interval, D_2 will be cut off because the signal voltage across the bottom half of the winding will reverse-bias the diode.

On the next half cycle, the top end of the secondary winding will be negative and the bottom end will be positive. This will bring diode D_2 into conduction, and now current will flow from the center tap through the microammeter to D_2 and then to the bottom end of the winding. During this interval, diode D_1 will be reverse-biased and will not conduct.

At the input to the system, a small 1:1 transformer is employed. An input signal of 20 mv will cause full-scale deflection of the null meter. If higher input voltages are encountered, a voltage-dividing network will be required.

Position Pick-off, or Proximity, Limit Switch

The circuit shown in Fig. 8·31 is a position pick-off, or proximity, limit switch. The system consists of a two-stage transistor amplifier in which the first stage acts as a linear amplifier and the second stage is biased to act as a detector stage. That is, the d-c collector current in transistor X_2 is almost linearly proportional to the a-c signal appearing between its base and emitter. An a-c voltage, obtained from the

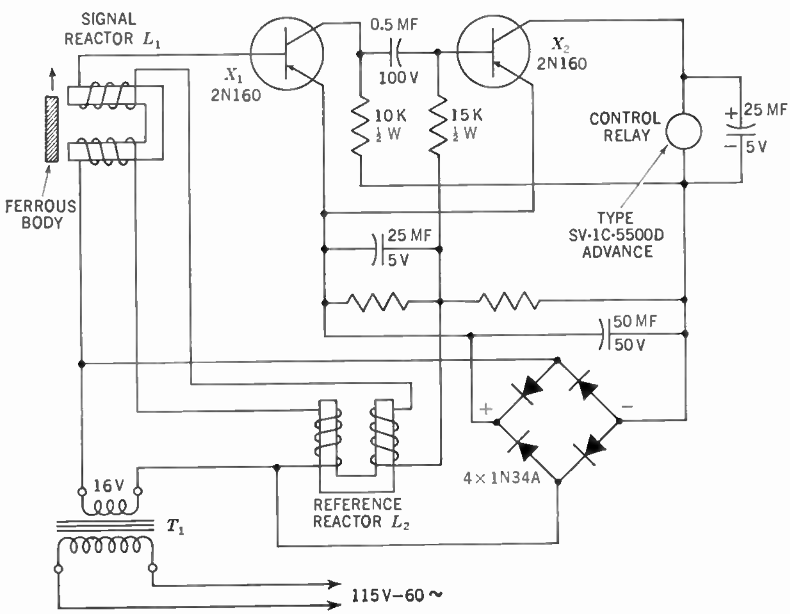


Fig. 8·31 The circuit of a ferrous-type proximity pick-off. (After J. R. Walker, Applying Transistors, *Automation*, February, 1958)

secondary of T_1, is applied in equal measure to one winding of signal reactor L_1 and one winding of reference reactor L_2. Coupled to each of these windings is a secondary winding and, in the absence of any ferrous body near L_1, the voltages induced in these secondary windings will cancel completely because they are connected in series opposition. Under this condition, no a-c signal will be applied to the base of X_1 and the control relay in the collector circuit of X_2 will not be energized.

When a ferrous body is brought close to L_1, an unbalance is created and a resultant a-c voltage will be fed to X_1. This signal is amplified and then applied to X_2, where the output current will energize the control relay.

It will be found that the amplitude of the a-c voltage present at the base of X_1 is inversely proportional to the spacing between L_1 and the ferrous body. If this spacing changes at a steady rate, i.e., linearly with time, the reactor voltage will change much more rapidly, and, as a result, the rise in amplified output at the base of X_2 will be steep in its change with time. Since the d-c collector current is proportional to the input at the base, this current will also possess a steep form. Repeated actuation of the control relay can be attained with a spacing

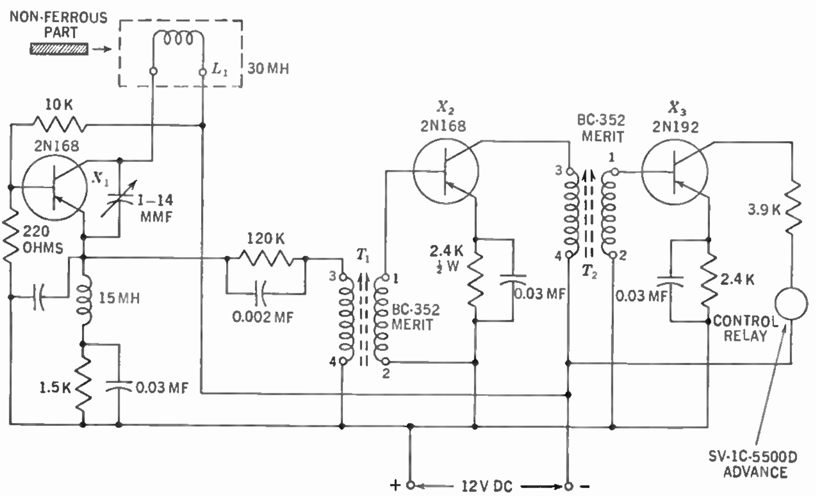


Fig. 8·32 An r-f proximity pick-off circuit. (After J. R. Walker, Applying Transistors, *Automation*, February, 1958)

between the pickup and the ferrous body of 0.5 in. and a longitudinal part travel of 0.015 in.

The foregoing circuit is useful where ferrous objects only are to be detected because of the use of 60 cycles as the a-c signal. For the detection of metallic bodies (nonferrous as well as ferrous), a circuit such as that shown in Fig. 8·32 can be employed. The initial stage is a 460-kc oscillator in which the operating frequency is determined by the inductance of L_1 and the variable capacitor in the circuit. The signal produced by the oscillator is transferred via tuned transformer T_1 to X_2. Here it is amplified and then transferred by means of tuned transformer T_2 to the linear detector stage X_3.

In the absence of a metallic body, the oscillator is tuned to 460 kc. Transformers T_1 and T_2, however, are tuned slightly off resonance, so that the amplitude of the 460-kc signal reaching the detector X_3 is small and the control relay is *not* actuated. However, if a metallic object is brought near pick-off coil L_1, its inductance will change. This will change the oscillator frequency to an on-resonance value for T_1 and T_2, and a large signal will be applied to X_3. This will increase the average value of the d-c collector current in X_3 and the control relay will be energized.

This circuit will operate on an object–to–pick-off spacing of ⅝ in. and a longitudinal part travel of 0.030 in.

Transistorized Counters

Extensive use is made in industry of devices which are capable of counting a number of events and either stopping the operation after a

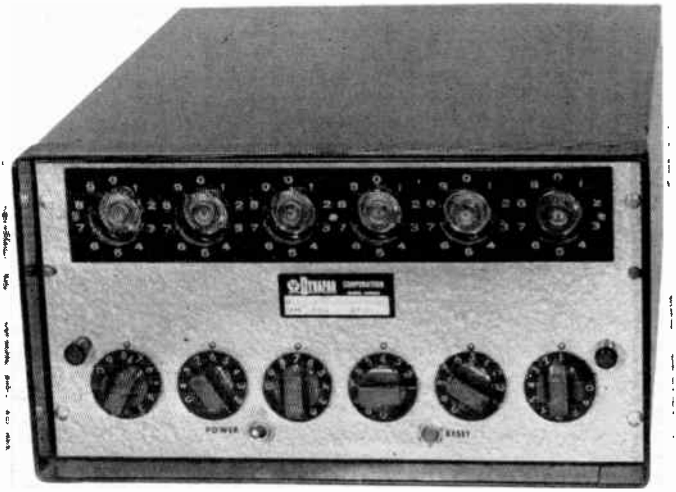


Fig. 8·33 A transistorized preset counter. This instrument contains two identical sets of counting circuits to enable it to perform a dual function. One counting circuit is shown in Fig. 8·37. (Dynapar Corp.)

certain number has been reached or else simply keeping track of the count so that a running total is always available. A preset counter which utilizes transistors throughout its design is shown in Fig. 8·33. This counter will respond to an incoming count rate up to 5,000 pulses per second with a sensitivity of 50 mv at 20,000 ohms input. Pulses with amplitudes of 150 volts (rms) can be handled, although a special potentiometer is required above 4 volts.

The major components of the circuit include cold-cathode decade counter tubes, multivibrators either for more power output or for pulse sharpening, and coincident circuits to trigger the necessary relays when the preset level has been attained. All transistors are type GT34HV, the HV designation signifying their ability to operate with higher than normal voltages. Up to six digits can be handled by the counter, either in single or dual arrangement, by the simple repetition of the basic coincident circuitry.

Before a discussion of the circuit is undertaken, it might be desirable to examine first the internal structure and operation of a cold-cathode decade counter tube. One such tube, shown in Fig. 8·34, op-

Fig. 8·34 A 6476 cold-cathode decade counter tube. (Sylvania Electric Products, Inc.)

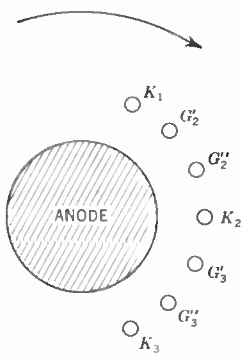


Fig. 8·35 Partial electrode arrangement in the counter tube of Fig. 8·34.

erates on the principle that the ionization or starting voltage of gas-filled tubes is lowered if ions or electrons are already present in the anode-cathode gap. Under these conditions, a glow discharge can be made to move from one cathode to an adjacent one by applying a relatively small negative voltage pulse to the new cathode, provided that electrons or ions are able to diffuse into this new anode-cathode gap. If a number of cathodes are positioned around a common anode, the glow discharge can be made to move in succession around the cathode circle by the successive application of negative voltage pulses to each cathode.

The tube shown in Fig. 8·34 has 30 cathodes equally spaced on a circle about a central anode disk. The cathodes are divided into 10 main, or output, cathodes and 20 intermediate guide cathodes—two of which are located between every main-cathode interval, Fig. 8·35.

The two guide cathodes are necessary to make certain that the glow always moves in the desired direction and does not return to the original cathode upon removal of the voltage from the new cathode. A description of the transfer mechanism from one main cathode to the next adjacent one will make this clearer. Assuming that a glow is present on K_1, a negative pulse on G'_2, Fig. 8·35, will move the glow to G'_2. At the end of this pulse, a second negative pulse is applied to G''_2, moving the glow on to G''_2. Because the guide cathodes are normally biased above the main cathodes, removal of the negative

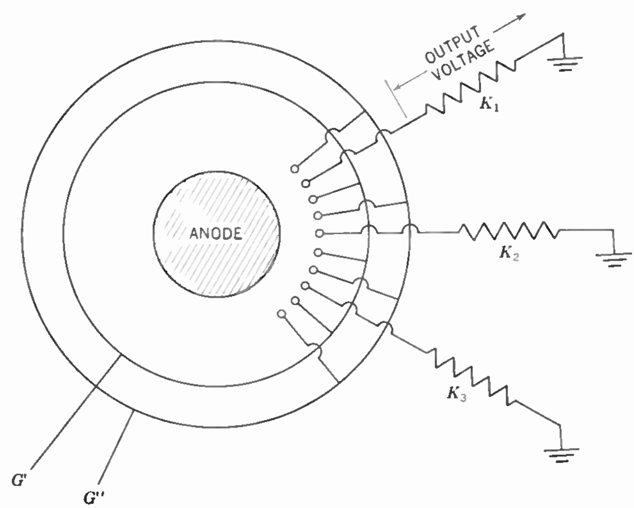


Fig. 8·36 Electrode arrangement in counter tube with circuit connections.

pulse from G''_2 means that the glow will go to a main cathode. The preferential influence of the ionization near K_2 causes the glow to settle on K_2 rather than return to K_1. The glow will remain on K_2 until the sequence is repeated for the next set of cathodes. In practice, all of the G' electrodes are connected together, as are also all of the G'' electrodes, Fig. 8·36. Although the negative transfer pulse is applied to all ten guide cathodes simultaneously, the priming influence of the discharge moves the glow to the guide cathode which is closest to the discharge.

A positive-output-voltage pulse can be obtained from each main cathode as the glow moves onto it. It should also be noted that reversal of the pulse sequence applied to G' and G'' will reverse the direction of the glow transfer.

The circuit diagram of the counter is shown in Fig. 8·37. Normal input signals from 50 mv to 4 volts can be applied to input terminal A; for higher voltages, up to 150 volts (rms), terminal B must be used and the 200,000-ohm potentiometer must be adjusted to the proper level. The incoming signal is amplified by X_1 and then applied to R_1 and, from there, to the base of X_2. The setting of R_1 is determined by the polarity of the incoming pulse, and both negative and positive polarities can be accommodated. This is accomplished in the following manner. The emitter of X_2 is connected to ground. Since this is a PNP transistor, the emitter-base junction will become forward-biased when the base is given a negative voltage. On the other hand, a positive or even a zero voltage at the base will cause the junction to be essentially reverse-biased and hence cut off.

Now, R_1 is part of a voltage-divider network containing R_2 and R_3 (as well as R_1). The top end of this network connects to +30 volts, while the bottom end of R_3 goes to —45 volts. When the center arm of R_1 is turned to point C, the base of X_2 receives a negative voltage and X_2 is driven into conduction. On the other hand, when the center arm of R_1 is turned to the end closest to point D, a positive voltage is applied to the base of X_2 and it becomes reverse-biased.

When positive pulses are fed to X_1, negative pulses appear at its collector. If R_1 is set near point D, X_2 is cut off. However, the arriving negative pulses at the collector of X_1 will overcome the positive bias at the base of X_2 and drive X_2 into conduction. Once current flows through X_2, it will also flow through X_3, since the base of this transistor is negative and the unit is waiting for current to flow in its emitter circuit. Thus, driving X_2 into conduction causes current to flow through X_2 and X_3, and a positive voltage pulse will appear across R_6.

If negative pulses are applied to the base of X_1, then positive pulses will appear at the collector of this transistor. If R_1 is set so that the base of X_2 is already positive, and hence nonconducting, then the positive pulses will have no effect. For the positive pulses to produce an effect, R_1 is set so that the base receives a negative voltage. This causes X_2 to conduct, together with X_3. When a positive pulse appears at the collector of X_1, it drives X_2 into cutoff. This also cuts off X_3 and forces the voltage at the collector of X_3 (at the top end of R_6) to go sharply negative.

The signal across R_6 is fed to a Schmitt oscillator circuit, X_4 and X_5. Here the pulses are sharpened in form so that their sides are steep and they are quite flat on top. Then the signal is fed to a one-shot multivibrator (X_6 and X_7), where they are further amplified.

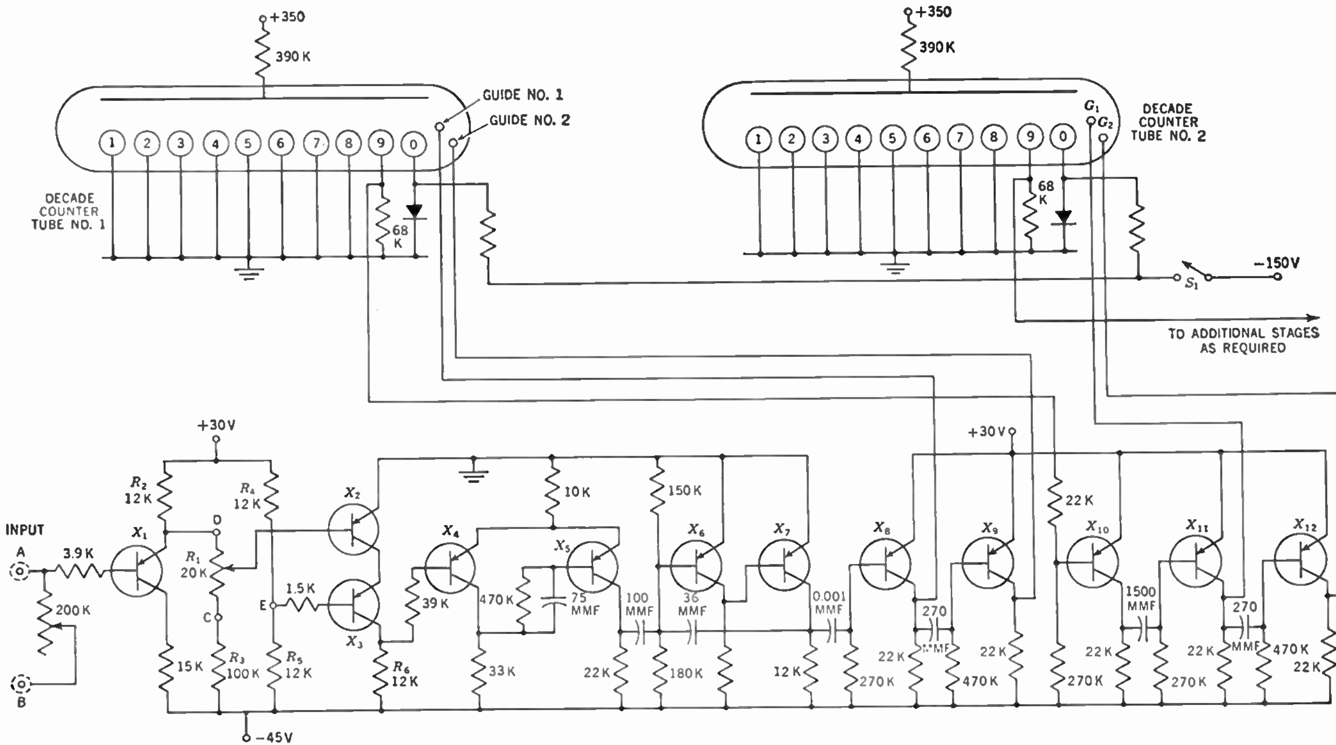


Fig. 8·37 The circuit diagram of a transistorized counter.

The reason for using both a Schmitt oscillator and a one-shot multivibrator stems from the fact that the pulses sent to the counter tubes (to be discussed in a moment) should possess a constant level and a fairly constant form. If we vary both the frequency and the amplitude of the pulses which the counter tubes receive, the counting action tends to be unstable. By using the Schmitt circuit to produce clean-cut pulses and the one-shot multivibrator to provide pulses of constant amplitude and duration, reliable counter operation is always obtained.

Note, though, that if the frequency of the incoming pulses varies, variation will be retained in the pulses reaching the counter tubes. This, of course, is important, for otherwise the system would not operate.

The output from X_7 is applied to X_8 for additional amplification. X_8 is ordinarily conducting. When a positive pulse is received from X_7, however, X_8 is driven sharply into cutoff. This causes the voltage at the collector of X_8 to drop from approximately +30 volts (when X_8 is conducting) to −45 volts when it is cut off. This negative pulse is fed to guide 1 of counter tube. 1. As indicated in the preceding discussion, it is necessary to apply a negative pulse to guide 1 and then to guide 2 in order to swing the glow discharge from one cathode to the next one. When X_8 is cut off, a negative pulse is sent to guide 1. This same negative pulse is applied to the base of X_9 through a 270-$\mu\mu$f capacitor. Since X_9 is already conducting strongly, the negative pulse is without effect. However, if we refer back to the base of X_8, we see that the positive pulse is applied to this transistor through an *RC* circuit consisting of a 0.001-μf capacitor and a 270,000-ohm resistor. The base of X_8 is held positive until the 0.001-μf capacitor has had a chance to discharge sufficiently. Then X_8 comes out of cutoff sharply, causing the voltage at its collector to rise to approximately +30 volts (from the −45 volts it possessed during cutoff). This positive pulse cuts X_9 off, producing a negative pulse at its collector which is transferred to guide 2, thereby completing the action of swinging the glow discharge from one main cathode to the next.

Initially, the counter-tube glow discharge is at the zero mark. When the action discussed above occurs and both guide 1 and guide 2 are pulsed negatively, the glow discharge swings to main cathode 1, etc. The counting action continues until the glow discharge reaches main cathode 9. With current flowing to this element, a positive voltage is developed across the resistor attached to this cathode. (This action is similar to the action which takes place across the cathode resistor of a conventional electron tube.) The base of transistor X_{10} is connected to cathode 9, and the positive voltage at the cathode tends to cut X_{10}

off. When the next incoming pulse arrives, however, the glow discharge swings to the zero starting cathode and the positive voltage across the cathode 9 resistor drops to zero. This permits X_{10} to come sharply out of cutoff, producing a positive pulse at its collector. This pulse, applied to the base of X_{11}, drives this stage into cutoff, and the resulting negative pulse at the collector of X_{11} is fed to guide 1 of counter tube 2. The action of X_{11} and X_{12} is identical to the action of X_8 and X_9, with the result that the glow discharge on the zero mark of counter tube 2 is swung over to the 1 position.

Thus, if we have counter tube 2 mounted to the left of counter tube 1 on the front panel, the visual indication will be 10. Every time the glow discharge of counter tube 1 swings around the tube and moves from position 9 to position 0, counter tube 2 moves from one main cathode to the next. When 99 is reached, a pulse from cathode 9 of the second counter tube is transferred to a third counter tube (recording the hundreds numbers) by means of a network similar to that of X_{10}, X_{11}, and X_{12}. As many similar circuits can be arranged sequentially as desired.

If it is desired to reset all the counter tubes to zero, a large negative voltage is placed on the zero cathode. This is done in Fig. 8·37 by closing manual switch S_1.

Many things can be done with this counting circuit. It can be made to trigger one or more relays after a certain number has been reached, producing a visual indication or automatically stopping an operation. Or it can be employed to start a second sequence of events after a preset number has been reached. The same circuit will measure frequency, control engine speed, regulate automatic feeds, cut material to length, and make continuous thickness measurements. Counting circuits are basic to many industrial processes, and the foregoing arrangement is an example of how such circuits can be designed wholly with transistors.

The foregoing represent some of the applications made with transistors outside the fields of radio and television. The range of such usage is so broad that only a few typical illustrations could be examined here. However, with an understanding of basic transistor operation and its use in the circuits which have been discussed, the reader is well fortified to unravel any other uses to which the transistor may be put.

QUESTIONS

8·1 Draw the diagram of a simple transistor switching circuit. Explain how it operates.

8·2 Why does it sometimes happen that the output pulse of a transistor amplifier is wider than the input pulse?

8·3 How can pulse broadening of the type described in Question 8·2 be prevented?

8·4 Explain how the circuit of Fig. 8·4 functions.

8·5 What is the difference between an AND and an OR circuit?

8·6 If transistor E in Fig. 8·8 became defective and ceased to conduct, how would this modify the operation of the circuit?

8·7 Describe the operation of the bistable multivibrator of Fig. 8·9.

8·8 What purpose do triggering amplifiers serve with bistable multivibrators?

8·9 Show another way of triggering such multivibrators.

8·10 What advantages and disadvantages do direct-coupled bistable circuits possess over a-c-coupled circuits?

8·11 Draw the diagram of a monostable multivibrator and explain how it functions.

8·12 What is a d-c to d-c converter? Where would it be employed?

8·13 Explain briefly the operation of a d-c to d-c converter.

8·14 What is a Zener diode? What characteristic makes it valuable for voltage regulating?

8·15 Show the use of a Zener diode in a transistor voltage-regulator circuit.

8·16 What is the difference between series and shunt regulation?

8·17 Draw the circuit of a simple series regulator and explain how it operates.

8·18 How does the circuit of Fig. 8·31 detect the presence of ferrous bodies?

8·19 What type of circuit is required to detect nonferrous metals?

8·20 How does the 6476 cold-cathode counter tube function?

8·21 Explain briefly the operation of the circuit of Fig. 8·37.

8·22 Draw the circuit diagram of a transistor chopper. Explain how the chopper operates.

8·23 What is offset voltage in a chopper? Offset current? What is their importance?

8·24 When would the chopper circuit of Fig. 8·15 work satisfactorily? When would the circuit of Fig. 8·18 be required?

CHAPTER 9

Additional Transistor Developments

THE VACUUM TUBE, at the outset of its development, contained only a filament and a plate. This was the diode. Eventually, Lee De Forest added a third element, the grid, and the triode came into being. As time went on and engineers became electronically more sophisticated, other elements were added, until today there is a vast assortment of tube types and structures capable of performing an almost incalculable variety of jobs.

In analogous fashion, the initial work on semiconductor electronic devices started with the diode, then followed with the triode and, more recently, the tetrode. In addition, variations, modifications, and new structural types of semiconductor devices are being constantly developed, and there is every reason to believe that after a number of years there will be as many different transistors as there are tubes, each designed to perform best within a certain range of applications.

It is the purpose of this chapter to examine the operation of those "other" transistors which have been developed to date. Some of them will bear a very close resemblance to the transistors already discussed; others will be entirely different. We shall also discuss a recent development of semiconductor manufacturers: the production of complete circuits using semiconductor elements. These circuits contain not only one or more transistors but also resistors, capacitors, and even inductors, each fabricated using basically the same manufacturing techniques as those employed for transistors. These new devices are known as solid-state circuits, and they are beginning to appear in increasing numbers.

Photosensitive Transistors

An application for which the transistor is receiving considerable attention from designers is its use in photoelectric equipment such as flame detection, automatic door openers, light dimmers, burglar alarms, counters, and the like. The interest in this particular direction is spurred on by the fact that whereas phototubes deliver microamperes of current, phototransistors deliver milliamperes. This means that intermediate amplifiers which are needed with phototubes to build up the minute currents to usable proportions can, with photosensitive transistors, often be eliminated. In addition, there is a further saving in power because of the lack of filament heating and a greater efficiency in power utilization where amplification is achieved.

A phototransistor that was developed by the Bell Telephone Laboratories is shown in Fig. 9·1. The heart of the device is a pill-shaped

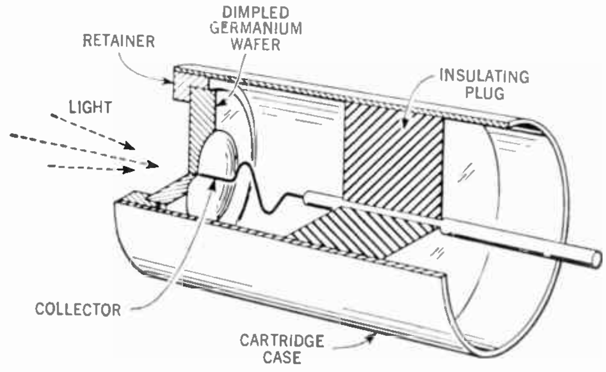


Fig. 9·1 A phototransistor that was developed by Bell Laboratories. (After J. N. Shive, The Phototransistor, *Bell Laboratories Record*, August, 1950)

wafer of germanium having a spherical "dimple" ground in one side so that the thickness of the wafer at the center is about 0.003 in. The wafer is forced-fitted into one end of a metal cartridge, and a pointed phosphor-bronze wire is brought into contact with the wafer at its center. This wire is called the collector. The far end of the wire fastens to a metal pin embedded in an insulating plug which is held in position at the opposite end of the cartridge. (The second electrical contact to this phototransistor is the cartridge case itself.)

The wafer is made of N-type germanium; hence, it contains an excess of electrons. The area directly beneath the collector wire becomes, under the action of a forming current, P-type germanium. This gives us a diode which will function in the same manner as other PN junction diodes previously described. That is, if it is biased in the forward direction, current will flow easily; if it is biased in the reverse direction, only a minute current will flow. In its present application, this phototransistor is biased in the reverse direction and the small current that flows is called the "dark current." When light is directed on the germanium wafer, the amount of current flowing increases in proportion to the light intensity. This is explained rather simply by the fact that the energy which the germanium wafer absorbs from the light quanta serves to break a number of covalent bonds, producing equal numbers of electrons and holes. Under the influence of the applied

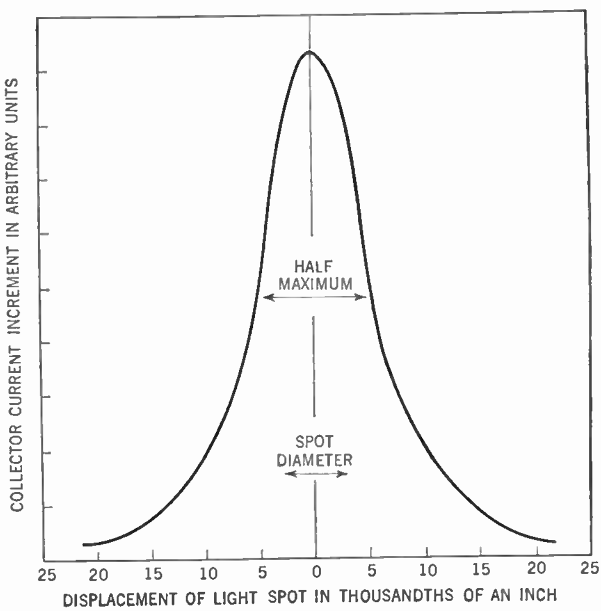


Fig. 9·2 Monner in which the current output will vory os o spot of light is moved ocross the responsive oreo of the phototronsistor shown in Fig. 9·1.

electric field, the electrons travel to the positive terminal of the battery (which connects to the case) and the holes go to the negative battery terminal (which is tied to the collector). Current flow through the circuit is thus increased.

A characteristic of this phototransistor is that maximum current response is obtained when the incident light rays fall in the immediate neighborhood of the point where the collector wire contacts the wafer. Figure 9·2 illustrates how the photocurrent varies as a tiny spot of light is moved across the center of the responsive area of a typical phototransistor. It is apparent that permitting light to fall uniformly over the germanium surface is wasteful of the light energy. For that reason, the phototransistor is combined with a small glass lens that focuses the arriving light rays into a narrow beam which is restricted principally to the desired area. The reason that this is desirable is that electrons and holes which the light quanta produce at points relatively far from the PN junction point recombine before they reach the junction. The greatest possibility of avoiding recombination occurs at the junction point; the longer the distance that the electrons and holes must travel, the greater the possibility of recombination with an opposite charge.

It was noted in Chap. 1 that light quanta must possess a certain minimum amount of energy in order to move an electron from a

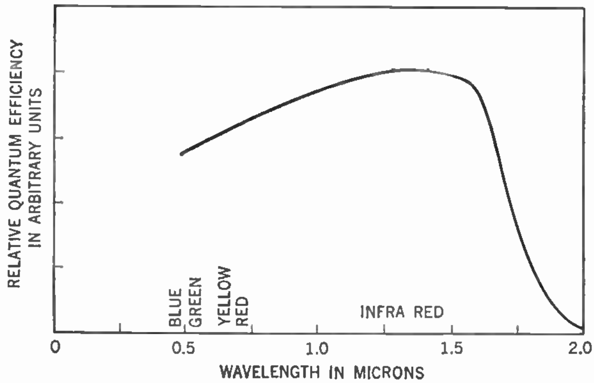


Fig. 9·3 The relative quantum efficiency (in arbitrary units) vs. the wavelength of the incident light for the phototransistor in Fig. 9·1.

lower to a higher level. Since the energy of a light quantum varies inversely with the wavelength of light, it follows that the photoconductivity response will depend upon the wavelength of the incident light. Figure 9·3 shows this dependence for the phototransistor of Fig. 9·1. The long-wavelength limit occurs in the vicinity of 2.0 μ. Thereafter, as the wavelength becomes smaller, the response rises sharply and then slowly starts to fall off.

An interesting sidelight is the observed behavior that when the quantum energy of the incident light becomes sufficiently great, electrons may be given enough of a "push" to enable them to escape completely from the germanium wafer. The substance now becomes photoemissive. To achieve this effect with germanium would require that ultraviolet light be employed. Since only the photoelectric effect is desired ordinarily, conventional incandescent light sources are employed.

The phototransistor just described is of point-contact construction. A similar unit possessing a PN junction assembly may also be devised with comparable results, Fig. 9·4. Here, too, response will be greatest when the light is directed at the junction; it will drop off gradually as the distance from the junction increases.

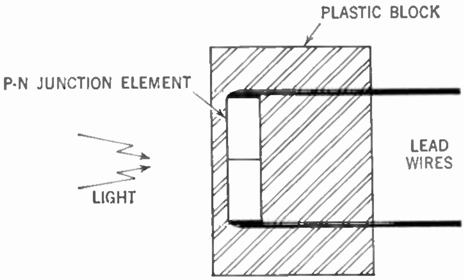


Fig. 9·4 A junction phototransistor.

Another form of phototransistor employs an NPN type of construction in which only the central base section is photosensitive, Fig. 9·5. No connection is made to the base. However, a voltage is applied across the two end sections. The end to which the negative battery is connected is the emitter; the other end section serves as the collector.

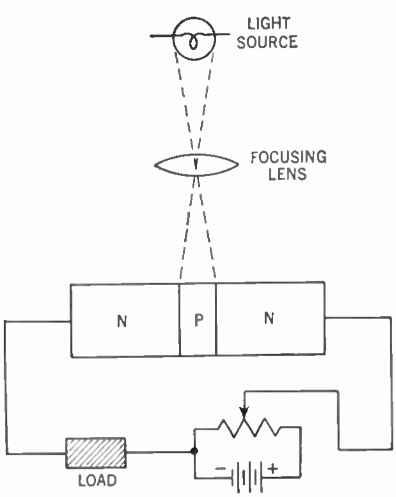


Fig. 9·5 An NPN phototransistor. Light is directed onto the photosensitive center section.

Without any external voltage at the base forcing the holes there to move toward the emitter junction, combinations at that junction with emitter electrons are quite low and only a minute current flows through the transistor. When light is focused on the photosensitive P section, however, holes are formed in sufficient quantity to produce as much as 4 to 6 ma of current through the load. With that amount of current, a relay can be operated directly.

The current-amplifying properties of a transistor are also utilized in this application. That is, a small change in emitter-base potential, produced here by light, results in a sizable change in collector current.

The P zone must be made exceedingly thin. Furthermore, each of

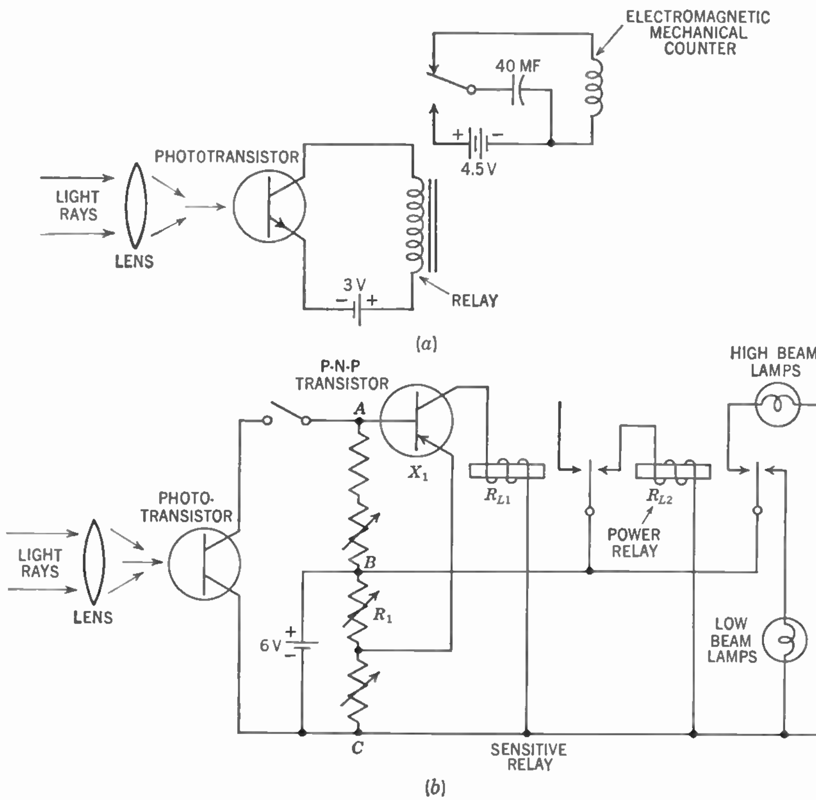


Fig. 9·6 (a) Circuit diagram of a phototransistorized counter. (b) The phototransistor employed to dim automobile headlights automatically.

the two N sections, especially the emitter, must have low resistivity. To raise the current flow for a specific light intensity, the applied voltage can be made greater within the permissible dissipation limits of the transistor.

Two applications of the foregoing phototransistor are shown in Fig. 9·6, and the simplicity of each circuit is immediately apparent. The first diagram, Fig. 9·6*a*, is that of a photoelectric counter. Light is collected and focused on the sensitive base region of the phototransistor by a collecting lens. The unit develops several milliamperes of current which flows through and actuates a sensitive relay. Closing the relay discharges a capacitor through a mechanical counter. Such a device will work on an ordinary pocket flashlight at distances up to 50 ft.

Figure 9·6*b* presents the diagram for an automatic automobile-headlight dimmer. The circuit contains the phototransistor with a single grounded-emitter amplifier in a d-c network. A 6-volt battery is connected across part of a voltage-divider network, between points *B* and *C*. Because of the battery polarity across these two points, the top end of R_1 is more positive than the bottom end. Since the top end of R_1 connects to the base of X_1 and the bottom end to the emitter, the base-emitter circuit is reverse-biased and the transistor is cut off. This is true with the switch open or with the switch closed and *no* light reaching the phototransistor. (With the phototransistor not illuminated, its internal resistance is very high, amounting almost to an open circuit as far as the 6-volt battery is concerned.) With no collector current flowing through X_1, relay R_{L1} is not actuated. Under this condition, the relay arm touches the right contact, closing the path for R_{L2} and actuating this relay. This pulls the contact arm of R_{L2} to the left, and the high-beam lamp is turned on. When an oncoming car appears, its light causes the phototransistor impedance to drop sharply. Now current from the battery can flow through this unit to point *A* and then down to point *B* and the positive terminal of the battery. The voltage drop across *AB* opposes the voltage across R_1, and X_1 is driven into conduction. Collector current now flows, the arm on R_{L1} is drawn to the left, and the current path through R_{L2} is opened. This releases the R_{L2} arm, and it swings to the right, completing the circuit for the low-beam lamp.

Still another phototransistor design is shown in Fig. 9·7*a*. The unit is basically an NPN transistor except that both N sections have been specially treated for photosensitivity. When light is directed at the left-hand PN junction, there is generated an internal voltage which causes a current to flow through the phototransistor and the external load without the need for a power supply. When the same light is focused on the right-hand PN junction, there is developed a similar voltage which causes current to flow through the phototransistor and the load in the opposite direction.

One application of this device is shown in Fig. 9·7*b*. Beams from lamps L_1 and L_2 are focused onto the phototransistor. L_1 may be a standard lamp with fixed intensity, while the intensity of L_2 may be variable. The lamps illuminate S_1 and S_2, two points which are equidistant from the two junctions. When the lamps are equally bright, they generate equal and opposite voltages and no current flows through the load. If L_2 grows brighter, the net voltage across the transistor load will be negative; less light from L_2 will produce a voltage which is positive.

These fluctuations in output voltage are fed to a control device. When the voltage is negative, L_2 is made dimmer; when the voltage is positive, L_2 is made brighter. Thus the circuit tends to maintain the lamps equally brilliant.

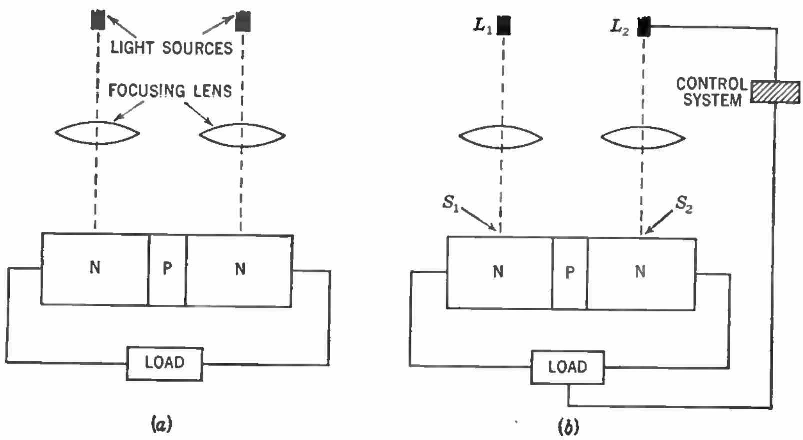


Fig. 9·7 (a) An NPN phototransistor in which both N sections are made photosensitive. (b) An application of the NPN transistor in (a) to control light. (Patent 2,641,712, R. J. Kircher, Summit, N.J.; assigned to Bell Telephone Laboratories, Inc.)

The circuit arrangement in Fig. 9·7*b* may be employed in photographic exposure-control processes, in stage light-intensity-control systems, and in infrared baking processes. It may also be employed in translating or transcribing systems for coding, decoding, and information-blending systems.

Tetrode Transistors

The tetrode transistor is basically an NPN junction unit of the type already described except that a fourth electrode, labeled b_2, is attached to the base layer at a point which is on the side opposite the original base connection b_1, Figs. 9·8 and 9·9. A potential which is considerably higher than the normal emitter-to-base potential is applied to this second base lead. The normal emitter-to-base voltage is generally on the order of 0.1 volt. On the other hand, b_2 is given a potential of about —6 volts. This voltage is fixed and will not vary with the signal, since the latter still is applied between the emitter and base b_1.

The presence of this large bias voltage at b_2 will modify the flow of current through the transistor. In the unit shown in Fig. 9·8 the emitter and collector sections are formed of N-type germanium and

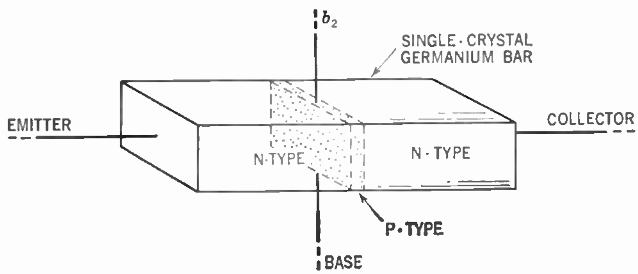


Fig. 9·3 A transistor tetrode.

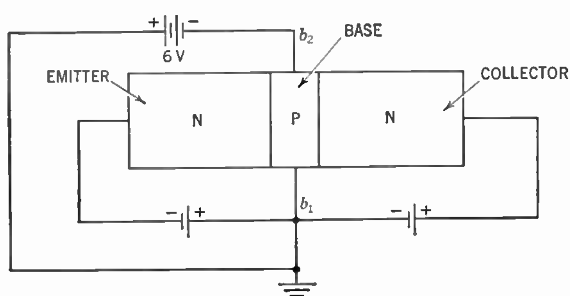


Fig. 9·9 The transistor tetrode with suitable d-c biasing voltages. Although not shown, the incoming signal would be applied to the emitter and the load resistor would be inserted in the collector circuit.

the base of P-type germanium. The application, then, of a relatively large negative potential to b_2 serves, in the base region under the influence of this voltage, to restrict the flow of electrons from the emitter to the collector sections.

It can be seen from Fig. 9·9 that the —6 volts is applied between b_2 and b_1, or, actually, across the length of the base region. Since the base has an internal resistance, the voltage decreases uniformly from —6 volts at the top to 0 volt at the bottom. The voltage is negative enough at all points except near the bottom edge of the base to prevent any flow of electrons from the emitter across the base to the collector. At the bottom edge, the 0.1-volt forward bias between b_1 and the emitter will permit electrons to travel from emitter to collector.

Thus, the addition of connection b_2 and the application of a negative voltage there alters the flow lines in the conventional NPN transistor to the extent shown in Fig. 9·10. The ability of this modification to improve the high-frequency operation of this transistor stems from two factors. First, the collector capacitance is reduced by decreasing the effective active area of the collector junction. Second, the effective

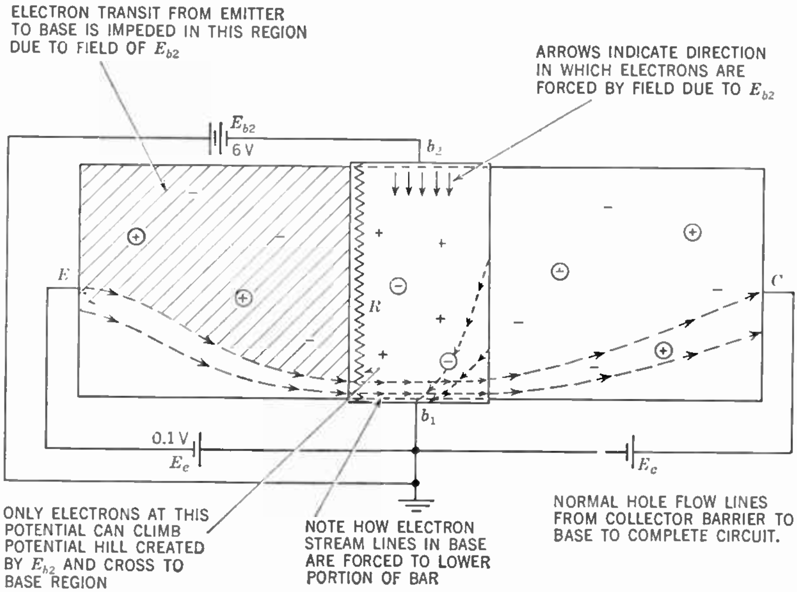


Fig. 9·10 The flow of carriers in the transistor tetrode. Note how little of the base section is actually used.

area of the base is reduced substantially, which means that the active base resistance is reduced. A low base resistance, we have previously seen, is conducive to improved high-frequency operation.

As the tetrode transistor is employed conventionally, b_2 has a fixed negative voltage. The other three elements are then used as they would be in a triode transistor. That is, the input signal is applied between emitter and base and the output signal is taken from the collector circuit. Circuit arrangements remain the same.

Active use of b_2, rather than the passive role indicated above, has also been made. For example, in Fig. 9·11*a*, the tetrode transistor is employed as a modulator. The r-f carrier signal is applied to the emitter, while the audio voltage is impressed at b_2. Both the emitter base b_1 and the base b_2 biasing voltages are obtained from the same bias battery, although each circuit is isolated from the other by resistors R_1 and R_2. The changing audio voltage affects the amount of carrier current flowing through the transistor and, in this way, alters or modulates its amplitude. The modulated signal appears across the load for transference to the rest of this system.

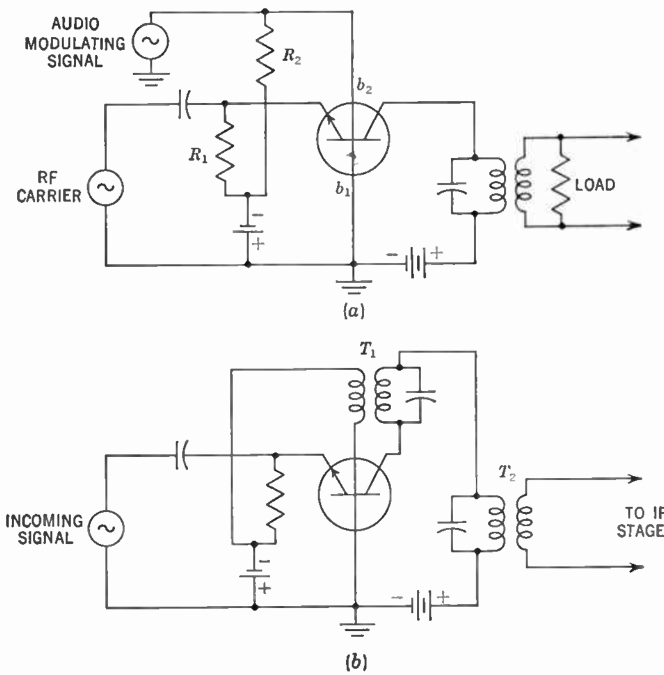


Fig. 9·11 (a) Tetrode transistor employed as a modulator. (b) Tetrode transistor connected as a frequency converter.

In Fig. 9·11*b* the tetrode transistor is shown connected as a frequency converter. Transformer T_1 feeds energy back from the output, or collector, circuit to the b_2 base terminal. If the polarity of the feedback voltage is correct, oscillations will occur. At the same time, the received signal is applied to the emitter. Interaction of this signal with that from the local oscillator produces sum and difference frequencies. Transformer T_2 is tuned to the i-f or difference-frequency signal, and this is then fed to several i-f amplifier stages, as in any conventional radio receiver.

Ultrahigh-frequency Transistors: The Coaxially Packaged Transistor

The search for useful high-frequency transistors has recently led to the development of transistors capable of performing at microwave frequencies. Improved performance at these ultrahigh frequencies has successfully been obtained by studying not only transistor geometry and physics but also transistor packaging and construction. One result of such an investigation is the coaxially packaged PNP microalloy diffused-base transistor. See McCotter, Walker, and Fortini, Coaxially Packaged MADT for Microwave Applications, *IRE Transactions on Electron Devices*, vol. ED-8, no. 1, January, 1961.

Some mention has already been made (Chap. 3) of the major factors which limit the high-frequency operation of a transistor. These were seen to be primarily the base width, base resistance, and emitter- and collector-junction capacitances. By providing graded doping of the base, with its resultant electric field which shortens the carrier transit time through the base, the drift, or diffused-base, transistor was found to offer a higher operating-frequency limit. The principal difference between the diffused-base transistor and this new microwave device is in the reduction of electrode size and in the use of a unique coaxial-package construction.

Although the basic transistor is fabricated by techniques common to diffused-base transistor manufacture, there are several important differences. To keep the base resistance to a minimum, the entire surface of the emitter side of the germanium blank is a heavily doped, low-resistivity material. This sheet of low-resistivity material, which extends up to the edge of the emitter etch pit, acts as a metallic (low-resistance) base connection. The emitter pit is etched deep enough into the diffusion layer to obtain a high hole (PNP) injection efficiency. However, it cannot be made too deep or the resistance in the emitter region will increase; also, the etch pit must be kept very small to keep the base resistance low and to minimize the emitter-junction capacitance. The collector is located in the high-resistivity region of the graded base, although when the collector voltage is applied, the depletion layer extends into the graded region, and this keeps the base transit time low. If the collector is too far from the graded region, the collector capacitance will be very low but the operating voltage will necessarily have to be high in order to have the depletion layer extend into the diffused-base region.

All of the above-mentioned techniques are aimed toward optimizing the internal transistor itself. In addition, there is the problem of making the capacitance of the transistor much smaller than the capacitance developed by the package in which the transistor is enclosed. For example, the collector-base junction capacitance of the transistor can be made as low as ½ $\mu\mu$f. However, the capacitance between external collector and base electrodes of a common transistor package is between ½ and 1½ $\mu\mu$f. This will overshadow the actual transistor capacitance. Thus, if transistors are to be useful in the microwave region, package capacitance must certainly be reduced.

Since it is extremely difficult to reduce package capacitance below 1 $\mu\mu$f, one may well ask, "Just what capacitance is the most important to reduce?" The impedance levels at input and output are fairly low in ultrahigh-frequency work; therefore, stray capacitance to ground at input and output is of a lesser importance. The big problem is feedback capacitance, i.e., capacitance between input and output. Feedback capacitance between input and output is detrimental in

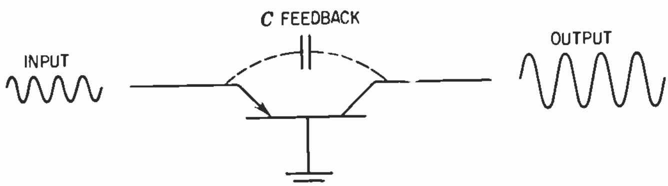


Fig. 9·12 The stray feedback capacitance between collector and emitter of a common-base amplifier.

three respects: first, it reduces the bandwidth of a low-pass amplifier; second, it reduces the power gain of a tuned amplifier; and third, it causes circuit instability, i.e., oscillation. Another consequence is a lowering in the maximum frequency at which the transistor can be used as an oscillator. From this we may conclude that a primary feature of a good high-frequency transistor package is a low feedback capacitance, or good isolation between input and output.

One useful method of reducing the feedback between input and output is to establish a ground plane between the input side and the output side of the transistor. The previous feedback capacitance then becomes stray input and output capacitance to ground. For example, consider Fig. 9·12, which shows a common-base amplifier with stray feedback capacitance. Now consider the same transistor with a ground plane between emitter and collector, as shown in Fig. 9·13. The feedback capacitance here is divided by 2, but, more importantly, the capacitance is now to ground and not between input and output. Figure 9·14 shows a simple way in which a ground plane is established between input and output.

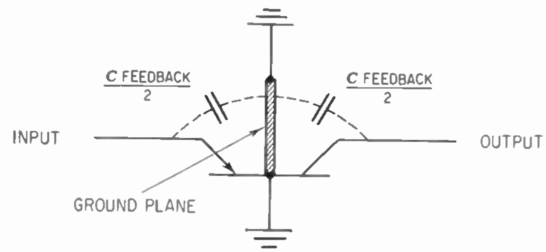


Fig. 9·13 A ground plane, inserted between emitter and collector, will reduce the effect of the feedback capacitance.

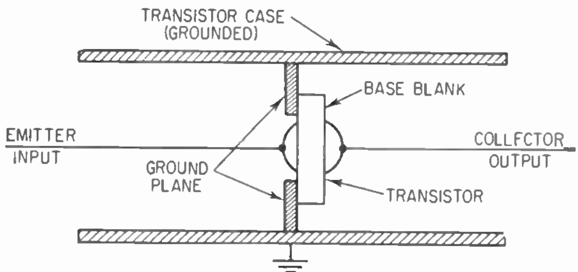


Fig. 9·14 A simple way of establishing a ground plane between input and output of a transistor. The transistor size is greatly exaggerated.

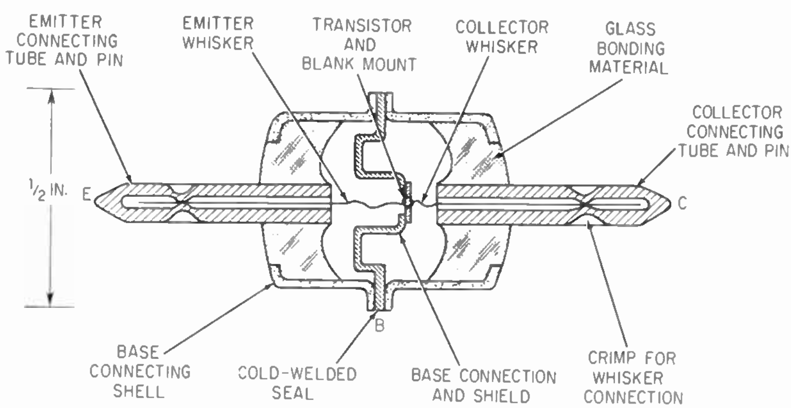


Fig. 9·15 Coaxial transistor package.

A more advanced package, the coaxial package, was developed to house kilomegacycle transistors. The coaxial package is presently available only in the common-base configuration; however, common-emitter packages should be available shortly.

A cross-sectional view of the transistor mounted in its coaxial package is shown in Fig. 9·15. The transistor blank is mounted on a copper base, and this serves two purposes. It provides a contact to the outside shell (base lead) and also acts as an electrical shield (ground plane) between the emitter and the collector (connection pins). Furthermore, it provides a good heat-sink connection between the transistor and the package. The copper shell is sealed with a glass insulator to the connecting pins. The two halves which make up the shell are cold-welded; hot welding leads to the formation of gases inside the package. The whiskers extend into the copper connecting pins where they are crimped for electrical contact. A dimensioned outline of the package is shown in Fig. 9·16. The diameters of the shell and pins are such that the characteristic impedance of the package is 50 ohms.

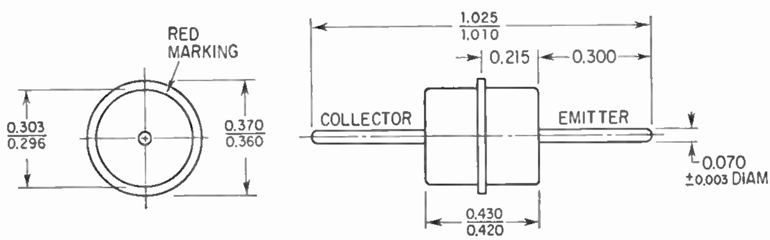


Fig. 9·16 Dimensioned outline of transistor package.

To gain an idea of the performance capabilities of this device, let us compare it to a typical high-frequency transistor, the 2N502. Except for the coaxial package and the emitter size, the devices are very similar; they are both MADT transistors and both have the same diffusion gradient and the same emitter and collector locations with respect to the diffused-base region. A few of the more important parameters are given for comparison in Table 9·1.

Table 9·1 Parameter Comparison

	2N502	Coaxial transistor
Collector voltage	−10 volts	−25 volts
Collector current	2 ma	2 ma
Base resistance	30 ohms	10 ohms
Emitter-junction capacitance	17 $\mu\mu$f	8.8 $\mu\mu$f
Collector-junction capacitance	0.45 $\mu\mu$f	0.16 $\mu\mu$f
f_{max}	894 Mc	3,500 Mc
Power gain	11 db (at 200 Mc)	11 db (at 1,000 Mc)

Note that all of the parasitic parameters are reduced, thus permitting a much improved maximum frequency of oscillation f_{max}. Finally, note that the coaxial transistor offers 11 db of power gain at 1,000 Mc, whereas the 2N502 yields the same gain at only 200 Mc.

PNPN Transistors

A class of four-layer semiconductor devices has been developed to serve primarily as switching devices. These devices, known as PNPN transistors, are available with two or three terminals.

The basic physical construction of a PNPN transistor, Fig. 9·17, consists of alternate layers of P-type and N-type germanium or silicon. Each P zone has an excess of holes; each N zone has an excess of electrons. In the two-terminal PNPN units, connection is made to the end

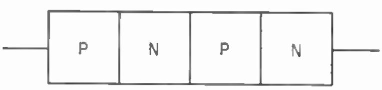


Fig. 9·17 Basic physical construction of a PNPN transistor.

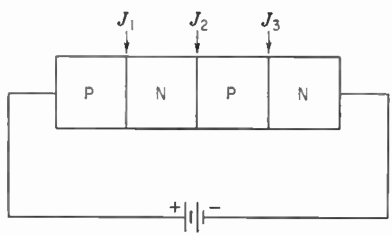


Fig. 9·18 A PNPN transistor with voltage applied.

semiconductor sections. This is shown in Fig. 9·17. In the three-terminal PNPN units, a third connection is made to the P section closest to the end N section. This third connection or terminal is known as the "gate."

Two-terminal PNPN transistors. The mode of operation of a two-terminal PNPN transistor can be understood by applying a voltage across the ends of the device as shown in Fig. 9·18. The positive side of the voltage goes to the P end section, and the negative side of the voltage connects to the N end section. This polarity voltage will cause junctions J_1 and J_3 to be forward-biased. The intermediate junction, J_2, will be reverse-biased because the voltage on the N side is relatively positive with respect to the voltage on the P side. Thus, junction J_2 will possess a considerably higher impedance than junctions J_1 and J_3. Nearly all of the applied voltage will appear across J_2, and the current that flows will be largely that characteristic of a reverse-biased diode. The overall impedance of the unit is very high, on the order of several hundred megohms.

As the voltage across the PNPN transistor increases, the electric field produced across J_2 will similarly rise. This, in turn, will cause more current to flow through the transistor and its external circuit. The current rise will follow the curve in region I of Fig. 9·19. At a certain value, the voltage will permit carriers from the two end sections to acquire enough energy at J_2 to dislodge additional carriers, producing a form of avalanche breakdown analogous to a Townsend discharge in gases. The action feeds on itself, quickly producing a current flow whose value is limited only by the resistance of the external circuit. This behavior is depicted by regions I and II of the characteristic curve, Fig. 9·19. The breakdown commences at V_B; at this point, current I_B is flowing. Thereafter, the voltage drop across the PNPN device

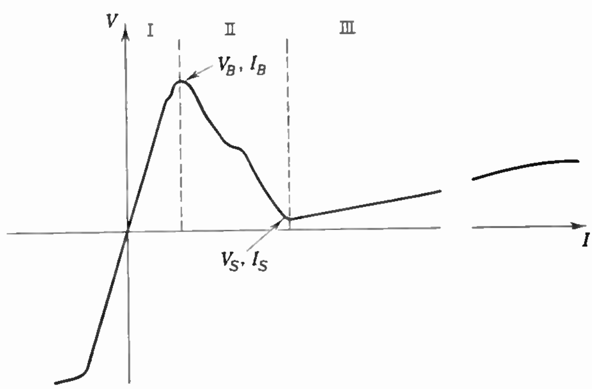


Fig. 9·19 Voltage-current characteristic of a PNPN transistor. Region I is the high-resistance region where little current flows. Point V_S, I_S is the low-resistance point.

decreases quite rapidly with current increase. Note that this is a condition of negative resistance because the voltage drop is decreasing while the current is increasing.

The foregoing action finally results in a condition of zero voltage drop across the center junction, at which time it acts as an emitter as well as a collector. The device is then "closed," the two center sections are flooded with charge, and the overall impedance is exceedingly low. We are now at the lowest point of the curve, between regions II and III. Further increase in current beyond this point will produce a slowly rising resistance, region III.

The PNPN transistor will remain in this highly conducting condition so long as the current through it remains greater than the value I_S. The latter value is called the holding current. If the current drops below this level, the unit will revert back to the high-impedance condition.

When the PNPN transistor is in its closed or conducting condition, each of the three junctions is forward-biased and the two central sections are heavily saturated with holes and electrons. To turn the device off, a reverse voltage must be applied. Upon application of this reverse voltage, the holes and electrons in the vicinity of junctions J_1 and J_3 will diffuse to these junctions and produce a reverse current in the external circuit. So long as the reverse current is appreciable, the voltage drop across the PNPN unit will remain small. After the electrons and holes in the vicinity of J_1 and J_3 have been removed, however, the reverse current ceases and the two junctions become reverse-biased. For all practical purposes, the switch is now open,

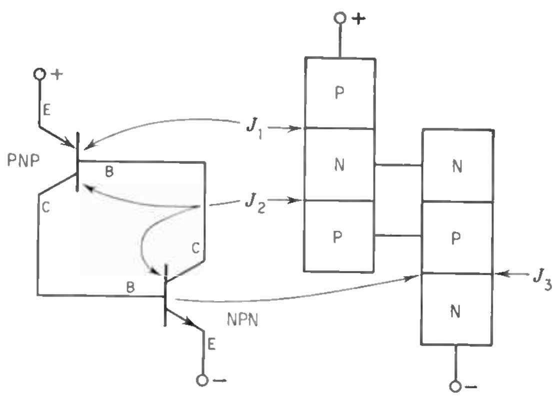


Fig. 9·20 The PNPN device can be regarded as essentially an NPN transistor and a PNP transistor that are interconnected.

although some recombination of holes and electrons must still take place around junction J_2. Once the action is complete, junction J_2 also assumes a blocking condition. The device is now ready to go through the entire process again.

It is comparatively simple to derive an expression for the current flowing through a PNPN transistor. We can depict the device as being composed essentially of a PNP and an NPN transistor interconnected as shown in Fig. 9·20.

Now, when we apply a voltage across the two end sections, with the polarity shown, junctions J_1 and J_3 are forward-biased while J_2 is reverse-biased. This provides each transistor with its conventional biasing arrangement and each has associated with it a current gain α. Let us call the current gain of the PNP unit α_1 and the current gain of the NPN unit α_2. Since α is, by definition, the fraction of the electron current (for NPN transistors) or hole current (for PNP transistors) injected into the base that reaches the collector, the two structures can be combined to derive the total current flowing.

In Fig. 9·20, J_2 is the common-collector junction, and it is affected by three components of current: I_{α_1}, the hole current from the P end region; I_{α_2}, the electron current from the N end region; and I_{CO}, the leakage current. The total current I flowing in the external circuit must be equal to the current at J_2. Thus

$$I_{J_2} = I = \alpha_1 I + \alpha_2 I + I_{CO}$$

or

$$I - \alpha_1 I - \alpha_2 I = I_{CO}$$

and

$$I(1 - \alpha_1 - \alpha_2) = I_{CO}$$

or

$$I = \frac{I_{CO}}{1 - (\alpha_1 + \alpha_2)} \tag{9·1}$$

From this expression, as $\alpha_1 + \alpha_2$ approaches unity, the current through the device becomes large and is limited only by the resistance of the external circuit. In short, as $\alpha_1 + \alpha_2$ approaches unity, the internal resistance drops to a very low value.

In deriving the above expression, the hole current from the P end region is added to the electron current from the N end region even though they are traveling in opposite directions. The two currents can be added because a hole current, being positive in charge, is equivalent in its electrical effect to an electron current flowing in the opposite direction.

With junctions J_1 and J_3 forward-biased and J_2 reverse-biased, very little current flows. From Eq. (9·1), this current is approximately I_{CO}, since α_1 and α_2 are close to zero. This is because at low emitter currents α is also low. This condition corresponds to the "off" state. To turn the device on, α_1 and α_2 must be increased, which means sending more current through the device. From Eq. (9·1), if $\alpha_1 = 0.49$ and $\alpha_2 = 0.49$, the current through J_2 and the external circuit is

$$I_{J_2} = \frac{I_{CO}}{1 - (0.49 + 0.49)} = \frac{I_{CO}}{0.02}$$

In this case, the current through J_2 is 50 times the normal leakage current. If $\alpha_1 + \alpha_2 = 1$, the current through J_2 is limited only by the resistance of the external circuit.

Three-terminal PNPN transistor. In the three-terminal PNPN transistor, a third connection is made to the P section closest to the end N section, Fig. 9·21. This third terminal is known as the gate. Its purpose is to switch the transistor from the open to the closed state without the need for exceeding the critical breakover voltage. In essence, what the gate does is inject enough current into the PNPN structure to cause junction J_2 to become forward-biased.

To produce the required gate current, some voltage-switching arrangement in which the gate lead is made positive with respect to the N end section (or cathode, as it is called) is employed. The arrangement may be a simple battery and mechanical switch, or a positive pulse for the gate circuit may be developed electronically. Once conduction through the PNPN device is initiated, the gate loses further control of the current flowing. To return the PNPN transistor to its "off" condition, either its current must be reduced below the

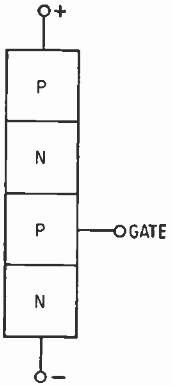


Fig. 9·21 A three-terminal PNPN transistor.

holding level or the voltage between the cathode and the anode (i.e., the other end connection) must be reversed. In this sense, the three-terminal PNPN transistor is equivalent to the thyratron tube.

A number of methods have been developed to turn PNPN transistors off. The unit in Fig. 9·22*a* is powered from an a-c source. On the positive half cycle, the unit can be triggered on by an appropriate gate current. When the a-c anode voltage goes negative, it automatically causes the current to turn off. The turn-off action is not instantaneous; rather, it must be held for a short period of time, on the order of 100 μsec or less; otherwise, the PNPN transistor will recover its ability to continue passing current.

PNPN devices may also be turned off by diverting enough current away from them for a similar short period of time. A suitable circuit is shown in Fig. 9·22*b*. Transistor *X* is shunted across the PNPN unit. Ordinarily, *X* is kept cut off. When it is desired to cut the PNPN unit off, the transistor is turned on by an external pulse. If the collector-to-emitter voltage drop of *X* is less than the forward voltage drop across the PNPN unit, enough PNPN current will be diverted to X to turn the PNPN device off.

The most widely used method of turning off a PNPN transistor is to connect a charged capacitor across it so that the cathode element is driven positive with respect to the anode. In a typical application, Fig. 9·22c, when PNPN unit A is gated on, capacitor C_1 will charge through R_1 to the polarity shown. Firing PNPN unit B will connect the left side of C_1 to ground, making the anode of unit A instantaneously

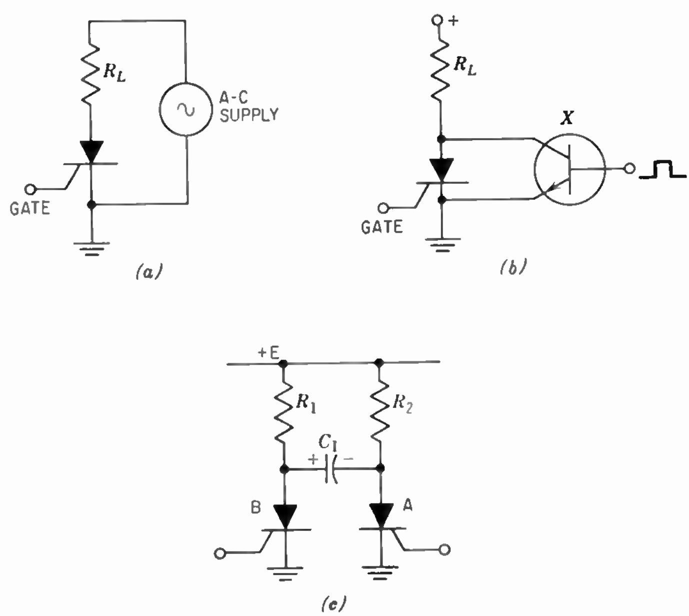


Fig. 9·22 Three different methods of turning controlled rectifiers off.

E volts below ground. If the time constant of R_2C_1 is long enough, unit A will remain back-biased long enough to shift into the open condition.

Three-terminal PNPN transistors are frequently called controlled rectifiers. If the semiconductor employed in their manufacture is silicon, then they are known as silicon controlled rectifiers. Commonly employed schematic symbols for the three-terminal PNPN transistor are shown in Fig. 9·23.

The Thyristor

There is a considerable need for a semiconductor device (transistor or otherwise) which is capable of functioning as a large-current switch. One answer to this problem is the Thyristor, developed by RCA. The name is derived from the thyratron-like properties of the device. That is, it can be turned on by applying a pulse to a control element, just as we can in a thyratron tube. However, unlike the thyratron, the Thyristor can also be turned off by the same element.

A cross section of the Thyristor is shown in Fig. 9·24. Arsenic is diffused into a P-germanium wafer to form an N-type base region. This

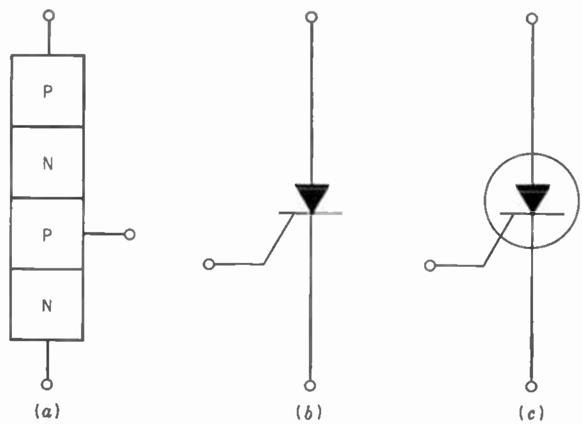


Fig. 9·23 Three commonly employed schematic symbols for controlled rectifiers.

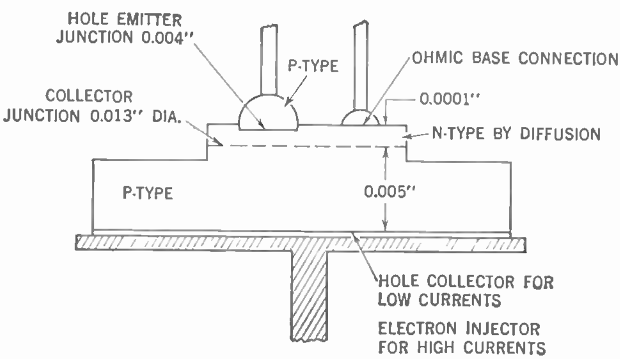


Fig. 9·24 Internal construction of the Thyristor. (After C. W. Mueller and J. Hilibrand, The Thyristor—A New High Speed Switching Transistor, *IRE Trans. Electron Devices*, January, 1958)

region is formed on a plateau having a diameter of about 0.013 in. An emitter junction and an ohmic base connection are soldered to this plateau. The emitter dot, which is about 0.004 in. in diameter, is an alloy of indium and aluminum.

Below the base, P-type germanium forms the collector. The collector connection is made by soldering, at a temperature of 300 to 400°C, a nickel tab to the germanium of the collector with an alloy of lead, tin, and indium. The connection may cover the entire area of the germanium wafer, and its physical nature has an important bearing on the way the Thyristor functions.

The Thyristor can be employed as a normal transistor switch or amplifier as long as the collector current is kept below a certain value, called the breakover current. However, if this current is allowed to increase, it will be found that the total current transfer ratio α increases also. When α exceeds unity, a large increase in current occurs for the grounded-emitter arrangement, and the voltage across the unit drops to a very low value.

The changeover from the normal state to this high-current, low-resistance state can be accomplished with a very short pulse applied to the base. Once the changeover has been effected, the pulse can be removed without reducing the collector current. However, by the use of a reverse voltage between the base and emitter, the high collector current can be turned off.

α can exceed unity because it consists of two parts. One part, the larger, is obtained from the normal relationship between the emitter, base, and collector. The emitter injects holes into the base, and over 90 per cent of these holes are received by the collector.

However, it is also found that the special collector connection injects electrons into the body of the collector. These electrons are swept to the emitter and therefore increase the emitter-collector current. This "electron α" is strongly dependent on collector current. It is very low when the collector current is low, but it increases about 100 times when the collector current rises from 1 to 5 ma.

Since the two α's add, a critical current at which the total α exceeds 1 is reached, and now a cumulative buildup in which the total current rises sharply occurs and the transistor resistance drops to a very low value (about 3 ohms between emitter and collector). The Thyristor remains in this condition for as long as the circuit remains intact. This is true whether or not the pulse which produced the initial current rise is retained.

The Thyristor can be turned off by applying a strong reverse-bias pulse to the base-emitter circuit, by interrupting the collector current, or by applying a strong pulse to the collector, which counteracts the normal collector bias.

Unijunction Transistor

The unijunction transistor is a development of the General Electric Company. It was originally called a double-base diode because of its similarity to a junction diode with two base connections instead of one. The name was changed subsequently to unijunction transistor because the device does possess a control feature; since there is only one junction (i.e., emitter to base), the prefix "uni" was added to the name.

Structurally, a unijunction transistor consists of a uniformly doped N-type silicon bar having ohmic contacts at each end and a PN junction at the center, Fig. 9·25*a*. The two ohmic contacts are called base 1 and base 2, and at room temperature the base-to-base resistance is 5,000 to 10,000 ohms. The PN junction at the center of the bar is formed with an aluminum wire. This junction is called the emitter.

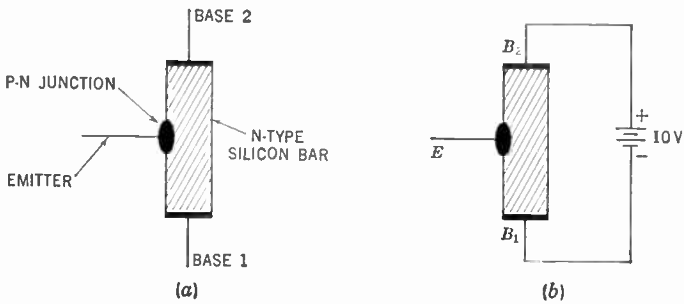


Fig. 9·25 The unijunction transistor.

If the two base leads are connected together, the entire unit behaves as a conventional diode, with the aluminum wire representing one terminal and the connected base leads the other. However, the unique properties of the unijunction transistor are obtained when the base leads are separated and a voltage is applied between them, Fig. 9·25*b*. This voltage may have values of 30 volts or more, but for this discussion, 10 volts will be used. This 10 volts, inserted between B_1 (base 1) and B_2 (base 2), produces a uniform drop across the 5,000 to 10,000 ohms internal resistance of the silicon bar. Since the emitter is placed halfway along the bar, it sees +5 volts (with respect to B_1) opposite it.

If, now, we apply anything less than +5 volts to the emitter lead, the emitter PN junction will be reverse-biased, and very little current will flow in the emitter circuit. However, the instant the emitter voltage rises above +5 volts, the junction becomes forward-biased, and holes are injected by the emitter into the silicon bar. These holes are drawn to B_1, increasing the current flow in the emitter-B_1 circuit. The increase is quite steep because the presence of these holes in the silicon bar attracts a considerable number of electrons to this region and, in consequence, causes the bar resistance between the emitter and B_1 to drop sharply.

Thus, the unijunction transistor has many of the characteristics of a gas thyratron. Until the control voltage reaches a certain value, determined by the potential between B_1 and B_2 and the silicon-bar temperature, the unit is reverse-biased and essentially cut off. The instant the critical value is exceeded, the emitter PN junction becomes forward-biased, the conductivity between the emitter and B_1 rises sharply (i.e., the resistance of the bar in this region drops to a low value), and the emitter current increases considerably.

A typical characteristic curve of a unijunction transistor is shown in Fig. 9·26. This curve has three distinct regions. Region 1 is the

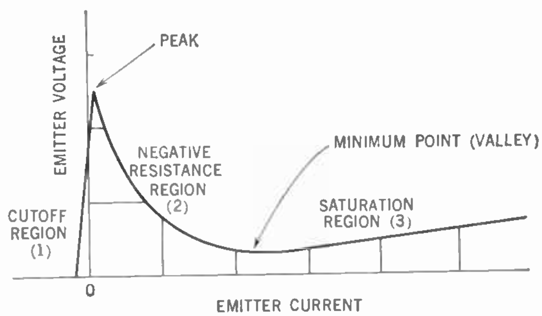


Fig. 9·26 A typical characteristic curve of a unijunction transistor.

cutoff region, where the emitter is reverse-biased. As the voltage applied between the emitter and B_1 rises, the current through the emitter circuit rises too, although the total current seldom exceeds 10 μa. In essence, this is the saturation current obtained whenever any PN junction is reverse-biased.

Region 1 ends when the applied voltage reaches the point marked "peak." At this point, the unijunction transistor "fires," with a large increase in current and a falling off of the voltage drop between emitter and B_1. This is the negative-resistance section, and it is this feature which gives the unijunction transistor its unique properties, as we shall see presently. Eventually there is reached a point, called the minimum, or valley, point, beyond which the device behaves as a positive resistor. That is, the current increases slowly with voltage. This region, called the saturation region, does not figure so prominently as the other two regions in the application of the unijunction transistor.

A typical circuit using this transistor is shown in Fig. 9·27*a*. Initially, capacitor C_1 is not charged and the emitter potential is zero. However, because C_1 is connected to the battery through R_1, it will slowly charge until enough voltage develops across C_1 to forward-bias the emitter junction; at this point the unijunction transistor "fires." C_1 now discharges rapidly through the low resistance of the emitter-B_1 path. With C_1 discharged, the transistor returns to its nonconducting state, and the cycle repeats itself.

The waveforms developed in this circuit are shown in Fig. 9·27*b*. The voltage across C_1 is a sawtooth wave, possessing a slow rise and a rapid descent. At the time of the firing, the current through the entire silicon bar rises, and if a small resistor is placed between B_2 and the

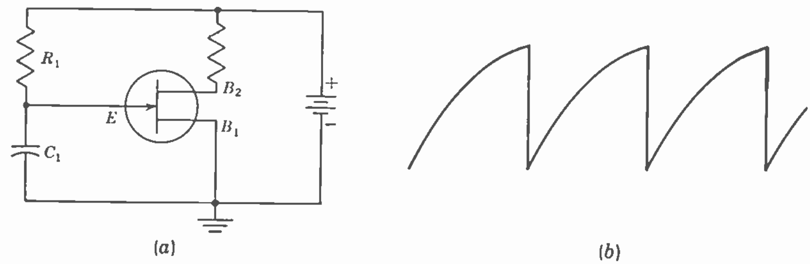


Fig. 9·27 (a) A sawtooth oscillator using the unijunction transistor. (b) Sawtooth wave developed across C_1.

battery, a negative pulse is obtained. By the same token, if a positive pulse is desired, a similar resistor can be inserted between B_1 and ground.

Note the extreme simplicity of this circuit, requiring only two resistors and a capacitor. As a matter of fact, if only the sawtooth wave is desired, the resistor can be dispensed with.

Another application of the unijunction transistor is as a multivibrator. However, before we consider its operation in this form, let us return briefly to the characteristic curve shown in Fig. 9·26. From this graph we see that once the transistor is "fired," it will remain conductive as long as the proper amount of current is maintained through the emitter-B_1 path. However, if the current drops to too low a value, the transistor relapses into its nonconductive state. This is what happened in the circuit of Fig. 9·27 when C_1 was unable to maintain the flow of current through the emitter-B_1 path.

Now, consider the multivibrator shown in Fig. 9·28*a*. Initially, C_1 is not charged; however, with the power on, it will charge through R_2 and diode D_1. The unijunction transistor is cut off at this time and remains cut off until the potential across C_1 becomes equal to or greater than the peak-point potential of the unit. At this instant, the transistor "fires" and the potential of point B (the emitter) drops to a value close to ground potential. This causes D_1 to cut off because now it is reverse-biased, with the cathode more positive than the anode.

With D_1 nonconducting, points A and B are isolated from each other. The transistor remains "fired" because R_2 is passing enough current to keep the transistor activated. At the same time, C_1 discharges through R_1, continuing to do so until the potential at point A is approximately equal to the potential between the emitter and ground. At this point, D_1 starts to conduct again. When this happens, enough

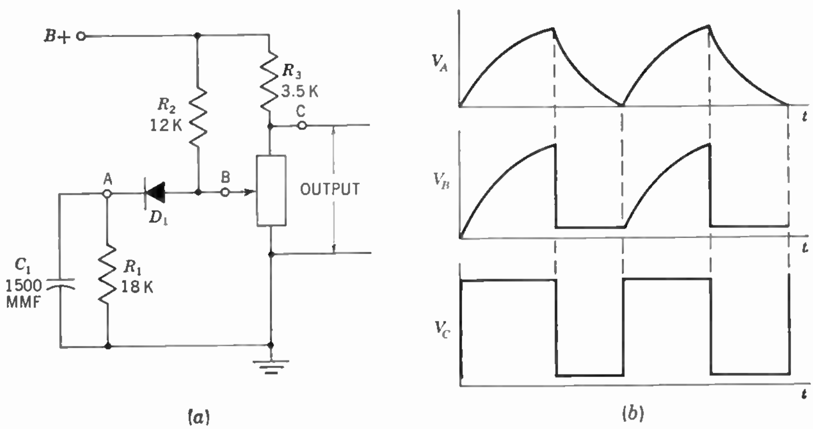


Fig. 9·28 (a) A unijunction multivibrator circuit and (b) the waveforms generated at several points in the circuit.

current is diverted away from the emitter to C_1 to cause the transistor to lapse again into cutoff. The transistor remains in this state until it is retriggered by a suitable voltage rise across C_1.

The waveforms generated at various points in this circuit are shown in Fig. 9·28*b*. The voltage across C_1 follows the usual charge and discharge pattern of RC networks. The waveform at point B follows that of C_1 on charge because D_1 is conducting. However, when D_1 is cut off and the transistor is conducting, the potential at B drops to a value near ground and stays there for as long as D_1 is cut off.

When the transistor is "fired," the current through it rises sharply, producing a square wave across R_3. It is this voltage which can be used as the multivibrator output.

There are a number of additional applications to which the unijunction transistor can be put, including pulse generators, frequency dividers, phase detectors, and one-shot multivibrators. All, however, rely on the basic operation outlined above.

The Field-effect Transistor

Still another form of transistor which is being investigated carefully and which apparently possesses sufficient commercial possibilities to warrant pilot-plant production is the field-effect transistor. Not only does the construction of this unit differ considerably from the construction of all previous transistors, but in it we encounter a new operational method of approach as well.

The basic structure of a field-effect transistor is shown in Fig. 9·29. The body of the device consists of a rectangular block of N-type

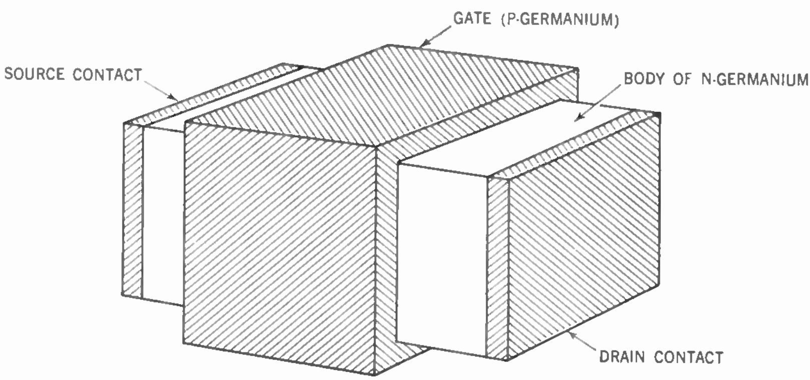


Fig. 9·29 Basic construction of a field-effect transistor.

germanium. Around the center of this block, a continuous P-type junction is formed. (Field-effect transistors have also been built with the gate just on two sides instead of all the way around. Operation remains the same.) In addition, there are two ohmic connections, one at each end of the N-type block. These contacts serve the purpose of bringing externally applied voltages (and signals) in contact with the unit.

In transistors discussed previously the three electrodes were always known as base, emitter, and collector. In the present instance, entirely new names are employed. The P section that is formed on the central block is called the "gate." The ohmic contact at the left is known as the "source," while its counterpart at the right is called the "drain." The reason for the choice of these particular designations will become clearer as we examine the operation of this device.

To start, let us apply a voltage between the source and drain electrodes with the drain terminal made positive with respect to the source.

This is shown in Fig. 9·30. At the same time, let us connect the P-type gate to the source terminal. Under these conditions the gate is said to possess zero potential, the source electrode serving as the reference point for the entire unit.

Electrons will travel from the source to the more positive drain electrode when a voltage is applied to these end terminals. Since a definite potential is being applied across the ends of the N-germanium block and since this material possesses a certain amount of resistance, the applied voltage will be distributed equally along the body of the

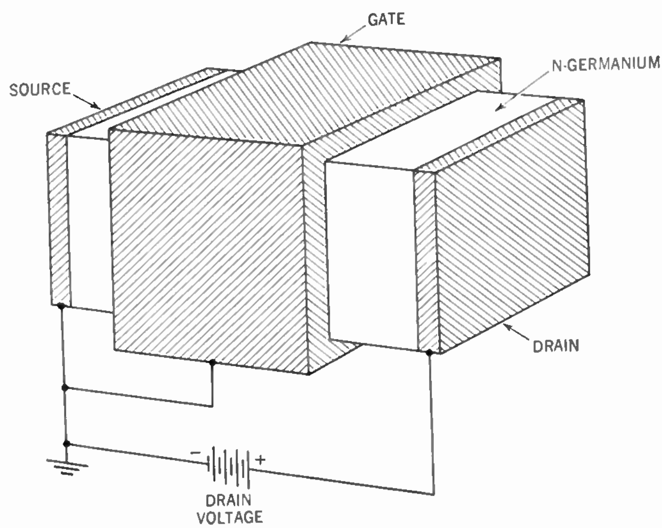


Fig. 9·30 Voltage applied to field-effect transistor. Gate is at same potential as source, while drain is positive with respect to source.

N germanium from the source to the drain. In the present arrangement, the potential will become progressively more positive as we travel from the source to the drain. (If the reader has any difficulty visualizing this distributed voltage drop, let him substitute a resistor for the germanium block. A point on this resistor which is closer to the positive end of the battery will be more positive than any point along the resistor to the left. The same situation holds for the germanium crystal.)

Now consider the gate. This forms a PN junction with the germanium block over the area in which the two are in contact. The end of the gate nearest the source will find the least amount of potential difference between it and the body of the germanium block just underneath. This is at point *A* in Fig. 9·31, and the reason is quite simple. The gate itself is at the same potential as the source, since the two are externally directly connected. Inside the body of the germanium block, however, at point *A*, there exists a small positive potential (with respect to the source) because of the above-mentioned voltage drop. This positive potential at *A* repels the holes in the end of the P-type gate just above it and also exerts an attractive force on the electrons in the germanium block. Thus, a small reverse bias is present here.

As we progress further down the block, the positive potential (with respect to the gate) increases, making the reverse bias across the PN

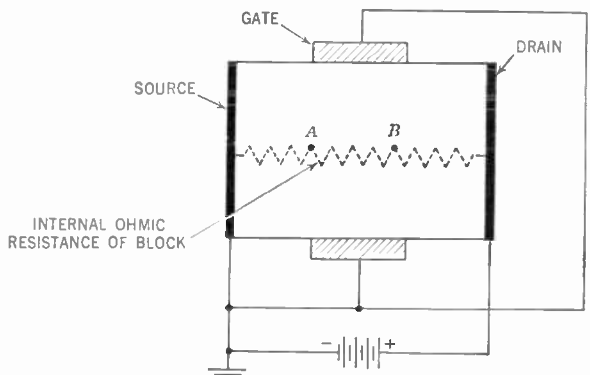


Fig. 9·31 A side view of field-effect transistor. The internal ohmic resistance of the germanium block is represented by the dotted resistor. Owing to the voltage drop acrcss this resistance, point *B* is more positive than point *A*.

junction even greater. Thus, there is no tendency on the part of the electrons in the block to flow to the gate or for the holes in the gate to cross the junction and move into the block.

However, holes do exist in the block. These come from three sources. (1) They may be thermally generated in the body of the semiconductor; (2) they may be developed at the surfaces of this block; or (3) they may come from the two end contacts, the source and drain metal electrodes. In any event, a certain number of holes exist in the N-type germanium, although the donor electrons exceed these by ratios as high as 10:1 or more.

Under the repelling effect of the positive voltage drop along the semiconductor body and the attractive force of the more negative gate, holes in the semiconductor will be drawn up to the gate. The number of holes which are drawn from any section of the block will be governed by the positive potential present in that area. Thus, more holes will be drawn from the right-hand section of the block than the left-hand section.

Now consider the P-type gate. It will possess a number of free electrons, possibly for some of the same reasons that the N-type body possesses holes. And these electrons will be repelled by the negative potential of the gate and attracted by the more positive potentials within the germanium block. Hence, there will be a movement of electrons out of the gate and into the block, with the greatest number of electrons leaving the gate at the right-hand side of the germanium body.

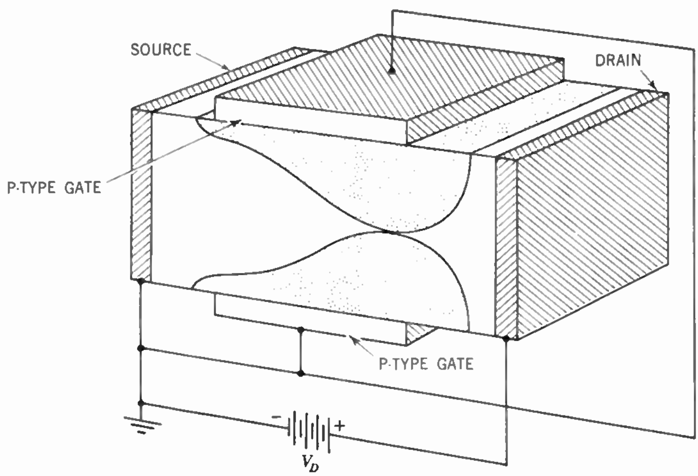


Fig. 9·32 The shape of the space-charge distribution within the body of a field-effect transistor. (IRE)

One result of this redistribution of charge is to make the PN junction more reverse-biased than before. Another result is to increase the negative space charge in the germanium block, with the concentration greatest at the right. The actual space-charge distribution is shown in Fig. 9·32. It rises to a maximum at the right-hand edge of the P gate and then decreases fairly rapidly. This space charge exerts a repelling force on those electrons traveling from the source to the drain electrodes because of the externally applied potential. What it actually does is channel, or direct, the current flow into the regions between the concentrated space charge. These regions are shown in white in Fig. 9·32. The dotted area represents the negative space charge.

As the drain voltage is increased, the current flow through the semiconductor will rise until the drain voltage reaches a certain critical limiting value which is referred to as the pinch-off voltage. Beyond this point, no further increases in drain current will result as the drain voltage is made greater. The only effect of higher voltages is to alter the shape of the channel through which the electrons flow.

Thus far, the gate has been held at the same potential as the source. If, now, we make the gate negative with respect to the source terminal by the insertion of an additional negative voltage, then the amount of external voltage needed between source and drain in order to reach the constant drain current value will be much less. This is because the negative gate voltage aids the space charge in the germanium body to

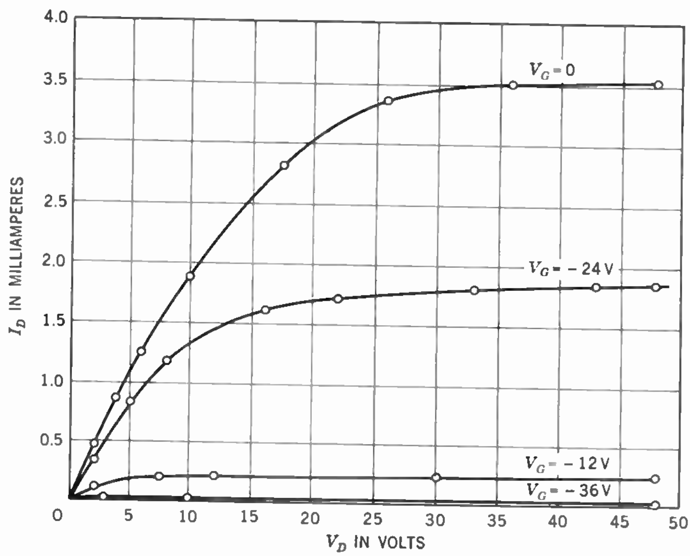


Fig. 9·33 The characteristic curves of a field-effect transistor. (IRE)

narrow the channel of the current. Eventually, if the negative potential of the gate is made high enough, current flow through the device will be halted completely.

The foregoing behavior is depicted by the curves in Fig. 9·33. Maximum saturation current flows when V_G (the gate voltage) is zero with respect to source. For each successively lower current curve, the negative bias on the gate becomes greater. When the gate potential is equal to the pinch-off voltage, the amount of current flowing through the semiconductor is very small.

The circuit in Fig. 9·34 is similar to that of Fig. 9·30 with the addition of a signal in series with the gate bias battery. As the signal varies, it will vary the total voltage applied to the gate. When the signal is negative, the gate will become more negative and serve to reduce the flow of drain current. On the next half cycle, when the signal goes positive, the overall gate voltage will be made less negative. This will reduce the space charge in the body of the semiconductor, and the drain current will rise. In short, a small voltage variation at the gate will produce a sizable current variation in the germanium crystal. In a vacuum tube, this characteristic is labeled "transconductance" and is responsible for the amplification which is obtained. Obviously the same results should be possible in the field-effect transistor, enabling us to use this device to achieve signal amplification.

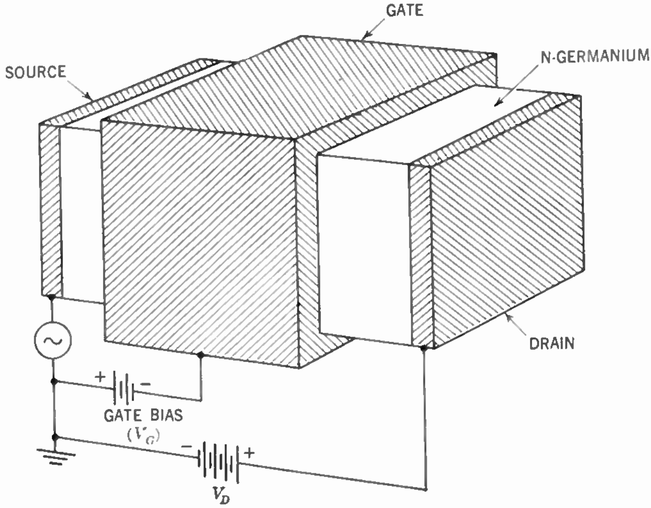


Fig. 9·34 The field-effect transistor with a negative gate bias voltage V_G. Also shown is the point where a signal would be injected.

The similarity between the field-effect transistor and triode vacuum tube is quite marked. In a vacuum tube, the grid potential regulates the space charge existing between cathode and plate and, by this control, determines the extent of plate current flow. In the field-effect transistor, the polarity of the gate governs the intensity of the negative space charge existing in the N material and, through this, the amplitude of the current flowing from source to drain.

Another interesting feature of this transistor is the fact that essentially only one type of carrier, here electrons, is involved in the process. (A field-effect transistor could be constructed by using a P-type main body and an N-type gate. Battery potentials would have to be reversed, and the current in the source-drain channel would be carried by holes. Similar results would be obtained, however.) This is in sharp distinction to the conventional transistor, where both types of carriers play roles. Hence, this device is called a unipolar transistor. It also results in a very significant difference in the way amplification is achieved. In the conventional triode transistor, the carriers must travel from the emitter through the base to the collector. This makes carrier transit time an important factor in determining frequency response. In the field-effect transistor, the signal at the gate serves to modulate the drain current and in this way produce the signal variations in the drain output circuit. Carriers do not have to transport the signal from the gate to the drain; hence, carrier transit time is not involved. We do not even have to worry about transit time from source to drain, because the signal voltage at the gate merely expands or contracts the current stream; it does not alter the rate of travel. This does not mean that the field-effect transistor is without frequency limitation. It is not, because there are certain shunting capacitances to be dealt with plus the variation of transconductance with frequency and other effects. But transit time does not possess the significance here that it does in other transistor structures.

Operating frequencies in the megacycle range have been obtained with experimental field-effect transistors. Another important feature of these devices is their high input and output impedances. In this respect they resemble pentodes. However, the units do possess a fairly high noise figure, which means that they are not suitable for low-power applications. They are also useful in various types of oscillators, such as multivibrators, Colpitts oscillators, and relaxation circuits.

The Tunnel Diode

The tunnel diode is a new semiconductor device capable of operation in the 1- to 10-kilomegacycle range. Although it is used in the conventional applications of amplifying, oscillating, and switching, its principle of operation is entirely different from the transistor or the vacuum tube. The name "tunnel diode" has been adopted because the physical mechanism by which the device functions is caused by a complex quantum-mechanical tunneling process. In actual practice, the tunnel diode is basically a very heavily doped PN junction, and thus, as one might expect, it possesses many of the properties of a conventional diode.

The tunnel diode has two outstanding properties: (1) extremely high frequency response and (2) very low power consumption. For example, where a vacuum tube may operate at 100 Mc and consume 1 watt of power and a conventional transistor may operate at 100 Mc and consume 10 mw of power, a tunnel diode can operate at 500 Mc and consume 1 mw of power. The tunnel diode is not without its problems, however. Instability and unwanted signal feed-through plague the device, and these difficulties are not easily resolvable. The instability arises because a tunnel diode is a negative-resistance device and negative resistance is always difficult to control. Feed-through arises because the tunnel diode is a two-terminal device and its input and output terminals are the same. Compare this with the vacuum tube and transistor, where the third lead offers isolation between input and output.

The discussion to follow will present the operation, physical characteristics, and applications of the tunnel diode. The important parameters of the device will be examined, and several figures of merit will be introduced.

Operation of the tunnel diode. The operation of the tunnel diode is dependent upon a quantum-mechanical principle called tunneling. To appreciate its significance, let us review semiconductor-diode operation. Each diode contains two sections, a P section and an N section. The P section possesses an excess of holes, while the N section possesses an excess of electrons. Between the two sections there is a depletion layer containing very few free electrons or holes. That is, there are no mobile charges in this region. This depletion region amounts to a potential barrier, and to pass current through the diode, sufficient external voltage must be applied, with the proper polarity, to overcome the potential barrier. The proper polarity, of course, is connection of the negative battery terminal to the N section and the positive battery terminal to the P section.

In the absence of any external voltage or in the absence of sufficient voltage, very little current will pass across the junction. However, some carriers on both sides of the junction will attain enough thermal energy to surmount the barrier presented by the depletion layer and reach the other side. This, however, occurs only to those relatively few electrons and holes capable of attaining a sufficiently high energy state.

Classical physical theory states that unless a carrier possesses enough energy to overcome a potential barrier, it will *never* cross or surmount that barrier. The more recent quantum mechanics, however, contradicts classical physics in this instance and states that a carrier can reach the other side of a potential barrier even though it does not have enough energy to surmount the barrier. The carrier does this by tunneling through the potential hill. In fact, quantum mechanics can predict the probability of this occurrence. The probability of a carrier tunneling through the potential hill is dependent on how far the particle must tunnel, i.e., the thickness of the potential hill. The probability of tunneling is nil unless the barrier is extremely narrow. This is proved by the fact that tunneling does not occur in a conventional PN junction.

To make a tunnel diode, the PN junction must be very heavily doped with impurities. The large number of impurities produces a very narrow depletion layer. With this thin layer, electrons can tunnel their way from the N region to the P region. This gives rise to an additional current in the diode at very small forward bias which disappears when the bias is increased. It is this additional current

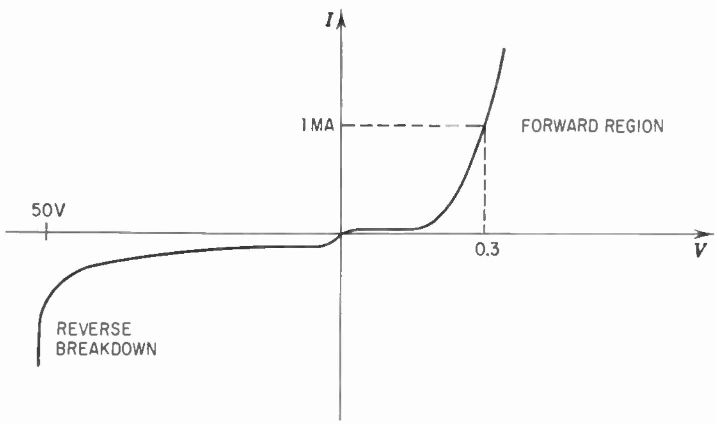


Fig. 9·35 The voltage vs. current characteristic curve of a conventional diode.

that produces the negative resistance in a tunnel diode. It has also been found that the electrons travel through the depletion layer at tremendously high velocities. This enables the tunnel diode to operate at far higher frequencies than conventional transistors. The theoretical frequency limit is in the neighborhood of 1 million Mc.

Figure 9·35 shows the voltage vs. current curve of a conventional PN junction diode. The forward voltage required to cause current flow is approximately 0.3 volt for germanium diodes and 0.8 volt for silicon diodes. The reverse breakdown usually occurs between 20 and 200 volts. If we superimpose on this characteristic curve the tunneling current which occurs at a forward bias of about 50 to 100 mv, we obtain the characteristic of Fig. 9·36. The composite characteristic is shown in Fig. 9·37. Note that the tunneling current appears at very small voltages. As we continue to raise the applied voltage, the carriers receive enough energy to surmount the potential barrier and cross the junction in the conventional manner.

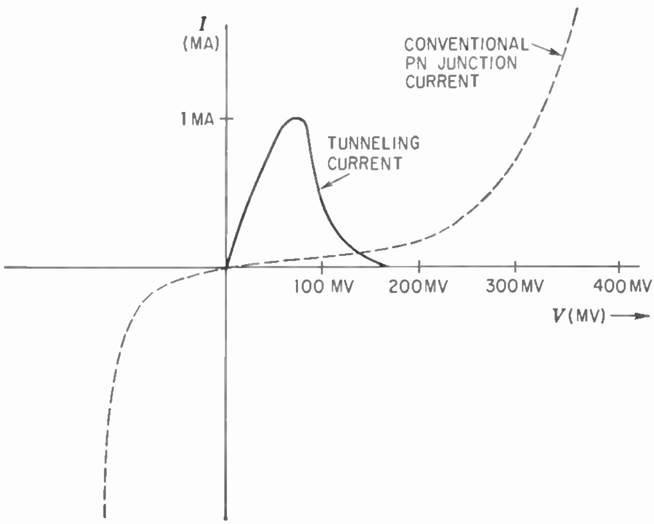


Fig. 9·36 Combination of tunneling current and conventional PN junction current.

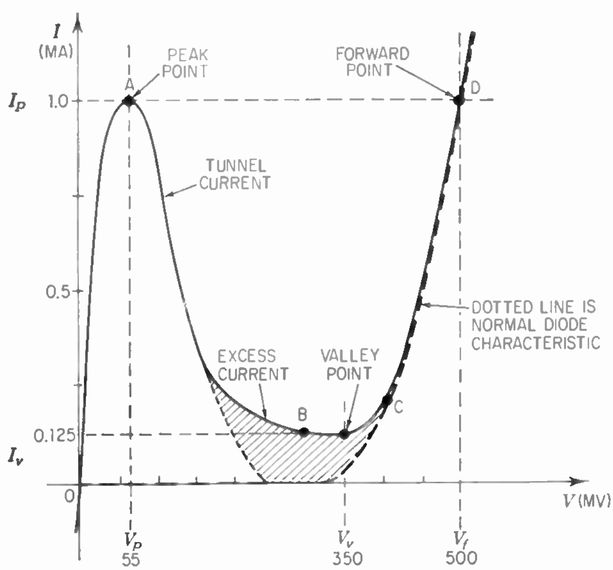


Fig. 9·37 Composite characteristic representing overall operational curve of a tunnel diode.

In the composite curve of Fig. 9·37, the dotted line in the valley region shows the curve one would expect to get if the two curves of Fig. 9·36 were superimposed. The solid line shows the actual curve which is obtained. The excess current which occurs in the valley region is as yet unexplained. The important characteristic to note is that the region between points A and B represents a negative resistance; i.e., as the voltage is increased, the current decreases (just the reverse of the behavior of an ordinary resistor). Before studying the tunnel diode as a circuit element, it will be useful to characterize the device on both an a-c and d-c basis and derive some of the important parameters which are usually specified on a tunnel-diode data sheet.

Parameters of the tunnel diode. Figure 9·37 shows the total voltage-current curve with various d-c parameters noted. Typical values of these parameters for a germanium tunnel diode are also given.

D-C Parameters. Let us study these various parameters and point out their significance. I_P is the peak current and is dependent upon the junction area and the doping that is used. The actual value of I_P is determined by the intended application. Common values are 1, 5, 10, and 100 ma. Of extreme importance in switching circuits is the variation of I_P from unit to unit. The tolerance of this parameter is usually held to within ± 5 per cent or less. V_P, the voltage at which the peak current occurs, is generally in the region of 40 to 100 mv for germanium diodes.

I_V, called the valley current, is the lowest current on the voltage-current curve. This current should be as low as possible. Actually, I_V should be small in comparison to I_P, because the largest possible current swing of the device is $I_P - I_V$.

V_V is the voltage at which the valley current occurs. Its value ranges from 200 to 350 mv for germanium diodes. V_V represents the cross-over point where the resistance goes from negative to positive.

V_f, called the forward voltage, is measured to a current equal to I_P. Typical values for germanium diodes range from 300 to 500 mv. The difference between V_f and V_P represents the largest possible voltage swing from the tunnel diode. If larger (0.5- to 1-volt) voltage swings are required, silicon or gallium arsenide tunnel diodes are used.

All of the above information refers only to d-c parameters of the tunnel diode; a discussion of the a-c characteristics of the tunnel diode follows.

A-C Parameters. In order to form an a-c equivalent circuit for the tunnel diode, we must consider what properties of this device would effect its a-c operation.

First, the tunnel diode exhibits a negative resistance. This negative resistance is given by the slope of the voltage-current characteristic shown in Fig. 9·37, and, for the curve shown in Fig. 9·37, is about -150 ohms. A moderate value of negative resistance is most desirable. If the negative resistance is too small, it limits the use of the tunnel diode to low-impedance circuits. If the resistance is too high, the problem of a-c instability (oscillation) arises. The value of negative resistance is virtually independent of frequency.

As in the conventional PN junction diode, the tunnel diode also has a barrier capacitance. This capacitance is analogous to the collector-junction capacitance in a transistor; typical values of it range from a few micromicrofarads for diodes of very small area to hundreds of micromicrofarads for large-area diodes. This junction capacitance is shown in the simple equivalent circuit of Fig. 9·38 shunting the a-c negative resistance.

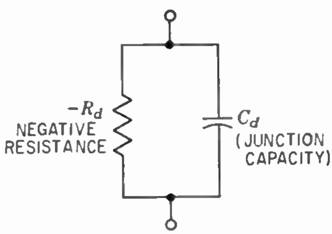


Fig. 9·38 Simple equivalent circuit of a tunnel diode.

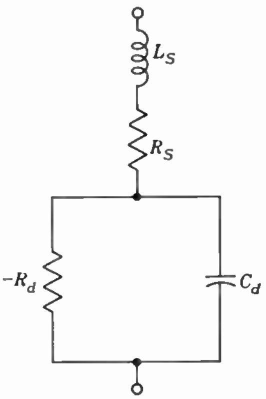


Fig. 9·39 A more complete equivalent diagram of a tunnel diode.

In addition to the foregoing, the tunnel diode has in series with it two parasitic elements, spreading resistance R_S and lead inductance L_S. The spreading resistance is a combination of the resistance of the bulk semiconductor material and of the leads. Its value is in the order of 0.1 to 10 ohms. The lead inductance is in the order of a few millimicrohenries; however, even this small inductance can cause problems with parametric oscillations. Both the series spreading resistance and the lead inductance should be kept as low as possible.

Figure 9·39 shows the equivalent circuit obtained when all of the foregoing factors are taken into consideration.

Figures of Merit. The equivalent circuit of Fig. 9·38 may be used to develop the various figures of merit associated with the tunnel diode. Let us consider the figures of merit which pertain to linear, small-signal circuits.

An important figure of merit for both video and tuned tunnel diode amplifiers is a quantity called the transducer gain-bandwidth product.

This is somewhat analogous to the gain-bandwidth product of a transistor f_T discussed in Chap. 12 and is a measure of the frequency response of the diode. It is

$$\sqrt{G_t}\,\mathrm{BW} = \frac{1}{2\pi R_d C_d}$$

where R_d is the negative resistance, C_d is the junction capacitance, and BW is the bandwidth. The important point to notice is that the square root of the power gain multiplied by the bandwidth is a constant that is dependent only upon the parameters of the diode. It also shows that as the gain G_t increases, the bandwidth decreases, so that at very high gains the bandwidth will be narrow, and vice versa. With presently available devices it is possible to achieve gain-bandwidth products on the order of 500 Mc.

Another important figure of merit is the cutoff frequency. Unlike the vacuum tube and transistor, the tunnel diode has two cutoff frequencies. One cutoff frequency, the resistive cutoff frequency, is

$$f_{COR} = \frac{1}{2\pi R_d C_d}\sqrt{\frac{R_d}{R_S} - 1}$$

This is the frequency at which the resistive part of the diode impedance, measured at its terminals, goes to zero. Above this frequency the tunnel diode has a gain less than 1 and is thus useless as an amplifier. The reactive cutoff frequency is given by

$$f_{COX} = \frac{1}{2\pi}\sqrt{\frac{1}{L_S C_d} - \left(\frac{1}{R_d C_d}\right)^2}$$

This cutoff frequency is sometimes called the self-resonant frequency, because it represents the frequency at which the tunnel diode will oscillate of its own accord. Actually, this frequency represents the point at which the reactive component of the diode impedance goes to zero.

In an actual application, both of the above frequencies are reduced by the effects of the external circuitry. Thus, the highest frequency of operation is very dependent upon the types of circuit, the component values, and the diode package. Lead inductance is especially important and extreme care should be taken to keep this element as small as possible. The newest tunnel diodes are packaged in microstrip, pill, and microwave cases in order to reduce lead inductance.

The resistive cutoff frequency is the maximum frequency at which the tunnel diode may be used as an oscillator. This is a natural consequence of the fact that this is the frequency at which the power gain falls below 1.

Another important figure of merit is the efficiency of the tunnel diode when used as an oscillator. The formula for the efficiency of a class A oscillator is

$$\text{Eff} = 25\left(1 - \frac{R_S}{R_d}\right)\left[1 - \left(\frac{f}{f_{COR}}\right)^2\right] \quad \text{per cent}$$

where R_S = series diode resistance
f = frequency of oscillation
f_{COR} = resistive cutoff frequency

Note that if $R_d \gg R_s$, a condition that is normally true, the maximum low-frequency efficiency is 25 per cent. This is just one-half that attainable when using transistors.

Still another important figure of merit is the noise figure of a tunnel diode. This is expressed by

$$\text{NF} = 10 \log\left(1 + \frac{IR_d}{52}\right) \quad \text{db}$$

where NF = noise figure
R_d = the negative diode resistance, ohms
I = the d-c bias current, ma

Unlike the noise figure of transistors, the noise figure NF of the tunnel diode is virtually independent of frequency over its entire operable frequency range. The noise figure of the tunnel diode is quite low, usually in the 3- to 5-db range.

A very useful area of application for tunnel diodes is in switching. Since the tunnel effect proceeds with the speed of light, the tunnel diode is not plagued with the problem of carrier transit time (across the base region) as the transistor is. A lowering factor on this switching speed is the shunt capacitance C_d of the tunnel diode structure. As a simple approximation, assume that the characteristic curve of the tunnel diode may be approximated as shown in Fig. 9·40.

If, now, we switch the tunnel diode from its low state to its high state or vice versa, the shunt capacitance C_d is being charged by a current of approximately $I_P - I_V$. The switching time is determined by the time it takes this current to charge the capacitance to voltage V_f. The resulting switching time is very closely predicted by the expression

$$t = C_d \frac{V_f - V_P}{I_P - I_V}$$

Both I_p and C_d are directly proportional to the PN junction area. But we wish I_p to be large and C_d to be small to minimize t. Therefore, the ratio I_p/C_d is a measure of the merit of the material used to fabricate the tunnel diode. It has the dimensions of milliamperes per micromicrofarad or volts per nanosecond. This figure of merit is also applicable to small-signal evaluations. It has been noted that for operation above 1,000 Mc, an I_p/C_d ratio of at least 0.1 ma per $\mu\mu$f is required.

Package and external inductances should be kept to a minimum, since increasing L_s will result in undesirable overshoots and delays, which increase the overall switching time.

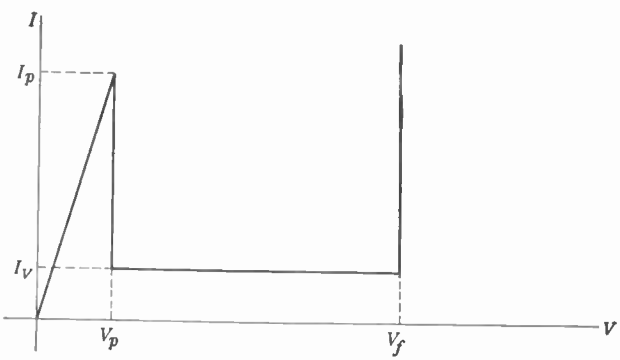


Fig. 9·40

Parameter Variation. One of the advantages of the tunnel diode is its environmental insensitivity. For example, the tunnel diode will survive intense nuclear radiation which would ordinarily alter transistor characteristics permanently. However, a more common environmental consideration is the effect of temperature variations on the tunnel diode parameters.

The key parameter in nearly every tunnel diode switching circuit is I_p. For reliable operation, the peak currents of the tunnel diodes employed must be matched to within 3 per cent to permit operation with practical supply and component tolerances. Fortunately, the I_p of tunnel diodes can be accurately controlled during manufacture. The peak current can be temperature-stabilized during the manufacture of the tunnel diode by the proper choice of materials and the temperature-time cycle of the alloying process. It is possible to obtain a zero coefficient of peak-current temperature dependence at room temperature. There is generally a reduction of I_p when the diode is operated at extremely cold or warm temperatures (outside the range of 0 to 55°C).

The valley current, often referred to as excess current which ideally should be zero, shows a nearly linear increase with temperature. A typical tunnel diode has an I_V of 0.09 ma at -25°C, 0.11 ma at room temperature, and 0.15 ma at $+75$°C.

The peak, valley, and forward voltages, V_P, V_V, and V_f, all decrease with increasing temperature. The change in V_f, approximately -1 mv per °C, is the largest of the three. This is lower than the typical 2-mv per °C decrease in the forward voltage drop of a conventional germanium diode or the forward V_{BE} of a germanium transistor.

Commonly used symbols for the tunnel diode are shown in Fig. 9·41. If the anode is positively biased with respect to the cathode, the

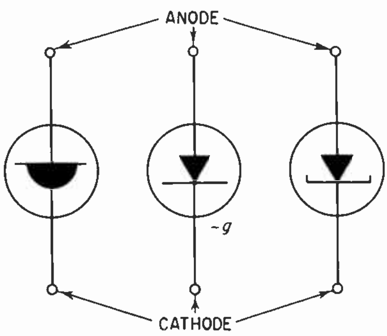


Fig. 9·41 Commonly used symbols for the tunnel diode. (After "Tunnel Diode Manual," General Electric Co., 1961)

tunnel diode will display its negative resistance characteristic. In the middle illustration, the $-g$ next to the symbol implies that the unit is a negative-resistance device. (g stands for conductance, the reciprocal of resistance.)

Backward Diode

Many of the circuits which utilize the tunnel diode also include a backward diode. The backward diode is essentially a tunnel diode in which the peak current I_P is as close to zero as possible. This is shown in the voltage-current (V-I) characteristic of Fig. 9·42.

In tunnel diodes, a large current will flow when a voltage is applied in the reverse direction. As a matter of fact, this current will increase continuously as the applied voltage is increased (up to the point where the current flow generates so much heat the diode is destroyed). Thus, with a forward voltage applied, we obtain a much smaller current than when a reverse voltage is applied. The backward diode is fashioned to emphasize this difference in currents. The term "backward" means that the diode conducts heavily with a negative rather than a positively applied voltage. It finds considerable application as a high-conductance coupling diode.

The backward diode can also be considered as a conventional diode with a very low reverse breakdown voltage. Its symbol, for the most part, is the same as that for diodes.

Tunnel Diode Applications

Three areas in which tunnel diodes have been applied are as oscillators, as amplifiers, and in switching. Of these, use as oscillators

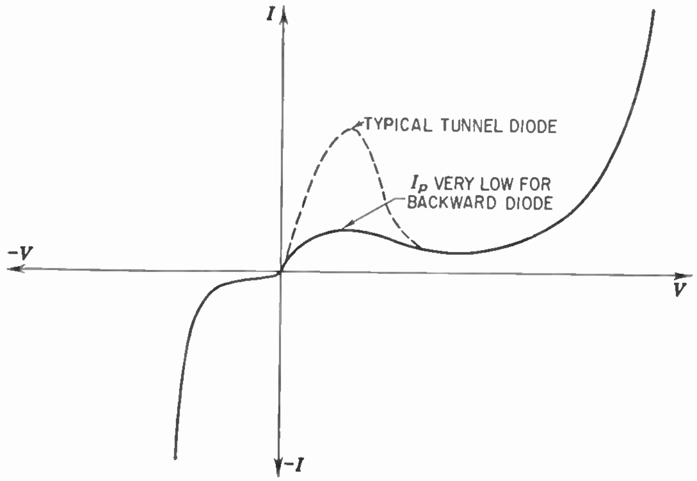


Fig. 9·42 Comparison of forward characteristics of tunnel and backward diodes.

and use in switching are of the most practical value because of the results achievable not only in regard to stability but also in comparison with other semiconductor devices.

Oscillators. To achieve oscillation with a tunnel diode, it must be set up so that the negative resistance it provides is greater than the positive resistance of the resonant components in the circuit. A typical tunnel diode has a negative resistance of approximately −100 ohms when it is biased at the center of its negative-resistance region. To bias the diode correctly, approximately 125 mv is required. At this voltage, the current drawn by the diode is approximately 0.5 ma.

There are several ways to set up this bias circuit. We can connect the tunnel diode in series with a resistor and a small battery. If we assume a battery voltage of 3 volts, then a series resistor of 6,000 ohms would be required to limit the current flow to 0.5 ma. But the 6,000 ohms would completely overshadow the -100 ohms of the tunnel diode and prevent us from properly utilizing this negative resistance.

This difficulty can be circumvented by the resistive arrangement shown in Fig. 9·43. The necessary bias voltage is here developed across a 20-ohm resistor, a value considerably less than the -100 ohms of the tunnel diode.

We can now connect a resonant circuit in series with the tunnel diode, and if this circuit possesses a resistance less than 80 ohms, then

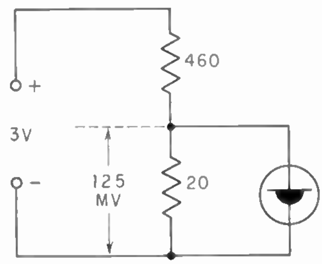


Fig. 9·43 A method of biasing a tunnel diode.

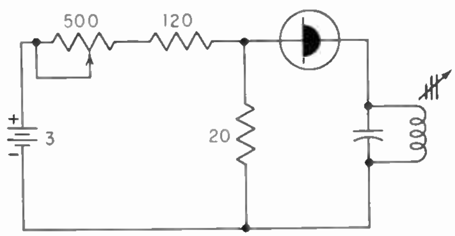


Fig. 9·44 A tunnel diode oscillator.

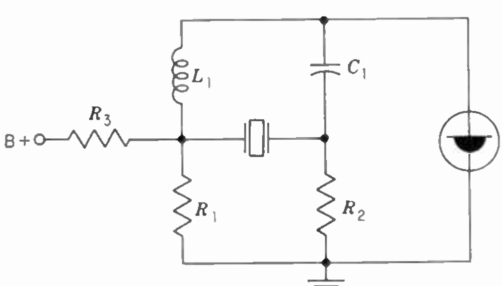


Fig. 9·45 A tunnel diode crystal-controlled oscillator. (After "Tunnel Diode Manual," General Electric Co., 1961)

the tunnel diode will completely balance out this resistive loss of the circuit and enable oscillations to take place. A suitable circuit capable of oscillating into the megacycle range is shown in Fig. 9·44.

A crystal-controlled oscillator using a tunnel diode is shown in Fig. 9·45. R_1 and R_2 are selected to each have about twice the value they should have to enable the negative resistance of the tunnel diode to predominate the circuit. A crystal is placed between these resistors. At all frequencies other than the resonance, the crystal impedance is high and the circuit is unable to function. At the resonant frequency, however, the crystal becomes a short circuit and R_1 is placed in parallel with R_2, thus reducing their total resistance value to half their individual values. This new value permits the circuit to oscillate freely at a frequency accurately governed by the crystal.

Amplifiers. If we increase the impedance of the external circuit connected across a tunnel diode until it equals the negative resistance of

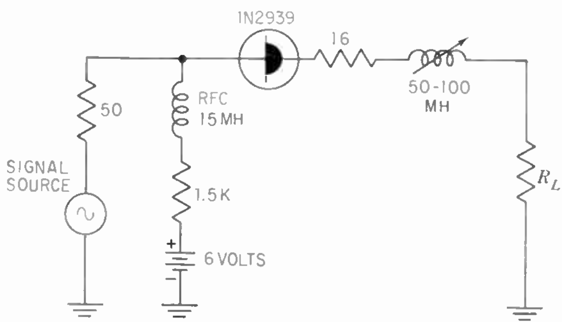


Fig. 9·46 A tunnel diode amplifier.

the tunnel diode, amplification rather than oscillation is obtained. This is done in the 100-Mc amplifier shown in Fig. 9·46. The 1N2939 tunnel diode has a negative resistance which is just counterbalanced by the circuit positive resistance. The latter, in this instance, is equal to 50 ohms from the signal source, 50 ohms from the load, 2 ohms from the internal lead resistance of the 1N2939, and 16 ohms from

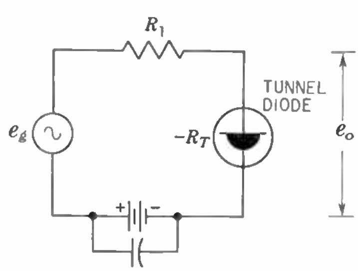


Fig. 9·47 Simple circuit to demonstrate ability of tunnel diode to amplify. (After W. L. Walters and J. V. Claeys, The Tunnel Diode Story, *Radio-Electronics*, July, 1960)

the series resistor. In this particular configuration, the 16-ohm resistor was added simply to achieve this counterbalancing.

While the actual mathematical justification for the above-indicated condition is quite complex, some inkling of how it is arrived at may be seen from an examination of the simple circuit shown in Fig. 9·47. Assume that the tunnel diode is biased to the center of its negative-resistance range and that e_g represents an incoming signal and e_o is the output signal. Then the voltage across the tunnel diode, with negative resistance $-R_T$, is

$$e_o = \frac{e_g(-R_T)}{R_1 + (-R_T)}$$

When R_1, the circuit positive resistance, exactly equals $-R_T$ in value, e_o becomes infinite. In an actual circuit, of course, this does not happen, but the largest output is obtained when $R_1 = R_T$. At this point, the ratio of e_o to e_g, or the voltage gain, is greatest.

Switching circuits. Perhaps the most important application of the tunnel diode is its use in switching circuits. Such switching circuits

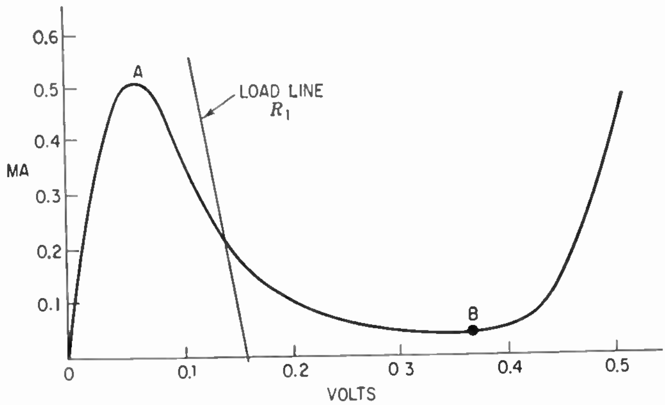


Fig. 9·48 Tunnel diode characteristic and a load line representing load resistor R_1. (After W. L. Walters and J. V. Claeys, The Tunnel Diode Story, *Radio-Electronics*, July, 1960)

may be simple AND or OR logic networks, relaxation or multivibrator oscillators, or bistable oscillators or flip-flops. In fact, a host of various logic circuits can be designed around the tunnel diode, all capable of switching from one state to the other in the order of nanoseconds (10^{-9} sec).

To fully appreciate the operation of a tunnel diode in a switching circuit, let us reexamine its characteristic curve, shown in Fig. 9·48. For amplification and sinusoidal operation, we are primarily interested in keeping the diode biased to the center of its negative-resistance segment. This is the region roughly extending from point A to point B. To further obtain the desired operation, the resistance of the circuit connected to the tunnel diode must be such that its load line intersects this negative-resistance segment with an angle greater than the angle or slope of this portion of the characteristic curve itself. Thus, load line R_1 satisfies this condition. *Note:* Load lines are discussed at length in Chap. 12. They are mentioned here because they present a clearer picture of tunnel diode operation. A load line represents the load resistance, or impedance, and its effect on a circuit. This will become more evident as we proceed.

The more vertical the load line, the less resistance it represents. By the same token, the more horizontal the line, the greater the load resistance in the circuit. Thus, the load line in Fig. 9·49 means that there is more resistance (of the positive variety) in the tunnel diode external circuit. Note, though, that this load line cuts across the tunnel

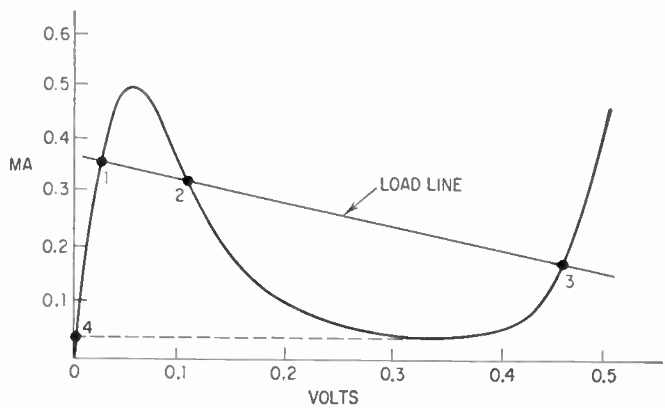


Fig. 9·49 A load line representing a higher load resistor than the load line of Fig. 9·48. (After W. L. Walters and J. V. Claeys, The Tunnel Diode Story, *Radio-Electronics*, July, 1960)

diode characteristic curve at points 1, 2, and 3. Points 1 and 3 are stable; point 2 is not. To demonstrate this, assume that the voltage corresponding to point 2 is being applied to the diode and the current represented by that point is flowing through the unit. If, for some reason, a slight increase in current should occur, then by looking at the characteristic curve we see that this would mean less voltage drop across the diode. If we assume the circuit of Fig. 9·47, then we see that if this happens, there is more voltage available to send current through the circuit. This current increase causes a further decrease in the voltage across the diode.

The foregoing action continues until point 1 is reached. At point 1, if there is to be any further increase in current, there must be an increase in voltage across the diode because now we are in the positive-resistance region. From Fig. 9·47 we see that the only way the voltage across the diode can increase is for the voltage across R_1 to decrease. This can happen only if the circuit current decreases. Thus, point 1 is a stable point and the circuit remains there.

If we had started from point 2 with a decrease in current instead of with an increase, the circuit would have ended up at point 3. This, too, is a stable point because it also is in a positive-resistance region. Thus, with a load line such as that shown in Fig. 9·49 there are two stable states that the circuit can assume. One of the points, point 1, is a low-impedance point possessing high current flow through the tunnel diode and low voltage across it, and point 3 is a high-impedance point. Here the diode current is low and the voltage across the diode is relatively high.

To operate the tunnel diode as a switch, let us initially bias it to operate at point 1. If we now inject a strong current pulse into the circuit, we essentially move the operative point up to the peak-current point and then down the other side (i.e., the negative-resistance side) to point 3. The speed of switching between these two points or states is exceedingly high.

The tunnel diode will remain at point 3 indefinitely. To shift it back to point 1, we simply inject a pulse of current of opposite phase into the circuit. This counteracts the current flowing and drops the current value below the valley point. This shifts the operating point from the valley point straight across to the left until point 4 is reached. After the injected current pulse has disappeared, the circuit current readjusts itself to the value indicated by point 1, and the tunnel diode is back to its initial state again. The switching speed is just as high in this changeover as it was in the first shift.

This switching behavior is readily put to work in a variety of circuits. For example, consider the simple arrangement shown in Fig. 9·50. The tunnel diode is biased to operate at point 1. If, now, a positive current pulse through R_2 raises the current through the diode above the peak of its characteristic curve, the diode will switch to point 3. If current pulses through R_3 can do the same thing, this arrangement becomes an OR circuit. If the diode is biased farther down on its characteristic curve, Fig. 9·50*c*, and it requires simultaneous pulse currents in R_2 and R_3 to exceed the peak, then this is an AND arrangement. With only one pulse present, no switching occurs because the operating point never is able to surmount the peak and move into the negative-resistance region.

In the above circuit, no output pulse is obtained so long as the tunnel diode is at point 1 because there its impedance is close to zero (low-voltage, high-current condition). However, when the tunnel diode is switched to point 3, an output pulse is obtained because now

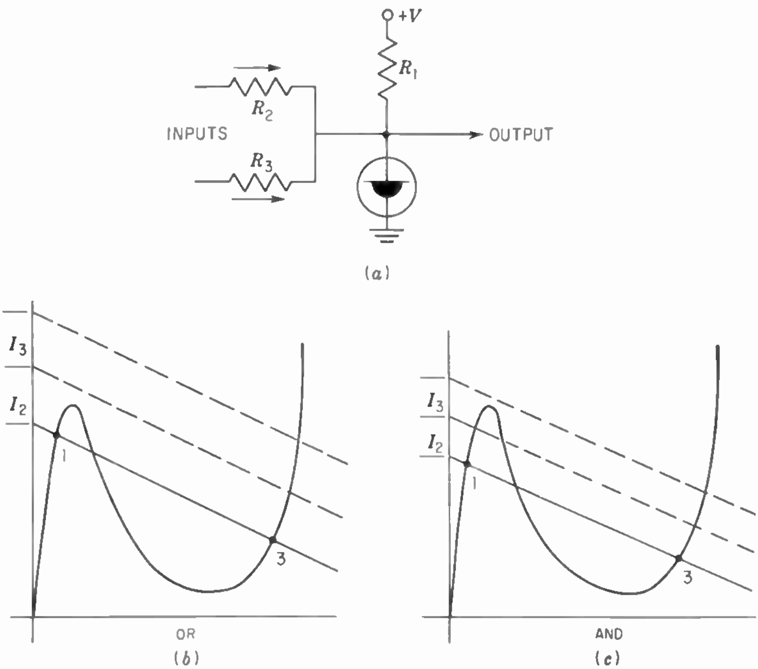


Fig. 9·50 A tunnel diode circuit for OR and AND operations.

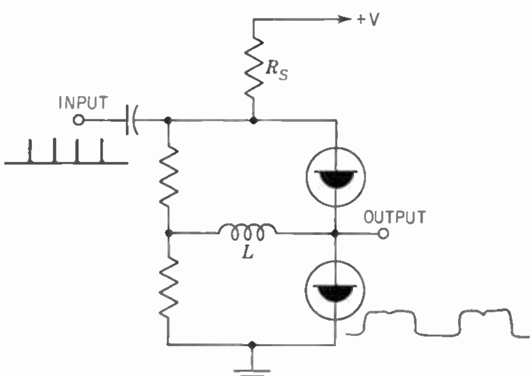


Fig. 9·51 A tunnel diode flip-flop circuit. (After "Tunnel Diode Manual," General Electric Co., 1961)

the diode impedance is high and a useful voltage drop appears across it.

A tunnel flip-flop circuit is shown in Fig. 9·51. A value of d-c supply voltage is chosen such that only one tunnel diode can be in the high-voltage state at any instant. The difference between the two tunnel diode currents flows through the inductance. When a positive trigger pulse turns the diode which is in the low-voltage state to the high-voltage state, the voltage induced in the inductance, because of the decreasing current through it, is of a polarity to reset the other tunnel diode to the low-voltage state. With two input pulses, a complete switching cycle is obtained.

A combination circuit employing a tunnel diode and a transistor as a multivibrator is shown in Fig. 9·52. Capacitor C_1 charges through resistors R_1 and R_L. As the voltage at point C rises, so does the current flowing through R_2 and the tunnel diode. When this diode current reaches the peak-current value of its characteristic curve, the diode switches to the high-voltage state. This forces the current flowing in

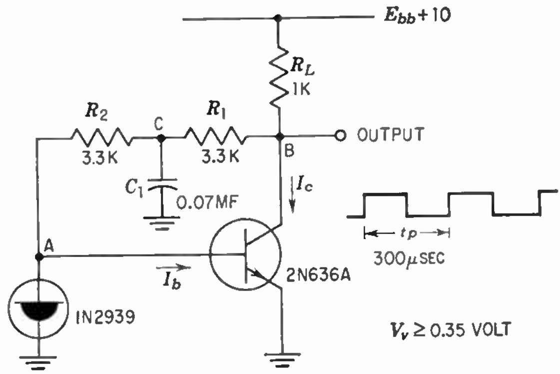


Fig. 9·52 A combination tunnel diode and transistor circuit generating a square-wave output.

the R_2 branch to travel through the base-emitter circuit of the transistor, and the transistor is driven to saturation. In this condition, the collector (and point B) are brought to ground potential. This causes C_1 to discharge until point A drops below the valley voltage of the tunnel diode. When this happens, the diode switches quickly to the low-voltage state, diverting the base current away from the transistor and cutting the latter off. The cycle now repeats itself, with C_1 again slowly building up a charge. Output of the multivibrator is a wave which is reasonably square. With the values of the components shown, the frequency is about 3 kc.

Finally, a bistable circuit using the tunnel diode and an NPN transistor is shown in Fig. 9·53. Initially, the diode is biased to the low-voltage high-current state by the current I_b flowing through R_1. This value of I_b is slightly less than the peak diode current. In this state, the diode presents essentially a short circuit to ground. Practically no current flows into the base of the transistor and the latter is cut off.

If, now, we inject a positive trigger pulse into the circuit, enough additional current will flow through the diode to cause it to switch to the high-voltage low-current state. In this high-impedance condition, I_b is diverted to the transistor, turning it on and bringing its collector voltage close to ground potential.

The diode will remain in the high-voltage state until a negative pulse at the input causes the diode current to fall below the valley-current value. This will switch the diode back to the low-voltage condition, again diverting I_b away from X_1 and turning the latter unit off. This now completes one full cycle. The 47-ohm resistor serves to keep the tunnel diode biased above the valley point when the latter

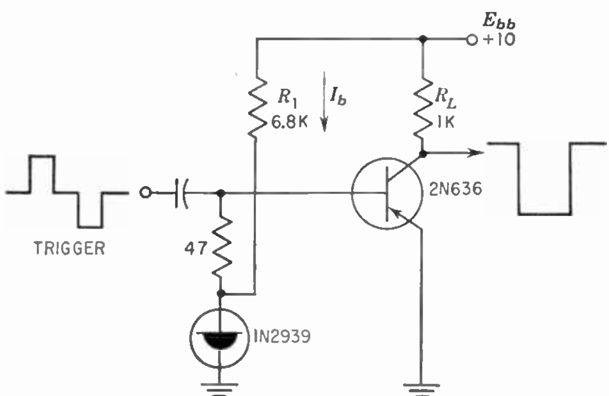


Fig. 9·53 A bistable circuit using a tunnel diode and an NPN transistor.

is in the high-voltage state. It also prevents the tunnel diode from loading the trigger pulse circuit.

The tunnel diode is thus seen to possess a range of application possibilities. Just how extensive its use will be will depend upon its stability (in a circuit) and its relative cost.

Other Semiconductor Materials

Although germanium and silicon have received the greatest amount of research and almost exclusively dominate the commercial transistor field, considerable interest and investigation are being directed to other semiconducting materials. Among the more promising of the newer types are the intermetallic compounds. They differ chemically from such semiconductors as silicon or germanium in that they are formed with two pure elements in place of one. Thus, a germanium transistor starts with pure germanium and then has added to it appropriate impurities to form the requisite P and N regions. The same is true of silicon. In an intermetallic compound, the basic crystal structure consists of *two* different metallic elements such as gallium and phosphorus, which together form gallium phosphide, GaP. Suitable impurities are then added to this compound to form the needed P and N regions.

In the introductory discussion on transistors, it was noted that germanium and silicon each have four valence electrons in their outer or chemically active rings. The crystalline structure is then formed by having the atoms share each other's outer electrons to form bonds. In the intermetallic compounds, one of the combining elements has three valence electrons per atom while the second element has five valence electrons per atom. Equal numbers of the two atoms are used, and these also share each other's valence electrons to form a crystal structure. This structure exhibits many of the same properties as germanium and silicon.

We have seen that a fundamental electrical property of a semiconductor is the energy needed to free an electron from the bond formed between two atoms. In silicon, more energy is required to liberate electrons, and this is a major reason why silicon can be employed at higher temperatures. With intermetallic compounds, by using different combinations of three valance atoms and five valence atoms, we can achieve a very wide range of energy gaps.

The mobility of electrons and holes in semiconductors (i.e., the speed with which these particles or charges move through a crystal) can also be regulated over a fairly extensive range in intermetallic compounds. All these variations make it possible to construct transistors or semiconductor diodes having a wide choice of such properties as current- or power-handling capacity, frequency range, and rectification ratio.

Some of the newer compounds which are being studied extensively include gallium arsenide, gallium phosphide, indium phosphide, indium antimonide, and indium arsenide. Gallium arsenide, as a semiconductor, can potentially combine the high-temperature capacity of silicon (at present, 150°C in practical devices) and the high-frequency capabilities of germanium. Higher-temperature capability will permit higher power applications and levels, since the heat energy can be more effectively handled. Gallium phosphide, used in conjunction with gallium arsenide in a device, is expected to extend the upper temperature limit anticipated from gallium arsenide alone to above 500°C.

Indium phosphide, with an upper temperature limit of approximately 400°C, is considered as a possible runner-up to gallium arsenide. Indium antimonide and indium arsenide are being experimentally tested in galvanomagnetic devices, where frequency and temperature do not play a primary role. Such semiconductors, as opposed to use in transistor or rectifier applications, have potential application in magnetometers, magnetic compasses without moving parts, and gyrators.

Many compounds are under investigation, and undoubtedly some of them will be employed commercially. At the present time, fabrication of these substances presents major problems because of the difficulty in purifying them to the degree necessary and then forming them into suitable transistors or diodes.

Microsystems Electronics

Semiconductor technology has advanced very far and very rapidly since the days of its beginning. The most significant aspects of its advancement may be seen in the areas of improved reliability, reduction in cost, and increased operating frequency. So far as operating frequency is concerned, computers operating at speeds in excess of 100 Mc are being contemplated; compare this with the fastest computer of the late 1950s, which operated at a speed on the order of ½ Mc. This increase in speed permits a tremendous increase in the amount of data that may be processed in a given time, providing a large saving in cost per computation. The importance of reliability is self-evident: a transistor failure could send a missile far from its destination or permit a computer to continue calculating incorrectly long before any error was detected. Although there are safety methods to prevent such errors, such as using two circuits to do the same job and then comparing their outputs before proceeding, guaranteed transistor and component reliability is a much more potent and much more economical solution.

Another area of advancement, now of prime concern but which has not received much attention until recently, is that of size and weight. This may be viewed with respect to component size, circuit size, functional operation size, and, finally, system size. Why are size and weight important? This is an interesting question in light of the small size and weight of the transistor compared to its predecessor, the vacuum tube. A computer, an aircraft, or, more obviously, a satellite has only so much room in which to store the required electronic circuitry. If this circuitry could be reduced in size, then a single satellite could send us information not just on temperature or on radiation, but on both these factors and much more. As another illustration, consider the relative size of a computer using vacuum tubes with a computer using transistors; the reduction in size per operation is greater than 3:1, not counting the reduction in the required air-conditioning equipment.

In addition to considerations of sheer size and weight, there are two other factors which lead to the conclusion that electronic components must become smaller. Consider a large computing system. In present high-speed computers, time delay per stage averages about 10 nsec. If this delay is to be reduced to 0.1 nsec, we must certainly reduce the distance between stages. It takes an electrical signal (traveling at 3×10^8 m per sec) almost 0.1 nsec to travel a distance of only 1 in. Thus, if we envision a kilomegacycle computer (where the total information must be contained in 1 nsec), we will certainly have to build a very small computer. However, a small computer implies a very high component density, where each component must dissipate less power to maintain the same heat-volume relationship.

The reader well may ask why such high computing speeds are desired. This brings us to the second factor: the relationship between the size of the memory and the logic section. For computers to have a higher information capacity, they must be able to store more information. We must be able to extract this information quickly, operate on it, and return it to storage. Thus, in general, a large amount of stored information implies a high-speed logic section. This high-speed requirement, coupled with the increased size of a larger computer, again implies a need for small package size and low power consumption. Considerable miniaturization has already taken place in conventional components such as resistors, capacitors, inductors, tubes, and transistors. To go beyond this point requires either new techinques in component miniaturization or entirely new concepts in circuitry. Both of these avenues have been explored with varying degrees of success.

If circuits are to be reduced in size, crammed into smaller volumes, and finally packed more densely to do more jobs in less space, there arises the immediate problem of heat. Heat, besides being an evidence of energy lost and unreclaimable, is probably the biggest contributor to semiconductor and component failure. Thus, any attempt to reduce circuit or system size must be accompanied by a reduction in operating power dissipation. This also has a desirable side effect: power supplies and other energy sources will need to provide less energy. This will enable them to be made smaller with a higher probability of long life.

Another obstacle to higher component density is solder joints and connections. One of the most important causes of poor reliability in computers is the failure of interconnections. Each transistor must have three external connections; each resistor, capacitor, coil, or wire needs two external connections. Consider now a large computer with 50,000 transistors, 50,000 capacitors, and 150,000 resistors. In such a computer, there are probably more than 500,000 component connections. If we assume a connection failure rate of only 0.002 per cent per thousand hours, we will have approximately one failure per week. Thus, even without component failures (0.01 per cent per thousand hours for transistors), the computer would still cease functioning once every week. It would be extremely useful to eliminate all external connections. It is a known fact that internal connections, such as those inside a transistor case, have a very low failure rate—much lower than that of external connections. This is yet another reason for pursuing microelectronics and packaged circuitry.

Research into the problem of microelectronics has produced a number of solutions. These various methods of producing microelectronic components may be grouped rather loosely by size into the following categories:

1. High-density packaging techniques—a circuit-reduction concept utilizing conventional miniaturized components
2. Micropackaging—a functional-operation-reduction concept whereby an entire operational circuit is packaged by using discrete microsize components or thin-film techniques
3. Solid-state or integrated circuits—a functional approach whereby a complete circuit is entirely formed, using semiconductor techniques, on a semiconductor slab and then packaged

It will prove worthwhile to study these three approaches in detail.

High-density packaging techniques. The number of components which can be fitted into a given volume, say a cubic foot, can be substantial by suitably reducing the physical dimensions of conventional components. (By conventional components, we refer here to the resistors, capacitors, inductors, tubes, and transistors that would be found in a pocket radio, for example. These retain the same basic appearance that we have for some time associated with these components.) Estimates as high as 1,000,000 parts per cubic foot have been made, and there are some engineers who have indicated even higher figures.

To appreciate some of the obstacles which must be overcome when a component is made more compact, consider the transformer for an example. A transformer generates a certain amount of heat as a result of the current flowing through its windings. This heat must be dissipated, but as we work toward miniaturization, we reduce the available surface radiating area. Furthermore, to attain a smaller volume in the first place, smaller-size wire must be employed in the windings, and this, in turn, possesses a higher resistance so that more heat is generated by the current flowing through the wire.

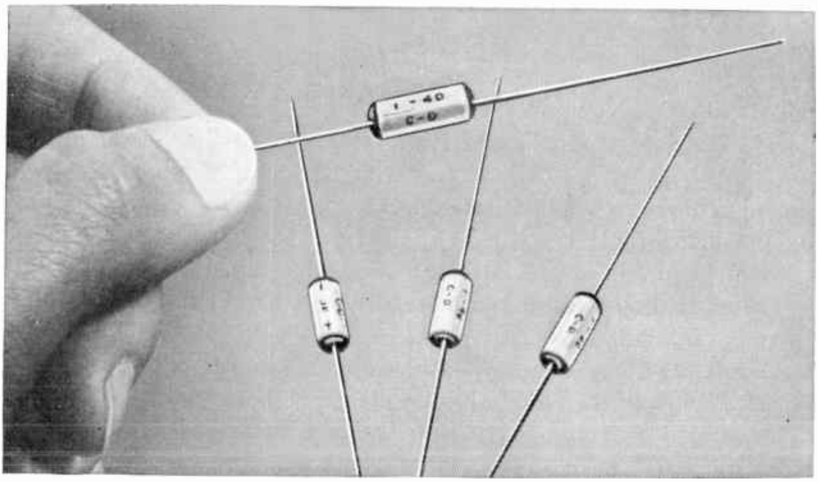
Both of the foregoing factors force the transformer to operate at a higher temperature, and if suitable insulating materials are not developed, component failure will be high.

Temperature also enters the picture in still another way. Overall reduction of equipment, a direct consequence of component miniaturization, means that the amount of heat generated per unit volume will be higher than in conventionally sized equipment, where the spacing between components can be made much greater. This leads to a higher ambient temperature, and this factor, added to the increased heat generated within each component itself, further aggravates the demands made upon the materials used.

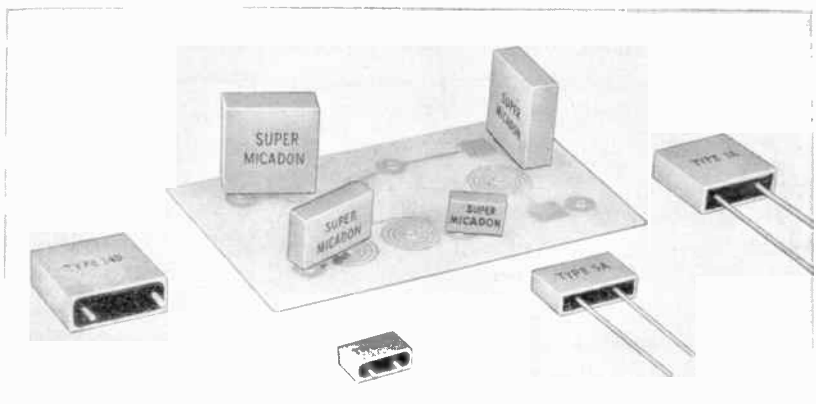
The solution resides in the development of new substances possessing greater heat-resistant properties than heretofore possible. In transformers, silicon-impregnated Fiberglas, Mylar, and adhesive Teflon tapes are used extensively to provide improved insulation between the various layers and windings. The shapes of cores used fall generally into the toroid or flat-laminated categories. Tape-wound toroids are a most convenient way of obtaining a miniature core using thinner material; air gaps are minimized with this construction, permitting utilization of almost the full permeability of the material. When flat laminations are employed, they generally possess an E or I configuration.

Cores are fabricated from such high-permeability materials as Ferroxcube, grain-oriented Hipersil steel, Moly-Permalloy powder, or grain-oriented Silectron steel. Formvar-type wire coatings, bobbin windings (in contrast to the layer windings common in larger transformers), and improved production techniques, especially in the handling of the very fine wire used, all have combined to permit the evolution of transformers which are truly miniature in size, Fig. 4·9.

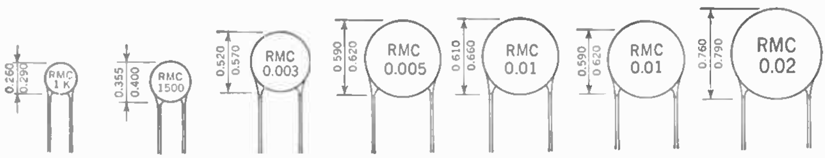
In another component, fixed capacitors, Fig. 9·54, the problem of miniaturization is being overcome in a number of ways. For nonelectrolytic capacitors, generally those less than 1 μf in value, extensive use is being made of plastic, impregnated paper, and ceramic dielectrics. Desirable characteristics in a dielectric are high insulation resistance, relatively low temperature coefficient capacitance (i.e., small change in capacitance with temperature), the ability to function satisfactorily over a wide range of temperatures, and a high dielectric



(a)



(b)



(c)

Fig. 9·54 Several illustrations of miniature capacitors employed in transistor circuits. (a) Tantalum-foil electrolytic capacitors. (b) Mica capacitors. (c) Ceramic capacitors.

constant. (Few substances possess all these characteristics, and the end product is usually a compromise.)

In plastic dielectrics, wide use is being made of Mylar, a product of E. I. du Pont de Nemours & Company. This substance, which is chemically a polyester, possesses a high resistivity to heat and excellent mechanical strength. In thin films it is frequently impregnated with silicon, mineral oil, or polystyrene. Dielectric constant of Mylar is about 3, and it remains fairly constant with frequency.

When paper dielectrics are employed, the paper is frequently impregnated with vegetable or mineral oils or with a synthetic compound such as Permafil (General Electric Company). Metallized paper capacitors are also rather widely used. In these units, the conventional separate layers of metallic foil are replaced with an extremely thin film of metal which is deposited directly on a lacquered surface of the paper dielectric by a high-vacuum vaporizing process. This lacquer coating considerably improves the dielectric strength and insulation resistance of the paper.

Within recent years, a wide variety of ceramic dielectrics such as the Erie Ceramicons have appeared. Because of their high dielectric constant, these dielectrics permit the fabrication of fairly high valued capacitors in small volumes. One disadvantage of some of the very high dielectric constant substances is their sensitivity to temperature. This means that the capacitance value can change markedly as the equipment warms up. However, in many applications, such as bypassing and coupling, considerable variation is tolerable. In fully transistorized equipment, where there are no vacuum tubes to generate large quantities of heat, this temperature dependence does not present any undue difficulties.

The trend toward miniature fixed capacitors of the type just described was established before transistors became commercially available. However, the story of miniature electrolytic capacitors is entirely different. For transistor operation, bypass capacitor values in the microfarad region are required because of the low impedances inherent in such circuits. Coupling capacitors of 2 to 10 μf are not uncommon, as we noted in earlier chapters, and for bypass functions, capacitance to 50 μf is employed. Fortunately, the voltages used in transistor circuits are extraordinarily low, and this does help to simplify the problem.

One dielectric which is being used widely in the fabrication of miniature electrolytic capacitors is tantalum. The tantalum anode is in the form of a wire which is completely surrounded by a porous, absorbent sleeve, effectively insulating this section from the base. A neutral electrolyte is used whose properties are so chosen that these capacitors show excellent capacitance stability and power characteristics over the temperature range -20 to $+55°C$.

The tantalum anode and the electrolyte are hermetically sealed into a solid-silver tubular case which serves as the cathode. Terminal leads are of solid tinned copper. Dimensions of a typical unit designed to possess a capacitance of 8 μf at 4 volts d-c are $\frac{1}{8}$ in. in diameter by $\frac{1}{2}$ in. long.

One of the most widely employed methods of mounting miniature components is the *module*. Components are clustered together and held in place by soldering or welding the component leads to printed-circuit boards or to supporting wires that extend from one end of the module to the other (see Fig. 9·55). Once the proper connections are made

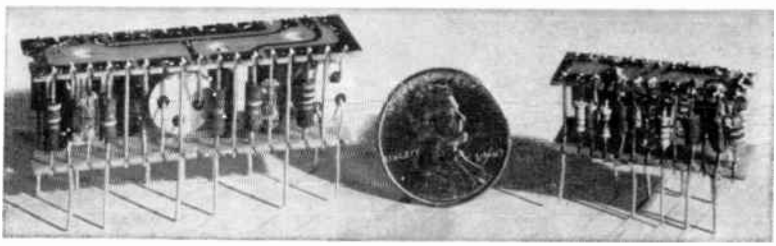


Fig. 9·55 Miniature modules showing how components are clustered together.

and the necessary input and output terminals are established, the body of the module may be encapsulated in a suitable resin. This not only provides additional physical support for the components but also protects the module from the atmosphere. The proper resin can also aid materially in removing heat from the interior of the module.

One new approach—new in that it does not represent a use of conventional miniaturized components—may be found in the "dot packaging" concept, Fig. 9·56. By this method, specially packaged components, in pill shape, are fitted into a hole structure on a nonconducting wafer. The pills generally have a 0.05-in. diameter and are 0.03 in. long. The flat faces of the "pills" are the connecting points; series or parallel connections are easily obtained by strapping the pills together with tiny conducting strips. Capacitors, resistors, and diodes are available in this package. The wafer is generally 1.1 in. square and 0.05 in. thick; this additional thickness represents a thin nonconducting film that covers both sides of the wafer and thus insulates the exposed pill faces. Small can-package transistors may also be sealed in the wafer, although a larger hole is required. All leads to the circuit are brought out one end of the wafer; a close-packed stack of many such wafers is then easily made and all connecting leads may terminate in a single connecting circuit board.

Micropackaging. The concept of micropackaging first began with the idea of combining two or more transistors or diodes within a single transistor case and providing the proper number of external leads. This is illustrated in Fig. 9·57. Many configurations are possible: common-cathode and common-anode diode arrays, common-element transistor

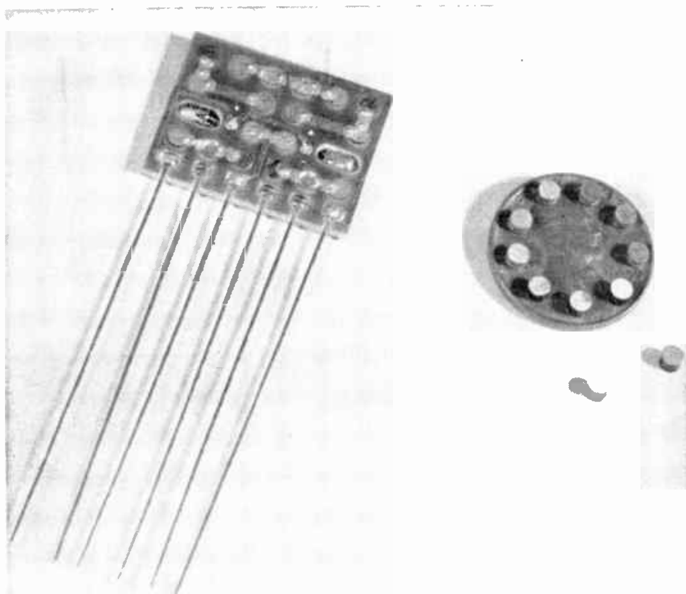


Fig. 9·56 Miniaturized dot components. (P. R. Mallory Co.)

configuration, and direct-coupled transistor circuits, as well as numerous transistor-diode combinations.

Such packaging brings many previously external connections inside the transistor housing. This is a step toward improving overall reliability. This packaging concept has now progressed to the point where an entire functional circuit is contained inside a transistor can. In this case, the circuit components are no longer the standard resistors and capacitors familiar to us; rather, they are formed by film or deposited-metal techniques. One typical example of a functional circuit packaged in this manner is shown in Fig. 9·58. This is an RC-coupled transistor flip-flop. In addition to four transistors, it also possesses four

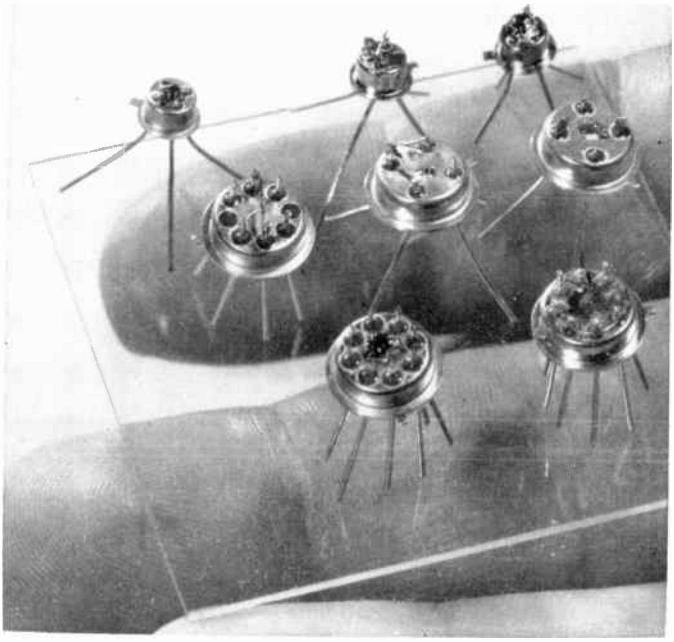


Fig. 9·57 Transistors and diodes packaged in transistor housings. (Philco Corp.)

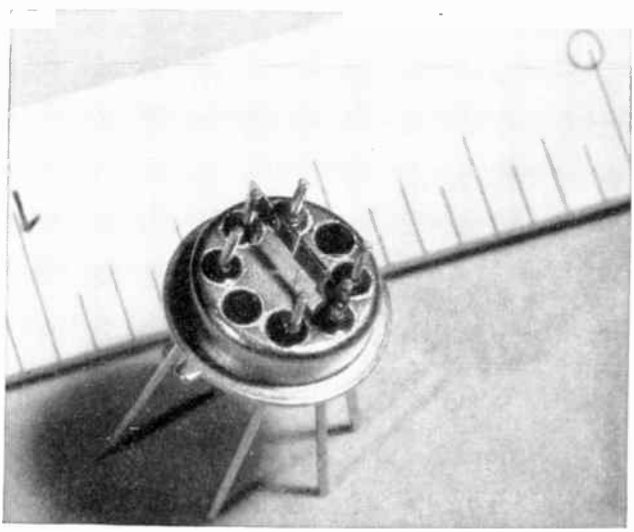


Fig. 9·58 A transistor case containing a complete flip-flop. (Philco Corp.)

resistors and two capacitors. Many such circuit combinations are commercially available, although for the most part they are circuits of prime use in computers. Because of the larger number of circuits which a computer employs, and because of the repetition of certain basic arrangements, computers lend themselves particularly well to this approach.

Thin-film techniques for fabricating components, notably resistors and capacitors, are becoming increasingly important in the manufacture of micro circuits. Present available methods for creating thin films with specific electronic characteristics are:

1. Physical methods such as sputtering and vacuum deposition
2. Chemical methods such as electroplating, chemical reduction (etching), and thermal decomposition
3. Mechanical methods such as spraying and rolling
4. Photographic methods such as photoengraving

Generally, more than one of these approaches are employed in any one process. The thin film is deposited on a nonconducting substrate

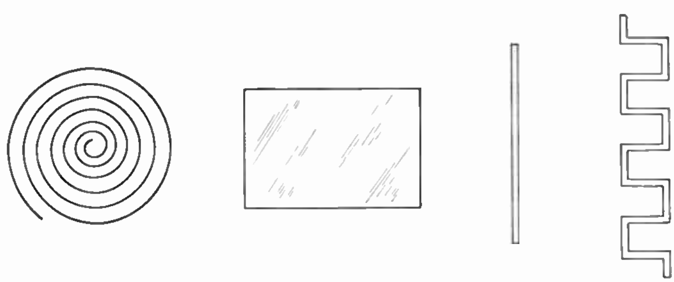


Fig. 9·59 Various geometrical shapes suitable for deposited resistors.

which is generally glass or ceramic. It is important that the surface of the substrate be extremely smooth so that the thickness of the deposited film is uniform. Resistances are obtained by simply depositing a long enough strip of a conductive (i.e., metallic) material. The geometrical shape may be a spiral; a square area; a straight, narrow strip; or some irregular pattern, Fig. 9·59. To obtain a high resistance, the thin film should be narrow, thin, and long. To obtain a low resistance, a wide, thick film pattern should be employed. A typical range of resistance is from 100 ohms to 1 megohm.

If the resistor is to dissipate much power, it must be made with a large area. Power dissipations as high as 100 mw are possible. Resistor tolerances of ± 5 per cent are common. If extreme care is taken during manufacture, tolerances of ± 1 per cent may be obtained. The amount of resistance variation with life and temperature is extremely small because the thin film is actually a metal.

Capacitors are obtained by depositing films on two sides of a substrate with a controlled thickness. Very small values may be obtained by paralleling two deposited strips on the same side of the substrate. Another method of making the capacitor is to first deposit the metallic thin film, then cause it to become oxidized, and then plate another metal film over the oxide. The metal films then act as the plates and the oxide serves as the dielectric. Thus a parallel-plate capacitor is formed. The value of the capacitor is then determined by the area of the metal film and the thickness of the oxide layer. Microminiature capacitors with values from 30 $\mu\mu$f to 0.05 μf can be readily formed in this fashion.

The same metal film employed to provide capacitors and resistors can also serve as a conductor to interconnect these components as well as the semiconductor diodes and transistors in the circuit. This single-metal system offers many advantages in the manufacturing process. Philco, for example, uses tantalum in microfilm circuits. Tantalum's high sheet resistivity, low temperature coefficient, stability, corrosion resistance, amenability to being formed into very fine lines, and reproducibility make it desirable for resistors. Its oxide provides the dielectric for capacitors. Finally, when overlaid with gold in selected areas, the resultant film forms the basic high-conductivity wiring pattern.

One of the advantages of the thin-film technique is that distributed CR networks may easily be obtained by depositing resistances on both sides of the substrate; this may, of course, also present a problem if distributed parameters are not desired. Another problem present in the thin-film approach is the limited power dissipation of the components. This fact may be used to point out that reduced power dissipation is not only desirable from the reliability standpoint but also a definite consequence of microminiaturization techniques and approaches. Finally, two problems are involved in thin-film fabrication at present: accuracy in depositing the films and the amount of time involved in the actual deposition.

Solid-state circuits. Solid-state circuits are another step in the evolution of microsystems electronics. In this case, the entire circuit is fabricated by using semiconductor techniques. The circuit is formed from a slab of semiconductor material, usually silicon. From this slab, resistances, capacitances, diodes, and transistors are formed. To illustrate the procedure, consider a block of P-type silicon possessing a fairly high resistivity (~ 200 ohms per cm). Figure 9·60*a* shows such a P-type block. If we now apply ohmic or nonrectifying contacts to the ends of the block, a resistor is formed. In general, the range of obtainable resistance values (200 to 100,000 ohms) is smaller and the tolerance not quite as good as that obtainable by using thin-film techniques. Furthermore, the resistance value will be found to change more with temperature and with age.

A capacitor can be formed by depositing metal films on either side of a diffused junction, Fig. 9·60*b*. By applying a reverse voltage across

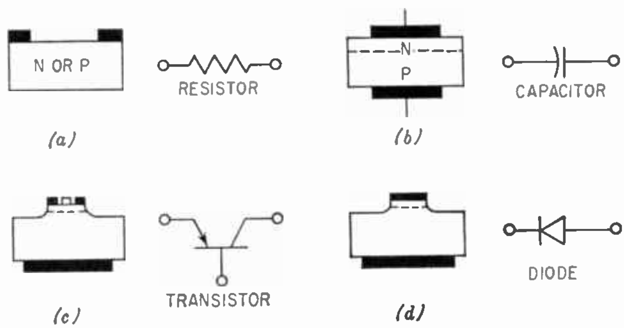


Fig. 9·60 Illustrations of how solid circuits can be employed to form a variety of components

the junction, capacitances of 20,000 to 200,000 $\mu\mu$f per cm^2 are achievable.

One important problem which arises when making a capacitor from a reverse-biased junction diode is that the leakage currents are high. Also, the capacitor can only be used for a signal of one polarity. Another problem with this type of capacitor is the resistance which the semiconductor itself possesses and which it essentially places in series with the capacitor. Typical values of series resistance range from 1 ohm to 100 ohms.

The same wafer or block of silicon can also serve as a transistor or a diode, Fig. 9·60*c* and *d*, with suitable diffusion of impurity materials. This technique has been fully covered in preceding chapters.

An example of a solid-state circuit, a diode-coupled transistor logic gate, is illustrated in Fig. 9·61. This circuit includes one transistor, two resistors, one capacitor, and four diodes. The transistor and diodes are formed in the slab, while the resistances are obtained by using the material of the slab. Such a circuit is easily housed in an eight-lead transistor package.

There are really few restrictions to the solid circuit except those imposed by size and the number of elements, both active and passive, that can be formed on a single slab of semiconductor material. For example, a large silicon wafer may be diffused to form a large planar diode. This wafer may then be photo-diffusion-processed further to obtain the desired number of diodes and their desired locations. Code-conversion matrices that will contain upwards of 1,000 diodes are contemplated.

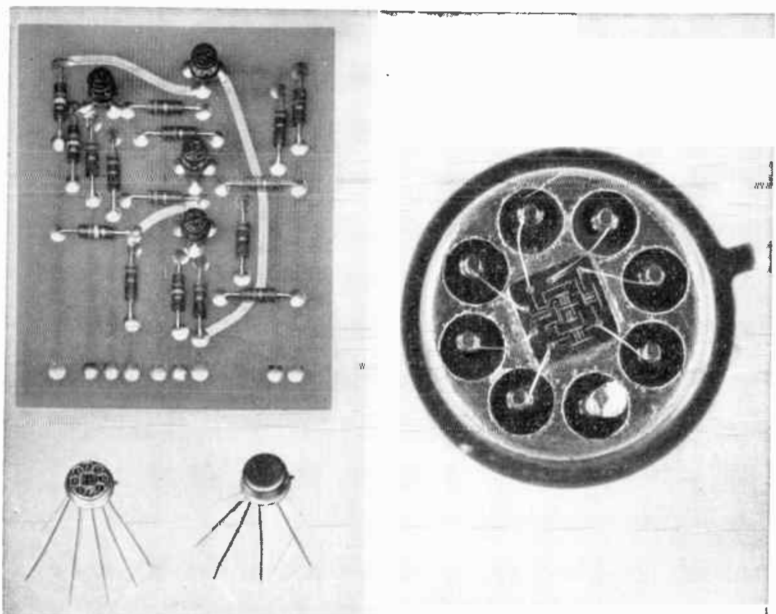


Fig. 9·61 A digital-computer logic circuit housed in a transistor package. The unit on the right performs the same logic function as the printed-circuit unit on the left.

These are but a few of the approaches to microsystem electronics. There are many problems yet to be solved, and probably a great number of solutions will be found. One of the most pressing problems lies in micropackaged-circuit performance; generally, performance capabilities, such as speed and number of possible loads, are related to power dissipation. Thus, at the present, speed and gain are necessarily sacrificed for a reduction in size and in power dissipation. A few transistors have been especially designed for microenergy use; such devices, tailored to give optimum performance at low power levels and incorporated in micropackages, may well be one immediate solution to this problem. The future will shortly find manufacturers of large electronic systems, such as computers, purchasing not individual transistors and components but entire functional packages, such as logic circuits. These will then be interconnected to form the computer.

QUESTIONS

9·1 Why are phototransistors frequently more desirable than phototubes? How does the phototransistor of Fig. 9·1 operate?

9·2 Explain the reason for the shape of the curve in Fig. 9·2.

9·3 Why does the response of a phototransistor depend upon the wavelength of the incident light?

9·4 How does the phototransistor in Fig. 9·5 function? Show one application.

9·5 What is a tetrode transistor? Show the structure of such a device and indicate the polarity of the biasing voltage applied to each element.

9·6 How can the tetrode transistor be employed as an amplifier?

9·7 Draw the circuit diagram of a transistor frequency converter using a tetrode transistor.

9·8 In what way does the coaxially packaged transistor permit higher frequency operation of transistors?

9·9 Describe the operation of a PNPN transistor.

9·10 Why can the PNPN transistor be employed as a switch? Explain.

9·11 Describe two methods for turning off controlled rectifiers.

9·12 What is a unijunction transistor? How does it operate?

9·13 Explain how a unijunction transistor can be employed to generate square waves.

9·14 Compare the field-effect transistor with a conventional PNP (or NPN) transistor.

9·15 Compare the field-effect transistor with a triode vacuum tube.

9·16 Why is carrier transit time not so important in the operation of a field-effect transistor as it is in a conventional PNP transistor?

9·17 Explain why the space charge in a field-effect transistor possesses the shape shown in Fig. 9·32.

9·18 Where is the signal applied in a field-effect transistor? How is an amplified version of this signal obtained? Draw a suitable circuit to illustrate your answer.

9·19 On what one feature does the operation of a tunnel diode depend primarily? Name three other electronic devices which depend upon the same characteristic for their operation.

9·20 Describe the operation of a tunnel diode.

9·21 Explain the following tunnel-diode parameters: V_V, I_V, V_f, I_P.

9·22 What is the principal frequency-limiting factor of a tunnel diode? How does this limit the upper frequency attainable by a tunnel diode?

9·23 Explain how the circuit of Fig. 9·44 functions.

9·24 Show how a tunnel diode can be employed to operate as a flip-flop circuit.

9·25 What is the difference between a tunnel diode used as an oscillator and a tunnel diode used as a bistable switch? *Hint:* Use a characteristic curve in your explanation.

9·26 What advantages do miniature circuits offer? What difficulties?

9·27 What obstacles are encountered when components are miniaturized? Name some of the ways these obstacles are being overcome.

9·28 What are thin-film components? How are capacitors, resistors, and conductors produced by thin-film techniques?

9·29 Differentiate between thin-film circuits and solid-state circuits.

9·30 How are resistors, capacitors, and conductors produced in solid-state circuits?

CHAPTER 10

Servicing Transistor Circuits

THE TRANSISTOR exhibits a curious combination of ruggedness and fragility. It is, for example, far more physically rugged than even the most powerfully built vacuum tube; it is capable of withstanding centrifugal forces with accelerations as high as 31,000 times the force of gravity and impact tests as great as 1,900 times. These are far in excess of the forces which will completely shatter any vacuum tube. On the other hand, a transistor is a fragile device with respect to heat or to the application of d-c biasing voltages possessing the wrong polarity. It is important, then, to be familiar with the physical handling limitations of transistors, so that transistor equipment can be built or serviced with a minimum adverse effect either on the transistors themselves or on the miniature components with which they are often employed.

Tools

Probably the first step to take in preparing yourself for transistor work is the acquisition of the proper tools. Since transistors and their associated components are extremely small, conventional-sized tools are frequently unsuitable for effective use. In their place, the technician requires tools which, because of their own reduced size, are better able to cope with the limited space encountered in compact, miniaturized equipment. In addition to the smallest cutting pliers that can be obtained, it is suggested that two or three shapes and sizes of tweezers be acquired. These will come in handy when fine wires must be soldered (or unsoldered) in the circuit. Another useful device is a soldering aid one end of which has a notch for gripping wires while the other end comes to a fine point for probing or cleaning away solder from small openings.

Servicemen have also found a large reading or magnifying glass to be useful, particularly one which is mounted on a holding stand so that both hands are left free. Other tools which should be available include needle-nose pliers and small- and long-shank screwdrivers having narrow blades.

Another change required by the transistor is the use of a small, low-wattage soldering iron (or pencil) possessing a narrow point or wedge. Wattage ratings on the order of 35 to 40 watts are satisfactory; anything larger than this could damage the transistor while it is being soldered into the circuit. (The same low-wattage iron is also required for the printed-circuit wiring of a transistor receiver.)

To provide the transistor with the maximum protection while it is being soldered or unsoldered, it is good practice to grasp the terminal

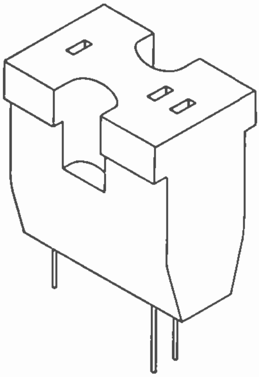


Fig. 10·1 A representative transistor socket.

lead tightly with long-nose pliers positioned between the transistor body and the lead end. With this arrangement, any heat traveling along the wire will be shunted away from the transistor housing. It is desirable to retain the pliers on the wire for a short time after the iron has been removed to make certain that all the heat has been dissipated. It is also good practice to provide such a heat shield when other wires are being soldered to any terminal lugs to which a transistor lead is attached.

Two helpful rules to follow are to keep the transistor leads as long as possible, consistent with the space available and the application, and to get whatever soldering that has to be done over with as quickly as possible. Helpful in this respect is 60/40 low-temperature rosin-core solder.

In some instances, transistors are constructed with leads stiff enough to permit plugging the transistor into a specially constructed socket, Fig. 10·1. In such cases, of course, soldering is no problem, and the only precaution to observe is to remove the transistor from the socket before the soldering iron is brought into contact with any of the socket terminal lugs.

As a final word concerning the use of any tools on transistors and their associated miniaturized components, always remember that because these units are small, their connecting wires are quite fragile. Handle these wires carefully and gently, both when the part is being installed and when it is being removed.

Battery Potentials

Two factors combine to make transistors particularly sensitive to applied bias voltages. First, there is the fact that the emitter-base junction is biased in the forward, or low-resistance, direction and the impedance of this circuit, under these conditions, is extremely low. Any voltage in excess of the required value could easily result in so large a current that the resultant heat would permanently damage the transistor. The correct operation of a transistor—any transistor—is intimately tied in with the maintenance of its crystal lattice structure and the distribution of certain impurity atoms throughout that structure. If enough heat is generated to disrupt the crystal structure, the effectiveness of the transistor to function as desired is seriously undermined. This is the reason for the oft-repeated warnings against applying too much heat or permitting the unit to become too warm during operation.

The second factor that makes transistors sensitive to applied bias voltages is the extremely minute dimension of the several elements and their very limited heat-dissipating ability. Collector current is important in this respect because this current, passing through the relatively high collector resistance, develops a certain amount of heat. If this heat, added to the ambient heat at which the transistor is operating, exceeds the maximum limits of the transistor, behavior becomes erratic. That is why the maximum collector dissipation is always specified at a definite ambient temperature. If the surrounding temperature is higher than specified, the collector-dissipation rating must be reduced proportionately. This is called derating and was discussed in Chap. 3. The maximum safe value of collector voltage is important also, since too high a value will lead to a reverse voltage breakdown.

Thus, because of the foregoing limitations, the value and the polarity of any voltages applied to the circuit must be scrutinized carefully. Make certain first that you have the right voltage, then check polarity before making final connection to the circuit. If you are at all in doubt about the latter point, check the type of transistors being employed.

PNP units require negative collector voltages and positive emitter voltages, both taken with respect to the base. In NPN transistors, the reverse situation holds.

Before the battery is connected to the circuit, the various transistors should be firmly in place. Never insert or remove a transistor when voltages are present. This is designed to prevent the development of surge currents which can damage a transistor. Always remove the voltage first. If you are experimenting with a new circuit or building a transistor kit, double-check all wiring before applying bias voltages. If you are doubtful about the outcome, insert a current meter in series with the collector circuit and then use a potentiometer arrangement to apply the collector voltage gradually. If the collector current begins to exceed the specified maximum, you know something is at fault.

To men who have gained all their radio and television experience on vacuum-tube circuits, all these precautions may appear somewhat excessive. However, experience has revealed that they are most certainly required. Transistors are extremely sensitive to heat, and anything that develops heat, such as current flow, must be watched with a wary eye.

Another source of potential danger lies in the signal generators which the technician uses to service radio and television systems. When a signal is injected into a transistor circuit, start with a very low amplitude signal and gradually increase the generator output until the desired indication is obtained. Never inject strong signals into a transistor circuit, particularly when it is a low-level stage. Frequently, indirect-rather than direct-coupling methods of signal injection are advisable. For example, clip the "hot" output lead from the generator across the insulated body of a nearby resistor or capacitor. The signal will then enter the circuit by radiation and capacitive coupling. This approach is widely practiced in television-receiver alignment when a marker signal must be brought into the system without swamping the sweeping signal.

It has also been suggested that signal injection can be achieved by connecting the output of the signal generator to a suitable coil and then aligning the axis of the coil with that of the input of the circuit under test. This will bring the signal into the circuit by inductive coupling. In using this method, the radiating coil should be geared to the signal frequency, i.e., a high-inductance coil for low frequencies and a low-inductance coil for high frequencies.

The sensitivity of a transistor to surge currents should be borne in mind when a voltmeter is being used to check voltages at various points in a transistor receiver. Because of the closeness with which components are placed, it is easy for the probe tip to touch two closely spaced terminals accidentally if the technician is not exceptionally careful. This simple slip may result in battery burnout or be responsible for a current surge through the transistor as, for example, when the probe makes simultaneous contact with the collector and base electrodes. Special emphasis is placed on this precaution because of the ease with which the mistake can be made. In vacuum-tube circuits, similar slips may occasionally cause a component to burn out, although they rarely affect tubes. In a transistor circuit, the transistor is usually the weakest link, and it becomes the victim.

Along these lines, here are some meter precautions which are issued by the manufacturer of the Regency pocket radio receiver discussed in Chap. 6. Some service ohmmeters utilize circuitry which necessitates an other-than-normal battery polarity inside the meter. With instruments of this type, the red test lead has a negative potential and the black lead has a positive potential. The technician should investigate his meter to determine the polarity of its leads. This can be done easily by connecting a voltmeter across the ohmmeter test prods. When measuring circuits which are critical with regard to polarity (such as those containing electrolytic capacitors), the technician should keep in mind the polarity of the meter leads and should connect them accordingly. The positive lead, whether it is red or black, should be connected to the positive lead of the electrolytic capacitor. The transistors in this receiver (i.e., the Regency model TR-1G) would not be ruined if an ohmmeter were to be connected into the circuit in the reversed polarity, but the electrolytic capacitors would give incorrect readings because they would be measured backwards. It is also imperative not to use an ohms range which utilizes a battery of more than 3 volts, because the transistors can be damaged if too much voltage is applied to them.

Before we leave the subject of heat and its effect on transistors, one word might be said about the precautions to observe when positioning transistors in electronic equipment. Keep transistors clear of any component, be it tube, resistor, or transformer, which passes enough current to develop a noticeable amount of heat. The ratings specified for a transistor are always given at a certain ambient temperature, generally 25°C. For every degree above this figure, a corresponding lowering of the transistor ratings must be made, thereby effectively reducing the operating range of the unit. It might be useful to remember this when you find that transistor equipment is not operating as it should and no component is apparently at fault. Measure the ambient temperature of the enclosure where the transistor is contained.

Make this measurement under the same conditions that prevail when the equipment is functioning normally, that is, with the chassis in the cabinet and all removable sections or panels in place. If you possess a Fahrenheit thermometer, the equation for conversion to centigrade is

$$C = \frac{5}{9}(F - 32)$$

where C = temperature, °C
F = temperature, °F

Transistor Testing

The transistor may be checked, in roundabout fashion, by substituting another unit known to be good. If the circuit operation returns to normal, the previous unit was defective; if the trouble persists, the transistor was not at fault.

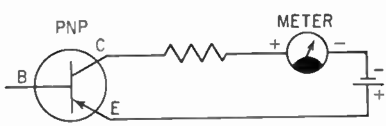
A more direct approach is to test the suspected transistor in a suitable tester. A number of such units are commercially available. Many of these testers are fairly simple in design, checking only the leakage current and current gain of a transistor. If the results of both checks are satisfactory, the unit is probably functioning normally.

In most testers, leakage is checked first because it is the more sensitive indicator of the condition of a transistor. Leakage will almost always drag the gain down with it, and the transistor that possesses more than the normal amount of leakage will have a greater distorting effect on the circuit than one with low leakage and gain. If a transistor passes the leakage test, it should then be checked for gain.

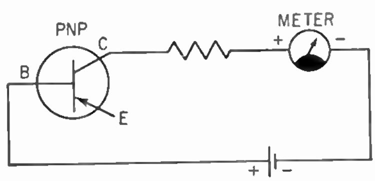
Leakage measurement. In nearly all commercial transistor testers, leakage tests are performed in one of two ways. The most widely employed method is the I_{CEO} test in Fig. 10·2*a*, in which a meter, resistor, and battery are connected in series. If a PNP transistor is being checked, the positive terminal of the battery is connected to the emitter and the negative terminal is connected to the end of the circuit terminating at the collector. The current through the circuit is a function of temperature, the resistivity of the germanium or silicon, and the applied voltage. Any contamination on the surface of the transistor or a short circuit within the device will produce a high reading on the meter.

The I_{CEO} notation of this leakage test indicates that current is flowing between the collector (C) and emitter (E) with the third element or base open (O). Thus, the first two letters after the I indicate the circuit in which the current is being measured. The disposition of the third element is then identified by the third letter, *O*, here standing for open circuit.

The foregoing test provides enough current to actuate a milliammeter, and this is the chief reason for its popularity. Perhaps a more sensitive indicator of the leakage condition of a transistor is the I_{CBO} current (frequently shortened to I_{CO}). As the notation indicates, current is measured between the collector and base with the emitter circuit open. This circuit is shown in Fig. 10·2*b*. The battery is connected between the collector and base, the negative battery terminal going to the collector of a PNP transistor and the positive terminal to the base. In essence, the transistor is reverse-biased, and only a very small current should flow (perhaps no more than 10 or 15 μa). Since



(*a*) HOOKUP FOR I_{CEO} TESTS



(*b*) HOOKUP FOR I_{CBO} TESTS

Fig. 10·2 Two leakage-current tests that are performed on transistors.

the transistor is reverse-biased, the reader may wonder why any current at all flows. The reason is the energy which the ambient, or surrounding, temperature supplies to the internal atoms of the transistor structure. This energy causes some of the atomic electrons to vibrate strongly enough to enable them to break away from their parent nucleus and leave behind an equivalent positive charge called a hole. The positive charges are attracted to the negative terminal of the battery, while the electrons are drawn to the positive terminal. It is the flow of these two opposite charges that produces a current flow. If a transistor is in good operating condition, however, this current should be very low.

Just what constitutes a very low reverse current for a transistor depends on its power rating. In milliwatt transistors such as are found in most signal stages except the output, 10 to 15 μa or less would be considered an acceptable I_{CO}. In a 5- or 10-watt power transistor, as much as 1 ma or so would be good. As a general rule, the larger the power capabilities of a transistor, the greater its acceptable I_{CO}.

Since the test for I_{CBO} current requires a highly sensitive meter movement, most low-cost transistor testers avoid this test and measure instead the larger I_{CEO} current.

Gain measurement. Once it has been established that the transistor leakage falls within a normally acceptable limit, the d-c β gain of the transistor is measured next. This gain is a measurement of the change in collector current for a small change in base current. A small current is introduced into the base circuit and its effect on the collector circuit is noted. There are several ways to do this, and they will be brought out as we examine the various commercial testers.

The most desirable condition for a low-frequency transistor is low leakage and high β. As the leakage current rises and the β value falls, the transistor becomes increasingly less desirable until its usefulness in the circuit is negligible. (Keep this in mind when checking any transistor.) If several similar transistors are available, it is possible, by the foregoing tests, to select the one which gives the best indication and to grade the others in terms of their indications.

For high-frequency transistors (say those above 10 Mc), low leakage is still desirable. The β value measured by most checkers, however, is not too indicative of the high-frequency behavior of the transistor. A β reading of 5 or more is still desired to ensure that the unit will amplify, but it gives no direct clue of what the transistor will do in the circuit. The latter operating condition can be determined only by an actual test at the frequency in question, and such a facility is not generally available to the technician.

Remember that transistors have d-c and a-c β values which, in high-frequency units, do not necessarily have any direct and simple relationship to each other. In low-frequency transistors, on the other hand, it is more likely that such dependence will exist. In general, high-frequency transistors have lower d-c and a-c β than low-frequency ones have.

Commercial Transistor Testers

General Electric transistor tester. This tester is shown in Fig. 10·3. The setup is for a PNP transistor. This unit performs the I_{CEO} leakage test, in which the base is left floating while a voltage is applied between the collector and emitter. A 680-ohm resistor is connected together with a 0- to 3-ma meter. When the transistor is inserted in the socket, the meter will indicate the leakage current flow. The meter scale is divided into several colored areas, and as long as the needle rests at some point other than "bad," the transistor leakage can be assumed to be within normal limits.

After this has been established, the gain switch is depressed. It connects a 200,000-ohm resistor to the base element, through which a current of 30 μa flows. With this current in the base circuit, a larger current should flow in the collector circuit. This larger current will be indicated by the meter. According to the instructions for the instru-

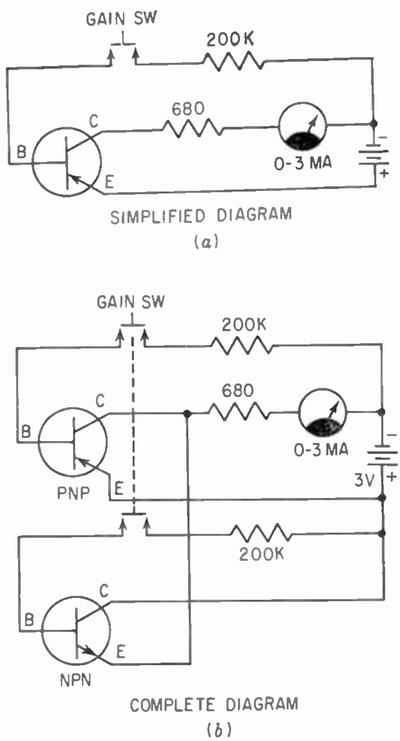


Fig. 10·3 Simplified and complete diagrams of a General Electric transistor tester.

ment, an increase of at least one division on the scale represents an acceptable current gain.

Note that this instrument does not directly give the value of β for the transistor under test. Rather, a relative reading is taken. So long as the needle moves a sufficient number of divisions for the gain test and the leakage does not exceed a certain value, the transistor is presumed to be satisfactory.

A complete diagram of the General Electric tester is shown in Fig. 10·3*b*. The connections for the NPN socket complement those of the PNP socket to provide the proper voltages for each.

Knight-Kit transistor tester. A second inexpensive transistor checker, the Knight-Kit transistor and diode checker, is marketed by the Allied

Radio Corporation. The internal circuit arrangement, Fig. 10·4, is quite similar to that in Fig. 10·3. The transistor is inserted into the tester and the switch labeled "Leakage-Gain" is shifted to "Gain." Calibration control R_2 is then rotated until the meter reads "1," or full scale. The switch is then returned to the leakage position and the meter reading is again noted. If the leakage current is less than the gain current (here, arbitrarily set at 1), the transistor can be presumed good. Actually, the greater the difference between the gain and leakage readings, the better the unit. Thus, although this instrument has the same basic circuitry as the preceding transistor checker has, the method of

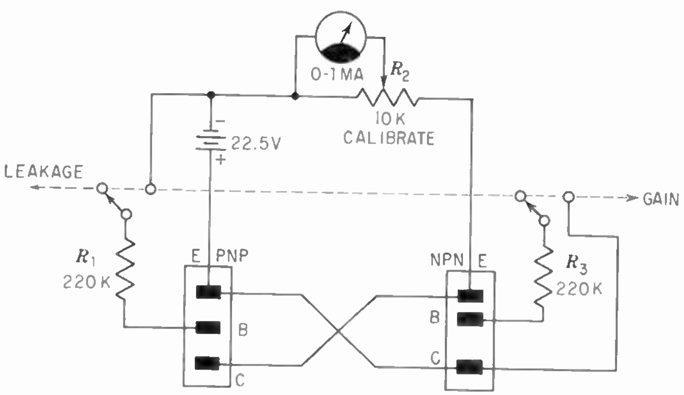


Fig. 10·4 The schematic diagram of the Knight-Kit transistor tester.

checking is somewhat different. The gain for every transistor is arbitrarily set at 1 by the calibration control while the switch is in the gain position. When the switch is returned to the leakage position, the current through the transistor should be considerably less than 1. (The leakage current being measured here is I_{CEO}.)

The foregoing arrangement permits a wide range of transistors to be checked, because no matter what the actual β of a specific transistor is, the calibration control always establishes it at 1. Thus, this checker does not directly measure β either; instead, it measures the ratio between β and leakage. The higher this ratio, the better the transistor.

Triplett transistor testers. The Triplett Electrical Instrument Company has two transistor testers, a Model 690-A and a Model 2590. Of the two, the Model 690-A is the simpler and will be considered first. For leakage, I_{CBO} is checked with the circuit in Fig. 10·5. The base-to-collector elements are reverse-biased by a 6-volt battery, and the emitter is left floating. A 200-μa meter, connected into this circuit, indicates any current. For small transistors (i.e., low-power units) the leakage current should be less than 15 μa. Otherwise, the transistor is defective. Power transistors, as previously noted, have higher leakage currents. Good silicon transistors should not produce any reading at all on the meter.

In order to measure β, the instrument must first be calibrated. The calibration circuit for the Model 690-A is shown in Fig. 10·6*a*. Here is how it operates. A 30-volt battery is connected in series with a 100,000-ohm potentiometer, a 33,000-ohm resistor, and a 200-μa meter movement having a 500-ohm shunt. From the diagram, it would appear that the meter movement is by itself in the base circuit and that the 100,000-ohm potentiometer and the 33,000-ohm resistor are in the

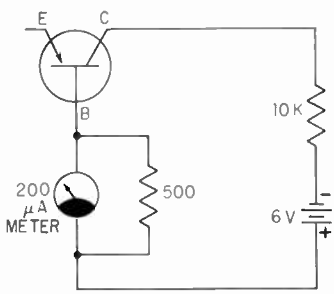


Fig. 10·5 Leakage test circuit in the Triplett Model 690-A transistor tester.

emitter circuit. Basically, this is so. However, because of the short circuit between emitter and base, no current passes through the transistor. Instead, all of it flows through the meter and its shunt. With this setup, the 100,000-ohm potentiometer is adjusted until the meter needle lines up with an indicated calibration point. This occurs when 500 μa is flowing in the circuit.

Once the calibration is established, the short circuit is removed, together with the 500-ohm shunting resistor, Fig. 10·6*b*. Now the current passes into the emitter element; part of it flows to the base circuit, but the bulk of it (about 95 to 98 per cent) flows to the collector circuit. In short, the transistor begins to function as an operating device. The current injected at the emitter passes, for the most part, through the collector circuit; only a nominal amount (3 to 5 per cent) flows in the base circuit. This causes the meter needle to decrease from the calibration line. The smaller the current in the base circuit, the higher the β value of the transistor. This causes the largest numbers to be at the left end of the gain scale. These markings can be read directly, and wherever the meter needle comes to rest, this represents the d-c β for the transistor under test.

In the second transistor tester, the Triplett Model 2590, leakage can be measured for I_{CEO} and I_{CBO}, Fig. 10·7. This is a departure from any of the preceding instruments. The initial step in gain measurement is to first calibrate the meter scale according to the transistor under test. The calibration circuit, Fig. 10·8*a*, is a simple one in which current is fed into the base. The resulting collector current is then measured by

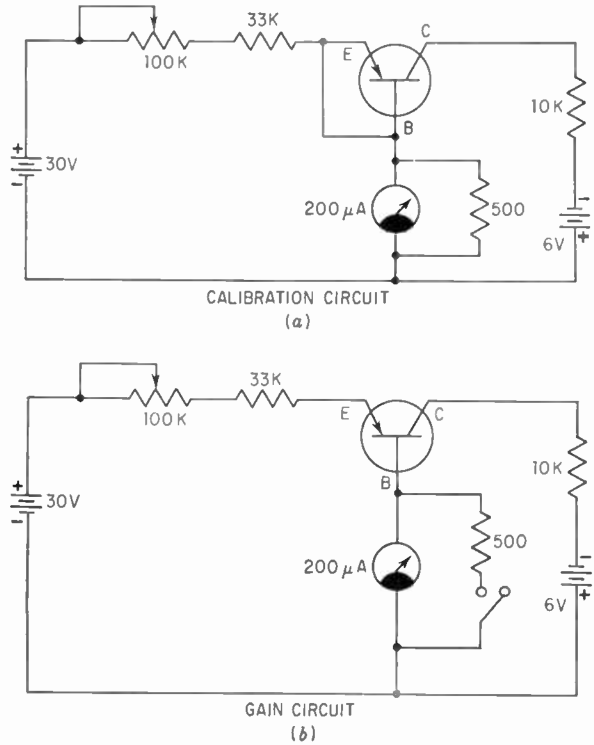


Fig. 10·6 The calibration and gain circuits in the Triplett Model 690-A transistor tester.

a 100-μa meter which has sufficient shunts to handle up to 500 μa. Calibration adjustment R_1 is then rotated until the meter needle rests at the extreme right edge of the dial, at the point marked CAL. Next, the central selector switch on the test instrument is switched to the BETA position. This produces the circuit in Fig. 10·8*b*. That is, the meter previously in the collector circuit has now been replaced by an equivalent resistor, and the meter movement itself has been shifted to the base circuit. Here, the base current is measured directly. Because of the previous calibration, the β value can be read directly from the scale. Note that here again, because the meter movement is shifted from a high- to a low-current circuit, the β values increase to the left. Obviously, the less current the base circuit requires to produce a certain collector current, the higher the β value for that particular transistor.

Servicing Transistor Circuits

In the servicing of transistor devices, much can be learned from a measurement of the voltages found in the circuits. This voltage-

Fig. 10·7 A transistor tester. (Triplett)

analysis method has long been employed with considerable success in vacuum-tube circuits; that it can be applied equally well to transistor circuits is indeed fortunate. There is, however, this very important difference: whereas electrons always flow through a tube in one direction, their path through a transistor is dependent on the type of transistor, PNP or NPN. Also, the relationship between the various electrode voltages is significantly different in transistors as compared with tubes. But once these obstacles have been overcome, voltage analysis of defective transistor circuits will pay off handsomely as a servicing tool.

The following are some general statements that apply to both PNP and NPN circuits. Since both have currents flowing through them, both will develop voltage drops across resistors in the base, emitter, and collector circuits. A typical audio amplifier stage using a PNP

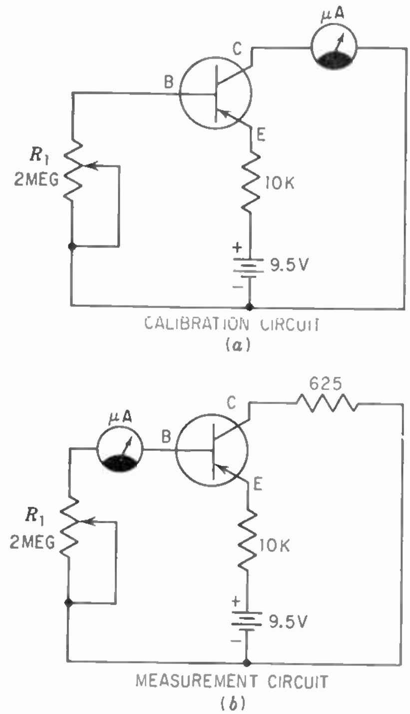


Fig. 10·8 Calibration and gain-measurement circuits employed in the transistor tester shown in Fig. 10·7.

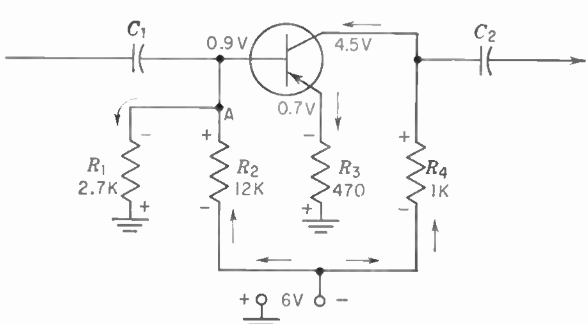


Fig. 10·9 Current flow in a PNP transistor circuit.

transistor is shown in Fig. 10·9. Resistors R_1 and R_2 provide the base with the proper voltage (or current). R_3 is an emitter-stabilizing resistor; R_4 is the collector load resistor. Since the battery potential applied to the circuit is negative, electrons will flow in the directions indicated by the arrows. In PNP transistors, electrons travel into the collector and base and out of the emitter element. In NPN transistor circuits, Fig. 10·10, electrons flow into the emitter and out of the collector and base. Thus, the NPN transistor comes closest to duplicating the electron flow of vacuum tubes.

In examining the voltages at the transistor elements, it will be found that the range of separation between emitter and base is generally on the order of 0.1 to 0.3 volt. This is because the emitter-base diode is forward-biased; hence, the internal resistance across this PN junction is very low and so is the voltage drop there. On

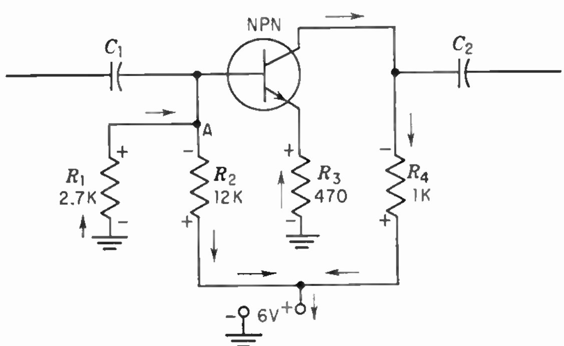


Fig. 10·10 Current flow in an NPN transistor circuit.

the other hand, the voltage difference between collector and base or collector and emitter is considerably greater.

The one exception to the above statements occurs when a transistor is normally cut off, conducting only when it is pulsed. In this case, the difference between base and emitter voltages can be greater than the range indicated. This situation does not occur in radio-receiver circuits, but it is found in television circuits and industrial equipment.

When the transistor is conducting, every resistor in its immediate circuit will have a voltage drop. The drops should be measured when a circuit is being investigated to establish the fact that current is flowing. If there is no voltage drop across one or more resistors, but voltages are present, then an open transistor is a possibility. On the other hand, if the voltage drops are higher than normal, the transistor may be leaky.

Of the voltages associated with a transistor, those at the base and emitter are the most critical. When both of these elements have the same voltage, the transistor is at cutoff. This will bring the collector voltage to the same value as the applied battery voltage to that stage.

Now, normally, cutoff will be difficult to achieve in a single stage if no outside influences enter the circuit. That is, in Fig. 10·9, R_1, R_2, and R_3 can normally change values within a wide range and the transistor will still conduct (although, perhaps, at a different value from its normal condition). This is because the emitter voltage follows the base and, as the base voltage changes, so will the emitter voltage. But if an outside voltage is brought into the circuit, such as a leaky C_1 might bring, then cutoff can occur.

Note, too, that the collector voltage can vary over a fairly wide range and still provide normal or close-to-normal operation. This is evident from the characteristic curves of transistors.

Voltage variations of 10 to 20 per cent from those stated by the manufacturer for the circuit or system are ordinarily permissible. Furthermore, since low voltages are the norm for transistor circuits, 10 or 20 per cent of these voltages will likewise be small. Thus, where we might be inclined to dismiss a variation of 5 to 10 volts (from a stated value) in a vacuum-tube circuit, in a transistor circuit this could represent a change considerably more than the permissible 10 to 20 per cent. So it is necessary to be more alert to voltage variations generally and always try to evaluate them from the percentage standpoint.

A circuit arrangement which is fairly common in power stages is shown in Fig. 10·11. The collector of this PNP transistor is d-c grounded, while a large positive voltage is applied to the emitter. This has the same effect as applying a similar negative voltage to the collector and grounding the end of the 470-ohm emitter resistor. The arrangement in Fig. 10·11, however, is advantageous with power transistors because it necessitates less stringent insulating precautions from the chassis on which the transistor case is mounted. In power transistors, the collector is usually connected directly to the transistor case.

Note that a similarly large positive voltage must also be applied to the base to establish the proper base-to-emitter bias. This is done in Fig. 10·11 by tying R_1 to the +9-volt line. The voltage-divider arrangement of R_1 and R_2 then provides the base with the necessary voltage.

Particular attention is called to the circuit setup of Fig. 10·11 not only because it enjoys widespread use but also because a zero collector voltage reading (or one close to it) will often be taken as an indication of a defect.

The effect of open elements and open leads. When an open element develops in a transistor or when an open circuit appears in one of the components (or connections) attached to a transistor, the voltages that are then measured at the transistor can easily prove confusing unless the situation is carefully analyzed. For example, in Fig. 10·11 the emitter voltage is ordinarily 8.4 volts, the base is 8.2 volts, and the collector is 0.6 volt. If, now, R_1 should open up, conduction through the transistor would effectively stop. In making a voltage check of the stage, however, it would be found that a meter would show 9 volts on the emitter *and* the base. That is, the base voltage would not show up as zero. The reason for this stems from the very low internal resistance between base and emitter. With this

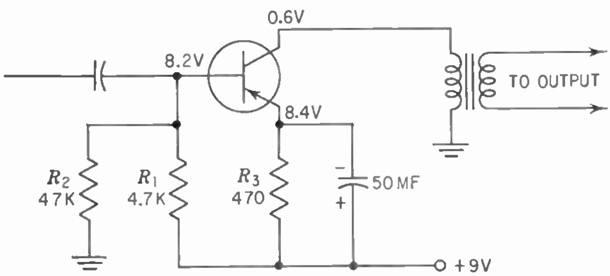


Fig. 10·11

low resistance, the base will show, to the voltmeter, the same voltage as the emitter. This could easily lead the serviceman to conclude that the base was receiving its voltage through its voltage-divider network and pass up the defect that exists in R_1.

Exactly the same result will be obtained when the open circuit develops in the emitter lead but the base is intact. Thus, if R_3 should open up, the base voltage (because of R_1 and R_2) would remain about 8.2 volts. Measuring the voltage at the emitter would reveal the same value, again because of the low resistance between the two elements.

On the other hand, if the emitter opened up internally so that no direct path then existed between base and emitter, the emitter voltage would be 9 volts, Fig. 10·11, while the base would be 8.2 volts. This difference of +0.8 volt between emitter and base in a PNP transistor would ordinarily cause excessive current flow through the circuit. Since it does not here and, further, because it reveals no voltage drop across R_3, suspicion is directed at the transistor. An open internal base lead will give precisely the same voltage readings.

One interesting effect appears when an open circuit develops in the collector external circuit. With the removal of the battery voltage, we would expect the collector potential to drop to zero. Instead, however, a voltage equal to the base voltage appears at the collector. To see why this occurs, let us consider the internal conditions in the transistor. Figure 10·9 will serve as an illustration. With the collector voltage gone, we find -0.9 volt on the base (which has not been affected by the change in the collector) and nothing on the collector. Since the base is an N-type semiconductor and the collector a P-type semiconductor, the foregoing voltages will forward-bias the base-collector diode and produce a low-resistance condition there.

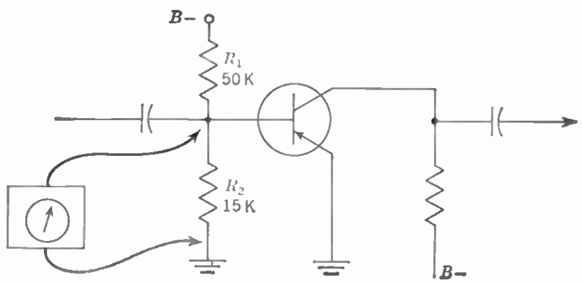


Fig. 10·12 The resistance value indicated by the ohmmeter will depend upon the way the internal battery of the instrument is connected.

Hence, when a voltmeter probe is touched to the collector electrode, the voltage at the base will be obtained. As a matter of fact, with the collector circuit open, the potentials at all three transistor elements will be fairly close to each other.

Finally, if the collector opens internally, no current will flow in the collector circuit. The collector voltage will then be at the full battery voltage.

Resistance measurements in a transistor circuit may pose a problem because of the presence of the transistors. To see how this can occur, consider the partial input circuit shown in Fig. 10·12. All power in the circuit has been turned off, and one of the resistance checks we wish to make is that of R_2. Now, depending on what ohmmeter lead we connect to the top end of R_2 and what lead we connect to the bottom end, the value indicated by the ohmmeter will be either 0 or 15,000 ohms, assuming the resistor to be good.

The reason for this behavior stems from the nature of transistors and the manner in which ohmmeters operate. An ohmmeter applies a voltage across the resistance to be measured, and from the resulting current flow the resistance value is determined. Let us suppose that the ohmmeter has an internal battery of 3 volts, this being a fairly common value. This voltage must make one lead positive and one lead negative. Ordinarily, the red lead is positive and the black lead is negative, although the reverse combination is also found. Now, if the positive lead goes to the top of R_2 and the negative lead to the bottom, the true value of R_2 will be indicated on the meter because, under these conditions, the base-emitter circuit of the transistor is reverse-biased. That is, the positive meter voltage is being applied to an N base here while the negative meter voltage goes to a P emitter (through the ground connection).

From this, it is readily apparent that if we reverse the ohmmeter leads, a negative voltage will be applied to the base and a positive voltage to the emitter. This will forward-bias the unit and throw an extremely small impedance across R_2. The measured result of this combination will also be quite small. (While it was indicated above that the value obtained under these conditions would be zero, this is not exactly true. Actually, under the conditions specified, it might be on the order of 100 ohms or less; but, with the ohmmeter set to read 15,000 ohms, the lower reading will appear to be quite close to zero on most conventional ohmmeters.)

Another possible side effect of the second reading is damage to the transistor because of the excessive current flow. To avoid damage, either the transistor should be removed prior to such measurements or care should be taken to see that the proper ohmmeter lead goes to the top end of R_2. If the polarity of the ohmmeter leads is not known, it can be determined readily by connecting the leads to a d-c voltmeter and noting whether the meter needle moves to the right or left.

What has been stated for resistance measurements in the emitter circuit is just as true in the collector circuit. Remember that a forward-biased collector has practically the same low impedance as a forward-biased emitter.

It is generally a good practice to inspect a transistor circuit visually as part of the servicing procedure. Look for such things as breaks in the printed circuitry or even in the printed board itself. A sudden twist, a sharp jar, or an inadvertent fall of a transistor device can readily damage any one of a number of miniature components common to transistor circuitry, including the mounting board containing the printed wiring. A careful visual inspection will frequently bring these defects to light and shorten what could otherwise be a lengthy service job.

Because of the wide variety of battery voltage values found in portable receivers, it is helpful to have a low-voltage power supply available for transistor-receiver servicing. Since battery voltages used in transistor sets seldom exceed 22½ volts, the power supply need not provide a higher voltage than that. Current requirements are in the milliampere range and are generally below 100 ma. The voltage output of the supply should be variable, preferably with a front-panel indicating meter to reveal the exact output voltage. Then, when the batteries in a unit are suspected of being low or dead, the power supply can be substituted directly for the batteries.

In selecting a power supply for testing purposes, make certain that a good d-c output is obtainable. Because of the low value of the d-c voltages required by a transistor circuit, even minute amounts of a-c ripple can produce annoying hum from the loudspeaker.

Printed Circuits

Transistors are used extensively with printed circuits, and so it behooves the technician to become familiar with the proper methods of removing or adding components to a printed-wiring chassis. The following discussion, from information furnished by the Admiral Corporation, will be helpful in this respect.

A printed circuit begins as a laminated plastic board with a sheet of thin copper foil bonded to one side. To form the necessary wiring, some of the copper foil is removed by a photographic and etching process. Holes through which various component leads are inserted are punched in the board. The leads of the various components are cut and bent over the copper-foil wiring. The wiring side of the board is then dipped in molten solder to make all solder connections at once. The copper-foil wiring also picks up solder, thus increasing its ability to carry current. Finally, a coat of silicone-resin varnish is applied to the wiring side of the board. This prevents dust or moisture from causing short circuits. The result is a circuit with uniformity of wiring, compactness, and freedom from wiring errors.

The foregoing method of producing printed wiring boards is known as the etched wiring method. It is more widely used today than any other form of printed wiring. This is largely because of the reliability, great flexibility, and low setup cost of the method. There are, however, other methods of manufacture such as embossed wiring, stamped wiring, and pressed-powder wiring. Since we are interested primarily in the end result, none of the other methods will be described here.

Circuit tracing of a printed-circuit board is usually simpler than that of conventional wiring owing to the uniform layout of the wiring.

Also, many boards are translucent, and a 60-watt light bulb placed underneath the side being traced will facilitate location of connections. Test points can frequently be located rather easily in this manner without the necessity of viewing both sides of the board.

Resistance or continuity measurements of coils, resistors, and some capacitors can be made from the component side of the board. In some cases a magnifying glass will assist in locating very small breaks in the wiring. Voltage measurements can be made on either side of the board. However, on the wiring side of the board (some printed circuit boards have component parts on both sides), a needle-point probe for circuit checking should be used, since the varnish coating must be pierced to make contact.

Be careful when removing components from the board. However, if the copper-foil wiring is damaged, a piece of wire can be used to replace the damaged foil. Small breaks can be "jumped" with molten solder. Larger breaks can be repaired with ordinary hook-up wire. It is seldom necessary to replace an entire board because of foil breakage.

Do not apply excessive pressure to the printed-circuit board or components. Although the board is sturdy in construction and mounting, it may crack or break if proper care is not taken when servicing. On extremely rare occasions, access to components on the board may be difficult. In that case the board may be removed from the chassis by removing the mounting screws around the edges and unsoldering a few leads between the board and the chassis. If this is done, a vise with protected jaws should be used to hold the board while servicing and care should be taken not to exert excessive pressure against the board.

In some areas on the printed board, the wiring is very closely spaced. When resoldering a new component, avoid excessive deposits of solder. Excessive solder may cause a short or intermittent trouble to occur later and be difficult to locate.

When using the soldering iron (35 watts or less), do not overheat the component terminals or the copper foil. Excessive heat (applying the soldering iron longer than necessary, using a higher-wattage iron than recommended, or using a soldering gun) may cause the bond between the board and foil to break. This will necessitate the replacement or repair of the foil connection.

Replacing Capacitors, Resistors, Couplates, and Peaking Coils

Defective resistors, couplates, and ceramic-disk and wax-encased capacitors can be replaced by either of the following two methods:

1. If the leads extending from the defective component are long enough for a replacement component to be soldered to it, cut the leads where they enter the defective component, Fig. 10·13.
2. If there is not enough length in the leads extending from the defective component to use the method described in (1), cut the defective component in half. Then cut through each half of the component until it is broken away from its lead. If this procedure is performed carefully, enough extra lead inside the component

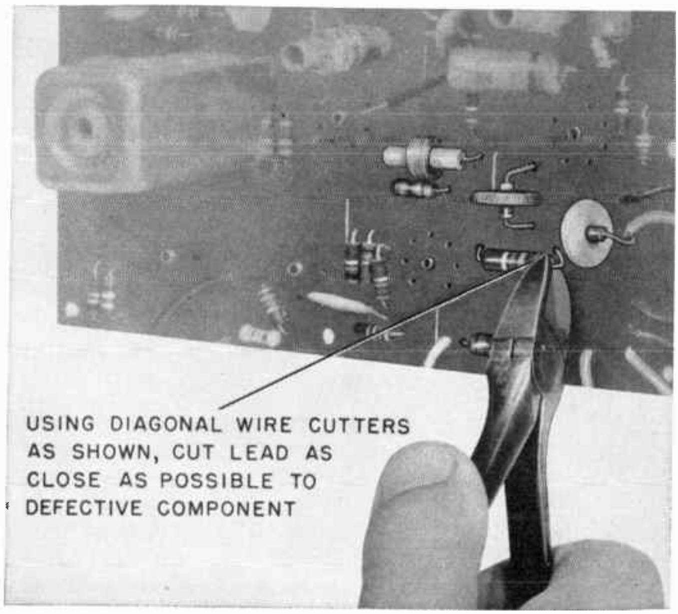


Fig. 10·13 Cutting a defective resistor free of the printed-circuit board. (Admiral Corp.)

will be gained to permit soldering the replacement component to it, Figs. 10·14 and 10·15.

Clean off the ends of the remaining leads, leaving as much of the leads as possible. Make a small loop in each lead of the replacement component and slide the loops over the remaining leads of the old component, Fig. 10·16. Caution should be observed not to overheat the connection, since the copper foil may peel or the original component lead may fall out of the board. This is possible because of heat transfer through the leads. The lead length of the replacement part should be kept reasonably short to provide some mechanical rigidity.

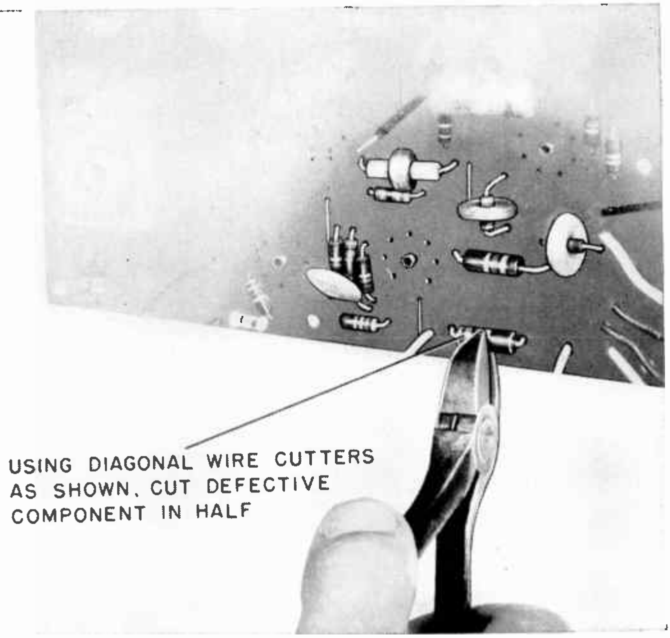


Fig. 10·14 Cutting a defective resistor apart so as to have maximum lead length left. (Admiral Corp.)

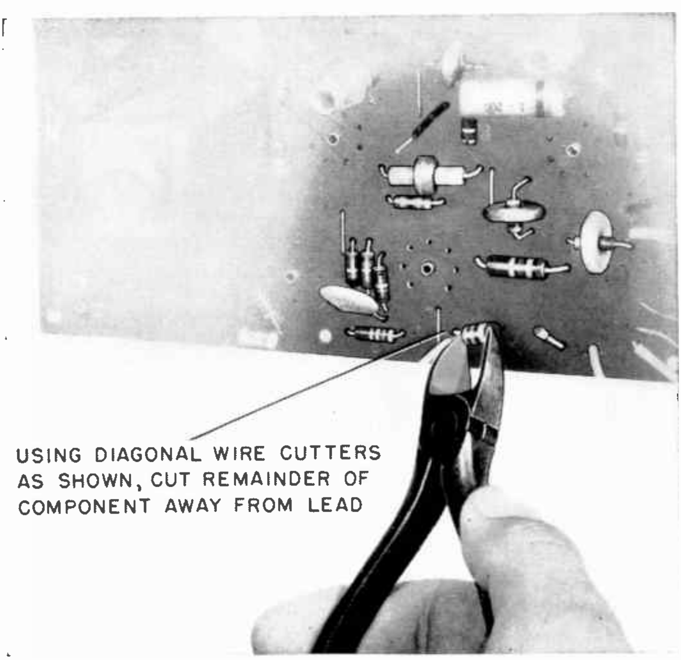


Fig. 10·15 Cleaning remaining leads of component that has been cut apart. (Admiral Corp.)

In some cases, components are mounted in such a manner that neither of the above methods can be used. In that event it is necessary to unsolder the defective component completely and replace it. The following procedure should be used whenever it is necessary to unsolder any connections to replace defective components.

1. Heat the connection on the wiring side of the board with a small soldering iron. When the solder becomes molten, brush away the

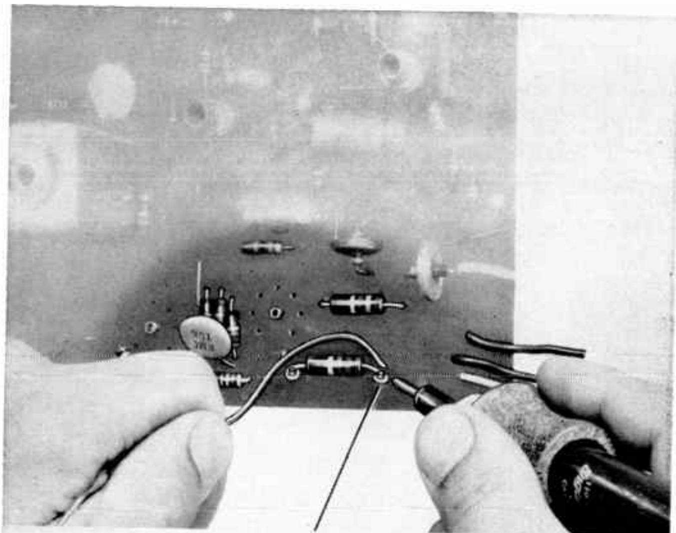


Fig. 10·16 Soldering replacement resistor in place. (Admiral Corp.)

solder, Fig. 10·17. A 60-watt bulb placed over the component side of the board will facilitate location of the connections on the wiring side if the board is translucent (and many boards are). In the process of removing the solder, caution is needed to prevent excessive heating. Therefore, do not leave the iron on the connection while brushing away the solder. Melt the solder, remove the iron, and quickly brush away the molten solder. (For this purpose, a small wire brush is suitable.) More than one heating and brushing process may be required to remove the solder completely.
2. Insert a knife blade between the wiring foil and the bent-over component lead, and bend the lead perpendicular to the board.

(It may be necessary to apply the soldering iron to the connection while performing this step, because it is sometimes difficult to break the connection completely by brushing.) Do not overheat the connection.

3. While applying the soldering iron to the connections, wiggle the component until it is removed.
4. Remove any small particles of solder embedded in the silicone resin (if such a coating is employed) by using a clean cloth dipped in solvent.

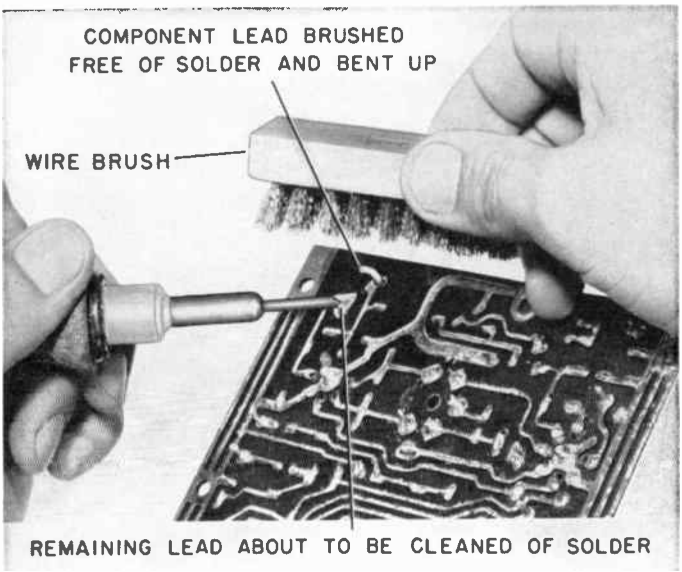


Fig. 10·17 All excess solder should be carefully brushed away. (Admiral Corp.)

5. A thin film of solder may remain over the hole through the board after removing the component. Pierce the film with the lead from the new component after heating the film with the soldering iron.
6. Insert the leads of the new component through the holes provided, cut to desired length, and bend over the ends against the copper foil. Resolder the connection with 60/40 low-temperature solder.
7. It is recommended that the cleaned area be recoated with clear lacquer or sprayed with Krylon for protection against shorts. If the Krylon spray is used, it will be necessary to cover the top of the tube sockets and chassis ground connections with masking tape to prevent the contact surfaces from becoming coated.

Replacing Coils

The terminal lugs of these components are not bent over against the foil in most cases; therefore, brushing is not necessary. Heat one connection until the solder becomes molten, and wiggle the coil back and forth until the connection is broken. Continue to wiggle and apply the soldering iron to the other connections, and lift the coil from the board while the solder is still molten. Insert the replacement coil in the exact same position, and solder the connections. Cover the connection points with a coat of lacquer or Krylon.

Transistor-receiver Servicing

Transistor-receiver servicing does not differ appreciably from the servicing of vacuum-tube-operated receivers. There are, however, certain differences of initial approach that are due to the use of batteries and transistors, and it is these differences (rather than the more familiar similarities) which will be considered here.

For example, when the output of a transistor receiver is distorted, weak, or completely dead, the prime suspect is the battery. The measurement is made with a vacuum-tube voltmeter or high-resistance voltmeter and is best taken with the battery in the receiver and the latter turned on. If the voltage reading is at or near the correct value, the battery can be presumed to be good. If the voltage reading is off by 20 per cent or more, then the receiver output may be weak or distorted, but it should not be dead. Since transistor characteristics are linear to very low voltages and currents, chances are that distortion will not occur until the battery voltage drops more than 20 per cent. There is, however, no set rule regarding this, and it is best to try a new battery when the voltage of the existing battery has decreased by this amount. If the distortion or weakness still persists, then some other defect is indicated.

Whenever a weak battery is found, it may be advisable to check the resistance of the circuit across the battery clips before a new unit is inserted. For example, in the Regency model TR-1G receiver, the manufacturer indicates that the resistance between the battery clips (with the battery removed and the receiver turned on) should be between 6,000 and 15,000 ohms as read by an ohmmeter with an internal battery of not more than 3 volts. A reading lower than 6,000 ohms will usually indicate a defective component somewhere in the receiver.

Some manufacturers indicate what the current drain on the battery should be instead of quoting the circuit resistance across the battery clips. In that case, a milliammeter must be inserted in series with the battery. For example, with the negative terminal of the battery touching the negative clip, a wire is connected from the negative terminal of the milliammeter to the positive battery clip. Then one end of another wire is connected to the positive terminal of the milliammeter while the other end of this wire is touched to the positive end of the battery. The value of current indicated on the meter should fall in the range specified by the manufacturer.

If the battery proves to be good, then the rest of the trouble-shooting procedure follows established practice. As an example of this, the method of attack for the Regency model TR-1G, as recommended by the manufacturer, is given below. Study this in conjunction with the schematic diagram of the set. The latter will be found in Fig. 6·3.

The alignment procedure for the same receiver is given also following the trouble-shooting outline.

Trouble-shooting Procedure for the Regency Model TR-1G Receiver

This section is reprinted from *PF Reporter* by permission of Howard W. Sams & Co., Inc.

Dead receiver—absolutely no output

1. Remove the battery and turn on the switch. Measure the resistance between the battery clips. (Make sure the positive meter lead is on the positive clip.) If the resistance is
   a. Approximately 10,000 ohms, the B+ circuit is normal.
   b. Less than 2,500 ohms, check the leads of capacitors C_{17} and C_{21}, and make sure they are not touching the battery clips or the frame of the output transformer. Check for a shorted condition in either C_{17} or C_{21}. Measure the resistances from the top ends of R_3 and R_7 to ground. These should be 2,200 ohms more than the reading across the battery clips.
   c. Infinity, check for an open switch.

2. Turn the volume control to maximum and insert the battery. If a click or noise is heard from the speaker, check X_4 by shorting its base to the frame of the output transformer. The audio stage is operating if a click is heard. If no click or noise is heard, proceed as follows:
   a. Check for an open or shorted jack. Indicative readings can be obtained by measuring the resistance from the fixed contact of the jack to ground. These readings are
      0 ohms—shorted jack
      2 ohms—normal

15 ohms—jack is open, or the ground between the wiring board and chassis is open

b. Check for an open condition in the speaker or in the output transformer.

c. Voltage at the base of X_1 (about +2 volts normal).

d. Voltage at the emitter of X_1 (should measure approximately 0.15 volt less than the base voltage).

3. Check capacitor C_{19} by paralleling it with a capacitor known to be good.

4. Measure the voltage at the output of the diode D_1 (should be approximately +0.1 volt).

a. If voltage is zero, check the resistance to ground with the positive meter lead on the output. This resistance should measure between 20 and 100 ohms. If the resistance is zero, check for a shorted condition in the diode circuit. If the resistance is 200 ohms or greater, check for an open in the diode circuit or for an open diode.

b. If voltage is negative when the receiver is tuned to a station, move the tuning dial so that no station is received. The negative voltage should decrease.

c. If voltage is negative by 1 volt or more and does not drop when the receiver is tuned off the station, the receiver is oscillating. Proceed to the section entitled Oscillating Receiver.

5. Make voltage and resistance measurements in the i-f stages.

Dead receiver—noise but no signal

Check the local oscillator in the receiver as follows: Tune another receiver to any station above 850 kc. On the receiver being serviced, rock the dial above and below a setting that is approximately 262 kc below the frequency of the station being received by the other receiver. If the local oscillator in the receiver being serviced is operating, a whistle will be heard from the other receiver as the radiation from the oscillator beats with the station frequency.

If the oscillator is dead, proceed as follows:

1. Check the voltage at the base of X_1. This should be between 3 and 10 volts.

2. Check the voltage at the emitter of X_1. This voltage should be within 0.1 volt of the base voltage.

3. Check the voltage at the top end of R_3. This should be measured from the B+ line, and it should be between 0.6 and 2 volts.

4. If any of the voltages measured in the three preceding steps are

incorrect, check for an open oscillator-coil primary or an open first i-f transformer.

5. Check resistances of
   a. The high side of the antenna coupling coil to ground (should be less than 1 ohm).
   b. The secondary of the oscillator coil (should be approximately 10 ohms).
   c. Stator of oscillator section of the tuning capacitor to ground (should be infinity).

Weak or distorted output

1. Turn volume control to maximum. Check capacitors C_{19} and C_{21} by paralleling a good capacitor across each.

2. Perform step 5 under section Dead Receiver—Absolutely No Output.

3. Measure voltages at
   a. Base of X_4 (should be approximately +2 volts).
   b. Emitter of X_4 (should be approximately 0.15 volt less than the base voltage).
   c. Base of X_3 (should be approximately 0.15 volt less than the voltage at the emitter of X_3).
   d. Top end of R_{11} (should be approximately —0.5 volt when receiving a signal of average strength).
   e. Avc line (should be from approximately 0 volt with signal to 0.5 volt with no signal).
   f. Emitter of X_2 (should be approximately 0.15 volt less than the avc line).

4. Check the alignment of the receiver.

Oscillating receiver

1. Measure the battery voltage. If it is below 15 volts, the battery should be replaced.

2. Check the local oscillator as in step 1 under the section entitled Dead Receiver—Noise but No Signal.

3. Check capacitors C_{17} and C_9 by paralleling a good capacitor across each.

4. Check ground connection between wiring board and chassis. This connection is the twisted lug near the negative battery clip and is the only lug which has been soldered to the board. Measure between an i-f transformer can and the metal chassis. These readings are

0 ohms—normal
15 ohms—ground lead is open

Alignment of Regency Model TR-1G Receiver

The alignment of the Regency receiver is quite simple. Signal injection is accomplished by connecting the signal generator to a loop formed of several turns of wire and situated close to the antenna coil of the receiver. Set the generator to 262 kc with 400-cycle modulation and reduce the output to as low a value as is usable. Connect an output meter (with a 0.1-volt scale) across the voice-coil terminals. (The high side of the voice coil is easily accessible at the spring of the phone jack in this set.) Set the volume control in the receiver to maximum. Adjust each of the cores of the i-f transformers for maximum indication on the output meter. Set the receiver dial to its maximum counterclockwise position, tune the generator to 535 kc, and adjust the core of the oscillator coil for maximum output. Tune the generator to 1,630 kc, set the receiver dial to its maximum clockwise position, and adjust the oscillator trimmer capacitor for maximum output. Repeat these last two adjustments alternately until no further improvements can be made. Then tune the generator to 1,500 kc, tune in this signal with the receiver dial, and adjust the antenna trimmer capacitor for maximum output. Turn the receiver dial to the high-frequency end and determine whether or not the range extends to 1,630 kc. If not, the oscillator trimmer capacitor must be readjusted and the alignment at 1,500 kc must be repeated.

QUESTIONS

10·1 In what respects are transistors sturdy? In what respects fragile?

10·2 What changes in working tools are necessary when dealing with transistors and their associated components?

10·3 Indicate several precautions to follow when soldering transistors into a circuit.

10·4 Why are transistors especially sensitive to applied voltages?

10·5 Outline the precautions to observe when injecting signals into a transistor circuit. Indicate a suitable safe method of bringing such signals into a circuit.

10·6 How could one determine whether a transistor was good or bad?

10·7 Draw a simple test circuit to measure I_{CBO}.

10·8 Indicate briefly how transistor gain is checked.

10·9 Indicate generally the precautions to observe when removing components from printed-circuit boards.

10·10 In what respects does the servicing of transistor receivers differ from the servicing of comparable vacuum-tube sets?

10·11 Is it better to check a battery (using a vacuum-tube voltmeter) while it is in the circuit or after it has been removed completely from the set? Give reasons for your answer.

10·12 How would you align the Regency model TR-1G receiver?

10·13 If it were determined that the oscillator in the Regency receiver was not functioning, how would you proceed to localize the defect?

10·14 What might cause a weak or distorted output from the Regency receiver?

10·15 Under what conditions would the battery in the Regency set not be the first component checked? (Assume that the receiver is not functioning properly.)

CHAPTER 11

Experiments with Transistors

IN THE PRECEDING chapters, the theory and application of transistors were covered in detail. The information contained there represents the first step toward the acquisition of a basic understanding of transistor operation. The next step for the man who is going to work with these units is actual physical contact so that he may become practically proficient in handling transistors and learning, first hand, of their characteristics and peculiarities. Toward this end, a series of experiments are presented, and all readers are urged to perform them prior to any work on transistors in commercial equipment.

The circuitry involved in these experiments has been kept as simple as possible. This serves the twofold purpose of making each experiment easy to perform and keeping component cost down. Furthermore, the same basic components are used over and over again. Because of this, caution should be exercised when leads are trimmed prior to soldering lest the amount of wire removed be so much that the unit will not be usable for as many times as required.

Another very important precaution to observe, one that was mentioned in Chap. 10 on transistor servicing, is the use of a low-wattage soldering iron when soldering transistor leads into the circuit. Keep the leads as long as possible, and grip the lead being soldered with a pair of long-nose pliers. Since the lead is held between the point where the heat is applied and the body of the transistor, any heat traveling along the lead wire will be shunted away from the transistor.

It is also important to observe battery polarity when connection is made to the circuit. If the collector terminal receives a forward-biasing voltage in place of a reverse-biasing voltage, the transistor can be damaged permanently. Double-check wiring before connecting any batteries. Do not use voltages higher than those indicated in the experiments. Also, make certain that the battery is disconnected before any wiring connections are altered. Finally, be especially careful to avoid circuit shorts between various wires. If necessary, cover all bare wires (or exposed ends of wires) with protective spaghetti.

No specific chassis form or size is recommended for the ensuing experiments. They may be performed on a breadboard or on any of the small metallic chassis that are obtainable at a parts jobber. In the latter instance, there are a number of standard base sizes ranging generally from 2 by 6 by 4 in. (height, width, depth) to 5 by 17 by 13 in. A recommended size of 3 by 7 by 5 in. was found to be entirely adequate for the experiments; however, any suitable dimensions may be used.

It would also be desirable to use terminal strips on which the components may be mounted. Choice of such strips is left to the reader.

Experiment 1. Adjusting Transistor Voltages and Currents

Note: NPN transistors can be used in place of the PNP units specified. If they are, the battery voltage as well as all electrolytic capacitors must be reversed. Otherwise, identical results should be obtained.

Object. To adjust the voltages on a transistor and to establish the principle of phase reversal in a grounded-emitter amplifier.

Material required

1 5-μf electrolytic capacitor
1 0.01-μf capacitor
1 1-megohm potentiometer
1 2N104 transistor (or equivalent) *Note:* Any low-frequency PNP transistor which operates with the voltages noted would be suitable in place of the recommended 2N104. Other high-frequency or higher-power transistors could also be employed, but these units generally cost more, and they would not provide any more useful information (in these experiments) than the inexpensive transistor
1 5,600-ohm resistor (¼ watt)
1 1,000-ohm resistor (½ watt)
1 4½-volt battery

Test equipment

1 vacuum-tube voltmeter or a good multimeter (preferably 20,000 ohms per volt)

Procedure

1. Wire the circuit of Fig. 11·1.
2. Check all connections before attaching the battery.
3. Connect the battery and adjust P_1 until the voltage between collector and emitter is approximately 3 volts.
4. Leave the voltmeter connected to the collector. Take a 10,000-ohm resistor and touch it between base lead and ground. Notice how the collector potential becomes more negative. When the 10,000-ohm resistor was touched from the base to ground, it caused the base to become less negative, or more positive. At the same time the collector became more negative.

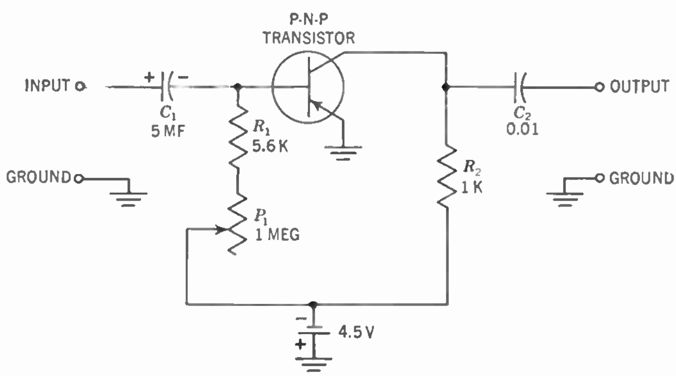


Fig. 11·1

Conclusions

1. The principal conclusion that can be drawn from the foregoing behavior is that a grounded-emitter amplifier reverses the phase of an applied signal.
2. As a secondary consideration, it was noted that as P_1 was rotated, it varied the base and collector voltages. We could also conclude that if a resistor had been inserted in the emitter lead, the voltage drop across the resistor would have varied also. Ninety-five per cent of the emitter current passes through the collector, and if we vary the collector current, we must also vary the emitter current.

Experiment 2. Distortion and Temperature Effects

Object. To observe the operation of a single-stage transistor with signal input and to demonstrate the effect of temperature change.

Material required

1 5-μf electrolytic capacitor
1 0.01-μf capacitors
1 1-megohm potentiometer
2 2N104 transistors (or equivalent)
1 5,600-ohm resistor ($\frac{1}{4}$ watt)
1 1,000-ohm resistor ($\frac{1}{2}$ watt)
1 $4\frac{1}{2}$-volt battery

Test equipment

1 audio signal generator or a filament transformer arrangement as shown in Fig. 11·2
1 oscilloscope

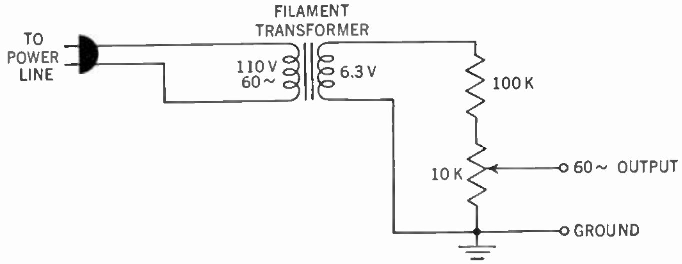


Fig. 11·2

Procedure

1. Apply a very low voltage signal, about 0.01 volt rms, between C_1 and ground in Fig. 11·1. An audio oscillator is desirable for this purpose. (Be careful not to use too strong a signal, because the transistor can be damaged by overdrive.) The signal lead goes to C_1, while the other generator lead goes to ground. If an audio oscillator is not available, a filament transformer can be used with the circuit shown in Fig. 11·2.

2. Connect the vertical input terminals of an oscilloscope between the output of C_2, Fig. 11·1, and ground.

3. Starting from zero, adjust the output of the audio generator until an undistorted sine-wave signal is just seen on the oscilloscope.

4. By varying P_1, the least distorted signal may be obtained. Note, when adjusting P_1, how the output signal reaches a peak and then starts clipping. The clipping occurs on the negative half cycle of the input signal. While this portion of the signal is active, the collector

voltage is actually becoming less negative because the collector current is increasing. At the negative input peak, the voltage drop across load resistor R_2 is almost equal to the battery voltage. Further increase in input signal cannot further lower the collector voltage, and the output signal flattens out or clips. This is similar to plate clipping in a vacuum tube.

If the input signal is increased beyond this point, clipping of the other half cycle also results. This occurs when the positive half cycle of the applied signal causes the emitter-to-base voltage to reach the reverse bias or cutoff region. Collector current is cut off too, at the same time.

5. Signal leakage may be observed by adjusting P_1 until an undistorted signal is observed and then disconnecting the battery. Notice how the input signal goes through the transistor and appears at the collector. It possesses the reverse phase from that normally seen on the collector; it is usually distorted; and, of course, there is no gain.

6. Reconnect the battery. With a vacuum-tube voltmeter or an oscilloscope, measure the input and output signals. The ratio of the output to input signals is the signal gain. After a numerical value is obtained, convert it to an equivalent decibel figure.

7. Readjust P_1 for the best possible gain with a fixed input signal. Without changing any of the instrument settings, note the height and waveform of the output signal. Now bring a hot soldering iron in the general vicinity of the transistor, thereby warming it. Be careful not to bring the iron too close. Note how the output signal changes as the transistor cools down. The amplitude of the input signal should be such that the output signal is just below distortion on top and bottom.

8. With the signal returned to normal and P_1 adjusted for maximum undistorted output, disconnect the battery and change transistors. Do not touch any of the other settings. Reconnect the battery and notice how the amplitude of the signal has changed. It may be even quite distorted. Readjust P_1 for the least distorted signal.

Conclusions

1. Clipping occurs when the signal is too strong.
2. Transistors are very sensitive to temperature variations.
3. The collector bias current may frequently have to be adjusted for each transistor.
4. The characteristics of transistors of the same type will often vary considerably. (In time it is expected that the variations among individual units will become less and less.)

5. A very definite voltage gain is obtained with grounded-emitter amplifiers.

Experiment 3. Compensation, Input Impedance, and Bias

Object. To establish temperature compensation, to note the effects of emitter degeneration on input impedance, and to note the effects of collector voltage on gain.

Material required

2 2N104 transistors (or equivalent)
2 5-μf electrolytic capacitors
1 0.01-μf capacitor
1 4,700-ohm resistor (¼ watt)
1 10,000-ohm resistor (¼ watt)
2 1,000-ohm resistors (¼ watt)
1 100,000-ohm potentiometer
2 4½-volt batteries
1 6-volt battery

Test equipment

1 audio signal generator or filament transformer arrangement as mentioned in Experiment 2
1 oscilloscope

Procedure

1. Note the gain, or output signal level, in Experiment 2.

2. Change the circuit to that shown in Fig. 11·3. Adjust P_1 until the voltage drop across R_2 is 1½ volts. (This occurs when the collector current is 1½ ma.)

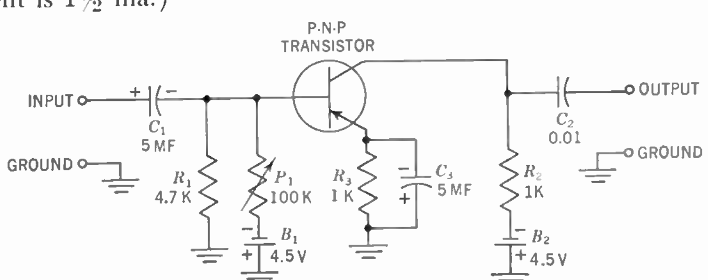


Fig. 11·3

3. Note how the gain has fallen. This may be explained as follows: Voltage measurements will show that the base and emitter voltages are higher than before but the base-to-emitter voltage is still about —0.1

volt. The collector voltage, on the other hand, is still the same. Consequently, the collector-to-emitter voltage is lower, and this accounts for the lower gain.

4. Try changing transistors. Note that transistors may now be changed without further adjustment. The circuit is much more stable because any variation in I_{co} (collector saturation current) or β is minimized by the degeneration introduced by R_3. This is also true for temperature changes.

5. Increase the collector battery voltage (B_2) to −6 volts. Notice how the gain increases. This is because of the high collector-to-emitter voltage. The transistor can also handle greater power with this increase. Do not go any higher than −6 volts.

6. Measure the gain of the stage by dividing the output signal voltage by the input signal voltage.

7. Note the amplitude of the output voltage and then remove the emitter bypass capacitor C_3. Notice how the gain drops markedly.

8. The input impedance will also increase if the bypass capacitor is removed. This can be shown as follows:

a. Reconnect the bypass capacitor C_3.

b. Place a 10,000-ohm resistor in series with the signal input generator and the input capacitor C_1, Fig. 11·4.

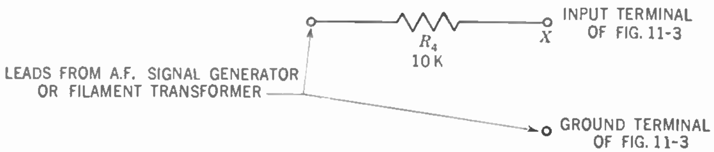


Fig. 11·4

c. Adjust the output of the signal generator for maximum undistorted output from C_2.

d. Note the height of the input signal at the junction of C_1 and R_4 (point X in Fig. 11·4).

Remove C_3 and note the increase in input signal at point X. This action can be explained as follows: With the bypass capacitor in the circuit, the base loading caused a voltage drop across R_4. When the bypass capacitor was removed, the loading was reduced and consequently more of the input signal became available to the transistor. Changes in input impedance of 5:1 or even 10:1 are common.

Conclusions

1. Emitter compensation improves transistor temperature stability.

2. Emitter compensation permits different transistors to be employed without separately adjusting the circuit for each.

3. The emitter bypass capacitor eliminates signal degeneration.
4. Removing the bypass capacitor increases the input impedance.
5. The higher the collector-to-base voltage, the greater the gain.

Experiment 4. Grounded-base, Grounded-emitter, and Grounded-collector Amplifiers

Object. To determine the relative differences among the grounded-emitter, the grounded-base, and the grounded-collector amplifiers.

Material required

1 2N104 transistor (or equivalent)
1 0.01-μf capacitor
2 5-μf electrolytic capacitors
1 4,700-ohm resistor (¼ watt)
1 10,000-ohm resistor (¼ watt)
2 1,000-ohm resistors (¼ watt)
1 100,000-ohm potentiometer
1 4½-volt battery

Test equipment

1 oscilloscope
1 audio signal generator or filament transformer arrangement

Procedure

1. Wire the circuit of Fig. 11·5. Adjust P_1 until the voltage drop across R_2 is 1½ volts.

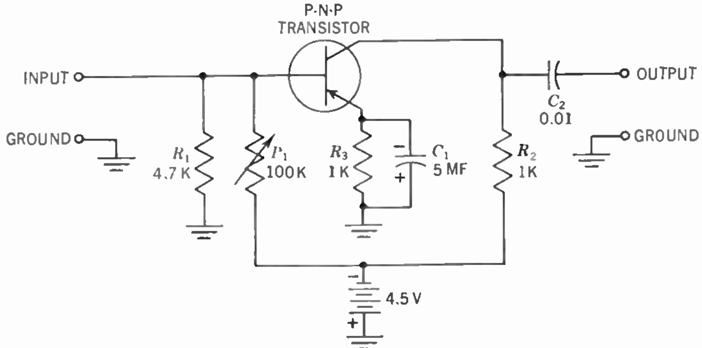


Fig. 11·5

2. Connect the vertical input terminals of an oscilloscope between the output (at C_2) and ground.

3. Feed an input signal (through a 5-μf electrolytic capacitor and a 10,000-ohm resistor in series, Fig. 11·6) to the input (which is the base lead).

4. Increase the signal input until maximum undistorted output is obtained. Check the magnitude of the output signal and also check the gain. The latter measurement should be taken as output voltage divided by the input voltage at point *A*, Fig. 11·6, and then computed again as output voltage divided by the generator voltage at point *B*.

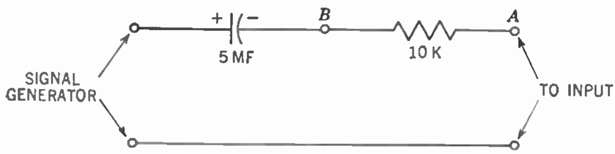


Fig. 11·6

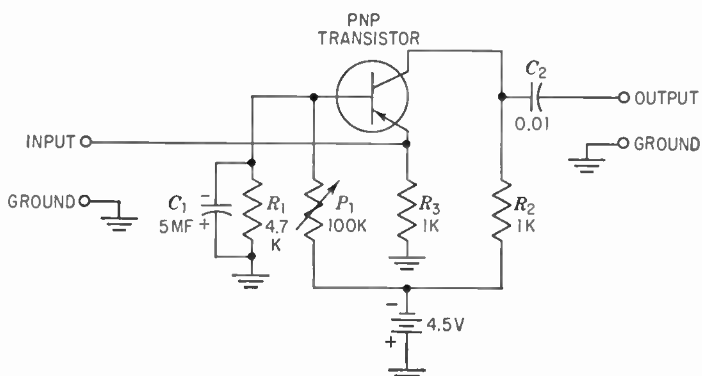


Fig. 11·7

There will be a difference between these two values because of the voltage drop across the 10,000-ohm resistor through which the signal is fed to the amplifier. This drop is caused by the input impedance of the transistor amplifier.

The true gain figure is the value obtained when the output voltage is divided by the input voltage at point *A*.

5. Connect a 0.01-μf capacitor from the collector of the transistor amplifier to the *external sync* terminal of an oscilloscope. When the instrument is properly synchronized, check the phase of the signal at the output and input terminals of the amplifier. The two waveforms should be 180° out of phase.

6. Repeat steps 1 to 5 using the circuit of Fig. 11·7, which is a grounded-base amplifier. Although the base does not go to ground di-

rectly, the bypass capacitor essentially brings it to ground so far as alternating current is concerned. (In setting up the circuit of Fig. 11·7, adjust P_1 for a voltage drop across R_2 of 1½ volts.)

7. Repeat steps 1 to 5 using the circuit of Fig. 11·8. This is a grounded-collector amplifier. The 5-μf capacitor from collector to

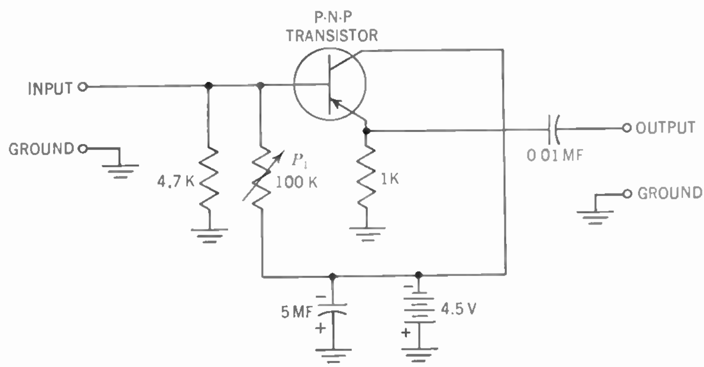


Fig. 11·8

ground places this element at a-c ground potential. (Adjust P_1 until the voltage drop across the 1,000-ohm emitter resistor is 1½ volts.)

Conclusions

1. The input and output signals of a grounded-emitter stage are 180° out of phase.

2. The input and output signals of a grounded-base stage are in phase. The same is true of a grounded-collector amplifier.

3. The grounded-base amplifier possesses a higher voltage gain than the grounded-emitter amplifier. However, the input impedance of a grounded-base amplifier is lower than it is in a grounded-emitter stage. This was shown by the greater signal drop across the series 10,000-ohm resistor. The grounded-collector stage possesses a voltage gain less than 1; input impedance is the highest of all three arrangements.

Actually, to measure the input impedance of any of these amplifiers, substitute a 100,000-ohm potentiometer for the 10,000-ohm resistor of Fig. 11·6. As an initial step, adjust this potentiometer for zero resistance (i.e., short it out of the circuit). Now feed in just enough signal for maximum undistorted output on the oscilloscope connected across the output terminals.

Note the amplitude of the signal on the scope screen. (The easiest way of doing this is by having the pattern cover a specific number of squares on a calibrated screen overlay mask.) Then gradually increase the resistance of the potentiometer until the amplitude of the output signal has been reduced to half. The resistance of the potentiometer at this point equals the input impedance of the amplifier circuit.

Experiment 5. Resistance-Capacitance and Impedance Coupling

Object. To observe a two-stage audio amplifier with *RC* and impedance coupling.

Material required

1 0.01-μf capacitor
4 5-μf electrolytic capacitors
1 50-μf electrolytic capacitor (25 volts)
2 4,700-ohm resistors ($\frac{1}{4}$ watt)
4 1,000-ohm resistors ($\frac{1}{4}$ watt)
2 100,000-ohm potentiometers
2 2N104 transistors (or equivalent)
1 interstage transformer. *Note:* Miniature interstage audio transformers designed especially for transistors are available at local parts jobbers. A suitable unit would possess a primary impedance of 20,000 ohms and a secondary impedance of 500 ohms
1 $4\frac{1}{2}$-volt battery

Test equipment

1 audio signal generator or filament transformer arrangement as described in Experiment 2
1 oscilloscope
1 vacuum-tube voltmeter

Procedure

1. Wire the circuit of Fig. 11·9. This is a two-stage *RC*-coupled transistor audio amplifier. Adjust P_1 for a voltage drop across R_3 of $1\frac{1}{2}$ volts. Adjust P_2 for the same drop across R_6.

2. Feed a very low level signal into C_1. The output is taken from C_4 and applied to the vertical input of an oscilloscope. Compute the overall gain of this system. Also compute the individual gain of each stage. This is done by measuring the amplitude of the signal at the base and collector of each transistor and then taking the ratio of collector signal to base signal.

3. Increasing the signal input will show clipping of the positive and negative cycles. Do not increase too much, since it is possible to damage the transistor with too strong a signal input.

4. If a microphone is available, feed its output into the amplifier instead of a signal generator. A considerably amplified version will be seen at the output.

5. Feed the signal generator into the input and note the maximum undistorted output. Normally this is about 1 volt rms, but it will vary with the type of transistor used.

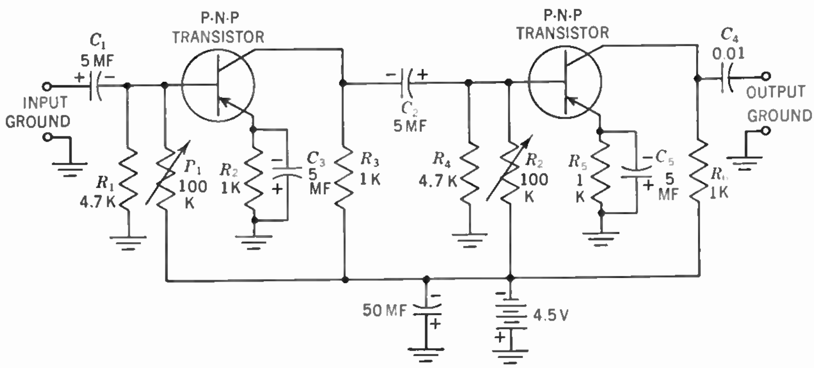


Fig. 11·9

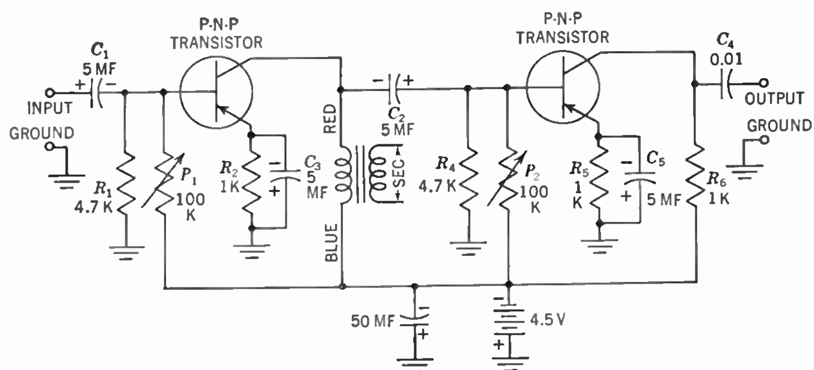


Fig. 11·10

6. Change to the circuit of Fig. 11·10 by removing R_3 and substituting the interstage transformer primary. This is impedance coupling. The blue lead goes to B— and the red lead to the collector. Do not connect the secondary. Notice how the signal clips about the same level. The gain is also the same. Adjust P_2 for a voltage drop of 1½ volts across R_6. Adjust P_1 for the same drop across R_2. (What we are doing here is setting each transistor to an operating point of 1½ ma.)

7. Remove the transformer, put back R_3, and remove R_6. Connect the primary of the transformer in place of R_6. The blue lead of the transformer connects to B— and the red lead to the collector, Fig. 11·11. Readjust P_1 and P_2 for 1½ volts across R_2 and R_5. Lower the input signal until maximum undistorted output is obtained. The gain will be found to have increased by 5 to 10 times.

What we have just seen may be summarized as follows: In Fig. 11·9, we had conventional resistance coupling. In Fig. 11·10, although the collector voltage of the first stage was increased, the output could not handle a larger signal than the resistance coupling of Fig. 11·9. In Fig. 11·11, however, the output voltage was much greater, since the

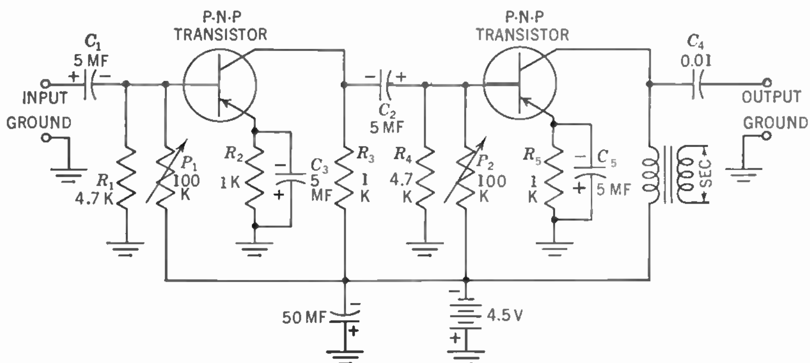


Fig. 11·11

collector voltage of the second stage was higher. It could, therefore, handle a larger signal from the first stage and also develop a greater output voltage. This indicates that higher outputs require higher collector voltages—a simple achievement with impedance coupling, since there is very little voltage drop through the low d-c resistance of the transformer primary. The a-c resistance, however, is quite high.

Conclusions

1. An *RC* amplifier has considerable gain.

2. Impedance coupling increases the collector voltage and therefore increases the gain and voltage-handling capacity of the stage. It is most valuable in the final stage where it is not hampered by any succeeding amplifiers.

Experiment 6. Audio-frequency Amplifier with Transformer Coupling

Object. To observe the performance of a transformer-coupled amplifier.

Material required

2 2N104 transistors (or equivalent)
4 5-μf electrolytic capacitors
1 50-μf capacitor
1 0.01-μf capacitor
2 4,700-ohm resistors ($\frac{1}{4}$ watt)
4 1,000-ohm resistors ($\frac{1}{4}$ watt)
2 100,000-ohm potentiometers
1 interstage transformer
1 $4\frac{1}{2}$-volt battery

Test equipment

1 audio signal generator or filament transformer arrangement as described in Experiment 2
1 oscilloscope
1 vacuum-tube voltmeter

Procedure

1. Wire the circuit of Fig. 11·9 again.
2. Compute the overall gain.
3. Change to the circuit of Fig. 11·12. Keep each collector current at $1\frac{1}{2}$ ma, as discussed previously.

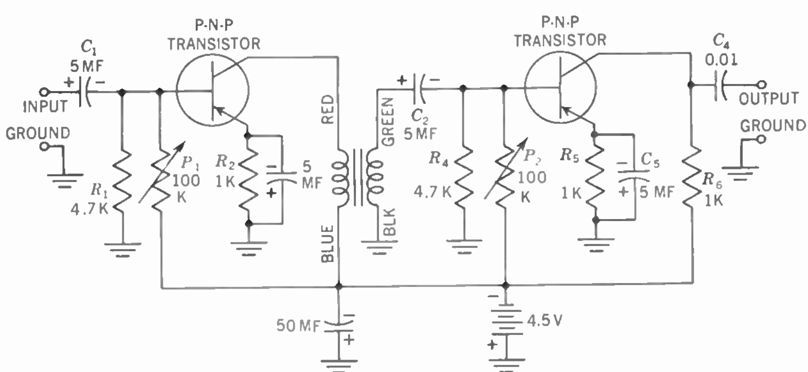


Fig. 11·12

4. Compute the overall gain.
5. An excellent example of the current-amplifying ability of the transistor can be observed by disconnecting C_2 from the transformer secondary and connecting it to the collector of the first transistor. (Be sure to reverse C_2 polarity when doing this step.) The circuit is now the same as that in Fig. 11·10. Note how the gain decreases when using impedance coupling. This is in spite of the fact that the signal

voltage across the secondary winding is very much lower than it is across the primary winding. The gain, however, increases when the input of the second stage is obtained from the secondary rather than from the larger primary. This is opposite to the action of interstage transformers in vacuum-tube amplifiers. The vacuum tube, of course, operates primarily on a voltage drive, while the transistor operates on a current drive. And the secondary of the transformer possesses a greater current than the primary.

Conclusions

1. Transformer coupling increases gain considerably by impedance matching.
2. The transistor is a current-amplifying device.
3. The transistor, in the grounded-emitter connection, has a much higher output impedance than input impedance. This can be more firmly established by rewiring the transformer in the circuit so that the secondary now goes to the collector and B— of the preceding stage while the primary connects to the base and ground of the following stage. Note the effect on distortion and overall gain.

If the reader possesses an audio generator, he might try checking the frequency responses of the *RC*-coupled, impedance-coupled, and transformer-coupled amplifiers. Look particularly for the points where the low-frequency response begins to fall off and where the high-frequency response begins to drop, and also note if the curve is flat between these two end frequencies. If desired, the reader might plot these responses on semilog paper.

Experiment 7. Complete A-F Amplifier and Supplementary Experiments

Object. To analyze audio amplifiers further.

Material required

3 2N104 transistors (or equivalent)
4 5-μf electrolytic capacitors
1 50-μf electrolytic capacitor
1 interstage transformer
2 1,000-ohm resistors ($\frac{1}{4}$ watt)
2 4,700-ohm resistors ($\frac{1}{4}$ watt)
2 100,000-ohm potentiometers
1 1-megohm potentiometer
1 speaker and output transformer
1 phonograph input
1 $4\frac{1}{2}$-volt battery
1 15-volt battery

Test equipment. None.

Procedure

1. Wire the circuit of Fig. 11·13. Adjust P_1 and P_2 for collector currents of $1\frac{1}{2}$ ma through each transistor. These will produce voltage drops of $1\frac{1}{2}$ volts across each emitter resistor. This is the same as Fig. 11·12 except that a loudspeaker and its output transformer have been added. High-impedance magnetic earphones (phones with low

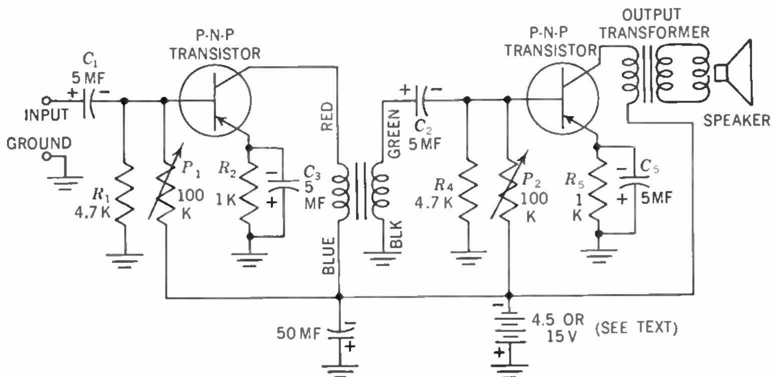


Fig. 11·13

d-c resistance) may be used in place of the loudspeaker and output transformer.

2. A phonograph pickup is fed into the input. If possible, use a pickup with a high output voltage (i.e., a crystal pickup).

3. By changing to a 15-volt battery, a much greater output can be obtained. If a 15-volt battery is employed in the circuit shown in Fig. 11·13, adjust P_1 and P_2 for collector-emitter voltages of $7\frac{1}{2}$ volts. The operating point will now be in the middle of the load line and the output signal is able to swing equally above and below this point.

4. Further output power may be developed by decreasing the values of emitter resistors R_2 and R_5 to 100 ohms each. Additional power output can be obtained by removing all base resistors and connecting a potentiometer (1 megohm) from the base of each transistor to B—. The collector current can then be adjusted for maximum output.

5. It is startling to measure the total current used with 15 volts as a supply. The drain is only about 4 ma, and yet enough volume is produced to serve an average living room. If more gain is desired, a phonograph step-down transformer may be used to match the relatively high phonograph impedance to the low input impedance of the transistor. The circuit is shown in Fig. 11·14.

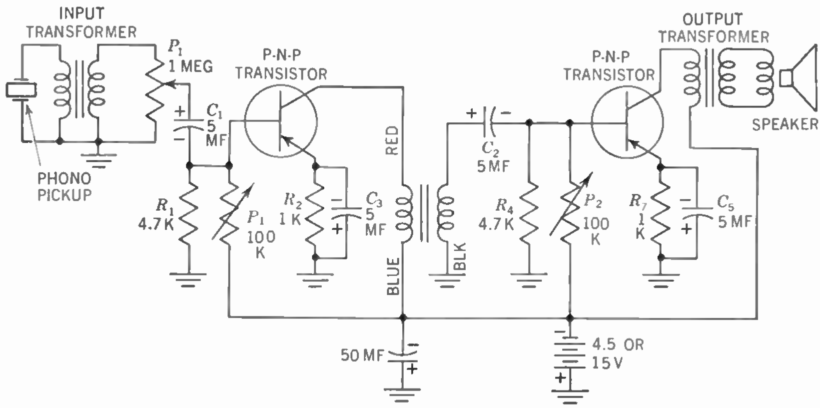


Fig. 11·14

6. An alternate circuit, Fig. 11·15, contains a class A push-pull output arrangement. The second-stage emitter bypass capacitor is omitted because the emitter currents are 180° out of phase, eliminating the need for a capacitor. (*Note:* This could not be done for class B push-

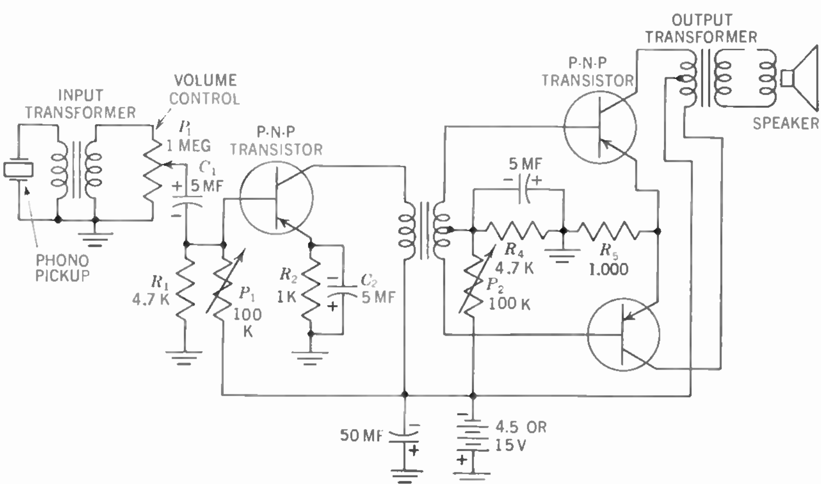


Fig. 11·15

pull output stages.) This circuit has the conventional advantage of push-pull operation. A 1-megohm potentiometer controls volume.

Experiment 8. Transistor Signal Tracer and Elementary Radio Receiver

Object. To construct a signal tracer and then tune it, thereby converting it to a tuned signal tracer or a trf radio receiver.

Material required

2 2N104 transistors (or equivalent)
1 loop antenna (with ferrite core)
4 5-μf electrolytic capacitors
2 50-μf electrolytic capacitors
1 0.01-μf capacitor
2 1,000-ohm resistors (¼ watt)
2 4,700-ohm resistors (¼ watt)
2 10,000-ohm resistors (¼ watt)
2 100,000-ohm potentiometers
1 interstage transformer
1 0.0015-μf capacitor
1 1N34-A or equivalent germanium crystal diode
1 4½-volt battery
1 15-volt battery
1 speaker and output transformer or earphone
50 ft of antenna wire
1 365-$\mu\mu$f midget variable tuning capacitor

Test equipment. None.

Procedure

1. Wire the circuit of Fig. 11·16. Adjust P_1 and P_2 for collector currents of 1½ ma through each transistor. Earphones may be substituted for the loudspeaker and its output transformer.

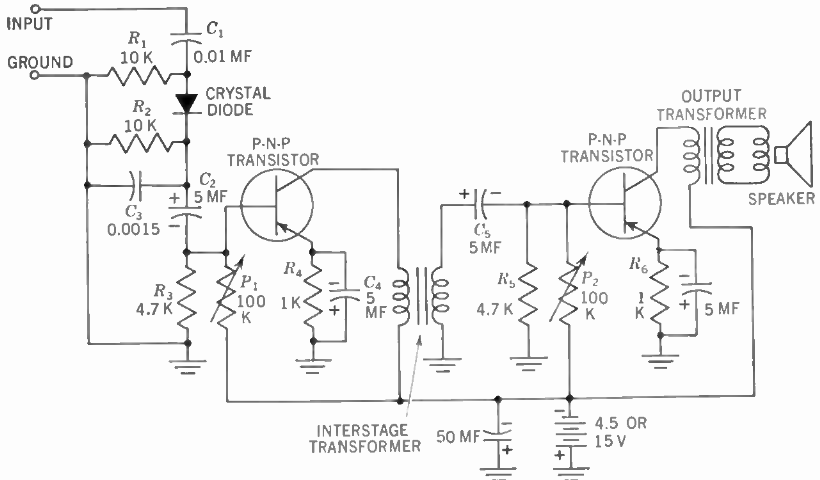


Fig. 11·16

2. Connect a short length of wire to the input capacitor and another length of wire to ground. This system may then be employed as a signal tracer in practically any circuit from audio through TV by connecting the ground lead to the chassis of the device being tested and probing with the other lead. This instrument can be used in any application where the well-known vacuum-tube signal tracer can be employed.

Louder signals can be obtained from the transistor tracer by using a 15-volt battery.

3. The signal tracer may be tuned by the addition of a loop antenna and a variable capacitor as shown in Fig. 11·17. An antenna 50 ft in

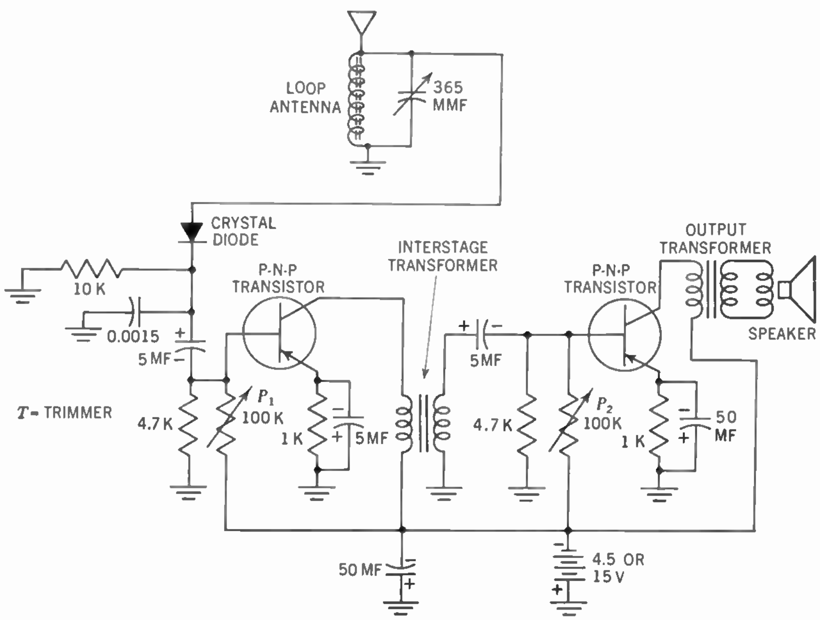


Fig. 11·17

length should be used. The chassis of the transistor receiver should be grounded to a radiator or water pipe. The importance of a good antenna and ground, particularly in an area somewhat remote from high-power broadcast stations, cannot be overemphasized.

Conclusions

1. The transistor audio amplifier may be converted to a signal tracer by the addition of a germanium diode.
2. Tuning the signal tracer converts it to a trf receiver.

Experiment 9. Relaxation Oscillator

Object. To analyze the mode of operation and the waveforms in a relaxation oscillator.

Material required

2 2N104 transistors (or equivalent)
2 10,000-ohm resistors (¼ watt)
2 1,000-ohm resistors (¼ watt)
2 0.01-μf capacitors
2 100,000-ohm potentiometers
3 5-μf electrolytic capacitors
1 0.0015-μf capacitor
1 4½-volt battery

Test equipment

1 oscilloscope

Procedure

1. Wire the circuit of Fig. 11·18. Adjust P_1 and P_2 for a collector-to-emitter voltage of 2 volts in each transistor.

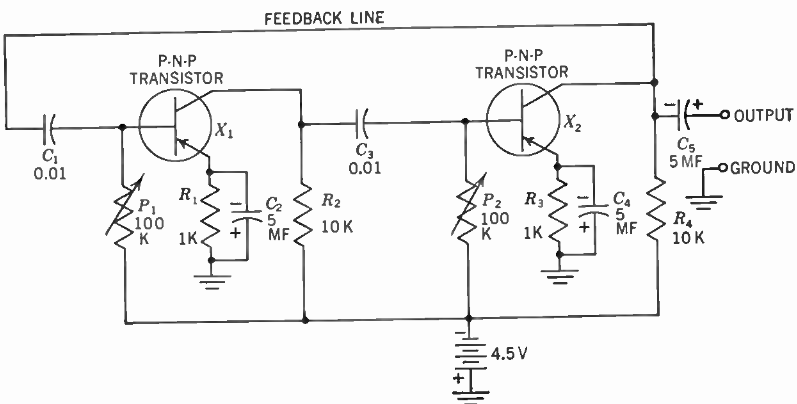


Fig. 11·18

2. When connected, the circuit will oscillate, producing the non-symmetrical waveforms shown in Fig. 11·19.

3. When observing waveforms with an oscilloscope, it is suggested that external synchronization be used. This may be accomplished by attaching a 0.0015-μf capacitor in series with a lead from the external sync terminal of the oscilloscope. The other end of this lead should be

connected to the collector of transistor 2. The vertical input probe of the oscilloscope may be used for observing the voltage waveform at various points in the circuit. If external synchronization is not used, the waveform will shift as the vertical lead is moved from one point to another.

4. To analyze the operation of the circuit, let us assume that the base of X_1 goes positive momentarily. This decreases the collector current in X_1, causing the voltage at the collector (and the top end of R_2)

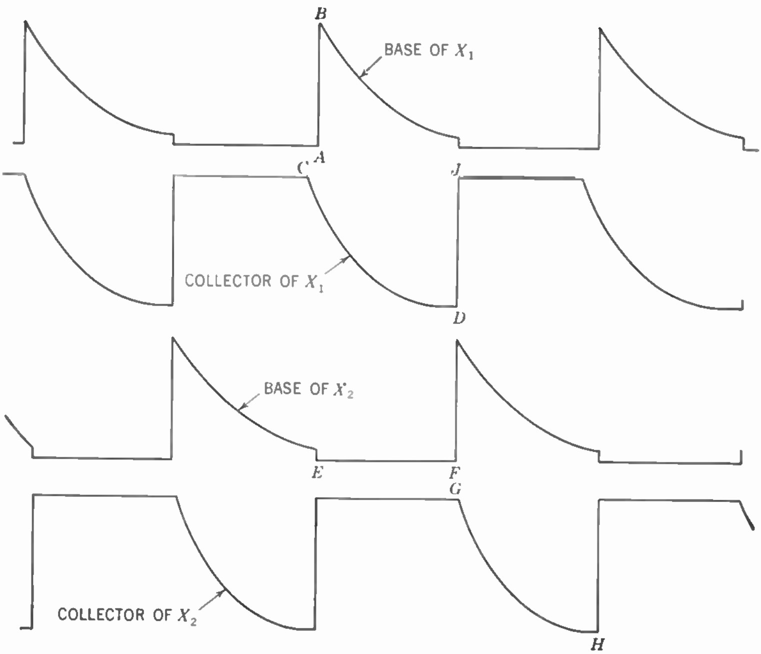


Fig. 11·19

to become more negative than it was before. This applies a negative voltage to the base of X_2. Since this acts to increase the forward bias between base and emitter of X_2, the collector current of X_2 will increase and the voltage at the collector will become less negative or more positive. This signal makes the base of X_1 go more positive, thereby aiding the initial positive signal. This is the fundamental requirement of oscillation—positive feedback with a gain greater than 1.

The buildup will continue with the current through X_1 decreasing and the current through X_2 increasing until X_1 is cut off. When the latter condition is reached, the base of X_2 no longer receives any driving signal from the collector of X_1 and the current through X_2 starts decreasing. This brings a negative signal to the base of X_1, gradually bringing this transistor out of cutoff. When X_1 starts conducting, the base of X_2 receives a positive-going signal from X_1, and its current is further decreased. Through this process, the current is built up in X_1 and reduced in X_2 until X_2 becomes nonconducting. In this fashion, control is passed back and forth between X_1 and X_2.

5. Waveform analysis. As the base of X_1 goes positive (points A and B of Fig. 11·19), it cuts off its collector current. Capacitor C_3 therefore starts charging through R_2. The charging is shown by points C and D. Since capacitor C_3 is charging at a constant rate, the voltage at the base of X_2 is maintained fairly constant (points E and F). When the charging of capacitor C_3 starts to round off slightly (approaching time-constant value), the base voltage of X_2 starts to decrease (go less negative, point F). This causes the collector of X_2 to go more negative (points G and H). When the collector of X_2 goes more negative, it drives the base of X_1 negative, which in turn makes the collector of X_1 go positive (see points D and J). The action is now the same as described previously for the base of X_1.

6. The frequency of oscillation may be changed by varying the coupling capacitors. The frequency may also be altered by varying the base resistors, but this may prove to be somewhat difficult because it will result in a change in the d-c base voltages.

7. By placing a potentiometer in the lead between the collector of X_2 and C_1, a square-wave output may be obtained.

Conclusions

1. A transistor multivibrator is feasible.

2. The frequency of oscillation is determined by the time constant of the RC circuits. In the circuit shown, it is easier to vary the frequency by varying the coupling capacitors than the resistors.

3. Square and sawtooth waves can be developed by this oscillator.

Experiment 10. Blocking Oscillator and Sawtooth Generator

Object. To study a blocking oscillator and sawtooth generator.

Material required

1 2N104 transistor (or equivalent)
1 1,000-ohm resistor (¼ watt)
1 100,000-ohm potentiometer
1 0.1-μf capacitor
1 5-μf electrolytic capacitor
1 transformer (use interstage transformer)
1 4½-volt battery

Test equipment

1 oscilloscope

Procedure

1. Wire the circuit of Fig. 11·20. Adjust P_1 for a 0.5-volt drop across R_3.

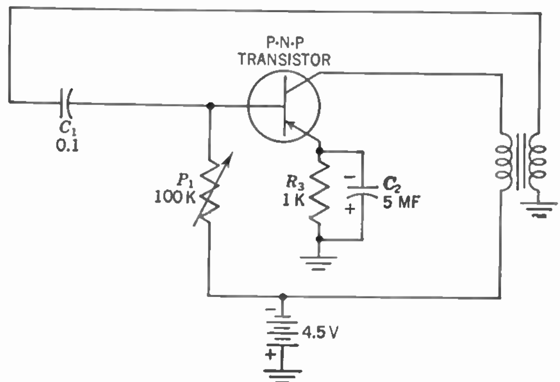


Fig. 11·20

2. Observe the waveforms at the base, collector, and secondary of the transformer. (Use external synchronization from the secondary of the transformer as described in Experiment 9.)

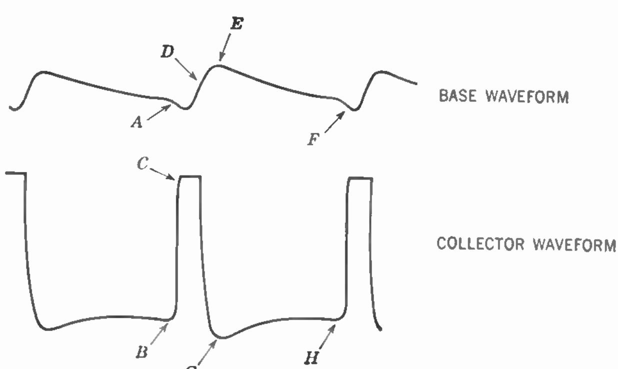


Fig. 11·21

3. To analyze the operation, at point A, Fig. 11·21, the base starts to go negative, and this drives the collector from point B to point C. At point C the collector current can no longer increase. This removes the feedback which the transformer has been supplying to the base.

The base, however, does not reach its most negative point until capacitor C_1 has been fully charged. Once the charging rate of C_1 ceases, the voltage across the base potentiometer P_1 starts to drop, becoming less negative, or more positive. This drives the collector more negative, which, because of the phase reversal of the transformer, causes the base to go more positive (point D). The gradual slope in the base and collector curves (points E and F and G and H, respectively) is the time constant of the capacitor and resistor.

4. If a capacitor is now placed from the collector to ground, a sawtooth wave will be obtained. Note the similarity between this and the blocking oscillator used in a television set. The frequency may be changed by varying C_1 or R_3.

5. It is suggested that two 10,000-ohm resistors be placed in parallel across the green and black leads of the transformer in waveform experiments. This will eliminate any transformer ringing, if present. By changing the value of this loading resistor you will see how various waveforms may be developed.

Conclusions

1. A transistor may be employed as a blocking oscillator.
2. The addition of a capacitor from the collector to ground converts the blocking oscillator to a sawtooth wave generator.
3. The frequency of oscillation is determined by the time constant of the circuit if the transistor has a high enough cutoff frequency. If it does not, then the transistor will determine the oscillation frequency.

Experiment 11. A Transistor Colpitts Oscillator

Object. To construct a variable-frequency Colpitts r-f oscillator

Material required

1 2N104 transistor (or equivalent)
2 0.01-μf capacitors
2 47-$\mu\mu$f capacitors
1 1,000-ohm resistor (½ watt)
1 coil, which may be the antenna coil of Experiment 8
1 100,000-ohm potentiometer
1 4½-volt battery

Test equipment

Oscilloscope or radio receiver

Procedure

1. Wire the circuit of Fig. 11·22. Adjust P_1 for a collector-to-emitter voltage of 2 volts.

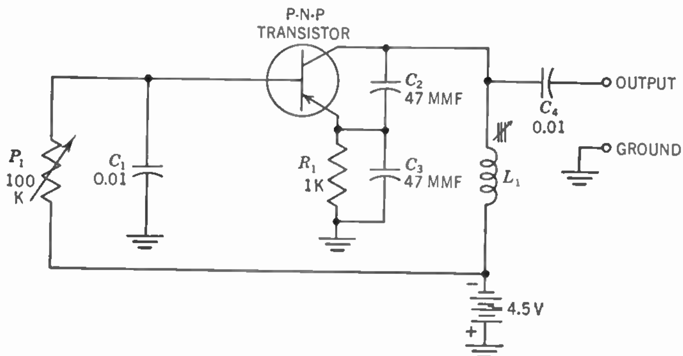


Fig. 11·22

2. Oscillation should be observed by connecting the vertical lead from the oscilloscope to the collector output capacitor C_4. The other lead from the oscilloscope goes to ground.
3. Vary the frequency of oscillation by varying the slug in the coil. The frequency can also be varied by varying capacitor C_2, C_3, or both.
4. Touch the vertical lead of the oscilloscope and note the change in oscillator frequency. This is because there is no intervening buffer stage.
5. The frequency of oscillation will probably not go much beyond 1 or 2 Mc with the transistor specified. However, with a high-frequency transistor, operation to 100 Mc could be obtained.
6. Check the amplitude of oscillations as the frequency is raised.
7. Try to reach the highest operating frequency by continually reducing the inductance of the tuning coil. Check the value of this frequency with the two transistors you have.

Conclusions

1. A Colpitts oscillator can be readily formed with a transistor.
2. The circuit is quite sensitive to "hand-capacitance" effects.
3. Oscillations tend to become weaker as the frequency is raised.
4. The highest operating frequency will tend to vary among similar transistors.

Experiment 12. Colpitts Oscillator and Buffer Using a Grounded-collector Connection

Object. To construct a buffer Colpitts oscillator arrangement.

Material required

2 2N104 transistors (or equivalent)
2 0.01-μf capacitors
2 47-$\mu\mu$f capacitors
1 coil, as in Experiment 11
1 3,300-ohm resistor
1 1,000-ohm resistor
2 100,000-ohm potentiometers
1 4½-volt battery

Test equipment

1 oscilloscope or radio receiver

Procedure

1. Wire the circuit shown in Fig. 11·23. Adjust P_1 for a collector-to-emitter voltage of 2 volts.

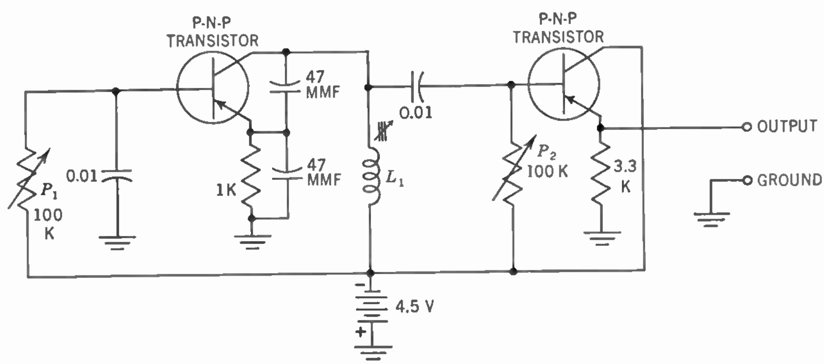


Fig. 11·23

2. The output is taken here from the emitter of the second transistor. To observe the waveform of the generated signal, connect the vertical input lead of an oscilloscope to the output terminal. Use a capacitor in series with the oscilloscope lead.

3. Observe the frequency of the waveform and then touch the output lead. The level of the signal may decrease, but its frequency remains essentially unchanged.

4. The grounded-collector stage acts very much as a vacuum-tube grounded-plate (i.e., cathode-follower) circuit. There is no voltage gain through the grounded-collector circuit, but there is a definite

buffer action. This stems from the change in resistance levels. The output is now a very low impedance and thus hand-capacitance effects will not be noted.

Conclusions

1. It is desirable to employ a buffer stage when operating transistor oscillators that may be subject to varying load conditions. The buffer here is a grounded-collector circuit. This is a good buffer circuit because load changes have less effect on the buffer input.

Experiment 13. Transistor Characteristic Curves

Object. To obtain the characteristic curves of transistors on the screen of an oscilloscope.

Material required

1 2N104 transistor (or equivalent)
1 100-ohm resistor (½ watt)
1 25-ma silicon rectifier
1 12-volt filament transformer (1-amp rating)
1 1½-volt battery
1 50,000-ohm potentiometer

Test equipment

1 oscilloscope
1 vacuum-tube voltmeter or VOM

Procedure

1. Wire the circuit shown in Fig. 11·24. The leads to the horizontal terminals of the oscilloscope provide a horizontal sweep which is di-

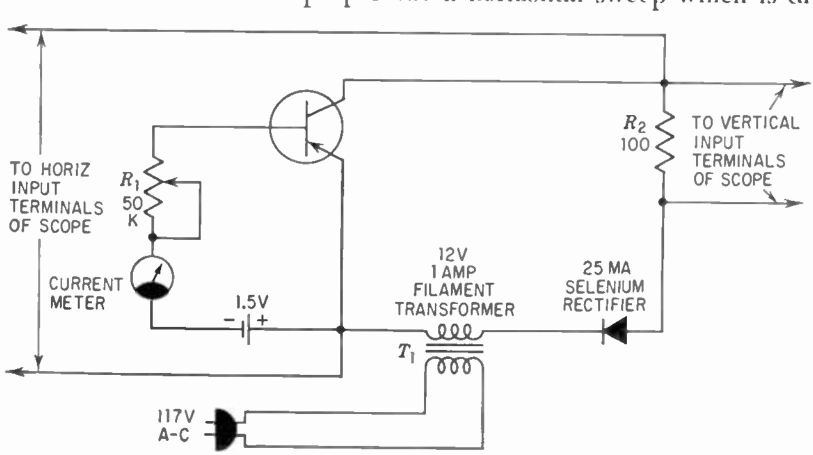


Fig. 11·24

rectly proportional to variations in collector voltage. The voltage across R_2 is applied to the vertical input terminals of the oscilloscope, giving a vertical deflection which is directly proportional to the variations in collector current. In this way, we obtain a collector-voltage–collector-current characteristic curve. Note that the internal sweep of the oscilloscope is not used; the horizontal amplifier is switched to "horizontal input."

2. The transistor is connected in Fig. 11·24 as a grounded-emitter amplifier. With this arrangement, a series of curves are desired for different base currents. (See Chap. 3, where typical transistor characteristic curves are shown.) To start, disconnect R_1, making the base current zero. Then, the vertical-gain and horizontal-gain controls of the oscilloscope are adjusted until the desired image size is obtained, Fig. 11·25. (Do not advance the vertical-gain control too far. The

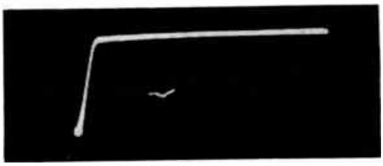


Fig. 11·25 Form of a single characteristic curve of a transistor.

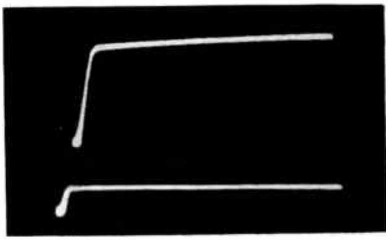


Fig. 11·26 Characteristic curves obtained with base current (above) of 100 μa and (below) of 20 μa.

curve obtained when the base current is zero possesses the lowest vertical amplitude. If this curve is made too high, some of the curves to follow will fall beyond the top of the screen.)

3. For each different value of base current, a different characteristic curve is obtained, Fig. 11·26. The base current is varied by adjusting potentiometer R_1.

If desired, a transparent scale can be placed over the scope screen and the vertical-gain and horizontal-gain controls then adjusted for a known amount of gain. Here is how to do this. Apply a known a-c voltage first to the vertical-input and then to the horizontal-input terminals of the scope and adjust the respective vertical-gain and horizontal-gain controls for a given deflection. For example, an a-c voltage having a peak-to-peak amplitude of 6 volts is applied to the horizontal-input terminals of the scope and the horizontal-gain control is adjusted for a deflection of 3 in. The horizontal sensitivity then becomes equal to 2 volts per in. This enables you to measure the collector voltage at any point on a traced-out curve by measuring its distance horizontally along the screen.

For the vertical-input terminals, a smaller voltage, say 0.6 volt peak to peak, is recommended. Then, if the vertical-gain control is adjusted for a deflection of 3 in., the vertical sensitivity becomes 0.2 volt per in. Since the vertical axis of the curves represents current, these voltage values must be converted into equivalent current values. This is readily accomplished, since whatever voltage is applied to the vertical-input terminals of the scope must come from the 100-ohm resistor of Fig. 11·24. Thus, if the transparent scale over the scope screen indicates a vertical amplitude of 0.4 volt, then

$$E = IR$$
$$0.4 = I \times 100$$
$$I = \frac{0.4}{100} = 0.004 \text{ amp} = 4 \text{ ma}$$

This is the collector current at the point where the above 0.4-volt amplitude was measured.

Note that once the oscilloscope gain controls are set, they are not touched again; otherwise, the calibration is disturbed.

4. The circuit shown in Fig. 11·24 is for a PNP transistor. To obtain the characteristic curve for an NPN transistor, reverse the connections to the battery and to the selenium rectifier.

5. To obtain the characteristic curves for a grounded-base arrangement, the connections to the base and emitter of Fig. 11·24 would be interchanged. Battery polarity, too, would be altered accordingly. Then the curve for a number of emitter current values would be obtained.

6. One curve of a family will be obtained for one value of base (or emitter) current. To obtain a composite or family group of curves, a series of successive photos would have to be taken, Fig. 11·27.

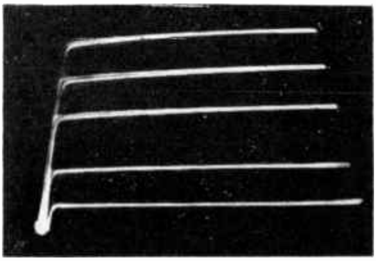


Fig. 11·27 Family of collector-current vs. collector-voltage curves for a 2N104 transistor at base currents of 25, 50, 150, and 200 μa (bottom to top).

7. If greater changes in collector voltage are desired, substitute another transformer for T_1 which will provide a greater secondary voltage. For example, a 25-volt transformer may be used here.

CHAPTER 12

Transistor-amplifier Design

THE TRANSISTOR, in order to be useful as an amplifier, must be employed in a circuit which is so designed that the transistor is operated over its most linear range. In this way, minimum distortion will be introduced in the signal passing through the stage. Another precaution that must be observed is to maintain the transistor within its heat-dissipation limits. Failure to observe these boundaries will lead to the destruction of the unit just as surely as a tube is destroyed when it is operated with abnormal voltages or currents. In the case of a transistor, there is far less leeway, necessitating even greater care in design.

Selection of a suitable operating point is one aspect of amplifier design, and it is governed almost completely by the d-c voltages applied to the transistor. The second aspect of amplifier design is the response of the stage to the applied signals. This is governed in part by the electrical characteristics of the amplifying device itself, in this case, the transistor, and in part by the characteristics of the coils, capacitors, and resistors attached to the transistor. It is these items which will frequently limit the bandpass of a circuit, at least for the low and medium frequencies. Beyond this, both the transistor and the circuit share the responsibility of degrading the frequency response.

In this chapter we shall analyze the problem of transistor-amplifier design from each viewpoint separately. Not only is this the conventional approach, but it is also an eminently practical one since it enables the circuit designer to deal with each variable separately. Both the a-c gain and the frequency response are functions of the operating point, however. This is possible because we can quite accurately represent the transistor by an equivalent circuit under small-signal conditions. Furthermore, the external circuit components (provided they are linear) will not affect the distortion produced by the transistor.

Selecting the Operating Point

The operating point for the transistor is established by the d-c voltages chosen for the base, emitter, and collector elements. In this respect, the transistor is identical to the vacuum tube; as a matter of fact, many of the steps followed in arriving at a bias point are similar to those employed in establishing the operating point for a tube. The biasing methods, however, will be frequently quite different because of the inherent differences between the two devices.

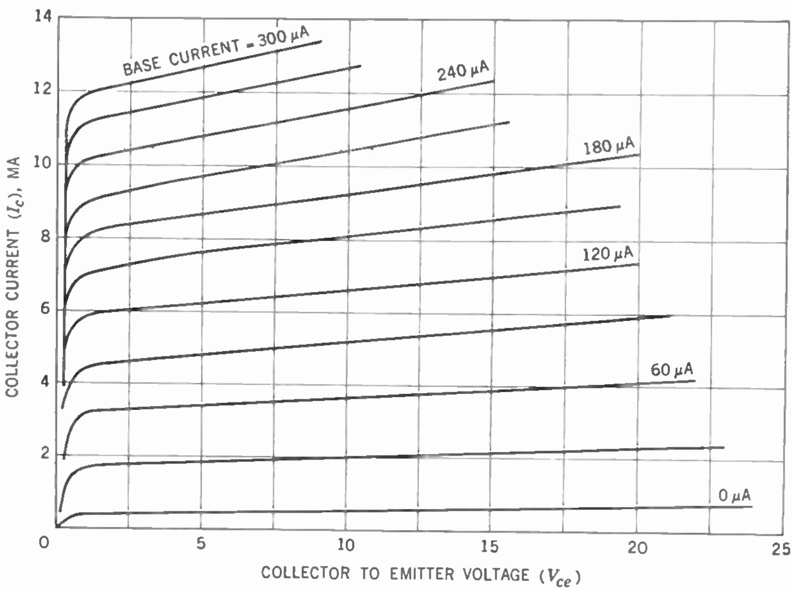


Fig. 12·1 The characteristic curves for a transistor connected in the common-emitter configuration.

The first items needed to select the proper bias point are a listing of the transistor's characteristics and a graph of its current and voltage behavior for the particular circuit arrangement chosen. In practically all instances, this means with the transistor connected as a grounded emitter, i.e., with the input signal fed into the base and the output signal taken from the collector.

The curves most useful for conventional design purposes show the variation in collector current for different collector voltages and with different base currents. A representative family of such curves is shown in Fig. 12·1.

Note specifically that these should be for the grounded-, or common-, emitter configuration. Frequently, curves are shown for the common-base arrangement, and the two are very similar in appearance. Both the listed and the graphical characteristics contain important information, and both are required to properly select a d-c working point. As a start, let us examine the listed characteristics to see what information they can provide that will help us choose a suitable d-c operating point for the transistor in question.

At the top of Table 12·1 the absolute maximum ratings are given. These include element voltages, collector current, collector dissipation, and junction temperature. Of specific interest is the collector-to-emitter voltage, because we are dealing here with a common-emitter configuration. The stated limit is —30 volts. (The minus sign simply means that the collector voltage is negative with respect to the emitter, this being a PNP transistor. For an NPN unit, the same voltage values would be given as positive.) Whatever d-c voltages are applied to the collector and emitter, their *difference* must not exceed 30 volts, or a voltage breakdown between these elements may occur. The word "difference" is important because the individual element voltages with respect to a common connection point such as ground may be much higher than 30 volts. However, it is only the voltage *between* emitter and collector that counts.

Table 12·1 Maximum Values and Average Characteristics

Ratings—Absolute Maximum Values	
Collector-to-emitter voltage	−30 volts
Collector-to-base voltage	−45 volts
Collector current	10 ma
Collector dissipation	60 mw
Junction temperature (maximum recommended operating temperature)	85°F
Average Characteristics—Design Center	
Collector voltage	−6.0 volts
Emitter current	1.0 ma
Current amplification	0.98
Collector-current cutoff I_{CO}	10 μa
Output capacitance	40 μμf
Frequency cutoff, f_{CO}	1.0 Mc
Maximum power gain	40 db

Collector current I_c possesses a maximum figure of 10 ma. If, now, we multiply this collector current by the maximum collector-to-emitter voltage (0.01×30), we obtain a power value of 0.3 watt. We might suppose that this is the maximum transistor dissipation. Just below the collector-current figure, however, maximum dissipation is given as 60 mw, or 0.060 watt. How is this possible?

The answer, of course, is that maximum collector current and maximum collector voltage do not occur at the same time. If they did, the transistor would be destroyed in short order. What the values do mean is that, within the limits of a transistor dissipation of 60 mw, the collector voltage may go as high as 30 volts (with respect to the emitter) or that the collector current may go as high as 10 ma. But the 60-mw value is the figure to watch and is the limiting factor on the choice of a suitable d-c operating point. More on this in a moment.

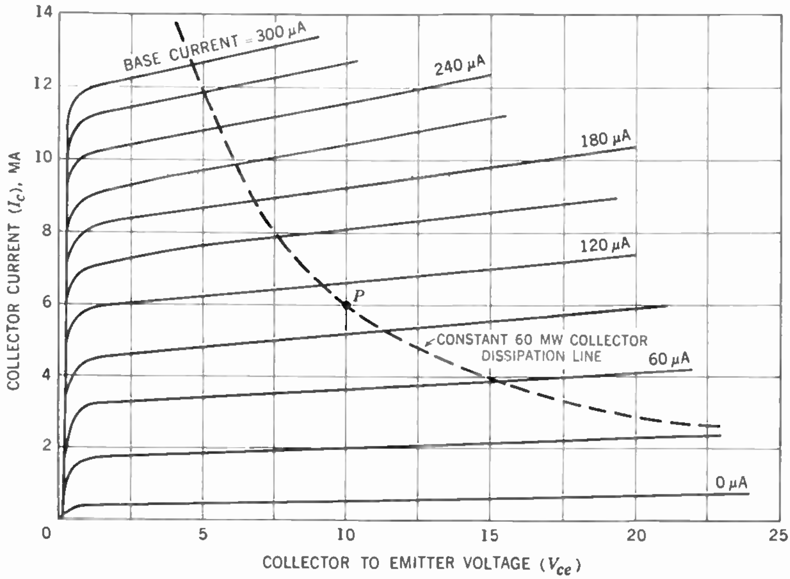


Fig. 12·2 The dotted curve indicates the maximum collector dissipation of the transistor represented by this set of characteristic curves.

Also shown in the characteristic listing are recommended design center values. These represent average values which will bias the transistor to a point at which it can be employed gainfully as an amplifier. Note that the recommended collector-to-emitter voltage is −6 volts and the recommended current is 1 ma. These values are far from the maximum figures quoted above.

Note, too, that with these values, collector dissipation is

$$6 \times 0.001 = 6.0 \text{ mw}$$

and this is well within the 60 mw given under the maximum ratings. In short, the operating figures provide considerable room for leeway, and this is as it should be.

An interesting way in which the maximum collector dissipation can be kept in front of the circuit designer is shown in Fig. 12·2. The dotted curve, at each point, equals 60 mw; so long as the operating condition of the transistor stays to the left of this line, the unit is within its maximum-dissipation range and no difficulty from burnout should be encountered. On the other hand, if this dotted curve is crossed, the chances are quite good that the useful life of the transistor will be materially shortened.

Some manufacturers provide these maximum-dissipation curves with the transistor characteristic graphs. When this is not done, the designer himself can draw such a curve by choosing, for each collector voltage, a value of collector current such that the product of the two equals the maximum-dissipation figure. For example, in Fig. 12·2 the point P, the collector voltage, is 10 volts. The dotted line crosses this point at 6.0 ma; the product of 10 and 6.0 ma is then 60 mw.

As a general rule, to determine the points on the dissipation curve, voltages may be selected among the horizontal axis and corresponding current values calculated ($I_C = P_C/V_C$), or current points may be selected along the vertical axis and corresponding voltage values calculated ($V_C = P_C/I_C$).

Load Lines

A collection of curves, such as that shown in Fig. 12·1, represents the behavior of a transistor over a wide range of collector currents and voltages and for a wide variety of base currents. When a transistor is connected with the emitter grounded, the amount of current which will flow from emitter to collector is determined by the base-to-emitter potential. Instead of specifying the latter values, which are very small and exceedingly difficult to measure, the resultant base current is given instead, since for each base-to-emitter potential a certain base current will flow. It is the latter values which are indicated for each of the curves in Fig. 12·1. As a matter of fact, base current values are more significant than voltage values because the transistor is a current-operated device. Furthermore, in the common-emitter configuration, the base is the element to which the incoming signal is applied.

Now, in order to determine how a transistor will operate with specific voltages applied to its elements, we must determine the section of the graph where this operation is to take place. For example, for the transistor represented by the curves in Fig. 12·2, we can place the operating point anywhere to the left of the maximum-dissipation line. The manufacturer recommends —6 volts, 1 ma, but we have selected —6 volts, 5 ma because it enables us to use a load resistance of 1,200 ohms, which is what we desire.

This point is identified by the letter *A* in Fig. 12·3. This is the central point of any design we may wish to attempt with the transistor; from this point, the variations of the signal current applied to the base will cause the collector voltage and current to vary above and below the conditions specified by point *A*.

With the location of the operating point, we are ready for the next step, i.e., the drawing of a load line through the operating point. This line reveals graphically what happens when a load resistor is

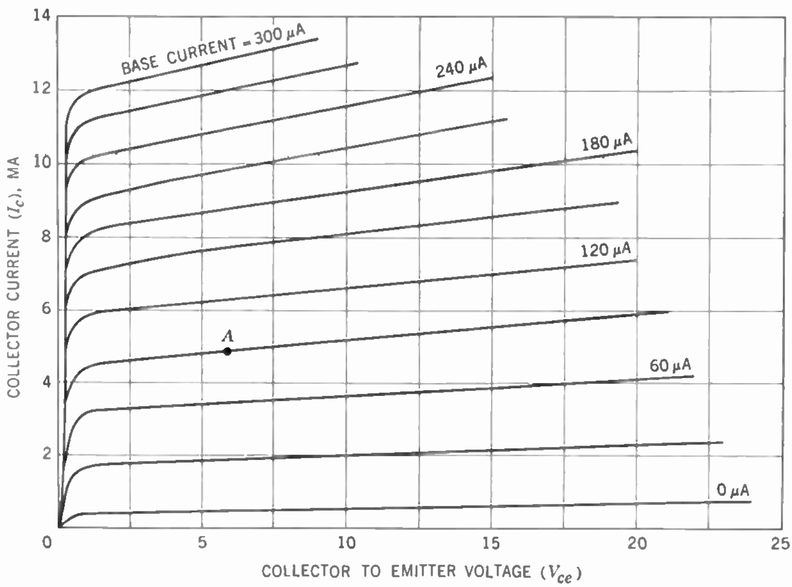


Fig. 12·3 Point A is the selected operoting point. A 1,200-ohm lood line, drown through point A, is shown in Fig. 12·4.

placed in the collector circuit and the input or base current varies. Now the question is, how do we know which line to draw? Several things are already known. First, we know that the load line must pass through point *A*. Second, it must be so drawn that it does not, at any point, cross the dotted line representing the maximum-dissipation line. These are the two most obvious restrictions on the load line. There are others which will become apparent in the ensuing discussion.

As noted above, the load line represents the load resistance and the value of this resistance will determine the slope of the line. For example, let us assume that the load resistance is 1,200 ohms. Such a line, passing through point *A*, is shown drawn in Fig. 12·4. To demonstrate that this represents 1,200 ohms, it is necessary to know that the slope of this line is given by the ratio of the change in collector voltage from minimum to maximum current and the current change itself,

$$\frac{\text{Maximum collector voltage} - \text{minimum collector voltage}}{\text{Maximum collector current} - \text{minimum collector current}}$$

with the load resistance in the circuit. From Fig. 12·4, we see that the collector voltage changes from a maximum of 12 volts when the

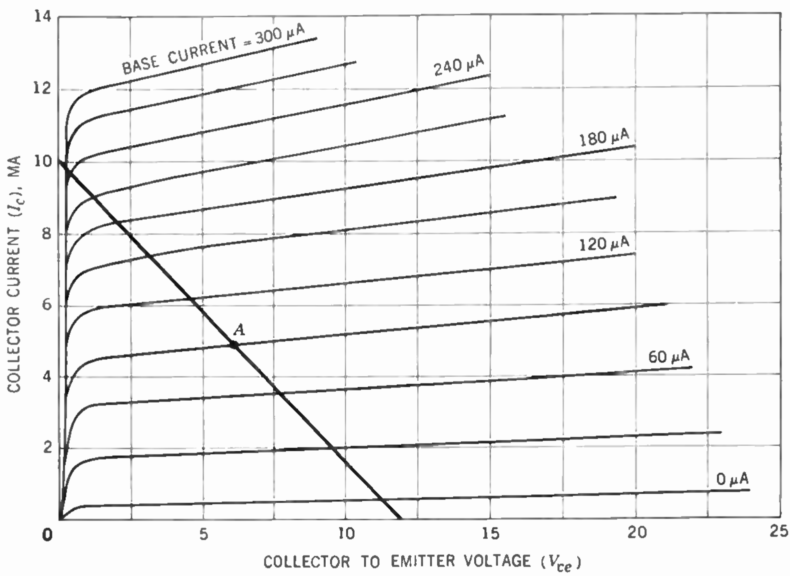


Fig. 12·4 A 1,200-ohm load line passing through point *A*. The base current, at point *A*, is 90 μa.

collector current is zero to 0 volts when the collector current is 10 ma. Substituting these facts in the above ratio, we obtain

$$\frac{12 - 0}{0.010 - 0} = \frac{12}{0.010} = 1{,}200 \text{ ohms}$$

For this value of load resistance, a battery voltage of 12 volts is needed. If only 10 volts is available and it is desired that the load line pass through point *A*, then another value of load resistance will be needed. This can be seen by actually drawing a line from 10 volts to point *A* and then extending the line out until it crosses the I_c axis. It does so at 12 ma. From the slope of this second line, the computed value of the required resistor is 833 ohms, Fig. 12·5.

If we examine the load line of Fig. 12·4, we note several things. At the operating point, the base current required is 90 μa. At the bottom end of the load line, it crosses the 0-μa base-current curve, while at the upper end of the line, the base-current curve crossed is 240 μa. Thus, the swing below point A extends only for 90 μa, while above point A it can go up 150 μa (240 − 90). This occurs because of the progressively closer spacing of the curves at the higher base-current levels because of a reduction in β. Utilization of the entire load line would

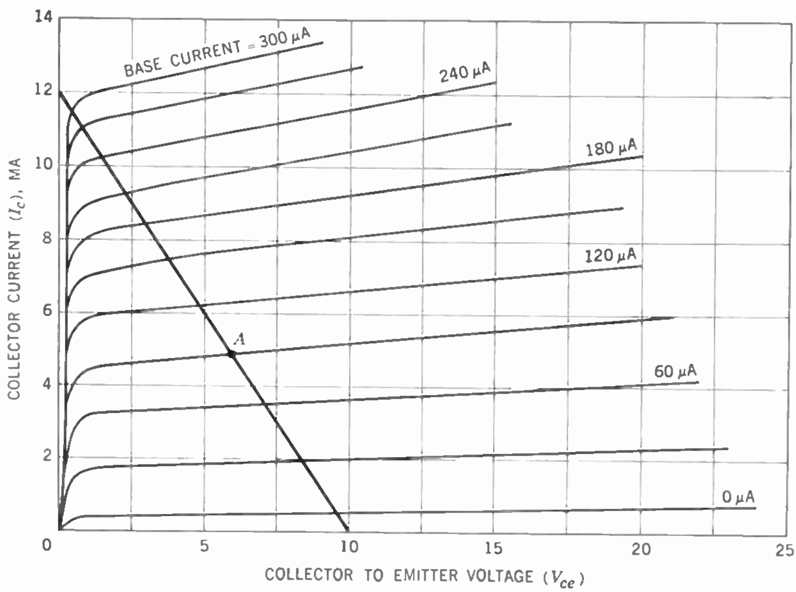


Fig. 12·5 A 833-ohm load line passing through point A.

thus lead to distortion, which is undesirable. In the present instance, it would be better to limit the swing of the signal to the region of more nearly equal base-current curve spacing, particularly when low distortion is more important than maximum output.

Other values of load resistances can be chosen to fit within the area to the left of the maximum-dissipation curve. Several are shown in Fig. 12·6, and the reader can see how the various factors influence the amount of distortion and output obtainable. For example, with the battery voltage fixed, larger and larger resistances permit less and less of an input signal swing between transistor cutoff (when the collector voltage is maximum) and maximum current (when the collector voltage is zero and the characteristic curves are highly distorted). However, with the larger load resistances, most of the swing over the load line is usable, except for the very extreme ends. With low loads, the extent of the curve is greater, but as we saw in Fig. 12·4, the entire curve cannot be used because of the distortion caused by the curves crowding together.

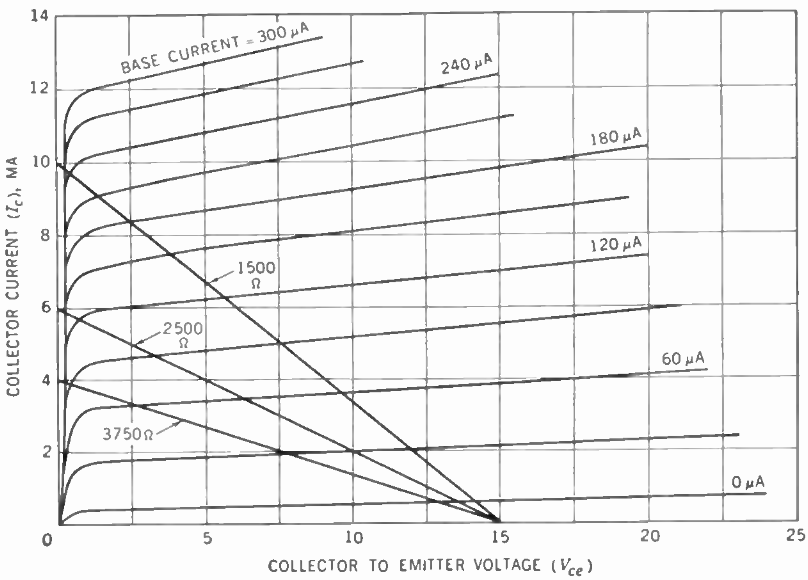


Fig. 12·6 Three different load lines using the same battery voltage.

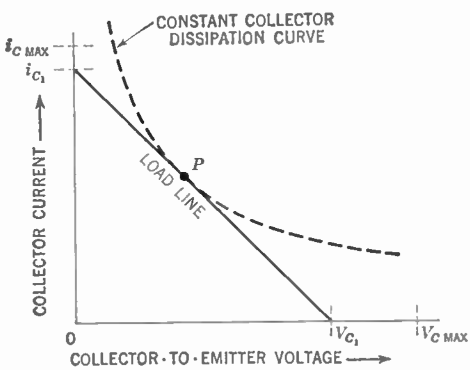


Fig. 12·7 The condition for maximum power output. The load line is tangent to the maximum collector dissipation curve at point *P*. The battery voltage is V_{c_1}, while the maximum current that flows with this load line is i_{c_1}.

If maximum power output is desired, the load line should be brought as close to the maximum dissipation line as possible, Fig. 12·7. This enables the line to encompass as wide an area as possible and this, in turn, means greater output power.

From the foregoing it is seen that there are a number of factors that determine the location of the load line, factors such as the operating point, the maximum collector dissipation permitted, the amount of battery voltage available, and the amount of distortion that can be tolerated. You start with those items that are fixed and adjust the remaining variable factors to fit the specified conditions.

Transistor Bias Circuits

Now that we have determined graphically the desired operating conditions for a transistor, let us see what we have to do circuit-wise

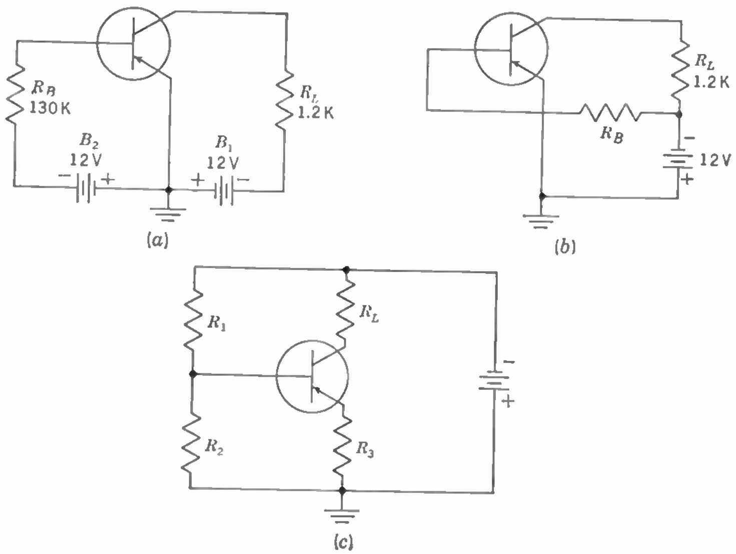


Fig. 12·8 Three different biasing methods.

in order to achieve this state. Fig. 12·8*a* is a very simple circuit which will provide the desired conditions. R_L is 1,200 ohms, B_1 is 12 volts, B_2 is also 12 volts, and R_B is 130,000 ohms. B_2 and R_B are chosen to provide a base current of 90 μa. If we assume that the voltage drop between base and emitter elements in the transistor is negligible compared to the voltage drop across R_B, then

$$R_B = \frac{E}{I_B} = \frac{12}{90 \times 10^{-6}} = 133{,}333 \text{ ohms}$$

The actual value of R_B is 133,333 ohms, but this resistance is close enough to the standard value of 130,000 ohms to permit its use.

A more economical arrangement, using only one battery, is shown in Fig. 12·8*b*. The conditions here are identical to those of Fig. 12·8*a*, and R_B is computed in exactly the same fashion. Still another circuit that is employed more widely than either of the preceding circuits is shown in Fig. 12·8*c*. The base current is supplied by the voltage-divider network of R_1 and R_2. R_L is the collector load resistor and R_3 is an emitter resistor which serves the same purpose as the cathode resistor in a vacuum tube. When this resistor is unbypassed, then it, together with R_L, receives the output signal of the transistor. That is, whatever collector current passes through R_L must also pass through R_3, and the sum of both resistors represents the total load resistance. It is this value which is used to plot the load line.

However, with R_3 unbypassed, degeneration is introduced, and this tends to reduce the gain of the stage. If the degeneration is not desired,

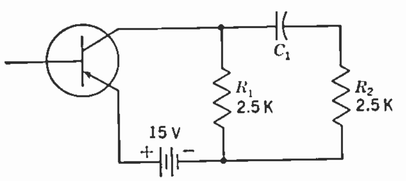


Fig. 12·9 To draw the a-c load line for this transistor, the effect of C_1 and R_2 must be considered. For the d-c load line, only R_1 need be considered.

then R_3 is bypassed. Now the question is, "Is R_3 still included in the computation of the load line?"

The answer is no if we wish to draw an a-c load line in distinction to the d-c line. The difference between these lines is this: For various fixed or d-c currents, transistor operation will take place along the load line computed using the sum of R_L and R_3. However, if a-c signals are applied to the base, then all current variations will be shunted around R_3 by the bypass capacitor, and the corresponding voltage variations will appear only across R_L. Hence, only the value of R_L would be used in computing the a-c load-line slope.

Since we are concerned ordinarily with the use of transistors to amplify a-c signals, we are more interested in the a-c load line than in the d-c load line. It is important that this distinction be recognized.

An interesting situation arises when we have an output circuit of the form shown in Fig. 12·9. If we disregard capacitor C_1 and resistor R_2, then the load line will be determined by the value of R_1 alone, in this case, 2,500 ohms. This line, *AB*, drawn with a 15-volt battery, is shown in Fig. 12·10. *Q* is the quiescent, or operating, point when the input signal is zero.

Now, attach C_1 and R_2. What an a-c signal, flowing in the collector circuit, now sees is the parallel combination of R_1 and R_2 (assuming the reactance of C_1 to be negligible with respect to R_2). Hence, a new load line should be drawn representing this new combination. This is CD in Fig. 12·10. Note that this second line is steeper than the preceding load line because the two resistors in parallel possess a total resistance which is less than R_1 alone. Further, the second load line must also pass through point Q because this is the resting position of the transistor when no signal is being received. It is this second line, too, which is carefully examined to see if it can accommodate the

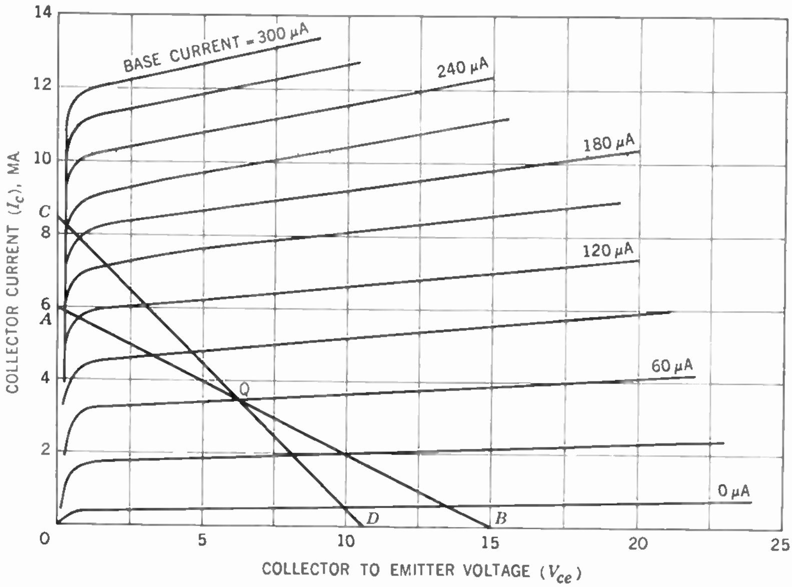


Fig. 12·10

desired range of input (i.e., base) current without distortion. And this line must not extend into regions where the power (product of collector voltage and current) exceeds the maximum power dissipation of the unit.

Determining the Best Bias Arrangement

We have just seen several methods of biasing transistor amplifier circuits in order to obtain a specific operating point and, with it, a desired mode of operation. With so many possible arrangements, it would be logical to ask: Is any one of these better than any other one? And if the answer is yes, how can we determine which circuit is best? It is toward the determination of a suitable answer to these questions that the following discussion is directed.

An important aspect of transistor behavior is its temperature dependence. Increase the temperature of a transistor and it will be found that, with all other conditions kept fixed, the collector current will increase. This rise stems from two causes: an increase in I_{CO} and a change in the input, or I_B-V_{EB}, characteristic. The collector-current rise from the latter source is important when the resistance in the base circuit is low. This occurs in stages that are transformer-coupled, where the base receives its d-c bias through the low-resistance winding of the transformer. In r-c-coupled stages, the base resistance is fairly high, and here the input characteristic change is not significant in its effect on the collector current. Only changes in I_{CO} need be considered. In the first part of the discussion to follow, we will assume fairly high base resistance and thus consider only the effect that changes in I_{CO} will have on the collector current.

I_{CO} is the current that flows through the collector-base sections when the emitter current is zero. It stems from the presence of minority carriers in the base and collector sections, and it gives rise to a small current when the collector is reverse-biased. I_{CO} is generally below 10 μa, and it is independent of emitter current or collector voltage when the latter is greater than a few tenths of a volt. Because of this it is called a saturation current. The value of I_{CO} is determined chiefly by the particular transistor being used and by the temperature. As a matter of fact, it is extremely sensitive to temperature, actually doubling in value with each 10°C rise in junction temperature. In short order, this can reach a value where it will have a disastrous effect on transistor operation.

To demonstrate this more clearly, let us determine exactly what effect a rise in I_{CO} has on the operation of a common-emitter amplifier. To start, we know that the emitter current is equal to the sum of the base and collector currents. In equation form, this is

$$I_E = I_B + I_C \tag{12·1}$$

The collector current, in turn, is composed of that portion of the emitter current I_E reaching the collector, plus the saturation current I_{CO}. Thus,

$$I_C = \alpha I_E + I_{CO} \tag{12·2}$$

Now, if we substitute Eq. (12·1) in (12·2), we obtain

$$\begin{aligned} I_C &= \alpha(I_B + I_C) + I_{CO} \\ &= \alpha I_B + \alpha I_C + I_{CO} \end{aligned}$$

or, rearranging terms,

$$I_C(1 - \alpha) = \alpha I_B + I_{CO}$$

and
$$I_C = \frac{\alpha}{1 - \alpha} I_B + \frac{I_{CO}}{1 - \alpha}$$

$$I_C = \beta I_B + (1 + \beta) I_{CO} \tag{12·3}$$

where
$$\beta = \frac{\alpha}{1 - \alpha}$$

Equation (12·3) tells us a number of things. If we ignore the I_{CO} term, it tells us that for every change in I_B, the effect on the collector current is β times as great. This is the basis for the wide popularity of the common-emitter arrangement. Since β is on the order of 30 or more, a small change in input current applied to the base will cause the collector current to change β times as much, giving us a sizable amplification or gain in signal.

Now, consider the I_{CO} factor in Eq. (12·3). As long as I_{CO} is small, even with a multiplying factor of 30 it still is considerably below the level of the normal collector current derived from the emitter. But, with I_{CO} doubling with each 10°C rise in temperature, it can quickly reach a point where the total collector current is appreciably above its normal operating value. This, in turn, will raise the wattage dissipated at the collector and cause the temperature there to increase. With a higher operating temperature, I_{CO} will increase, further raising the collector current and the collector temperature. The process is cumulative, and if no current limiting is provided, the temperature and current at the collector can reach destructive values.

By way of contrast, consider the situation when the common-base arrangement is employed. Under these conditions,

$$I_C = \alpha I_E + I_{CO}$$

A change in emitter current, caused by an input signal, will produce an α change in collector current. Note, however, that I_{CO} stands alone, and even if it doubles with each 10°C rise in temperature, it is still quite small in total value. Consequently, I_C is hardly affected at all. Thus, the common-base arrangement is less disturbed by temperature changes than the common-emitter circuit. However, because of the marked gain characteristics of common-emitter circuits, they are used more widely.

Now that the role played by temperature in transistor-circuit operation is understood, the next thing to uncover is which bias arrangement previously shown provides the best defense against such changes.

For this determination, a special stability factor S is useful. This is defined as

$$S = \frac{\Delta I_C}{\Delta I_{CO}} = \frac{dI_C}{dI_{CO}} \tag{12·4}$$

This expression reveals the change in collector current dI_C for a change in collector saturation current dI_{CO}. The lower the value of S, the more stable the arrangement because changes in I_{CO} have only a small effect on I_C. We can apply Eq. (12·4) to the two situations already discussed, i.e., common base and common emitter. For the former, we saw that

$$I_C = \alpha I_E + I_{CO}$$

from which we obtain

$$dI_C = dI_{CO}$$

or

$$\frac{dI_C}{dI_{CO}} = 1$$

For common-emitter configurations,

$$I_C = \beta I_B + (1 + \beta)I_{CO}$$

and

$$dI_C = (1 + \beta)dI_{CO}$$

or

$$\frac{dI_C}{dI_{CO}} = 1 + \beta \tag{12·5}$$

Obviously, the latter result possesses a higher value than the former, corroborating the conclusion reached previously, i.e., that the common-base arrangement is less affected by changes in I_{CO} than the common-emitter arrangement.

Note: For those readers who are not familiar with the calculus and differentiation, the following modified procedure will serve equally well. To the original expression

$$I_C = \alpha I_E + I_{CO} \tag{A}$$

add a small increment ΔI_C to the left-hand side and a small increment ΔI_{CO} to the right-hand side. This is the mathematical way of stating that increasing I_{CO} by ΔI_{CO} will cause I_C to increase by ΔI_C.

Making the addition just indicated, we obtain

$$I_C + \Delta I_C = \alpha I_E + I_{CO} + \Delta I_{CO} \tag{B}$$

Now, substitute Eq. (A) for I_C in Eq. (B):

$$\alpha I_E + I_{CO} + \Delta I_C = \alpha I_E + I_{CO} + \Delta I_{CO}$$

Cancel similar terms on opposite sides of this equation. The result is

$$\Delta I_C = \Delta I_{CO}$$

which is equivalent to

$$dI_C = dI_{CO}$$

as shown above. The same procedure can be followed whenever differentiation is indicated.

To recapitulate the process, whenever you find I_C, substitute $I_C + \Delta I_C$. Whenever I_{CO} appears, substitute $I_{CO} + \Delta I_{CO}$. Then, for I_C, substitute the original equation and cancel out similar terms found on opposite sides of the equals sign. The resulting ratio of ΔI_C to ΔI_{CO} (or dI_C to dI_{CO}) is the stability factor S.

The stability factor of specific circuits. The foregoing generalized method of determining the thermal stability of a particular network can be applied to specific circuits if we develop the equation for I_C in terms of I_{CO} for the particular amplifier. The object is to obtain as low a stability factor S as possible consistent with the other requirements

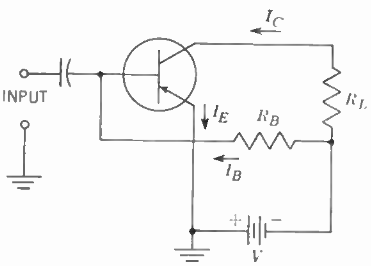


Fig. 12·11

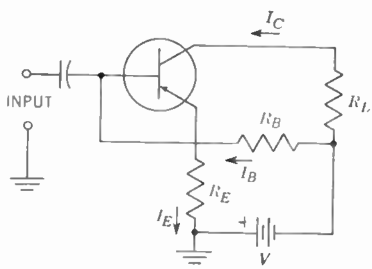


Fig. 12·12

of the circuit. For example, consider the circuit shown in Fig. 12·11. The expression for I_C in this circuit is

$$I_C = \beta I_B + (1 + \beta) I_{CO}$$

I_B, however, is equal to V divided by R_B, assuming that the voltage drop between base and emitter is close enough to zero to be disregarded. Substituting this into the above equation, we have

$$I_C = \beta \frac{V}{R_B} + (1 + \beta) I_{CO}$$

If, now, we differentiate I_C with respect to I_{CO}, we have

$$S = \frac{dI_C}{dI_{CO}} = 1 + \beta \tag{12·6}$$

The first term after the equals sign, V/R_B, becomes zero because all its components are fixed and do not change with change in I_{CO}. Thus, the stability factor of Fig. 12·11 is $1 + \beta$, and since β is generally large, the stability of this circuit is quite poor.

An interesting feature of Fig. 12·11 is that the stability (actually, the instability) is governed only by the value of β and not at all by the values of R_B or R_L. While this is generally true, β is somewhat dependent on the choice of the operating point selected for the transistor, and so it will change with different values of R_B and R_L. This is particularly true when the transistor is near cutoff or when the current flow is high.

Let us now obtain the stability factor for the circuit of Fig. 12·12. This is similar to the preceding circuit with the addition of a resistor in the emitter leg. To set up the necessary I_C and I_{CO} expressions, we proceed as follows.

The battery voltage V is equal to the sum of the voltage drops across R_B and R_E (again assuming that the base-emitter voltage drop is zero). Thus,

$$V = I_B R_B + I_E R_E$$

or, transposing,

$$I_B R_B = V - I_E R_E \tag{12·7}$$

The object now is to express I_B in terms of I_C and I_{CO} and to do the same for I_E. This is achieved as follows. I_C, we know, is equal to

$$I_C = \alpha I_E + I_{CO}$$

or

$$I_E = \frac{I_C - I_{CO}}{\alpha} \tag{12·8}$$

This gives us one expression we need. Now, going back to fundamental transistor action,

$$I_E = I_C + I_B$$

Further,

$$I_C = \alpha I_E + I_{CO}$$

We can substitute this expression in the equation just above it. This gives us

$$I_E = \alpha I_E + I_{CO} + I_B$$

or

$$I_B = I_E(1 - \alpha) - I_{CO} \tag{12·9}$$

Now, we substitute expression (12·8) into (12·9):

$$\begin{aligned} I_B &= \frac{I_C - I_{CO}}{\alpha}(1 - \alpha) - I_{CO} \\ &= \frac{I_C}{\beta} - \frac{(1 - \alpha)}{\alpha} I_{CO} - I_{CO} \\ &= \frac{I_C}{\beta} - I_{CO}\left(\frac{1 - \alpha}{\alpha} + 1\right) \\ &= \frac{I_C}{\beta} - \frac{I_{CO}}{\alpha} \end{aligned} \tag{12·10}$$

This is the second equation we need. Now, let us substitute Eqs. (12·8) and (12·10) into Eq. (12·7).

$$\left(\frac{I_C}{\beta} - \frac{I_{CO}}{\alpha}\right) R_B = V - \left(\frac{I_C - I_{CO}}{\alpha}\right) R_E$$

Rearranging terms,

$$I_C\left(\frac{R_B}{\beta} + \frac{R_E}{\alpha}\right) = V + I_{CO}\left(\frac{R_B}{\alpha} + \frac{R_E}{\alpha}\right)$$

Converting β back to its equivalent, $\alpha/(1 - \alpha)$, and dividing both sides of the equation by R_B, we obtain

$$I_C\left(\frac{1-\alpha}{\alpha} + \frac{R_E}{\alpha R_B}\right) = \frac{V}{R_B} + I_{CO}\left(\frac{1}{\alpha} + \frac{R_E}{\alpha R_B}\right)$$

$$I_C = \frac{\alpha V/R_B + I_{CO}(1 + R_E/R_B)}{1 - \alpha + R_E/R_B}$$

and

$$S = \frac{dI_C}{dI_{CO}} = \frac{1 + R_E/R_B}{1 - \alpha + R_E/R_B} \qquad (12\cdot11)$$

Here is the stability factor for the circuit of Fig. 12·12. If R_E is set equal to zero, we obtain Eq. (12·6) again, thus verifying the correctness of the derivation. To see what happens as we increase the value of R_E, let us assume a value for α of 0.98 and a value for R_B of 50,000 ohms. R_E, in the first computation, will be 1,000 ohms. Then,

$$S = \frac{1 + 1{,}000/50{,}000}{1 - 0.98 + 1{,}000/50{,}000} = \frac{1.02}{0.04} = 25.5$$

This, already, is considerably less than $1 + \beta$ for the circuit of Fig. 12·11, assuming that the same transistor is employed. In the latter instance, $1 + \beta = 1 + 48 = 49$.

If we increase R_E to 5,000 ohms in the above example, we obtain

$$S = \frac{1 + 5{,}000/50{,}000}{1 - 0.98 + 5{,}000/50{,}000} = \frac{1.1}{0.12} = 9.1$$

This demonstrates how rapidly S decreases as R_E increases. The reason, of course, is that R_E provides negative feedback. When I_{CO} increases, it produces an amplified current through R_E. This raises the voltage drop across R_E and reduces the voltage across R_B according to Eq. (12·7). Less voltage across R_B means less bias current, and this reduces the collector current. Thus, R_E functions here in much the same way as a cathode resistor in a vacuum-tube circuit.

It may be instructive to examine the stability of one final circuit because of its wide use. This circuit is shown in Fig. 12·13, and it is seen to differ from the circuit of Fig. 12·12 in having a voltage divider, R_1 and R_2, supply the d-c voltage (actually, current) to the base. To develop the necessary I_C equation for this arrangement in terms of I_{CO}, let us first note what conditions hold here.

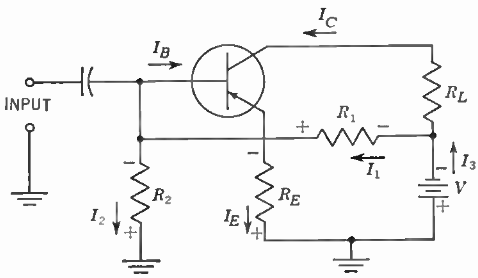


Fig. 12·13

If we assume, as we did before, that there is negligible drop between base and emitter elements of the transistor, then we can write

$$I_E R_E = I_2 R_2$$

Also

$$I_2 R_2 + I_1 R_1 = V$$
$$I_2 + I_B = I_1$$
$$I_1 + I_C = I_3$$

plus the two standbys

$$I_E = I_B + I_C$$

and

$$I_C = \alpha I_E + I_{CO}$$

We have six equations here, and by manipulating them, we can eliminate I_1, I_2, I_3, I_B, and I_E to give us finally the expression

$$I_C\left(\frac{R_E}{\alpha} + \frac{R_E R_1}{\alpha R_2} + \frac{R_1(1-\alpha)}{\alpha}\right) - I_{CO}\left(\frac{R_E}{\alpha} + \frac{R_E R_1}{\alpha R_2} + \frac{R_1}{\alpha}\right) = V$$

Differentiating I_C with respect to I_{CO}, we obtain

$$S = \frac{dI_C}{dI_{CO}} = \frac{R_E/\alpha + R_E R_1/\alpha R_2 + R_1/\alpha}{R_E/\alpha + R_E R_1/\alpha R_2 + (R_1/\alpha)(1-\alpha)}$$

All of the α's in the denominators of the various terms can be canceled out. Furthermore, by dividing numerator and denominator by R_1, we have

$$S = \frac{1 + R_E(1/R_1 + 1/R_2)}{1 - \alpha + R_E(1/R_1 + 1/R_2)} \qquad (12\cdot12)$$

If we compare this expression with Eq. (12·11), we see that we have essentially replaced $1/R_B$ in (12·11) by $(1/R_1 + 1/R_2)$. In short, in so far as the circuit of Fig. 12·13 is concerned, R_1 and R_2 act as if they were in parallel with each other.

By substituting appropriate values for R_1 and R_2 in Eq. (12·12), it will be found that S decreases as R_1 and R_2 become less. Offsetting this, of course, is the fact that the lower R_1 and R_2 are made, the greater the current drain on V. Also, R_1 and R_2 are shunted across the input to the stage, and by lowering their overall value, we reduce the input impedance. This could adversely affect the interstage gain.

One further word concerning Eqs. (12·6), (12·11), and (12·12). Each possesses α (or β, which is directly related to α), and hence we can expect a change in stability as α varies. This is an important consideration in present transistor practice because of the fairly wide variation in α among transistors of the same type. It will be found, however, that the better stabilized a circuit is against increases in I_{co}, the less affected it will be for changes in α or β. This is very fortunate, because variations in α or β present as much difficulty to transistor-amplifier design as instability due to increases in I_{co}.

The effect of input characteristic changes. Now that we have seen the effect of I_{co} variations on the collector current, let us turn our attention to the input circuit. Here we find that a considerable sensitivity to temperature exists, as demonstrated by the curves in Fig. 12·14. Note that if the base-emitter voltage is kept constant, the base current rises quite sharply with temperature. In a common-emitter amplifier, every increase in I_B produces an increase in I_C which is β times as great. Hence, it would not take a very large increase in I_B with temperature to completely alter the operating conditions of any transistor circuit.

Now, what the curves in Fig. 12·14 indicate is that as the temperature rises, the base resistance decreases sharply. By maintaining a fixed V_{EB}, we produce a sharply rising I_B. This behavior is not important when a fairly large resistor is employed in the base circuit, because then this external resistor dominates the circuit and maintains a fairly constant d-c base current. (Because the base-emitter junction is forward-biased, its resistance is quite low.)

Consider, however, the situation when an input transformer is employed, Fig. 12·15. Now, the d-c resistance in the external base circuit is low, and changes in the base-emitter resistance with temperature will markedly alter the base current. This, in turn, will affect I_C.

To stabilize I_B under these conditions, either thermistors or compensating diodes have been inserted in the base-emitter circuit. Figures 12·16 and 12·17 demonstrate typical circuits. In Fig. 12·16, the thermistor stabilizes I_B in the following manner. As the temperature rises, the base current also tends to rise. However, the resistance of

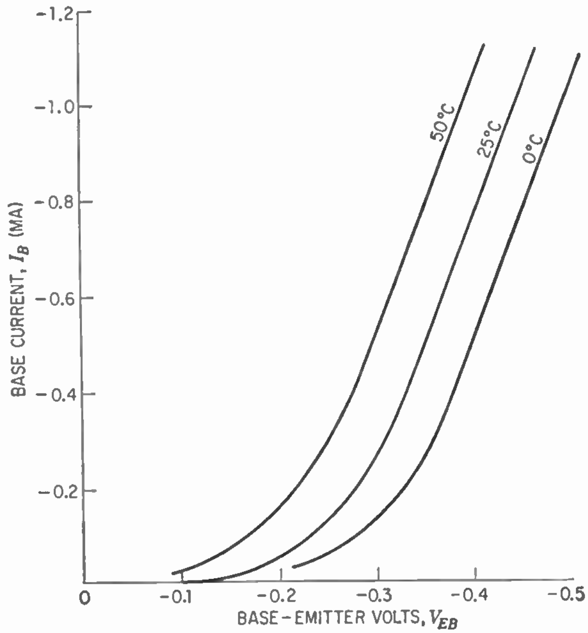


Fig. 12·14 I_B-V_{EB} curves at different temperatures.

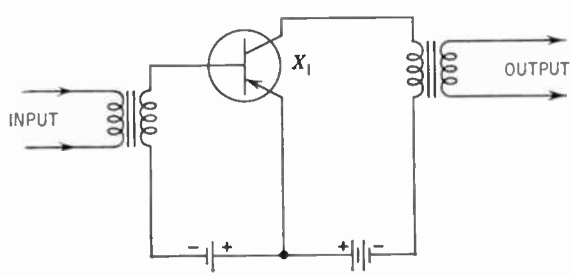


Fig. 12·15

thermistor T_1 decreases with an increase in temperature, causing more current to flow through it and R_1. The increased current will produce a larger voltage drop across R_1, leaving less battery voltage for T_1. Thus, the available voltage for forward bias, that across T_1, is reduced and this reduces the base current.

In Fig. 12·17, a diode is employed in the same manner, with practically the same results. If the resistance of D_1 can be matched to the base-emitter resistance of X_1 and if the two vary similarly, I_B can be d-c stabilized over a wide range of temperatures.

Transistor Equivalent Circuits

In selecting a proper operating point for a transistor, as we just did, we were concerned with the unit solely from a d-c point of view. The bias was so chosen that it would cause the transistor (1) to provide

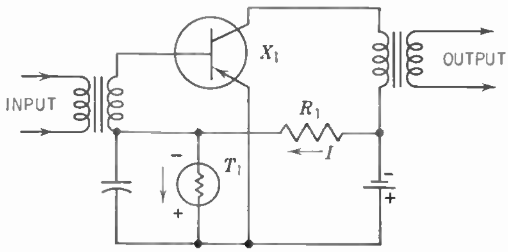


Fig. 12·16 A thermistor-stabilized transistor amplifier.

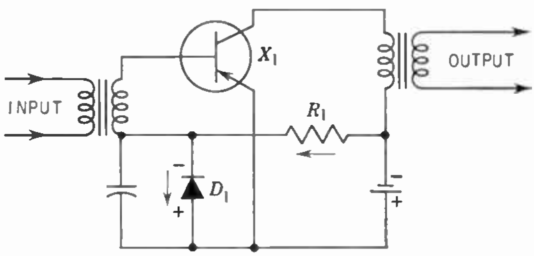


Fig. 12·17 A diode-stabilized transistor amplifier.

operation as distortion-free as possible, (2) to develop a certain amount of power, or (3) both. We did not know, however, how the circuit would behave over a range of frequencies. Would the gain be the same at 2,000 cycles as it is at 200 cycles? How high could we raise the frequency before the gain became negligible? What effect does the transistor itself have on the frequency response of the circuit? These and other questions show that for a complete picture of transistor circuit operation at various signal frequencies a more detailed examination is required. Specifically, the information we seek is contained in the equivalent circuit of a transistor. In this section we shall examine such equivalent circuits to see what form they take and how they can be manipulated to provide us with the information we seek.

Vacuum-tube equivalent circuits. Prior to the introduction of the transistor, the technique of equivalent circuits was applied most intensively to vacuum-tube circuits, specifically in depicting the vacuum tube as a circuit element. In its most common form, the equivalent circuit for a triode is as shown in Fig. 12·18. The voltage applied to the grid (and cathode) is e_g and is arbitrarily taken at the instant the grid is positive with respect to the cathode. With this voltage at the grid, an amplified version of the input, μe_g, appears in the plate circuit. Furthermore, μe_g possesses the polarity shown, indicating that when the grid is driven in the positive direction, the plate tends to go in the negative direction. (This, incidentally, is all that the two sets of polarity signs mean.)

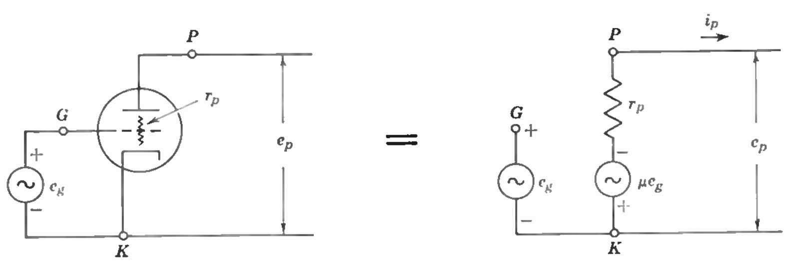


Fig. 12·18 The equivalent circuit of a triode; r_p is the internal plate resistance of the tube, and e_g is the a-c applied signal.

In Fig. 12·18, r_p represents the internal plate-cathode resistance of the tube, i_p is the plate current, and e_p is the output voltage, the voltage that would be applied across a load resistor. If we were to write the governing equation for this circuit, it would be

$$e_p + r_p i_p = \mu e_g$$

If a load resistor is connected between plate and cathode, e_p will be replaced by $i_p R_L$, but this is of only incidental importance. The equation is the significant item, since it enables us to deal with a tube in a rather simple manner.

Now, in drawing the equivalent circuit of the tube, as we do in Fig. 12·18, one tacit assumption is made. That is, we assume that while a variable voltage on the control grid produces a voltage in the plate circuit, the converse is not true. That is, if we start with a variable voltage in the plate circuit, no voltage will appear in the grid circuit. In short, there is no internal feedback path in the tube. This agrees with our experience with tubes at low and medium frequencies, but it is not completely true at high frequencies. At high frequencies, some energy does feed back, and we can say that now the tube possesses inverse (or reverse) amplification. It will be found that the value of this inverse amplification is less than 1, and it is generally not desirable. Still, when it exists, it is just as real as the normal forward amplification factor μ, and recognition is taken of this by assigning it the special symbol of μ_r (r stands for reverse).

It is obvious that when μ_r is not zero, the equivalent diagram of Fig. 12·18 no longer holds, since it makes no provision for μ_r. To represent the new conditions, the equivalent circuit of Fig. 12·19 can be drawn. The right-hand side of the circuit remains the same, since nothing has changed there. At the left, however, we now have r_g,

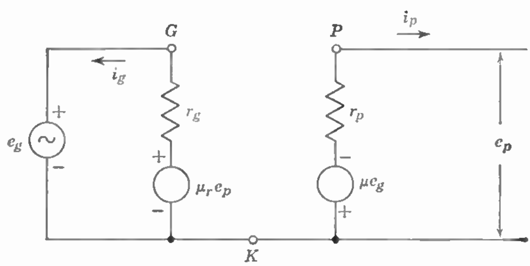


Fig. 12·19 The equivalent circuit of a triode vacuum tube when signal feedback exists between plate and grid elements.

representing the grid-to-cathode resistance, and $\mu_r e_p$, representing the voltage which is transferred from the plate to the grid. (r_g is present in the circuit of Fig. 12·18 also and, if shown, would be in parallel with e_g. However, since it does not enter into the governing equation for this equivalent circuit, it was omitted.)

To completely describe this new equivalent circuit mathematically, two equations are needed. One equation, that for the output circuit, remains unaltered, since the circuit is the same. That is,

$$\mu e_g = i_p r_p + e_p$$

The second equation represents the input circuit and its form is

$$e_g = i_g r_g + \mu_r e_p$$

Thus, we now have two equations where before we had one, and the second equation stems solely from the fact that energy (i.e., a signal) is fed from the plate back to the grid. The tube is no longer a unilateral device. Current can flow in both directions, although not with equal facility.

Before we continue, a word should be said of the fact that d-c voltages are never shown in equivalent circuits. In an actual circuit, d-c voltages are needed to bias the tube to the desired operating point. There is a d-c voltage in the grid-to-cathode circuit and a d-c voltage in the plate-to-cathode circuit. However, once the tube is placed at the proper operating point, we are interested only in its behavior to a-c signals, and, thus, only a-c voltages are shown in the equivalent circuit. Direct-current voltages are always understood as being in the background, ready to be utilized once the equivalent circuit has yielded the desired information and the actual circuit is to be designed.

Transistor equivalent circuits. Now that we have examined the equivalent circuit of a tube, let us turn our attention to the transistor.

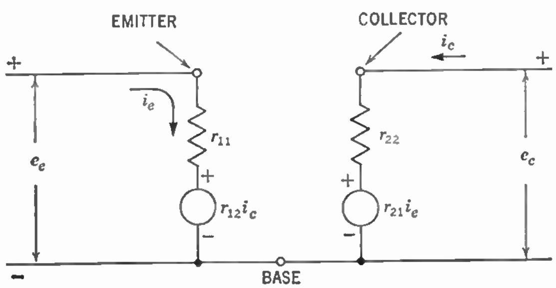


Fig. 12·20 Equivalent circuit of a transistor. This is one of several forms which will be discussed

We have already learned that a transistor, like a tube, provides gain and further that in a transistor there is internal feedback from the collector to the emitter. (This is demonstrated graphically on page 113 for a junction transistor, where it is shown how the input impedance varies as the load impedance changes.) Since such feedback exists, the equivalent circuit best suited for transistors should resemble the circuit of Fig. 12·19 more closely than the circuit of Fig. 12·18. This turns out to be correct, because the equivalent circuit of Fig. 12·20 is eminently suitable for transistors.

In Fig. 12·20 e_e represents the a-c input signal applied between emitter and base. e_c is the output voltage, and it would appear across a load resistor connected between collector and base. $r_{12}i_c$ is the voltage (because $e = ir$) transferred from the output circuit to the input circuit. $r_{21}i_e$ is the signal voltage appearing in the output circuit because of a voltage applied to the input circuit.

To set up the equations for the circuit of Fig. 12·20, each branch is considered separately. Thus, for the left-hand branch, we have

$$e_e = r_{11}i_e + r_{12}i_c \tag{12·13}$$

And for the right-hand side, the equation is

$$e_c = r_{21}i_e + r_{22}i_c \tag{12·14}$$

Here, then, are the equations that govern the behavior of the transistor equivalent circuit, and if this combination correctly represents an actual transistor, manipulation of the two equations will tell us what we need to know about the workings of a transistor when a signal is applied. We are considering the operation of a transistor at low and medium frequencies only in this discussion. For high-frequency operation, emitter and collector shunting capacitances have to be added, plus several other components, and the analysis becomes more involved. A discussion of the equivalent circuit of a transistor at high frequencies is given at the end of this chapter.

The next step is to determine what the various items in these two equations represent. i_e, the current in the left-hand circuit, stands for the emitter current. By the same token, i_c is the collector current. If we temporarily set i_c equal to zero in Eq. (12·13), indicating that the output circuit is open, then

$$e_e = r_{11}i_e + 0$$

or

$$r_{11} = \frac{e_e}{i_e} \tag{12·15}$$

If i_e is the input current and e_e the input voltage, then e_e divided by i_e represents the input resistance r_{11}.

If, now, we permit i_c to flow, but prevent i_e from flowing by opening the input circuit, Eq. (12·13) becomes

$$e_e = 0 + r_{12}i_c$$

or

$$r_{12} = \frac{e_e}{i_c} \tag{12·16}$$

This equation tells us that a current i_c flowing in the output circuit produces a voltage in the input circuit. r_{12} is the feedback or reverse impedance common to both input and output circuits, and it is the voltage drop developed by i_c flowing through r_{12} that produces the feedback voltage in the input circuit.

Now, turning to Eq. (12·14), let us consider it when the output circuit is open and i_c is equal to zero. Substituting this value of i_c in (12·14), we obtain

$$e_c = r_{21}i_e + 0$$

or

$$r_{21} = \frac{e_c}{i_e} \tag{12·17}$$

Equation (12·17) reveals that even though the output circuit is open, a voltage e_c equal to $r_{21}i_e$ is present there. Since i_e is the input current, it means that r_{21} is common to both input and output circuits, and when i_e flows, the voltage drop across r_{21} caused by i_e appears in the output circuit. r_{21} could thus be called the forward-transfer impedance. Note that if r_{21} did not exist, an incoming signal could never reach the output circuit. Thus, r_{21} is the element through which the input circuit affects the output circuit.

For the final step, the input circuit is opened, reducing i_e to zero. When this substitution is made in Eq. (12·14),

$$e_c = 0 + r_{22}i_c$$

or

$$r_{22} = \frac{e_c}{i_c} \tag{12·18}$$

That is, r_{22} is the output impedance or, here, the impedance of the collector circuit. Note that the e_c value in (12·18) would not be the same as the e_c value in (12·17), since the equations are obtained under different conditions. The same is true of e_e in Eqs. (12·15) and (12·16).

As an interesting sidelight, observe that r_{12} and r_{21} are not equal to each other. In a passive network, such as that formed by resistances, inductances, and capacitances only, the impedance in one direction would be the same as the impedance in the opposite direction. In a circuit containing a nonlinear device such as a tube or a transistor, this is not true, a fact we know from our own experience.

Second transistor equivalent circuit. The equivalent circuit of Fig. 12·20 represents the transistor electrically, but it is not the only equivalent circuit. (In similar fashion, Fig. 12·19 is not the only equivalent circuit of a vacuum tube in which feedback from the output to the input circuit is taken into account. For the present discussion, it is the most convenient and, for that reason, it is used.) There are a number of others, one of the most widely used being shown in Fig. 12·21. This circuit is popular because its electrical form ties in closely with the three sections of transistor. For example, r_e is the resistance of the emitter section, r_b is the resistance of the base section, and r_c is the resistance of the collector section; $r_m i_e$ is a generator placed in the collector arm to represent the effect of the input circuit on the output circuit; and r_m is the mutual resistance between the input and output circuits by which the input current i_e produces a signal in the output. Here this output signal is called $r_m i_e$. It is essentially equivalent to $r_{21}i_e$ of Fig. 12·20, differing from r_{21} only by the value of the base resistance r_b. (r_{21}, however, is so much greater than

r_b that the latter may be disregarded in any equation involving r_{21} and r_b.) By the same token, we make provision for the feedback of signal from output to input by the presence of the common resistor r_b. The backward voltage transferred by r_b is equivalent to the generator $r_{12}i_c$ in Fig. 12·20. By separating the voltage that the input delivers to the output from the voltage fed in the opposite direction, we tacitly recognize that they are unequal in size and, furthermore, that they may vary in different fashions. This, of course, again jibes with our experience of tube and transistor circuits. Thus the form of the equivalent circuit in Fig. 12·21 conforms to the general requirements of a transistor as revealed by what we know about it.

Now, if this equivalent circuit and the preceding one represent the same transistor, there should exist an electrical correspondence

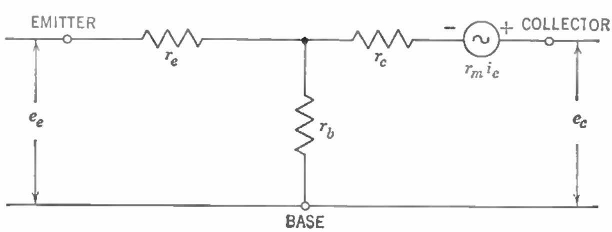


Fig. 12·21 Another equivalent circuit of a transistor.

between them. To ascertain what this similarity is, let us first set up the equations for the circuit of Fig. 12·21 and then equate them properly to the two equations governing the preceding equivalent circuit.

The second equivalent circuit has been redrawn in Fig. 12·22, this time divided into two loops by the two arrows. The arrows show the assumed directions of flow for current i_e and current i_c. Before we write out the equation for the left-hand and right-hand sides of the circuit, it should be noted that the following rules are being applied.

1. When current flows through a resistor, the end entered first is given a positive sign, while the exit end becomes negative. This is the sign notation associated with the conventional flow of current (i.e., from positive to negative) in distinction to the electron flow. The conventional notation is being employed in the discussion because it is widely used in the literature and will enable the reader to follow other presentations more easily.
2. When there is a generator or a source of voltage in the circuit, the sign to be given this item will depend on its polarity. For

example, if the loop arrow encounters the positive side of the generator first, this voltage is added to the loop equation with a plus sign. However, if the arrow meets the negative sign first, the quantity is added to the equation preceded by a negative sign.

3. We will use Kirchhoff's law which states that the sum of the voltages in a closed loop adds up to zero. Previously, we had applied Kirchhoff's other law, which says that the sum of all the currents flowing toward a point is zero.
4. When two currents flow through the same element, such as r_b, the effect of each current is considered separately. If the currents flow in the same direction through the component, their voltage drops add; if they flow in opposite directions, the voltage drops subtract.

With these rules, we are ready to start the analysis. Consider loop I first. Current i_e flows through resistors r_e and r_b, producing voltage drops across each. i_c also flows through r_b, and since its direction is the same as i_e, their voltage drops add. Finally, i_e meets the negative terminal of e_e first; hence, this voltage is given a negative sign in the equation. Thus

$$r_e i_e + (i_e + i_c)r_b - e_e = 0 \tag{12·19}$$

In the second loop, conditions are somewhat the same except for the presence of $r_m i_e$. This is a voltage generator; furthermore, the arrow encounters the plus sign first. Consequently, $r_m i_e$ is added to the equation bearing a positive sign. Thus

$$r_m i_e + r_c i_c + (i_c + i_e)r_b - e_c = 0 \tag{12·20}$$

Now, let us reshuffle the items in Eqs. (12·19) and (12·20) so that e_c and e_e are each on the other side of the equals sign and all the currents are properly segregated. Doing this, we obtain for Eq. (12·19)

$$e_e = (r_e + r_b)i_e + r_b i_c$$

and for Eq. (12·20)

$$e_c = (r_m + r_b)i_e + (r_c + r_b)i_c$$

Here are the two governing equations for the equivalent circuit of Fig. 12·22. If these are compared with Eqs. (12·13) and (12·14), they are seen to possess the same form with the only difference being the resistance factors of i_e and i_c. And since the circuit of Fig. 12·22 represents the same transistor as the circuit of Fig. 12·21, their governing equations should be equivalent to each other. This will occur when

$$r_{11} = r_e + r_b \tag{12·21}$$
$$r_{12} = r_b \tag{12·22}$$
$$r_{21} = r_m + r_b \tag{12·23}$$
$$r_{22} = r_c + r_b \tag{12·24}$$

These are the relationships between r_{11}, r_{12}, r_{21}, and r_{22} in one equivalent transistor circuit and r_e, r_b, r_c, and r_m in the other equivalent

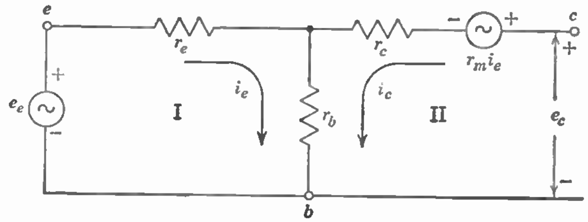


Fig. 12·22

circuit. Sometimes, the same information is given in the following form:

$$r_e = r_{11} - r_{12} \tag{12·25}$$
$$r_b = r_{12} \tag{12·26}$$
$$r_c = r_{22} - r_{12} \tag{12·27}$$
$$r_m = r_{21} - r_{12} \tag{12·28}$$

These relationships follow directly from the preceding set by a slight manipulation.

Typical values for r_e, r_b, r_c, and r_m in commercial transistors fall within the following ranges:

$$r_e = 20 \text{ to } 30 \text{ ohms}$$
$$r_b = 400 \text{ to } 700 \text{ ohms}$$
$$r_c = 1.0 \text{ to } 2.0 \text{ megohms}$$
$$r_m = 0.95 \text{ to } 1.9 \text{ megohms}$$

We are now ready to examine the behavior of the equivalent circuit of Fig. 12·22 with the addition of a load resistor across the output and a signal source across the input. We shall do this briefly with the common-base equivalent circuit, Fig. 12·22, and then swing over to an equivalent circuit for a common-, or grounded-, emitter arrangement.

Figure 12·23 is the circuit we shall use. A load resistor R_L has been placed across the output terminals, and a signal generator e_g and its internal resistance R_g have been connected across the input terminals. In accordance with conventional practice, the common base has been grounded, but whether this point actually is so connected in the circuit or not will have no effect on the derivation.

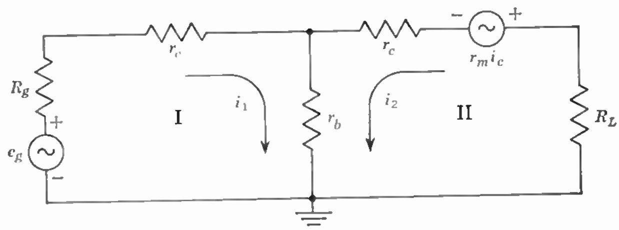


Fig. 12·23

Proceeding as we did when Eqs. (12·19) and (12·20) were set up, we can write for loop I,

$$-e_g + R_g i_1 + r_e i_1 + r_b(i_1 + i_2) = 0 \tag{12·29}$$

Rearranging terms,

$$e_g = (R_g + r_e + r_b)i_1 + r_b i_2$$

For loop II, we have

$$i_2 R_L + r_m i_e + r_c i_2 + r_b(i_1 + i_2) = 0$$

The term i_e represents the current in the emitter circuit. In the present instance, i_1 is the emitter-circuit current and consequently $i_1 = i_e$. Making this substitution in the last equation and rearranging terms, we have

$$0 = (r_m + r_b)i_1 + (R_L + r_c + r_b)i_2 \tag{12·30}$$

We now have two variables, i_1 and i_2, and two simultaneous linear equations. Hence, we can solve them for i_1 and i_2. Doing this gives us

$$i_1 = \frac{e_g(R_L + r_c + r_b)}{(R_L + r_c + r_b)(R_g + r_e + r_b) - r_b(r_b + r_m)} \tag{12·31}$$

$$i_2 = \frac{-e_g(r_b + r_m)}{(R_L + r_c + r_b)(R_g + r_e + r_b) - r_b(r_b + r_m)} \tag{12·32}$$

These equations can be employed directly to provide the current gain for a grounded-base amplifier. This is denoted by the symbol α

and is equal to the ratio of i_2 to i_1. Thus, dividing (12·32) by (12·31), we have

$$\text{Gain} = \frac{i_2}{i_1} = \frac{r_b + r_m}{R_L + r_c + r_b}$$

The gain, then, is dependent upon the load resistance, becoming smaller as R_L becomes larger. The maximum current gain is achieved when R_L is set equal to zero (i.e., a short circuit across the output). Under these conditions

$$\alpha = \frac{i_2}{i_1} = \frac{r_b + r_m}{r_c + r_b} \tag{12·33}$$

Since r_b is numerically much smaller than either r_m or r_c, Eq. (12·33) reduces to

$$\alpha \cong \frac{r_m}{r_c} \tag{12·34}$$

Numerical values of α vary from 0.95 to 0.99 for commercial junction transistors. This information, of course, is supplied by the manu-

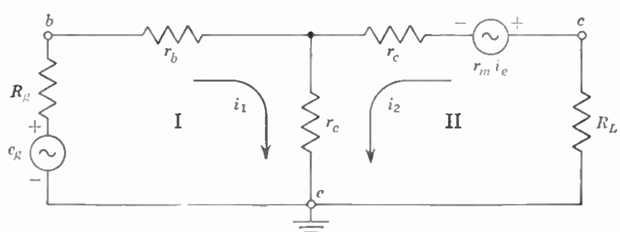


Fig. 12·24 Equivalent circuit of a transistor connected with its emitter common to both input and output circuits.

facturer, although if the values for r_m and r_c are given, α could be computed. Additional manipulations could be performed with i_1 and i_2 to obtain other information concerning the grounded-base amplifier, but rather than spend time on a circuit that is not widely used, let us turn to the equivalent circuit of a common-emitter amplifier and examine it in detail.

Common-emitter Amplifiers

To obtain the equivalent circuit of the common-emitter arrangement, we employ the circuit of Fig. 12·23, changing only the relative positions of the base and emitter arms, Fig. 12·24. In all other respects, the two circuits are alike.

We would proceed as before to set up the equations for the two loops of Fig. 12·24. Thus, for loop I we have

$$-e_g + R_g i_1 + r_b i_1 + r_e(i_1 + i_2) = 0 \qquad (12\cdot35)$$

For loop II

$$R_L i_2 + r_m i_e + r_c i_2 + r_e(i_1 + i_2) = 0 \qquad (12\cdot36)$$

Before we consolidate each equation, the question of $r_m i_e$, specifically i_e, must be resolved. In Fig. 12·23, i_e was seen to be equal to i_1 because only i_1 passed through r_e. Furthermore, i_1 passed through r_e going toward the junction of r_e and r_b. The latter point is important because direction must be maintained; otherwise the signs between i_e and i_1 will require reversal.

Now, examine Fig. 12·24. Not only does i_1 pass through r_e, but i_2 as well. Consequently, i_e is now equal to $i_1 + i_2$. With respect to polarity, it is seen that for i_e to remain unchanged, it should be traveling through r_e toward r_b, as in Fig. 12·22. i_1 and i_2, however, are traveling in the opposite direction. Hence, $i_e = -(i_1 + i_2)$, and this is the substitution to be made for i_e in Eq. (12·36). Thus

$$R_L i_2 - r_m(i_1 + i_2) + r_c i_2 + r_e(i_1 + i_2) = 0 \qquad (12\cdot37)$$

Now segregate similar terms in Eqs. (12·35) and (12·37).

$$e_g = (R_g + r_b + r_e)i_1 + r_e i_2 \qquad (12\cdot38)$$

and

$$0 = (r_e - r_m)i_1 + (R_L + r_c + r_e - r_m)i_2 \qquad (12\cdot39)$$

Solving these equations simultaneously for i_1 and i_2, we obtain

$$i_1 = \frac{e_g(R_L + r_c + r_e - r_m)}{(R_L + r_c + r_e - r_m)(R_g + r_b + r_e) + r_e(r_m - r_e)} \qquad (12\cdot40)$$

$$i_2 = \frac{e_g(r_m - r_e)}{(R_L + r_c + r_e - r_m)(R_g + r_b + r_e) + r_e(r_m - r_e)} \qquad (12\cdot41)$$

Current gain. Let us determine next the current gain for the common-emitter arrangement, just as we did previously for the common-base circuit.

$$\text{Current gain} = \frac{i_2}{i_1} = \frac{r_m - r_e}{R_L + r_c + r_e - r_m} \qquad (12\cdot42)$$

While this expression differs from that of Eq. (12·34), we see that here, too, current gain will vary with load resistance and in the same manner as before. For maximum current gain, R_L is set equal to zero, giving

$$\text{Current gain} = \frac{r_m - r_e}{r_c + r_e - r_m} \qquad (12\cdot43)$$

Divide numerator and denominator by r_c.

$$\text{Current gain} = \frac{r_m/r_c - r_e/r_c}{1 + r_e/r_c - r_m/r_c} \tag{12·44}$$

Numerically, r_e/r_c is very small compared with r_m/r_c and could be disregarded. Hence

$$\text{Current gain} = \frac{r_m/r_c}{1 - r_m/r_c} \tag{12·45}$$

But r_m/r_c is equal to α, as evidenced by Eq. (12·34). Consequently,

$$\text{Current gain} = \frac{\alpha}{1 - \alpha} = \beta \tag{12·46}$$

This, of course, agrees with the definition given for the current gain of a common-emitter circuit, and here we see how it is actually derived. Values of β for most commercial transistors fall generally between 30 and 50, but units with higher values are available.

Input impedance. Another result that we can derive rather easily from Eqs. (12·40) and (12·41) is the input impedance of the circuit of Fig. 12·24. As a first step, it should be recognized that R_g in Fig. 12·24 is not part of the input impedance. Consequently, if we divide e_g by i_1 (giving us the total impedance seen by the generator) and subtract R_g from this, we shall obtain the value we seek. In equation form, this is

$$R_{\text{in}} = \frac{e_g}{i_1} - R_g \tag{12·47}$$

Substituting the expression for i_1 from (12·40) in (12·47), we have

$$R_{\text{in}} = \frac{e_g}{\dfrac{e_g(R_L + r_c + r_e - r_m)}{(R_L + r_c + r_e - r_m)(R_g + r_b + r_e) + r_e(r_m - r_e)}} - R_g$$

$$= \frac{(R_L + r_c + r_e - r_m)(R_g + r_b + r_e) + r_e(r_m - r_e)}{(R_L + r_c + r_e - r_m)} - R_g \tag{12·48}$$

$$R_{\text{in}} = r_b + r_e + \frac{r_e(r_m - r_e)}{R_L + r_c + r_e - r_m} \tag{12·49}$$

If we neglect r_e with respect to r_m in the numerator of the last term, we obtain

$$R_{\text{in}} = r_b + r_e + \frac{r_e r_m}{R_L + r_c + r_e - r_m} \tag{12·50}$$

Again, we see quite clearly that R_L plays a part in determining the input resistance of the transistor. Graphically, this is demonstrated on page 113. Note, too, that the highest input impedance is achieved when R_L is zero. Thereafter, R_{in} decreases as R_L rises, although after R_L reaches 1 megohm or so, R_{in} remains fairly constant for all practical purposes.

It is interesting to note that when R_L in Eq. (12·49) or (12·50) is small compared to r_c (which is generally the case), the input impedance equation reduces to

$$R_{in} = r_b + r_e + \beta r_e$$

using Eq. (12·45). Combining terms, we have

$$R_{in} = r_b + (\beta + 1)r_e$$

or

$$\cong r_b + \beta r_e$$

because β is so much greater than 1. This relationship reveals that the input resistance for the common-emitter configuration is approximately β times as large as the input resistance of the common-base arrangement. The latter is $r_b + r_e$.

Output impedance. We can derive an expression for the output impedance of the equivalent circuit by utilizing essentially the same techniques as above, modified only for the right-hand side of the circuit instead of the left. To see how this is done, consider the circuit of Fig. 12·25. This is identical to the circuit of Fig. 12·24 except that

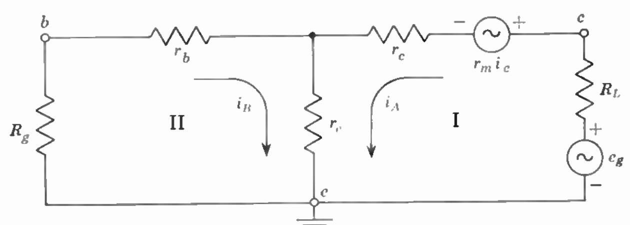


Fig. 12·25 The equivalent circuit arrangement for determining the output impedance of a transistor.

e_g has been moved from the input to the output circuit. In this move, R_g was left untouched so that its effect on the circuit would remain the same. R_L has also been left unchanged, but now, because it is in series with e_g, it assumes the role of the generator impedance. The output impedance of this circuit is the impedance seen by e_g and R_L looking back into the network. Thus, all we need do is find i_A and then apply it in the equation

$$R_{out} = \frac{e_g}{i_A} - R_L \tag{12·51}$$

Note that we have altered the labeling of the preceding i_1 and i_2 currents to i_B and i_A. This is because the equations to be set up differ from the preceding equations and consequently the two currents will have other values.

Following the same setup procedure as we did for Fig. 12·24, the circuit equations obtained for loops I and II are

$$i_B(R_g + r_b + r_e) + i_A r_e = 0 \qquad (12\cdot52)$$

$$i_B(r_e - r_m) + i_A(R_L + r_e + r_c - r_m) - e_g = 0 \qquad (12\cdot53)$$

Solving these simultaneously for i_A, we obtain

$$i_A = \frac{e_g(R_g + r_e + r_b)}{(R_g + r_b + r_e)(R_L + r_e + r_c - r_m) - r_e(r_e - r_m)} \qquad (12\cdot54)$$

Substituting this equation in (12·51),

$$R_{\text{out}} = \frac{e_g}{\dfrac{e_g(R_g + r_e + r_b)}{(R_g + r_b + r_e)(R_L + r_e + r_c - r_m) - r_e(r_e - r_m)}} - R_L \qquad (12\cdot55)$$

$$R_{\text{out}} = R_L + r_e + r_c - r_m - \frac{r_e(r_e - r_m)}{R_g + r_e + r_b} - R_L \qquad (12\cdot56)$$

$$R_{\text{out}} = r_e + r_c - r_m + \frac{r_e r_m - r_e^2}{R_g + r_e + r_b} \qquad (12\cdot57)$$

The presence of R_g in the formula indicates that changes in R_g will affect the output impedance. Thus, a transistor exhibits effective bilateral behavior, with input and output impedances reacting on each other. Incidentally, if we examine the various terms in (12·57), we see that r_e and r_c will always be greater than r_m and that $r_e r_m$ will always be greater than r_e^2, so that the entire expression can never become negative, no matter what value R_g is.

Equation (12·57) also reveals an interesting fact if we assume that R_g is very large compared to r_e and r_b, a situation that is generally true. Under this condition

$$\frac{r_e r_m - r_e^2}{R_g + r_e + r_b}$$

can be ignored, since a large value of R_g reduces the expression to a low value. Equation (12·57) then becomes

$$R_{\text{out}} = r_e + r_c - r_m \qquad (12\cdot57a)$$

However, from Eq. (12·34), we have

$$\alpha \cong \frac{r_m}{r_c}$$

Substituting this into (12·57a) gives

$$R_{out} = r_e + r_c - \alpha r_c$$
$$= r_e + r_c(1 - \alpha) \qquad (12 \cdot 57b)$$

The expression $r_c(1 - \alpha)$ can be readily shown to equal

$$\frac{r_c}{1 + \beta}$$

Substituting this into (12·57b), we have

$$R_{out} = r_e + \frac{r_c}{\beta + 1}$$

or

$$\cong r_e + \frac{r_c}{\beta} \qquad (12 \cdot 57c)$$

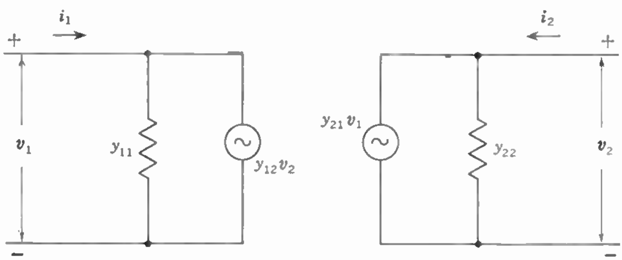
Since it can be shown that the output impedance of a common-base amplifier is r_c, Eq. (12·57c) reveals that the output impedance of a common-emitter stage is approximately β times *lower*.

The foregoing analysis has demonstrated how the equivalent circuit is established and has detailed some of the information which can be derived from it. This does not exhaust the possibilities by any means, and for those readers interested in pursuing the subject further, there are a number of engineering texts available, in addition to various technical magazines in which many articles will be found.

Hybrid Parameters and Their Equations

In the preceding development of the transistor equivalent circuit, the internal structure was represented by resistances r_e, r_b, r_c, and r_m. (Actually, to be more general, impedance symbols Z_e, Z_b, Z_c, and Z_m should have been used. However, for low and medium frequencies, the internal structure can be characterized by resistances. Under this assumption, the bandpass of a transistor amplifier is limited only by the external coupling circuit. This is made evident by the fact that none of the formulas derived above had any frequency-dependent terms in them.) To the circuit designer, the value of these resistances for each different transistor he deals with is important and, ideally, they should be values that can be measured readily with equipment that is not too complex. This, however, is not possible with some of the measurements that have to be made. To see why, let us determine what must be done to measure each parameter. For this discussion, we shall use the equivalent circuit of Fig. 12·20 and its components, r_{11}, r_{12}, r_{21}, and r_{22}. This will cause no difficulty because once these values are obtained, they can be made to yield r_e, r_b, r_c, and r_m by Eqs. (12·21) to (12·24).

In Fig. 12·20 the value of r_{11} is obtained by setting i_c equal to zero in Eq. (12·13) and then measuring the current through the input circuit and the voltage across it. The chief difficulty here lies in making a suitable measurement with $i_c = 0$. This is because all parameter measurements on a transistor should be made with the normal operating d-c bias voltages applied. To achieve $i_c = 0$, the collector circuit should be opened for a-c signals, yet closed and complete for d-c voltages. This condition can be met by feeding the d-c collector voltage to its element through a parallel resonant circuit which would



$$i_1 = y_{11}v_1 + y_{12}v_2$$
$$i_2 = y_{21}v_1 + y_{22}v_2$$

Fig. 12·26 An equivalent transistor arrangement in which conductances y's and current generators yv's are employed in place of resistance and voltage generators.

possess negligible d-c resistance but a very high a-c impedance. Because the collector resistance itself is high, making the output impedance high, it can be seen that the Q of the resonant circuit would have to become impractically high in order to approximate the open circuit needed for the measurement.

The same difficulty arises again for the measurement of r_{21}, because here, too, i_c must be reduced to zero. Obviously, the equivalent circuit, formed by using the resistance parameters, lends itself well to design analysis but not to the measurement of those values on which the design computations would depend.

As a solution to this problem, other forms of equivalent circuit arrangements have been fashioned. For example, there is a configuration that employs conductances and current generators in place of resistances and voltage generators. Such an arrangement is shown in Fig. 12·26, together with the equations which govern it. If this configuration appears strange to the reader, he has only to recall the equivalent circuit which is commonly used for pentodes, Fig. 12·27.

Here, $g_m e_g$ is a current generator representing the plate current of the tube and flowing into the parallel combination of r_p and R_L. When the plate resistance is high, the circuit of Fig. 12·27 representing a tube is easier to deal with than the series equivalent circuit of Fig. 12·20. However, when the internal resistance is low, as it is in the emitter circuit of a transistor, a series arrangement is preferable.

Herein lies the difficulty of using wholly either the resistance or the conductance equivalent circuits to represent the transistor, particularly when measurements are to be made to determine the internal

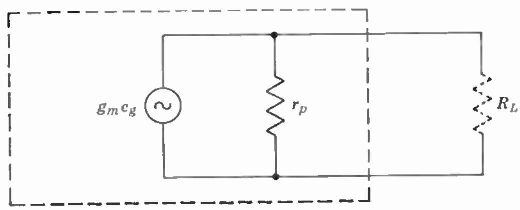


Fig. 12·27 Equivalent circuit for a pentode vacuum tube.

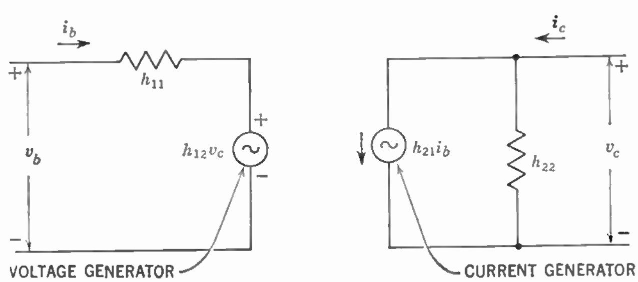


Fig. 12·28 The hybrid equivalent circuit for a transistor.

parameter values. The low-resistance input circuit requires one type of approach, the high-resistance output circuit another. What is needed is a combination circuit.

It was in response to this need that the hybrid parameters, labeled h, were developed and are now widely employed by transistor manufacturers to indicate the values of the internal elements of transistors. An equivalent circuit for such hybrid parameters is shown in Fig. 12·28, and its equations are set up in the same manner as they are for Fig. 12·19. That is, for the input loop we have a series circuit, and the equation consists of the voltage drops (and voltage generators), each properly combined with due regard to its polarity. In the present instance,

$$v_b = h_{11}i_b + h_{12}v_c \tag{12·58}$$

In the output loop, we are concerned with currents, and here we assume that the current through h_{22} is in the same direction as the current produced by the current generator $h_{21}i_b$. Hence, the governing equation for this circuit is

$$i_c = h_{21}i_b + h_{22}v_c \tag{12·59}$$

Note that the two equations differ in form from each other because the circuits they are derived from differ. In one instance we are concerned with loop voltages; in the other, loop currents. In both cases, however, the same two independent variables, i_b and v_c, are employed.

The next step is to evaluate the various h parameters such as h_{11}, h_{12}, h_{21}, and h_{22}. This will give us an indication of what they represent.

To obtain h_{11}, v_c in Eq. (12·58) is set equal to zero. That is, the output would be shorted. For this condition,

$$v_b = h_{11}i_b$$

or

$$h_{11} = \frac{v_b}{i_b} \tag{12·60}$$

In this equation h_{11} reveals itself to be the input resistance of the circuit. Consequently, its unit of measurement would be ohms.

h_{12} is determined similarly by setting i_b equal to zero. Equation (12·58) then becomes

$$v_b = h_{12}v_c$$

or

$$h_{12} = \frac{v_b}{v_c} \tag{12·61}$$

Since v_b and v_c have similar units, the units cancel out, leaving h_{12} simply as a number. Now, had v_c been in the numerator and v_b in the denominator, h_{12} would represent the forward voltage gain of the transistor. However, because of the reversal in position, this ratio is sometimes called the feedback voltage ratio.

The next hybrid parameter h_{21} occurs in Eq. (12·59). To find what it represents, let us set v_c to zero. Equation (12·59) now becomes

$$i_c = h_{21}i_b$$

or

$$h_{21} = \frac{i_c}{i_b} \tag{12·62}$$

h_{21} is seen to be the ratio of collector current to base current, or the current amplification of the circuit. Since we are dealing here with a common-emitter arrangement, h_{21} is equal to β.

The final hybrid parameter is h_{22}, and it is obtained by setting i_b in Eq. (12·59) equal to zero. This gives us

$$i_c = h_{22}v_c$$

or

$$h_{22} = \frac{i_c}{v_c} \tag{12·63}$$

This is the equation for the output conductance; hence, h_{22} possesses the unit of mhos.

An alternate designation for the four h parameters which is being used increasingly by manufacturers is

$$h_{11} = h_i \qquad h_{21} = h_f$$
$$h_{12} = h_r \qquad h_{22} = h_o$$

The i of h_i stands for input resistance, the r of h_r for reverse voltage ratio, the f of h_f for forward current ratio, and the o of h_o for output conductance. An additional letter is then added to denote the circuit configuration; for example, h_{ie} denotes the common-emitter arrangement, h_{ib} the common-base arrangement, and h_{ic} the common-collector arrangement.

In arriving at each of the hybrid parameters, we set either $v_c = 0$ or $i_b = 0$. Now, suppose we had an actual transistor whose hybrid parameter values we wished to measure. How could we reduce the a-c base current (that is, i_b) or the a-c collector voltage v_c to zero without disturbing the d-c bias voltages applied to that transistor? (It was on such measurements that the resistance and conductance parameters fell down.)

In the case of $i_b = 0$, the base circuit should be an open circuit for signal currents, yet closed and complete for the d-c bias voltage. This can be achieved by feeding the d-c voltage to the base via a very high resistor, say 1 to 10 megohms. Since the transistor input impedance is quite low, constructing such a circuit is easily accomplished by using conventional components.

For $v_c = 0$, we must maintain the correct bias voltage at the collector while bypassing all alternating currents at the collector to ground. This, too, is readily done with a moderate-sized capacitor because the collector impedance is so high that an a-c short circuit (for all practical purposes) is attained with a nominal value of shunting capacitance. Thus, use of the hybrid factors provides us with an arrangement whose parameters (such as h_{11}, h_{12}, h_{21}, and h_{22}) are readily measured. This is the outstanding advantage of this configuration.

Common-emitter Hybrid Equations

Since the hybrid parameters are employed so extensively, it may be instructive to derive suitable equations governing current gain, voltage gain, power gain, and input and output impedance for a common-emitter amplifier by using hybrid parameters. The basic circuit of Fig. 12·28 will be employed, with the addition of a load resistor R_L, an input signal e_g, and its internal resistance R_g. The corresponding circuit is shown in Fig. 12·29.

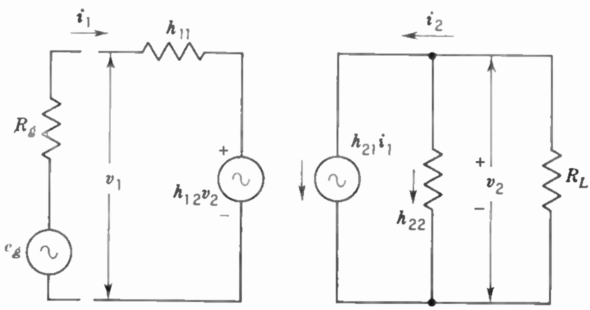


Fig. 12·29 Hybrid equivalent circuit for a transistor.

The basic relations governing the behavior of this circuit can be written in four equations. The first two equations will relate to the transistor only, the second two will relate to the external circuit. The first two equations are (12·58) and (12·59) previously given:

$$v_1 = h_{11}i_1 + h_{12}v_2 \tag{12·58a}$$
$$i_2 = h_{21}i_1 + h_{22}v_2 \tag{12·59a}$$

Note that the only change we have made is to use v_1 for v_b, i_1 for i_b, v_2 for v_c, and i_2 for i_c. The change was made to enable these equations to represent *all three* possible configurations of a transistor amplifier.

The second two equations needed are

$$v_1 = e_g - i_1R_g \tag{12·64}$$
$$v_2 = -i_2R_L \tag{12·65}$$

Equation (12·64) simply states that voltage v_1 is equal to the input signal voltage e_g minus the voltage drop across R_g, the internal resistance of the signal source. Equation (12·65) indicates that v_2 is equal to the voltage developed across R_L. The negative sign is needed before i_2R_L because the current flows up through R_L, while it flows *down* through h_{22} and $h_{21}i_1$. Obviously then, if v_2 possesses the polarity indicated because of the voltage drop across h_{22}, the voltage drop across R_L must be opposite in sign (because of the opposite flow of current). This balancing of the voltages must be made; otherwise, the resulting equations will be incorrect.

Current gain. To derive the equation for the current gain of this arrangement, we eliminate i_2 between Eqs. (12·59*a*) and (12·65). This gives us

$$0 = h_{21}i_1 + v_2\left(h_{22} + \frac{1}{R_L}\right) \tag{12·66}$$

or

$$v_2 = -\frac{h_{21}i_1}{h_{22} + 1/R_L} \tag{12·67}$$

or

$$v_2 = -i_1\frac{h_{21}R_L}{h_{22}R_L + 1} \tag{12·68}$$

Substituting for v_2 from Eq. (12·65), we obtain

$$-i_2R_L = -i_1\frac{h_{21}R_L}{h_{22}R_L + 1} \tag{12·69}$$

or

$$A_i = \frac{i_2}{i_1} = \frac{h_{21}}{h_{22}R_L + 1} \tag{12·70}$$

Input resistance. To derive the input resistance, we can substitute Eq. (12·68) into Eq. (12·58*a*). This will remove v_2 and leave only v_1 and i_1 in the equation. The ratio of v_1 to i_1 then is R_{in}. Performing the substitution, we have

$$v_1 = i_1\left(h_{11} - h_{12}\frac{R_Lh_{21}}{R_Lh_{22} + 1}\right) \tag{12·71}$$

$$v_1 = i_1\frac{h_{11}h_{22}R_L + h_{11} - h_{12}R_Lh_{21}}{R_Lh_{22} + 1} \tag{12·72}$$

$$v_1 = i_1\frac{R_L\Delta + h_{11}}{h_{22}R_L + 1} \tag{12·73}$$

The input resistance, then, is

$$R_{\text{in}} = \frac{v_1}{i_1} = \frac{R_L\Delta + h_{11}}{h_{22}R_L + 1} \tag{12·74}$$

where Δ is equal to $h_{11}h_{22} - h_{12}h_{21}$

Voltage gain. We can derive the voltage gain for the circuit by using Eqs. (12·68) and (12·73).

$$\text{Voltage gain} = A_v = \frac{v_2}{v_1} = -\frac{R_Lh_{21}}{R_L\Delta + h_{11}} \tag{12·75}$$

Power gain. Equally easy to derive is the equation for the power gain. This is simply the product of the current and voltage gains.

$$\text{Power gain} = A_iA_v = \left(\frac{h_{21}}{h_{22}R_L + 1}\right)\left(\frac{R_Lh_{21}}{R_L\Delta + h_{11}}\right)$$
$$= \frac{h_{21}{}^2R_L}{(1 + h_{22}R_L)(R_L\Delta + h_{11})} \tag{12·76}$$

Output resistance. To obtain the governing equation for output resistance, we have to transfer the signal generator to the output terminals. At the same time, we place only R_g across the input terminals, Fig. 12·30. The equations for this arrangement are

$$0 = (R_g + h_{11})i_1 + h_{12}v_2 \tag{12·77}$$
$$i_2 = h_{21}i_1 + h_{22}v_2 \tag{12·78}$$

Note that while Eq. (12·78) does not make any overt mention of R_L, this quantity is contained in v_2, as indicated in Eq. (12·65). It is

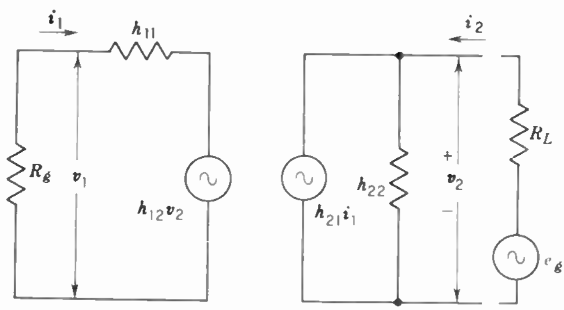


Fig. 12·30

tacitly assumed in Eq. (12·78) that v_2 is developed by generator e_g which is now in the output circuit.

The required information needed is the ratio of v_2/i_2, since this equals the output impedance seen when looking back into the network from the output terminals.

The first step is to rearrange Eq. (12·77) to the form

$$i_1 = -\frac{h_{12}v_2}{R_g + h_{11}} \tag{12·79}$$

This is then substituted into Eq. (12·78):

$$i_2 = -\frac{h_{12}h_{21}v_2}{R_g + h_{11}} + h_{22}v_2 \tag{12·80}$$

$$i_2 = v_2\left(-\frac{h_{12}h_{21}}{R_g + h_{11}} + h_{22}\right) \tag{12·81}$$

The ratio, then, of v_2 and i_2 is

$$\frac{v_2}{i_2} = \frac{R_g + h_{11}}{(R_g + h_{11})h_{22} - h_{12}h_{21}} \tag{12·82}$$

$$\frac{v_2}{i_2} = \frac{R_g + h_{11}}{R_g h_{22} + \Delta} \tag{12·83}$$

where $\Delta = h_{11}h_{22} - h_{12}h_{21}$.

It is interesting to see that the input impedance, from Eq. (12·74), is dependent upon R_L, while the output impedance, from Eq. (12·83), is dependent on R_g. This, of course, is in keeping with the results previously derived by using resistance parameters.

In deriving Eqs. (12·70), (12·74), (12·75), (12·76), and (12·83), the general circuit of Figs. 12·29 and 12·30 was used. This can represent a common-base, common-emitter, or common-collector arrangement by simply using the proper parameters for each circuit arrangement. This can be better shown formula-wise by using the notation

$$h_{11} = h_i \qquad h_{21} = h_f$$
$$h_{12} = h_r \qquad h_{22} = h_o$$

Thus, the input resistance from Eq. (12·74) can be expressed as

$$R_{\text{in}} = \frac{R_L\Delta + h_{ie}}{h_{oe}R_L + 1} \tag{12·84}$$

This is the input resistance for the common-emitter circuit. By using h_{ib} and h_{ob} in place of h_{ie} and h_{oe}, the same equation will represent the input resistance for a common-base amplifier. And finally, by substituting h_{ic} and h_{oc}, we obtain the input resistance for a common-collector arrangement. (Δ changes in value, too.)

The remaining equations can be written (for the common-emitter configuration) as

$$A_i = \frac{h_{fe}}{h_{oe}R_L + 1} \tag{12·85}$$

$$A_v = -\frac{R_L h_{fe}}{R_L\Delta + h_{ie}} \tag{12·86}$$

$$\text{Power gain} = \frac{h_{fe}^2 R_L}{(1 + h_{oe}R_L)(R_L\Delta + h_{ie})} \tag{12·87}$$

$$\text{Output resistance} = \frac{R_g + h_{ie}}{R_g h_{oe} + \Delta} \tag{12·88}$$

Relationship between Hybrid and Resistance Parameters

Now that we have the hybrid h factors, the next question might be, "How can these be converted into the resistance parameters, r_e, r_b, r_c, and r_m?" The answer follows directly if we remember that Eqs. (12·58) and (12·59) can represent a transistor in the common-, or grounded-, emitter configuration. Hence, all we need do is set up the resistance parameters in a similar common-emitter equivalent circuit, obtain the governing equations, and then manipulate (12·58) and (12·59) to the same form. Once this is done, the equivalence between the h and r factors becomes quite evident. The following discussion indicates how this is carried out.

The equivalent circuit of a transistor, in a common-emitter configuration, using resistance parameters, is shown in Fig. 12·31. A signal generator is not being used at the input, nor a load resistor at the output, in order to place this circuit on an equal footing with the circuit of Fig. 12·28. Actually, the generator impedance R_g and the

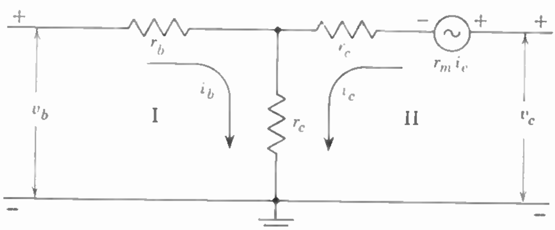


Fig. 12·31

load resistance R_L have no significance so far as the conversion of h to r factors is concerned. Hence, they can be omitted.

The governing equations for Fig. 12·31 are derived by methods identical with those previously explained for Figs. 12·23 and 12·24. Thus, for loop I we have,

$$-v_b + r_b i_b + i_b r_e + i_c r_e = 0$$

or

$$v_b = (r_b + r_e)i_b + r_e i_c \tag{12·89}$$

For loop II the equation is

$$-v_c + r_m i_e + r_c i_c + r_e i_c + r_e i_b = 0$$

But

$$i_e = -(i_b + i_c)$$

Substituting this in the above equation, we obtain

$$v_c = -r_m i_b - r_m i_c + r_c i_c + r_e i_c + r_e i_b$$

or

$$v_c = (r_e - r_m)i_b + (r_c + r_e - r_m)i_c \tag{12·90}$$

Now, if we compare Eqs. (12·58) and (12·59) with (12·89) and (12·90), we see that their forms are not comparable. The next step is to rearrange Eqs. (12·58) and (12·59) so that all the variables, such as v_b, v_c, i_b, and i_c, are assembled in the same order as that possessed by Eqs. (12·89) and (12·90). Here is how this is done. Equation (12·58) is

$$v_b = h_{11}i_b + h_{12}v_c \tag{12·58}$$

Since we are dealing here with a common-emitter arrangement, h_{11} will be converted to h_{ie}, h_{12} will be converted to h_{re}, h_{21} will be changed to h_{fe}, and h_{22} will be written as h_{oe}. Doing this, we have (for the equation just written)

$$v_b = h_{ie}i_b + h_{re}v_c \tag{12·91}$$

This contains v_c, which can be replaced from (12·59).

$$i_c = h_{fe}i_b + h_{oe}v_c \tag{12·59}$$

Rearranging, we have

$$v_c = \frac{i_c}{h_{oe}} - \frac{h_{fe}i_b}{h_{oe}} \tag{12·59b}$$

Now, we substitute this for v_c in Eq. (12·58). Doing this gives us

$$v_b = h_{ie}i_b + \frac{h_{re}}{h_{oe}} i_c - \frac{h_{re}h_{fe}}{h_{oe}} i_b \tag{12·92}$$

Or, collecting similar terms,

$$v_b = \left(h_{ie} - \frac{h_{re}h_{fe}}{h_{oe}}\right) i_b + \frac{h_{re}}{h_{oe}} i_c \tag{12·93}$$

This is one of the equations needed. The other one is (12·59*b*) transposed:

$$v_c = -\frac{h_{fe}}{h_{oe}} i_b + \frac{i_c}{h_{oe}} \tag{12·59c}$$

These are the hybrid equations rearranged to the same form as (12·89) and (12·90), and a direct comparison is now possible. Thus, (12·89) and (12·93) are similar in v_b, i_b, and i_c, and since they deal with the same transistor in the same configuration (here, common emitter), the coefficients of similar terms should be equal. Hence

$$h_{ie} - \frac{h_{re}h_{fe}}{h_{oe}} = r_b + r_e \tag{12·94}$$

and

$$\frac{h_{re}}{h_{oe}} = r_e \tag{12·95}$$

The same comparison can be made between Eqs. (12·90) and (12·59c). This gives us

$$-\frac{h_{fe}}{h_{oe}} = r_c - r_m \tag{12·96}$$

and

$$\frac{1}{h_{oe}} = r_e + r_c - r_m \tag{12·97}$$

Further simplification follows from the fact that $r_e = h_{re}/h_{oe}$ by Eq. (12·95). Using this information in (12·94), we can readily determine the hybrid equivalent of r_b.

$$h_{ie} - \frac{h_{re}h_{fe}}{h_{oe}} = r_b + \frac{h_{re}}{h_{oe}}$$

or

$$r_b = \frac{h_{ie}h_{oe} - h_{re}(1 + h_{fe})}{h_{oe}} \tag{12·98}$$

Also, substitution of $r_e = h_{re}/h_{oe}$ in (12·96) will give r_m. Thus

$$-\frac{h_{fe}}{h_{oe}} = \frac{h_{re}}{h_{oe}} - r_m$$

or

$$r_m = \frac{h_{re} + h_{fe}}{h_{oe}} \tag{12·99}$$

Now, with all this information, we can derive the value for r_c alone by suitable substitution in (12·97).

$$\frac{1}{h_{oe}} = \frac{h_{re}}{h_{oe}} + r_c - \frac{h_{re} + h_{fe}}{h_{oe}}$$

$$r_c = \frac{h_{fe} + 1}{h_{oe}} \tag{12·100}$$

Similar comparisons can be made among all of the parameters so long as the precautions outlined above are kept in mind.

Coupling and Bypass Capacitors

The discussion thus far has dealt with the various aspects of transistor-amplifier design at low frequencies. Before leaving this subject for a consideration of high-frequency transistor-amplifier design, special mention should be made of the problem of selecting the coupling capacitor and the emitter bypass capacitor. Both of these capacitors determine the low-frequency cutoff of the transistor amplifier. The transistor itself has a response which is flat down to and including direct current (zero frequency). However, the coupling and bypass capacitors, acting with the collector and bias resistors and the transistor input resistance, set a lower limit to the attainable frequency response of the overall amplifier.

Figure 12·32 illustrates a two-stage *RC*-coupled amplifier, together with typical component values. The bypassed emitter resistor stabilizes the operating point, as previously discussed. R_C is the collector resistor of the stage, while R_{B1} and R_{B2} form a voltage divider which, in conjunction with R_E, establishes the operating point. The coupling capacitor C_K should have such a value that the signal current leaving the collector of the first stage is adequately coupled into the base of the second stage. The emitter bypass capacitor C_E should possess such a value that emitter degeneration caused by R_E is negligible at the lowest frequency of interest; that is, R_E should be adequately bypassed at all frequencies of interest.

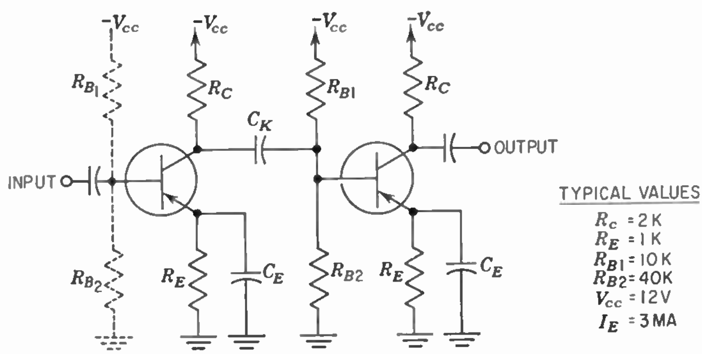


Fig. 12·32 Two-stage RC-coupled amplifier.

It is beyond the scope of this book to derive the formulas for the required values of C_K and C_E; however, a discusssion of these formulas is in order. The equation for C_K (in farads) is given by

$$C_K = \frac{1}{2\pi f_{3\text{db}}\left(R_C + \dfrac{R_{B_t}R_{\text{in}}}{R_{B_t} + R_{\text{in}}}\right)} \tag{12·101}$$

and for C_E by

$$C_E = \frac{\beta_0 + 1}{2\pi f_{3\text{db}}\left(R_{\text{in}} + \dfrac{R_C R_{Bt}}{R_C + R_{B_t}}\right)} \tag{12·102}$$

where $f_{3\text{db}}$ = the desired 3-db low-frequency cutoff

R_{B_t} = the parallel combination of R_{B_1} and $R_{B_2} = \dfrac{R_{B_1}R_{B_2}}{R_{B_1} + R_{B_2}}$

R_{in} = the input resistance of the transistor, h_{ie}

β_0 = the low-frequency, small-signal current gain of the transistor, h_{fe}.

There are some very important design considerations to be noted here, so let us consider Eqs. (12·101) and (12·102) separately.

The coupling capacitor. The value of the coupling capacitor is seen to depend upon the collector resistor of the preceding stage and the bias resistors and input resistance of the following stage. The dominant factor, however, is the collector resistor of the preceding stage. The input resistance is dependent upon the operating point of the transistor; R_{in} will decrease as the emitter bias current increases. Consider the typical component values indicated in Fig. 12·32 and assume a typical input resistance at 3 ma emitter current of $R_{in} = 500$ ohms. In order to obtain a low-frequency 3-db cutoff of 10 cps, Eq. (12·101) predicts that a coupling capacitor of

$$C_K = \frac{1}{2\pi \times 10\left(2{,}000 + \dfrac{8{,}000 \times 500}{8{,}000 + 500}\right)} = \frac{1}{20\pi(2{,}000 + 471)} = 6.45\ \mu\text{f}$$

is required. Note that this value depends primarily on the value of R_c (2,000 ohms) of the first stage. This differs considerably from vacuum-tube circuit practice, where the required capacitance depends primarily upon the input resistance of the following stage. Note also that a rather large value of capacitance is required for the transistor amplifier. Whereas coupling capacitors on the order of 0.01 to 0.1 μf are generally quite adequate at this frequency for vacuum-tube amplifiers, transistor amplifiers require values from 1μf to 10 μf.

The emitter bypass capacitor. In a similar manner, let us calculate the required value of emitter bypass capacitor C_E. Again using 10 cps as the lower cutoff frequency, we see from Eq. (12·102) that a transistor having a small-signal current gain β_0 of 50 requires an emitter bypass capacitor of

$$C_E = \frac{50 + 1}{2\pi \times 10\left(\dfrac{2{,}000 \times 8{,}000}{2{,}000 + 8{,}000} + 500\right)} = \frac{51}{20\pi(1{,}600 + 500)} = 387\ \mu\text{f}$$

This is quite large compared with capacitor values required in vacuum-tube circuits where the bypass capacitor is totally dependent upon the size of the cathode resistor. A capacitor of only 25 to 75 μf would be required with a vacuum-tube circuit to obtain the same low cutoff frequency.

The value of C_E is primarily dependent on the current gain of the transistor β_0 and secondarily dependent upon the collector resistance of the preceding stage. It is also influenced by the input resistance of the transistor, which, in turn, is operating-point dependent. Finally, note that the value of C_E is entirely independent of the size of the emitter resistance R_E, quite the opposite from vacuum-tube theory.

The reader will find it well worth his time to study Eqs. (12·101) and (12·102) in detail, especially from a practical viewpoint. For example, consider the case where stages 1 and 2 are transformer-coupled and stage 2 is a power-output stage with an emitter bias current of 24 ma. Then the transistor input resistance R_{in} will typically be 100 ohms and the value of R_L reflected through the transformer may well be 10 ohms or less. Now, let us calculate the value of C_E required for a low-frequency cutoff of 10 cps.

$$C_E = \frac{50 + 1}{2\pi \times 10\left(\dfrac{8{,}000 \times 100}{8{,}000 + 100} + 100\right)} = \frac{51}{20\pi(99 + 100)} = 4{,}080\ \mu\text{f}$$

This value is so large it is impractical.

High-frequency Response

Knowledge of Eqs. (12·101) and (12·102) gives one the ability to calculate the low-frequency cutoff of an *RC*-coupled transistor amplifier. The other frequency response of interest is the high-frequency cutoff. The high-frequency response of a transistor amplifier is determined primarily by the transistor itself and to some extent by the external circuitry. Early transistors were limited in frequency response to the kilocycle range; however, present-day transistors are used in FM and TV tuners in the 100- to 400-Mc range and useful operation into the kilomegacycle range has been achieved.

In the case of the vacuum tube, the upper-frequency response limitations are set by the interelectrode capacitances. This is not so with the transistor because, while the vacuum tube is basically a voltage-operated device, the transistor is essentially a current-operated device. To change the current through a transistor from one value to another requires the movement of charge to or from the base region of the transistor. This movement of charge takes a finite amount of time. For example, consider the illustration of Fig. 12·33. V_{EE}, V_{CC}, R_{E1}, R_{E2}, and R_C are adjusted to provide the desired operating point. V_G and R_G constitute a sine-wave signal source.

If the applied sine-wave signal goes positive, the emitter current and the collector current will increase. This means a movement of charge into and through the base region of the transistor. This movement takes time. Another way of regarding this action is to note that the base of the transistor acts like a capacitor. In order to increase the current through the transistor, one must increase the total charge in the base region (capacitor). When the applied sine wave goes negative, the reverse process occurs. This implies the movement of charge out of the base region of the transistor.

At low frequencies, the output sine wave will be an amplified version of the input and will be exactly in phase with the input. However, if the signal frequency is steadily raised, the output amplitude will not only decrease but will also shift in phase with respect to the input signal. In fact, if the input signal is held constant and a plot is made

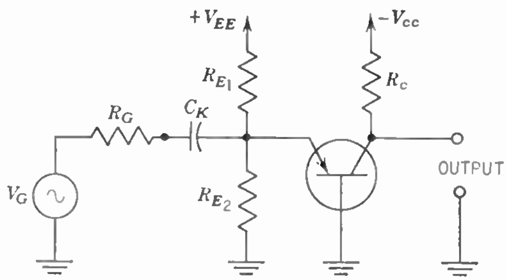


Fig. 12·33 Common-base amplifier.

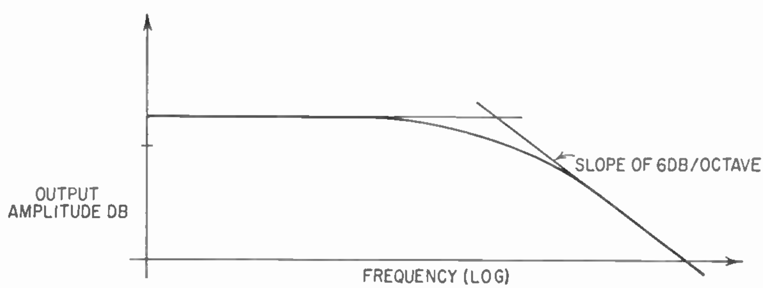


Fig. 12·34 Output amplitude vs. frequency.

of the output signal vs. frequency, the curve shown in Fig. 12·34 will be obtained.

Let us consider why the gain of the transistor falls off with increasing frequency. First, observe what happens at lower frequencies. As the signal amplitude is increased, i.e., as the sine wave goes positive, charge moves into the base region and the collector current increases. When the signal decreases, i.e., the sine wave goes negative, charge moves out of the base region and the collector current decreases correspondingly. This seems quite simple; however, one must realize that charge flows into and out of the transistor base region by the process of diffusion. Current flow by diffusion, unlike the normal current flow in a wire due to an applied emf, is a random, undirected process. Therefore, the time required to move charges into and out of the base region by diffusion can become important. As the frequency is increased, the time between the positive and negative portions of the sine wave becomes shorter and shorter. Eventually, the frequency reached is such that the positive portion of the sine wave persists for a time which is shorter than that required to move the appropriate amount of charge into the base region. Thus, when the sine-wave signal begins to go negative, the charge in the base is still trying to increase. At this point, the transistor is not responding

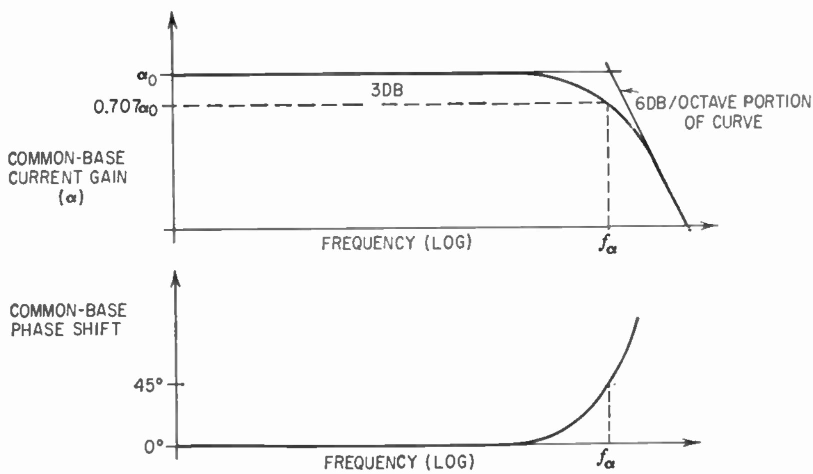


Fig. 12·35 Common-base current gain and phase shift vs. frequency.

properly to the input signal, resulting in a decrease in transistor gain and a definite phase shift between the input and output signals.

A plot of the gain and phase characteristics of the transistor in the common-base connection is shown in Fig. 12·35. This curve is obtained with an a-c open-circuited input (current signal source) and an a-c short-circuited output.

The 3-db upper cutoff frequency in the common-base arrangement is referred to as the α cutoff frequency f_α. The gain above this frequently falls off at 6 db per octave; i.e., the gain drops 6 db every time the frequency is doubled. f_α is the frequency at which the small-signal current gain is equal to 0.707 of its value at low frequencies α_0; it is also the frequency at which the phase shift introduced by the transistor is 45°.

A similar set of curves for the common-emitter configuration is shown in Fig. 12·36. Here again, the curves are for an a-c open-circuited input and an a-c short-circuited output.

The 3-db upper cutoff frequency in the common-emitter amplifier is called the β cutoff frequency f_β. The gain at frequencies above f_β falls off at 6 db per octave. Another important parameter shown in Fig. 12·36 is f_T. This is the frequency at which the common-emitter current gain is equal to 1. The value of f_T is referred to as the gain-bandwith product for the following reason: If the current gain is measured at any frequency along the 6-db per octave roll-off, then

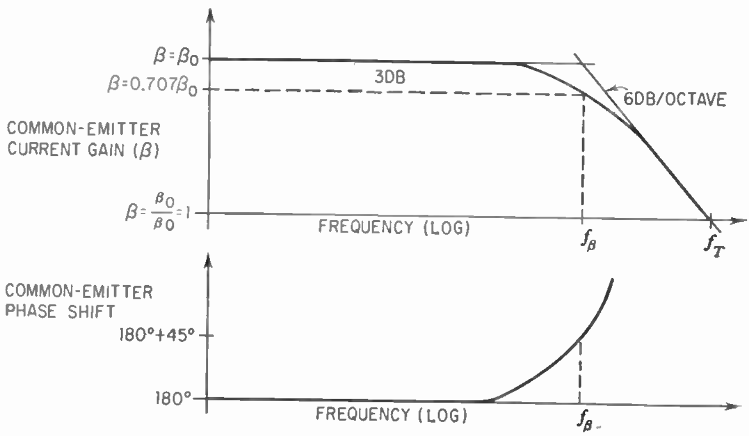


Fig. 12·36 Common-emitter current gain and phase shift vs. frequency.

the product of the gain and the frequency of measurement will be a constant, namely, f_T. In fact, an expression for the common-emitter current gain at any frequency above f_β is

$$\beta = \frac{\beta_0}{\sqrt{1 + \left(\frac{\beta_0 f}{f_T}\right)^2}}$$

where β_0 = low-frequency current gain, h_{21}
f = frequency of interest
f_T = gain-bandwidth product

Note that the current gain will be 0.707 β_0 when $f = f_T/\beta_0$; thus, this frequency is the value of f_β.

For the common-emitter configuration, the primary criterion in determining the frequency response of the transistor is f_T, the gain-bandwidth product. However, note that the upper cutoff frequency

f_β is a function of the low-frequency current gain, as shown in Fig. 12·37.

Thus, if three transistors are obtained, each with the same value of f_T (gain-bandwidth product) but with different values of β_0, and if each unit is inserted into a transistor amplifier, the upper cutoff frequency of the amplifier will be lowest for the high-β unit and highest for the low-β unit. Furthermore, it is interesting to note that in the common-base configuration, the current gain α_0 is less than 1 but the

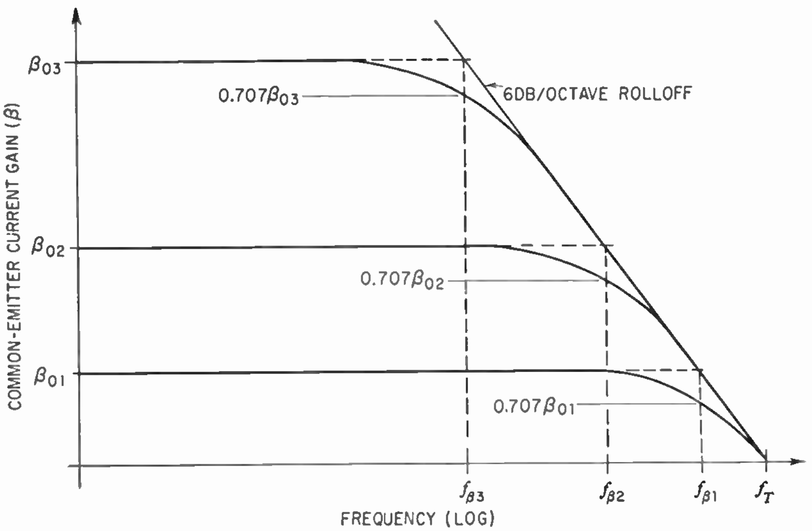


Fig. 12·37 Common-emitter gain vs. frequency for three low-frequency transistors with the same f_T.

high-frequency cutoff f_α can be high; in the common-emitter configuration, the current gain β is high but the high-frequency cutoff f_β can be low. For example, consider the 2N207, for which the following typical values hold:

Common base	$\alpha = 0.99$	$f_\alpha = 2$ Mc
Common emitter	$\beta = 100$	$f_\beta = 19$ kc

One additional point is of interest. In transistor-amplifier design, the high-frequency figure of merit is f_α in the common-base connection and f_T in the common-emitter connection. For any given transistor, values of f_α and f_T are not equal. For homogenous-base transistors such as alloy-junction types, $f_T \approx 0.85\ f_\alpha$, and for diffused-base transistors such as MADT and mesa types, $f_T \approx 0.5\ f_\alpha$. One must be sure to use f_T when designing a common-emitter amplifier. f_α should be used only when designing a common-base amplifier.

The foregoing facts are very useful in estimating the high-frequency cutoff of a transistor amplifier under the specified conditions, namely, that the input appears open-circuited to the signal and the output is a-c short-circuited to the signal. However, in a practical case, the respective input and load resistances need not be infinite and zero, as previously stated. The above estimates will hold so long as the source resistance feeding the base is very large and the load resistance in the collector circuit is small. Let us be more specific in what we mean by "large" and "small." In stating that the source should be large,

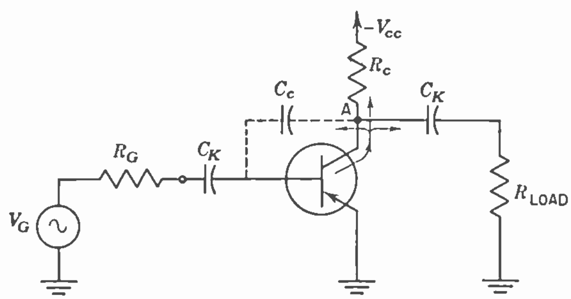


Fig. 12·38 Circuit showing feedback caused by C_c.

we mean that it must be large compared to the input resistance of the transistor. Usually, a value of source resistance of 3,000 ohms or greater will suffice. By the same token, the load resistance can be considered small if it is small in comparison to the output resistance of the transistor. Usually a value of 1,000 ohms or less will suffice. For example, the circuit of Fig. 12·32 almost meets these conditions. The source resistance of the stage is the collector resistor of the preceding stage, approximately 2,000 ohms. The load resistor of the stage is the input resistance of the following stage, approximately 500 ohms. Thus, for the transistor-amplifier stage shown in Fig. 12·32, the upper cutoff frequency is approximately $f_\beta = f_T/\beta$.

At this point the reader might well ask what happens to the upper cutoff frequency when the source resistance is not large, the load resistance is not small, or both. This question will be answered in detail presently, but first let us consider other transistor parameters which will affect the upper cutoff frequency of a transistor amplifier.

Collector-junction capacitance. In the circuit of Fig. 12·38, a capacitance C_c is shown between collector and base. The presence of this capacitance comprises a feedback path between the output (collector) and the input (base) of the transistor. Since the common-emitter amplifier exhibits a 180° phase shift from input to output, the resulting feedback signal is 180° out of phase with the input signal. The effect of C_c, then, is to introduce negative feedback which subtracts from the input signal and reduces the output signal. Furthermore, since the capacitive reactance of C_c decreases with increasing frequency, the effects of the feedback become more pronounced at the higher frequencies.

There is another important fact concerning the effect of the collector-junction capacitance C_c. The output current which leaves the collector travels to point A and there divides: part goes to the load and part goes through C_c back to the base. The current at point A

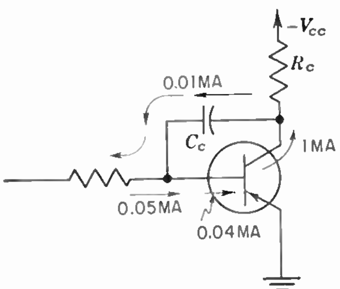


Fig. 12·39 Effects of C_c on output signal.

will divide according to the impedance of the respective current paths. If the load resistance R_c is small compared to the reactance of C_c at the frequency of interest, most of the current will flow into the load, the desirable effect. In this case, the effect of C_c is small and the frequency response is set by f_T, as previously explained. However, if the load resistance is large, much of the output current will be fed back through C_c and a loss in gain will occur. That is why f_T is a valid measure of the high-frequency response only when the output is a-c short-circuited or when the load resistance is small. The effect of C_c is quite important because the magnitude of its feedback current is actually multiplied by β, the current gain of the transistor. For example, consider the case of a transistor with a β of 20 and an output signal of 1 ma, Fig. 12·39. If 1 per cent of this output signal, 0.01 ma, is fed back to the base (which has a signal current of I_c/β, or 0.05 ma), the net base current after the feedback is subtracted is 0.04 ma. The collector current will then decrease from 1 ma to 0.04β, or 0.8 ma. In this case, a 1 per cent feedback factor reduces the output of the amplifier by 20 per cent. Thus, the effects of C_c can become quite devastating.

High-frequency equivalent circuit. In the preceding section, we discussed the two major parameters f_T and C_c involved in determining the high-frequency response of a transistor. In this section, a more exact method is presented, and it will enable us to calculate the gain and the bandwidth of a transistor amplifier for any values of source and load resistance. In order to do so, a high-frequency equivalent circuit of the transistor is developed. The components of this equivalent circuit are readily lumped with external circuit components. The resulting composite circuit is then useful in obtaining expressions for gain and frequency response.

As a first step in obtaining a useful equivalent circuit for the transistor. let us first isolate all of the internal resistances and capacitances associated with the device. Since the transistor is nothing more than two semiconductor diodes connected back to back, an equivalent circuit must certainly contain the resistances and capacitances associated with each diode. However, transistor action is not obtained by simply interconnecting two semiconductor diodes; the very close spacing of the base region is most essential.

One method of acknowledging this mutual coupling between the diodes circuit-wise is to include a current generator in the output, where the amount of current generated depends on the input voltage or current. Another important fact which must be considered is that the base region itself stores charge when the transistor is in operation. This was previously mentioned in our discussion of effects which limit the frequency response of the transistor. Figure 12·40 shows the physical picture of a typical alloy-junction transistor. Within this picture are included all the resistances and capacitances affecting the gain and frequency response of the transistor. Although they are shown as lumped components in the model, it must be realized that these resistances and capacitances are actually distributed throughout the emitter, base, and collector regions.

Assuming that the transistor shown in Fig. 12·40 is normally biased for common-emitter operation, let us discuss each of the lumped parameters indicated in both their physical and operational aspects. Starting with the input to the base, the first component which the signal encounters is a resistor $r_{bb'}$. This resistance is referred to as the base-spreading resistance; it is the physical resistance of the bulk semiconductor material which exists between the external base connection and the active base region. The active base region is defined as the region directly between the emitter and collector electrodes.

All of the carriers which flow between the emitter and the collector must flow through this region. Since we wish to control the flow of these carriers, the signal voltage or current must be applied between the active region and the common terminal (emitter in this case). In order to obtain a high value of β, the resistivity of the base region is usually kept high compared to the resistivity of the material in the emitter and collector regions (commonly 10 to 1,000 times higher). Also, the base region is very thin; a typical value for base width is ¼

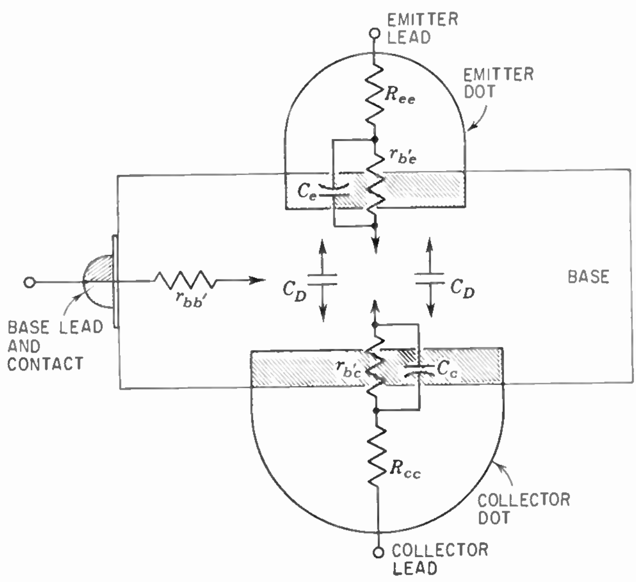


Fig. 12·40 Physical representation of a typical junction transistor.

mil to 2 mils. Thus, even though the length of this resistance path is small, its high resistivity in conjunction with its small cross-sectional area results in a value of $r_{bb'}$ which may well exceed 100 ohms. In a homogeneous-base transistor, typical values of $r_{bb'}$ range from 100 to 500 ohms. In a diffused-base transistor, where part of the base is purposely made of low-resistivity material, typical values of $r_{bb'}$ range from 10 to 200 ohms depending upon the intended application. A low value of $r_{bb'}$ is especially essential in transistors which are to be used in oscillator applications. This can be seen from Eq. (12·103), which is the expression for the maximum frequency at which a transistor can oscillate.

$$f_{\max} = \frac{\alpha_0 f_T}{8\pi r_{bb'} C_c} \tag{12·103}$$

where α_0 = common-base low-frequency current gain
f_T = gain-bandwidth product
$r_{bb'}$ = base-spreading resistance
C_c = collector-junction capacitance

If $r_{bb'}$ could be reduced to zero, the transistor could theoretically oscillate at any high frequency. Although this would be true for the isolated transistor, the highest frequency of oscillating would still be limited by the stray capacitances and resistances of the external circuit. In fact, some transistors that are available today require specially constructed circuits and housings in order to attain their high-frequency capabilities. In amplifier applications, $r_{bb'}$ is important only when the source resistance is small; generally, if the source

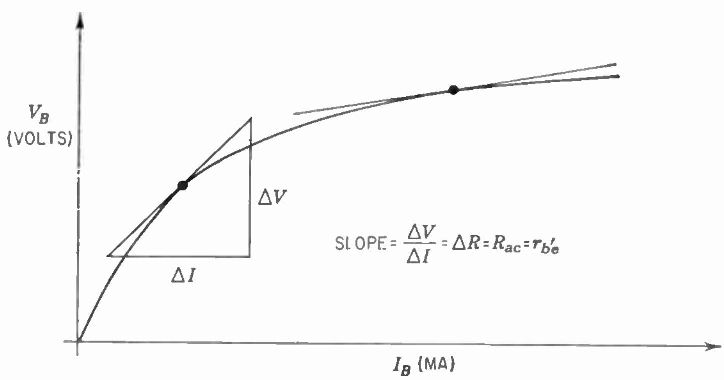


Fig. 12·41 Input *VI* characteristic. (Common emitter.)

resistance is above one or two thousand ohms, $r_{bb'}$ may be neglected.

After the signal passes $r_{bb'}$, it arrives at the active base region where it sees a host of resistances and capacitances. Both the emitter and collector junctions have small-signal a-c resistance and capacitance associated with them. Since the emitter-base junction is forward-biased and the collector-base junction is reverse-biased, it is reasonable to expect that the values of these junction resistances and capacitances will be different, and such is the case. Let us consider each junction separately.

At the forward-biased emitter-base junction, the small-signal a-c resistance is low. This resistance $r_{b'e}$ exists in the equivalent circuit, as it does in the physical transistor, between the active base region and the grounded emitter. Its value can be determined by finding the slope of the input *VI* characteristic of the grounded-emitter transistor at the correct operating point. An example of this is shown in

Fig. 12·41. Note that the slope of the characteristic curve changes for different values of base current. The slope is given by V_{BE}/I_B, and this is $r_{b'e}$. Generally, the size of this resistance ranges between 100 and 3,000 ohms.

The junction capacitance associated with the emitter-base diode depends upon the area of the junction and the voltage applied to the junction. The junction capacitance can be compared to a familiar parallel-plate capacitor, whose capacitance is given by

$$C = \frac{\epsilon A}{d} \tag{12·104}$$

where A = area of the plates
d = spacing between the plates
ϵ = dielectric constant of the material between the plates

Capacitance is proportional to the plate area and inversely proportional to the plate spacing. In the single-transistor model, the two

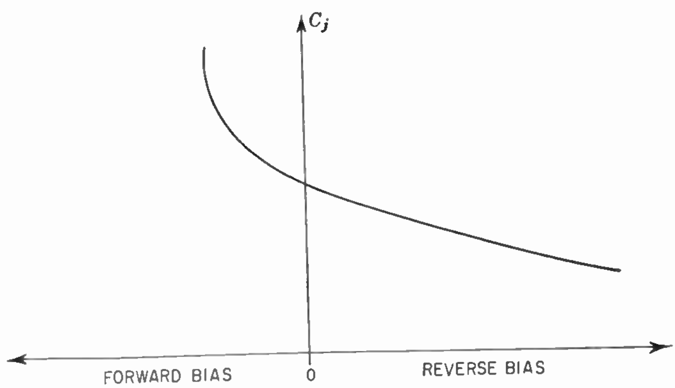


Fig. 12·42 Diode junction capacitance C_j vs. bias.

plates are the emitter region and the active base region. Between them is the depletion layer whose width is governed by two things: the resistivity of the base in the vicinity of the junction and the applied voltage. A curve of junction capacitance vs. voltage for both a forward- and reverse-biased diode is shown in Fig. 12·42, and it can be seen that the junction capacitance decreases for increasing reverse bias. This is because the width of the depletion layer increases with increasing reverse bias. Since the emitter-base junction is forward-biased, its capacitance is larger than that of any reverse-biased diode such as the collector-base junction.

Now consider the collector-base junction. As in the emitter-base junction, there is a small-signal resistance and capacitance. However, since the collector-base junction is reverse-biased, the resistance $r_{b'c}$ is quite large, typically varying from 10,000 ohms to 10 megohms. In many applications, this is so large it can be ignored.

Because the width of the depletion layer for a reverse-biased junction is large, the collector-base junction capacitance is small. Even though the area of the collector junction is usually larger than that of the emitter junction, the effects of the depletion layers are such that the collector-junction capacitance is generally one-sixteenth to one-half that of the emitter-junction capacitance. Typical values of collector-junction capacitance for an alloy-junction transistor are 5 to 25 $\mu\mu$f. For diffused-base transistors, such as the MADT, PADT, and mesa, the collector-junction capacitance is typically ½ to 5 $\mu\mu$f. This is because the collector area is quite small and the resistivity of the material on the collector side of the base is high. However, in a diffused-base transistor, the resistivity of the base material on the emitter side is quite low. This leads to a small depletion-layer width and a large value of emitter-junction capacitance. In such transistors, the size of the emitter-junction capacitance in the forward bias region may well run 25 to 100 $\mu\mu$f. As previously mentioned, both the emitter-junction capacitance and the collector-junction capacitance limit the high-frequency response of the transistor.

There is still another capacitance associated with the transistor: the diffusion capacitance C_D. C_D is not a physical capacitance in the same sense as a parallel-plate capacitor, but it is an equivalent capacitance representing the charge stored in the active base region. Recall that charge within the active base region must be changed in order to change the collector current of a transistor. As one might expect, the amount of charge stored within the active base region, and thus the equivalent value of diffusion capacitance, is dependent upon the base width of the transistor. As the base width is made smaller, less charge can be stored within the active base region. Therefore, for smaller base widths, the diffusion capacitance will be smaller and the frequency response of the transistor will be higher.

Finally, there are two resistances R_{cc} and R_{ee}. These are spreading resistances, that is, the physical resistance of the bulk material comprising the collector and emitter electrodes. It was previously mentioned that the resistivity of the material within the emitter and collector electrodes was quite low compared to the resistivity in the base region. Therefore, one would expect the value of R_{ee} and R_{cc} to be quite small. In fact, in most practical situations, these resistances are neglected; however, the reader should be aware that they are present even though their value is usually under 10 ohms.

Now let us combine all of the physical resistors and capacitors into a lumped-parameter equivalent circuit. Such a circuit is shown in Fig. 12·43. This circuit is called the hybrid pi because its shape is similar to that of the Greek letter π. It includes all of the physical resistors and capacitors shown in Fig. 12·40. In addition, the equivalent circuit includes a current generator at the output between the emitter and collector. The signal from the current generator is equal to the negative of a constant g_m times a voltage $V_{b'e}$.

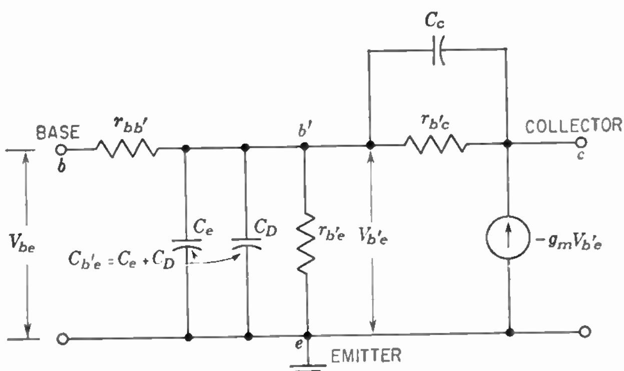


Fig. 12·43 Lumped small-signal-parameter equivalent circuit for normally biased common-emitter transistor.

The minus sign represents the 180° phase reversal between the input signal and the output signal. That is, if the input signal goes positive, the output voltage goes negative, and vice versa. The voltage $V_{b'e}$ is the internal base-to-emitter voltage. This is the voltage between the active base region and the emitter, and it differs from the external base-to-emitter voltage V_{be} by the drop across $r_{bb'}$. The point b' in the equivalent circuit represents the internal base of the transistor: the active base region. It is the voltage across this point, between b' and the grounded emitter, that determines the amount of signal current which will flow out of the collector.

The constant g_m is the transconductance of the transistor. It is quite similar to, and circuit-wise is analogous to, the transconductance of the vacuum tube. g_m relates the small-signal output current of the device to the small-signal input voltage and is given as

$$g_m = \frac{I_E}{27} \qquad \text{ma per volt} \tag{12·105}$$

where I_E = emitter current, ma

27 = temperature-dependent constant

At temperatures other than 25°C, the constant, 27, changes slightly. Since the transconductance of the transistor is dependent upon the emitter bias current I_E, it is dependent on the operating point.

Before attempting to derive expressions for amplifier gain and bandwidth by using the equivalent circuit shown in Fig. 12·43, let us review the components of this equivalent circuit and discuss their numerical values. $r_{bb'}$ is the base-spreading resistance. Its value varies with the type of transistor but, in general, is low and on the order of 100 ohms. C_e is the emitter junction capacitance, and C_D is the diffusion capacitance of the transistor. When the transistor is connected as an amplifier, the input signal to the base sees the two capacitors in parallel and cannot distinguish between them. From a circuit designer's viewpoint, it is their combined value which is of interest; this combination is referred to as $C_{b'e}$, the capacitance between the internal base region and the emitter, and is given by

$$C_{b'e} = \frac{g_m}{2\pi f_T} = C_e + C_d \qquad (12\cdot106)$$

where f_T = gain-bandwidth product of the transistor

g_m = transconductance

Since g_m and f_T are operating-point dependent, $C_{b'e}$ is dependent on the bias point. Since this capacitor is inversely proportional to f_T, the optimum operating point occurs where f_T takes on its highest value.

$r_{b'e}$ is the small-signal resistance of the internal base-emitter diode. Its value is low because the base-emitter diode is forward-biased. It is given by

$$r_{b'e} = \frac{\beta_0}{g_m} \qquad (12\cdot107)$$

where β_0 = low-frequency common-emitter current gain

g_m = transconductance

There is a subtle point of interest here. Note that the product of $r_{b'e}$ and $C_{b'e}$ forms a time constant

$$r_{b'e}C_{b'e} = \frac{\beta_0}{g_m}\frac{g_m}{2\pi f_T} = \frac{\beta_0}{2\pi f_T} = \frac{1}{2\pi f_\beta} \qquad (12\cdot108)$$

which is the inverse of the frequency response $2\pi f_\beta$ (or f_β). If all other components of the equivalent circuit were neglected, the frequency response would be f_β. It will be shown later that this case holds when the source resistance feeding the stage is very large and the load resistance of the stage is very small; these are the identical conditions mentioned during our definition of f_T and f_β. $r_{b'c}$ is the small-signal resistance of the reverse-biased collector-base diode. This resistor is so large that it is usually neglected in the calculation of gain and bandwidth of a transistor amplifier. C_c is the collector-junction capacitance which is dependent upon the collector-base voltage. In general, it is wise to minimize this capacitance. This is done in high-frequency transistor amplifiers by using supply voltages above 3 volts.

From the preceding discussion, it is seen that in order to properly design a high-frequency transistor amplifier, one must obtain the following values from the transistor data sheet (the values, of course, must be obtained at the operating point of interest): $r_{bb'}$, the base-spreading resistance; β_0, which is used to calculate the value of $r_{b'e}$; f_T, which is used to calculate the value of $C_{b'e}$; C_c, the collector-junction or barrier capacitance; and $r_{b'c}$, the collector-base resistance. The value of g_m, the transconductance, is set by the emitter bias current.

Before proceeding to the problem of obtaining expressions for the gain and bandwidth of a transistor-amplifier stage, let us compare the equivalent circuits of the transistor and the vacuum tube. The primary difference is noted on the input side of the two devices; whereas the input resistance of the tube is high and its input capacitance is low, the transistor has a low input resistance and a high input capacitance. It is interesting to note that the resultant *RC* time constant may be the same for the two devices.

Gain and Bandwidth Expressions

Now that an understanding of the equivalent circuit of the transistor has been established, it is possible and meaningful to derive the gain and bandwidth expressions for a common-emitter amplifier stage. In deriving these important expressions, we shall completely neglect the effects of $r_{b'c}$ by essentially removing it from the equivalent circuit; this is valid when dealing with load resistances of less than 10,000 ohms, a common condition.

Low-frequency current and voltage gain. To derive the expressions for the current and voltage gain of a single-stage common emitter, we shall first consider the equivalent circuit of Fig. 12·43 at low frequencies. We shall neglect all capacitances by assuming an operating frequency at which neither the high- nor the low-frequency effects come into play.

Let us start first with the current gain. The equivalent circuit under consideration is shown in Fig. 12·44. Current gain is defined as the ratio of the output signal current I_{out} to the input signal current I_{in}. The input current, which may be the signal from another stage or any other signal source, is represented by the current generator I_g and its parallel source resistance R_g. The output current is given by

$$I_{out} = -g_m V_{b'e} \tag{12·109}$$

where $V_{b'e}$ may be derived as

$$V_{b'e} = \frac{r_{b'e}}{r_{bb'} + r_{b'e}} \frac{I_g R_g (r_{bb'} + r_{b'e})}{r_{bb'} + r_{b'e} + R_g} = \frac{I_g R_g r_{b'e}}{r_{bb'} + r_{b'e} + R_g} \tag{12·110}$$

In spite of the formidable appearance of Eq. (12·110), it is simply the Ohm's law expression for the voltage developed across $r_{b'e}$ by the

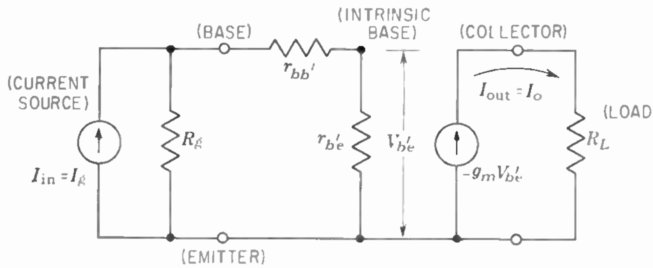


Fig. 12·44 Common-emitter low-frequency equivalent circuit with current source.

input current I_g. The circuit is complicated by the fact that R_g is in parallel with the series combinations of $r_{bb'}$ and $r_{b'e}$.

The current gain is

$$G_i = \frac{I_{out}}{I_g} = \frac{-g_m R_g r_{b'e}}{r_{bb'} + r_{b'e} + R_g} \tag{12·111}$$

The minus sign indicates that there is a 180° phase shift from input to output. If R_g becomes very large, this expression reduces to

$$G_i = -g_m r_{b'e} = \beta_0 \tag{12·112}$$

which has previously been shown to be the small-signal low-frequency current gain under conditions of large source resistance.

To derive the expression for the low-frequency voltage gain, consider the equivalent circuit shown in Fig. 12·45. Voltage gain is defined as the ratio of output signal voltage V_{out} to input signal voltage V_{in}. The output voltage is developed across the load R_L, and the input signal, which may be the signal from a preceding stage or any other source, is represented by the voltage generator V_g and its series resistance R_g. The output voltage is given by

$$V_{out} = I_{out}R_L = -g_m V_{b'e} R_L \tag{12·113}$$

where $V_{b'e}$ may be derived as

$$V_{b'e} = \frac{r_{b'e}V_g}{r_{bb'} + r_{b'e} + R_g} \tag{12·114}$$

Thus, the voltage gain is found to be

$$G_v = \frac{V_{out}}{V_g} = \frac{-g_m r_{b'e} R_L}{r_{bb'} + r_{b'e} + R_g} \tag{12·115}$$

Notice that the expressions for current and voltage gain differ only by the factor R_L/R_g; thus, the following relationship may be seen to exist as

$$G_v = G_i\left(\frac{R_L}{R_g}\right) \tag{12·116}$$

Upper cutoff frequency. It was previously explained that two effects control the upper cutoff frequency of the transistor: the gain-band-

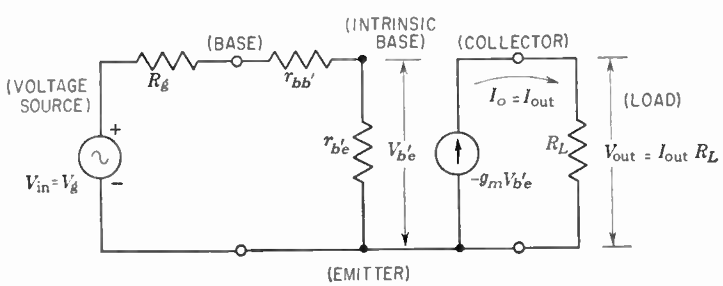


Fig. 12·45 Common-emitter low-frequency equivalent circuit with voltage source.

width product f_T, whose effect is shown by the diffusion capacitance C_D, and the collector-junction capacitance C_c. We wish to derive an expression for the upper cutoff frequency of the amplifier which will include the effect of both C_D and C_c. In order to accomplish this, a rather clever circuit trick is employed whereby the collector capacitance C_c is reflected back to the input side of the transistor equivalent circuit. This reflection of the feedback capacitance into the input is called the Miller-effect transformation, and it is similar to the transformation employed with the grid-to-plate capacitance in a vacuum tube. In order to visualize the transformation and its simplicity, consider the circuit of Fig. 12·46*a*.

If one breaks the equivalent circuit at points b' and e and looks to the right, he will see the circuit of Fig. 12·46*b*. The following equations may readily be written. The output voltage is

$$V_{out} = I_{out}R_L = -g_m V_{b'e} R_L \tag{12·117}$$

This may be rewritten as

$$\frac{V_{out}}{V_{b'e}} = -g_m R_L \tag{12·118}$$

Now, let us hold Eq. (12·118) in abeyance while we obtain an expression for the input impedance of the circuit of Fig. 12·46*b*. The

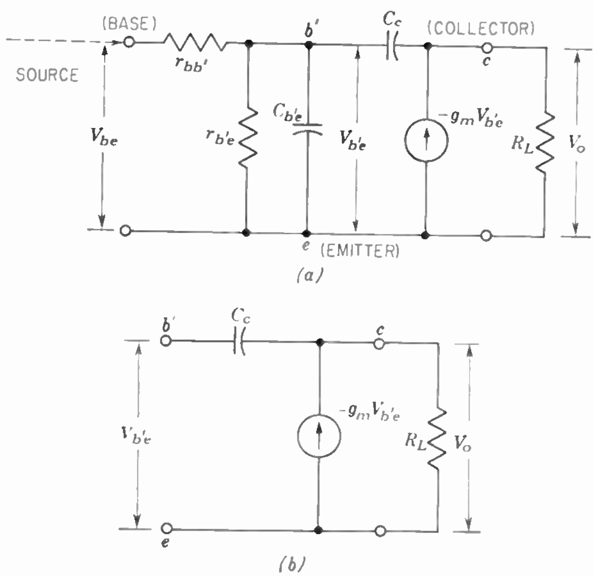


Fig. 12·46 (a) Equivalent circuit including capacitances. (b) Circuit to right of points b' and e.

input impedance would be Z_{in}, and this value, divided into $V_{b'e}$ would provide the current flowing into the circuit. Thus,

$$I = \frac{V_{b'e}}{Z_{in}} \tag{12·119a}$$

This same current must also flow through C_c. To obtain an expression for this capacitor current, note that C_c has voltage $V_{b'e}$ on one side and voltage V_{out} on the other. It is the difference between these voltages that will produce a current through C_c. Hence, we can write

$$I_c = \frac{V_{b'e} - V_{out}}{X_{c_c}} \tag{12·119b}$$

Since the currents of Eqs. (12·119a) and (12·119b) are equal, we may equate them:

$$\frac{V_{b'e}}{Z_{in}} = \frac{V_{b'e} - V_{out}}{X_{c_e}} \tag{12·119c}$$

or
$$Z_{in} = \frac{X_{c_e} V_{b'e}}{V_{b'e} - V_{out}} = \frac{X_{c_e}}{1 - V_{out}/V_{b'e}} = \frac{X_{c_e}}{1 + g_m R_L} \tag{12·120}$$

This may finally be written as

$$Z_{in} = \frac{1}{j\omega C_c} \frac{1}{1 + g_m R_L} \tag{12·121}$$

which implies an equivalent, or transferred, capacitance of

$$C = C_c(1 + g_m R_L) \tag{12·122}$$

From this it is seen that the effect of the feedback of C_c is to increase its size by $1 + g_m R_L$ when reflected across the input at points b' and e.

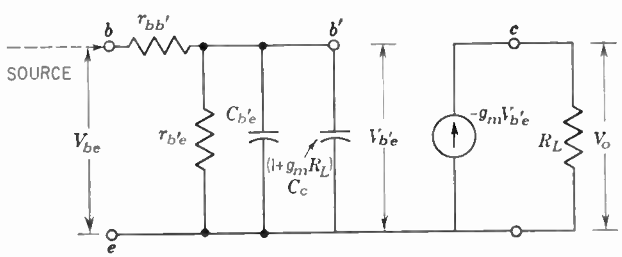


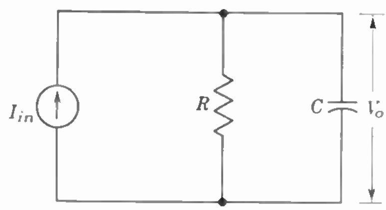
Fig. 12·47 Equivalent circuit with C_c reflected to input.

The result of this transformation is illustrated in Fig. 12·47, where the new equivalent circuit is shown. This circuit shows the effect of C_c transformed to the input; at the same time, the feedback path between input and output has been removed. The transformed equivalent circuit has many advantages, the most important of which is that the input circuit now takes the form of a simple parallel RC circuit. This gives rise to a simple expression for the upper cutoff frequency. Since the output current of the transistor is directly proportional to the internal base-to-emitter voltage $V_{b'e}$, the value of the output current will be down 3 db (0.707 of its low-frequency value) when the voltage $V_{b'e}$ is also down 3 db. Let us consider the simple circuit shown in Fig. 12·48. It is easily shown that the output voltage of this circuit will be down 3 db at an upper frequency which is given by

$$f_{3db} = \frac{1}{2\pi RC} \tag{12·123}$$

The input of the equivalent circuit of Fig. 12·47 is almost of this simple form. If one combines the effect of $r_{bb'}$ and R_g, the source resistance, an equivalent circuit which is a condensed form of Fig. 12·47 is obtained. This is shown in Fig. 12·49. The input to this equivalent circuit is of the same form as the simple circuit shown in Fig. 12·48;

Fig. 12·48 Single RC network.



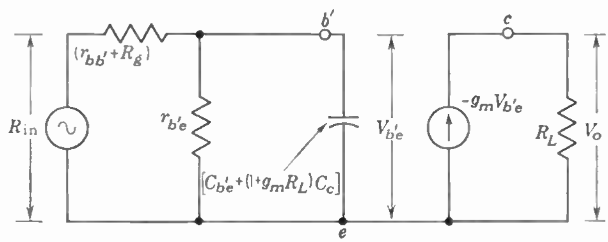


Fig. 12·49 Condensed equivalent circuit.

thus, the upper cutoff frequency of this circuit is given by the expression

$$f_{3db} = \frac{1}{2\pi \dfrac{(r_{bb'} + R_g)(r_{b'e})}{R_g + r_{bb'} + r_{b'e}} [C_{b'e} + C_c(1 + g_m R_L)]} \tag{12·124}$$

which may be rewritten as

$$f_{3db} = \frac{r_{bb'} + r_{b'e} + R_g}{2\pi(r_{bb'} + R_g)(r_{b'e})[C_{b'e} + C_c(1 + g_m R_L)]} \tag{12·125}$$

Notice that the upper cutoff frequency of the transistor-amplifier stage is dependent upon the transistor parameters and upon the source and load resistance. Consider the ideal case when R_g becomes very large and R_L very small. Under these conditions,

$$f_{3db} = \frac{1}{2\pi r_{b'e}(C_{b'e} + C_c)} \tag{12·126}$$

and since $C_{b'e}$ is much larger than C_c, this reduces (recalling preceding equalities) to

$$f_{3\text{db}} \approx \frac{1}{2\pi r_{b'e}C_{b'e}} = \frac{1}{2\pi \dfrac{\beta_0}{g_m}\dfrac{g_m}{2\pi f_T}}$$

$$\approx \frac{f_T}{\beta_0} = f_\beta \tag{12·127}$$

This is the result one would expect, since we previously demonstrated that a large source resistance and a small load resistance were required for the upper cutoff frequency to be defined as it is in Eq. (12·127) [see discussion concerning Eqs. (12·106) to (12·108)].

We might pause and note that now we have examined, in varying detail, four different equivalent circuits of transistors. Actually, many more equivalent circuits have been suggested by numerous engineers, and each of these undoubtedly had some reason for favoring their particular arrangement. Equivalent circuits are designed to represent a transistor electrically, and since a transistor can perform a multiplicity of functions, it is logical to expect a number of different arrangements. The most desirable circuit is the one that is simplest in form and yet which represents the transistor for the widest possible range of application. Obviously, with a device as complex as a transistor, these two aims are in conflict and some compromise must be struck. At the present time, the equivalent circuit of Fig. 12·43 is favored by many circuit designers; whether it will continue to be popular will depend on the trend in future developments in transistor fabrication and application.

Complete Design of a Transistor Amplifier

In the preceding sections, we derived the formulas for the gain and the upper cutoff frequency of a transistor-amplifier stage. These equations, together with those developed for the coupling and emitter bypass capacitors and for the stability factor, provide the necessary expressions to design a complete transistor amplifier. In order to illustrate the utility of these expressions, let us perform an actual design of a typical amplifier.

The information necessary to complete the design and the desired performance of the amplifier are listed as follows (keep in mind that this is a typical application and that the requirements will vary from one application to another):

1. Lower cutoff frequency, $f_{3\text{db}}$ low, less than 100 cps
2. Upper cutoff frequency, $f_{3\text{db}}$ high, greater than 2 Mc

3. Stability factor S less than 10
4. Voltage gain G_v greater than 500
5. Source resistance R_g and load resistance R_L equal to 1,000
6. Collector supply voltage V_{cc}, 32 volts

The first step is selection of a suitable transistor. Since the upper cutoff frequency must be above 2 Mc, the choice is limited to transistors having an f_T greater than 100 Mc. Transistors such as the MADT, PADT, or mesa could be used; for our application we shall select an MADT, the 2N502. This is a high-frequency amplifier-type diffused-base transistor with an f_T on the order of 200 Mc. Since the frequency-determining characteristics f_T and C_c are dependent upon the device and the operating point, choice of the operating point is generally concurrent with the choice of transistor. The 2N502 data sheet indicates a typical f_T of 260 Mc at a collector-emitter voltage V_{ce} of -10 volts and an emitter current I_E of 2 ma. The maximum C_c is specified as 1 $\mu\mu$f at $V_{ce} = -10$ volts. In choosing the operating point, the maximum power dissipation, 60 mw for the 2N502, must not be exceeded; in our case $P_c = V_{ce}I_c = 20$ mw.

With the operating point determined, it is necessary to determine the additional parameters of interest; these are usually included on the data sheet, or they may be measured or calculated from other available data. The available data are listed as follows (at $I_c = 2$ ma, $V_{ce} = -10$ volts):

$$f_T = 200 \text{ Mc (degraded for worst-case transistor)}$$
$$\beta_0 = 40 \text{ (degraded for aging)}$$
$$C_c = 1\ \mu\mu\text{f}$$

$$r_{bb'}C_c = 60 \text{ psec} \qquad \therefore r_{bb'} = 60 \text{ ohms}$$

From these parameters, we may calculate the following:

$$\alpha_0 = \frac{\beta_0}{\beta_0 + 1} = \frac{40}{41} = 0.975$$

$$g_m = \frac{I_E}{27} = \frac{2 \text{ ma}}{27} = 0.074 \text{ mho}$$

$$C_{b'e} = \frac{g_m}{2\pi f_T} = \frac{0.074}{2\pi \times 200} = 58.9\ \mu\mu\text{f}$$

$$r_{b'e} = \frac{\beta_0}{g_m} = \frac{40}{0.074} = 540 \text{ ohms}$$

The next step in the design is to devise a suitable circuit that will provide the desired operating point and the desired stability factor.

The general form of our circuit is shown in Fig. 12·50. The two base biasing resistors form a voltage divider which, in conjunction with the emitter resistor, establishes the required bias. The emitter resistor serves to aid stability but cannot be too large or gain will be lowered;

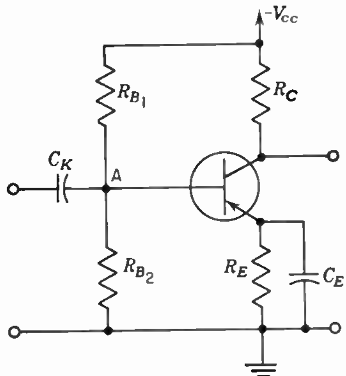


Fig. 12·50 General amplifier circuit.

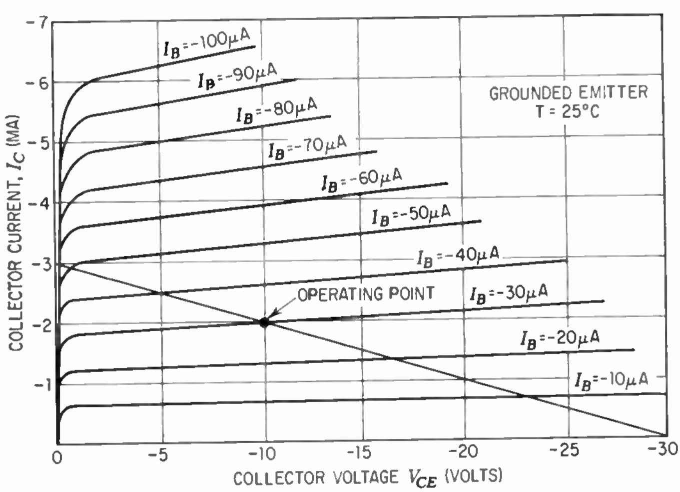


Fig. 12·51 Typical collector characteristic curves for a 2N502 transistor.

the collector resistor, on the other hand, may be large to avoid interstage loss and to fix the operating point. Before obtaining design values for these components, consider the collector characteristics for the 2N502 shown in Fig. 12·51. The supply voltage is 32 volts and the operating point is 2 ma, 10 volts. This point corresponds to a base current of 30 μa. A load line may be constructed through the two points to obtain the total series resistance $R_c + R_E$ as

$$\frac{32}{3 \text{ ma}} = 10{,}700 \text{ ohms}$$

Thus, we shall choose $R_E = 2{,}200$ ohms and $R_c = 8{,}700$ ohms.

Experience is probably the best design tool in the matter of selecting the values of R_{B_1} and R_{B_2}. Lack of experience may be compensated for by making an operating circuit and experimentally adjusting the values of R_{B_1} and R_{B_2} until the desired operating point is achieved. However, we can "zero in" on the appropriate values by a few approximations. Referring to Fig. 12·50, the current I_1 that travels down through R_{B_1} divides at point A into two currents. One segment of this current, I_2, passes through R_{B_2}, while the other segment flows into the base, becoming I_B for the transistor. In equation form, this relationship is expressed as follows:

$$I_1 = I_2 + I_B \qquad (12\cdot128)$$

From a design standpoint, it is more convenient to express Eq. (12·128) in terms of the resistances and voltages in the circuit. I_{co} is neglected and I_B, since it is a design parameter, remains. Doing this gives us

$$\frac{V_{CC} - (V_{BE} + I_E R_E)}{R_{B_1}} = \frac{V_{BE} + I_E R_E}{R_{B_2}} + I_B \qquad (12\cdot128a)$$

If we neglect V_{BE}, which will be on the order of 0.05 to 0.6 volt, this becomes

$$\frac{V_{CC} - I_E R_E}{R_{B_1}} = \frac{I_E R_E}{R_{B_2}} + I_B$$

Substituting the assumed value for I_E, R_E, V_{cc}, and I_B, we have

$$\frac{32 - (2 \text{ ma})(2.2 \text{ kilohms})}{R_{B_1}} = \frac{(2 \text{ ma})(2.2 \text{ kilohms})}{R_{B_2}} + 0.03 \text{ ma}$$

or

$$R_{B_1} = \frac{27.6}{4.4/R_{B_2} + 0.03 \text{ ma}}$$

If we assume a value of 10,000 ohms for R_{B_2}, then

$$R_{B_1} = 59{,}000 \text{ ohms, approx}$$

Allowing for I_{co} and V_{BE}, R_{B_1} finally comes down to a value of 52,000 ohms.

We are now in a position to calculate the stability factor, which is given by

$$S = \frac{1 + R_E\left(\dfrac{R_{B_1} + R_{B_2}}{R_{B_1}R_{B_2}}\right)}{1 - \alpha_0 + R_E\left(\dfrac{R_{B_1} + R_{B_2}}{R_{B_1}R_{B_2}}\right)} \tag{12·129}$$

Substituting the appropriate values given,

$$S = \frac{1 + 2.2(62/520)}{1 - 0.975 + 2.2(62/520)} = 4.4$$

which indeed meets the requirement of $S \leq 10$.

The next step is to calculate the voltage gain available from our single stage. From this, we may calculate the number of stages required.

$$G_v = \frac{g_m r_{b'e} R_L}{r_{bb'} + r_{b'e} + R_g} = \frac{0.074 \times 540 \times 1{,}000}{60 + 540 + 1{,}000} \tag{12·130}$$
$$= 25$$

This indicates that two stages will be necessary. The bandwidth for a single stage is [using Eq. (12·125)]

$$\mathrm{BW} = \frac{r_{bb'} + r_{b'e} + R_g}{2\pi(r_{bb'} + R_g)(r_{b'e})[C_{b'e} + C_c(1 + g_m R_L)]} \tag{12·131}$$

$$= \frac{60 + 540 + 1{,}000}{2\pi(60 + 1{,}000)(540)[58.9 + 1(1 + 0.074 \times 1{,}000)]}$$

$$= 3.32\ \mathrm{Mc} \tag{12·106}$$

We shall design the second stage in the same fashion as the first stage. However, we must calculate not only the individual gains and bandwidths but also the overall gain and bandwidth of the two-stage amplifier.

Before we make these calculations, bear in mind that the load on the first stage is the input resistance of the second stage and the source for the second stage is the output resistance of the first stage. In finding the load on the first stage, the biasing network may be neglected, since the resistors are large compared with $r_{b'e}$; however, the biasing network does become important when determining the source for the second stage.

Thus, the load that stage 1 works into is the input resistance of stage 2. This is

$$
\begin{aligned}
R_{in} \text{ of stage 2} &= r_{bb'} + r_{b'e} \\
&= 60 + 540 \\
&= 600 \text{ ohms}
\end{aligned}
$$

The source resistance for stage 1 is the series resistance of the input signal, be it a generator or a preceding stage. In Fig. 12·52, this is indicated to be 1,000 ohms.

The input or source resistance for stage 2 consists of the 8,700-ohm collector load resistor of stage 1 in parallel with the 52,000-ohm resistor and the 10,000-ohm resistor of the base biasing network of

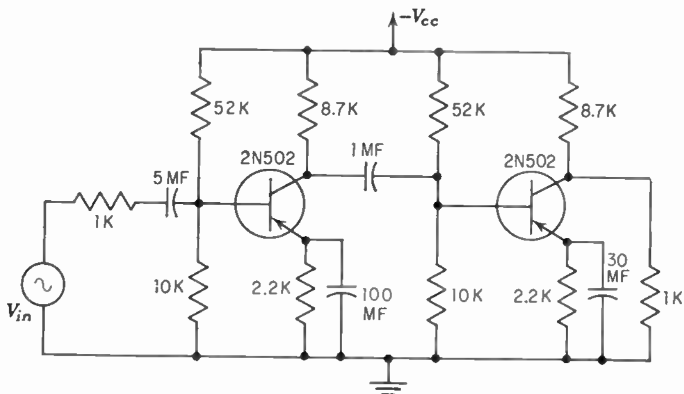


Fig. 12·52 Final circuit design.

stage 2. Hence R_{g_2} possesses a value of 4,260 ohms, obtained by performing the indicated calculation. Finally, the load resistance of stage 2 is 1,000 ohms, as revealed by Fig. 12·52.

We must use the expressions for current gain and then convert into voltage gain because two stages are involved, not one. The current gains are calculated as

$$G_{i_1} = \frac{g_m r_{b'e} R_g}{r_{bb'} + r_{b'e} + R_g} = \frac{(0.074)(540)(1{,}000)}{60 + 540 + 1{,}000} = 25$$

$$G_{i_2} = \frac{g_m r_{b'e} R_g}{r_{bb'} + r_{b'e} + R_g} = \frac{(0.074)(540)(4{,}260)}{60 + 540 + 4{,}260} = 35$$

Recalling that voltage gain is equal to $(R_L/R_g)G_i$, we obtain

$$G_v = G_{i_1} G_{i_2} \times \frac{R_{L_2}}{R_{g_1}} = 25 \times 35 \times \frac{1{,}000}{1{,}000} = 875$$

which satisfies our requirement of $G \geqq 500$.

The individual bandwidths may be calculated as in Eq. (12·131). The results are

$$\mathrm{BW}_1 = \frac{60 + 540 + 1{,}000}{2\pi(60 + 1{,}000)(540)[58.9 + 1(1 + 0.074)(600)]}$$
$$= 4.26 \text{ Mc}$$
$$\mathrm{BW}_2 = \frac{60 + 540 + 4{,}260}{2\pi(60 + 4{,}260)(540)[58.9 + 1(1 + 0.074 \times 1.000)]}$$
$$= 2.48 \text{ Mc}$$

From vacuum-tube theory, one recalls that the bandwidth of a two-stage amplifier is smaller than the smallest bandwidth of either stage.

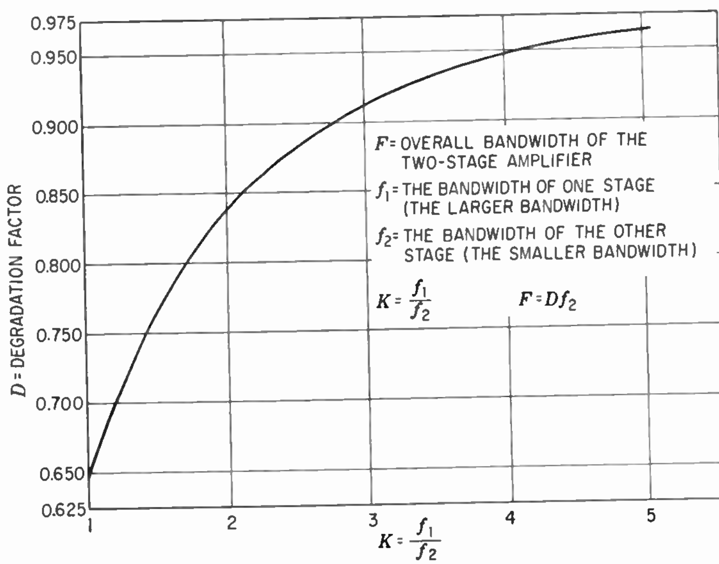


Fig. 12·53 Bandwidth degradation curve for a two-stage amplifier.

The calculation of the overall bandwidth is easily accomplished with the aid of the bandwidth-degradation curve shown in Fig. 12·53. To use this curve, we must find the ratio of the larger individual bandwidth to the smaller individual bandwidth as $K = f_L/f_s$. This ratio is used to obtain the degradation factor D from Fig. 12·53. The overall bandwidth is then given by the product of the degradation factor and the smaller individual bandwidth as $\mathrm{BW} = Df_s$. In our case $f_L = f_1 = 4.26$ Mc and $f_s = f_2 = 2.48$ Mc. Thus the ratio is

$$K = \frac{4.26 \text{ Mc}}{2.48 \text{ Mc}} = 1.72$$

From Fig. 12·53 we obtain the value of the degradation factor as $D = 0.81$. Thus the overall bandwidth is

$$\text{BW} = (0.81)(2.48\text{ Mc}) = 2.01\text{ Mc}$$

which satisfies the requirements of BW $\geqq$ 2 Mc.

The only calculations that remain are those required to find the values of the coupling capacitors and the emitter bypass capacitors. These capacitors will be different for each stage because of the different source impedances involved. The coupling capacitor is given by

$$C_K = \frac{1}{2\pi f_{3\text{db}}\left(R_C + \dfrac{R_{B_t}R_{\text{in}}}{R_{B_t} + R_{\text{in}}}\right)}$$

where

$$R_{B_t} = \frac{R_{B_1}R_{B_2}}{R_{B_1} + R_{B_2}}$$

We may calculate R_{B_t} as

$$R_{B_t} = \frac{(10\text{ kilohms})(52\text{ kilohms})}{10\text{ kilohms} + 52\text{ kilohms}} = 8.4\text{ kilohms}$$

and it is the same for both stages. Similarly for both stages, $R_{\text{in}} = r_{bb'} + r_{b'e} = 60 + 540 = 600$. Thus, $\dfrac{R_{B_t}R_{\text{in}}}{R_{B_t} + R_{\text{in}}} = 560$. For the first stage the R_C term is the source resistance $R_{g_1} = 1$ kilohm, and for the second stage it is the collector resistor $R_C = 8.7$ kilohms. We may now calculate the coupling capacitors as

$$\text{First-stage } C_{K_1} = \frac{1}{2\pi(100\text{ cps})(1{,}000 + 560)} = 1\ \mu\text{f}$$

$$\text{Second-stage } C_{K_2} = \frac{1}{2\pi(100\text{ cps})(8{,}700 + 560)} = 0.17\ \mu\text{f}$$

The emitter bypass capacitor is given by

$$C_E = \frac{\beta_0 + 1}{2\pi f_{3\text{db}}\left(R_{\text{in}} + \dfrac{R_C R_{B_t}}{R_C + R_{B_t}}\right)}$$

We may calculate $\dfrac{R_{B_t}R_C}{R_{B_t} + R_C}$ for the first stage as

$$\frac{(8.4\text{ kilohms})(1{,}000)}{8.4\text{ kilohms} + 1{,}000} = 895\text{ ohms}$$

and for the second stage as

$$\frac{(8.4 \text{ kilohms})(8.7 \text{ kilohms})}{8.4 \text{ kilohms} + 8.7 \text{ kilohms}} = 4.28 \text{ kilohms}$$

We can now determine the emitter bypass capacitor values

$$\text{First-stage } C_{E_1} = \frac{40 + 1}{2\pi(100 \text{ cps})(600 + 895)} = 43.8\ \mu\text{f}$$

$$\text{Second-stage } C_{E_2} = \frac{40 + 1}{2\pi(100 \text{ cps})(600 + 4{,}280)} = 13.4\ \mu\text{f}$$

In order to further ensure the low-frequency response, let us increase each of these capacitors to the following values:

$$C_{K_1} = 5\ \mu\text{f} \qquad C_{E_1} = 100\ \mu\text{f} \qquad C_{K_2} = 1\ \mu\text{f} \qquad \text{and } C_{E_2} = 30\ \mu\text{f}$$

The capacitors required for the second stage are smaller than those for the first stage; this is because the source resistance feeding the second stage is larger than that feeding the first.

The completed design circuit is given in Fig. 12·52. We have so designed this circuit that it offers a stability factor of 4.4, a voltage gain of 875, a bandwidth of 2.01 Mc, and a low-frequency cutoff of less than 100 cps.

As a summary of the foregoing design, let us again list the steps involved and the measures that have to be taken if the design requirements cannot be met in all instances. The first step is to select the operating point. Perhaps the only critical limitation we might encounter here would be transistor power dissipation or voltage rating limits. The remedial action would be to go to a less desirable operating point or to change the supply voltage. The next step is to obtain the derived stability factor. If that were not possible, we would have to reduce the base biasing resistor or would have to increase the emitter resistor. In the former case, we would sacrifice gain. Gain may be increased, within limits dictated by device linearity, by adding more stages. However, as indicated by the bandwidth-degradation chart, an increase in gain is accompanied by a decrease in bandwidth. Unique and complex feedback methods are available for increasing the bandwidth with less sacrifice in gain, but such techniques are beyond the scope of this text. An example of one such feedback technique is shown in Fig. 12·54. R_e and R_f are feedback resistors which increase the bandwidth at the expense of gain. This two-stage amplifier employing the 2N502 yields a voltage gain of 10 with a bandwidth of 50 Mc if $R_g = 50$ and $R_L = 50$. A voltage gain of 40 is possible with a bandwidth of 15 Mc if $R_g = 1{,}000$ ohms and $R_L = 1{,}000$ ohms. Finally, the low-frequency cutoff is attained by using large coupling and emitter bypass capacitors. There are limits on how large these capacitors may be both electrically and physically, however.

As a final word, keep in mind that there may be design requirements that are unattainable. For instance, it would be a difficult problem indeed to design our amplifier for a voltage gain of 500 and a bandwidth of 20.0 Mc. Such a requirement might be met, but only after

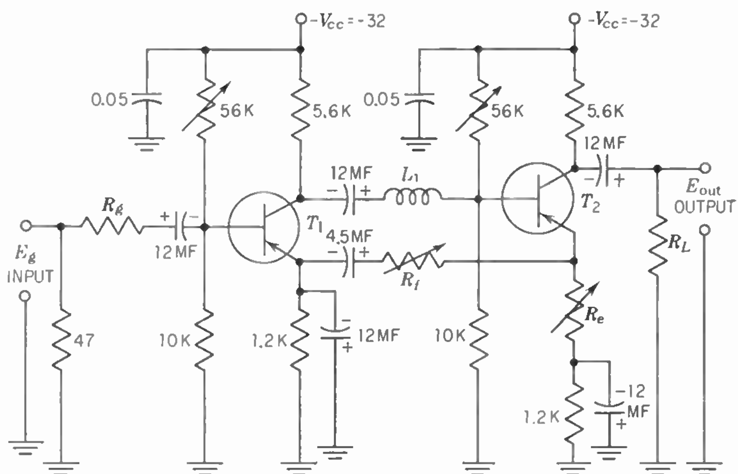


Fig. 12·54 Band amplifier employing feedback.

resorting to very complex feedback techniques such as those shown in Fig. 12·54.

QUESTIONS

12·1 What factors must be considered when selecting a d-c operating point for a transistor amplifier?

12·2 Obtain the collector-voltage–collector-current characteristics for a transistor to be operated as a grounded emitter. Plot the maximum collector dissipation curve for that transistor on the characteristic chart.

12·3 How is the position for a load line determined?

12·4 Draw a suitable load line for the transistor represented by the characteristics selected for Question 12·2. Explain why the line drawn is suitable.

12·5 What is the difference between d-c and a-c load lines? Which is more important to the transistor in its role as an amplifier?

12·6 Why is the behavior of the saturation current I_{co} with temperature more critical in grounded-emitter amplifiers than in grounded-base circuits?

12·7 Derive the relationship between β and α of a transistor.

12·8 What do we mean by the stability factor S?

12·9 Derive the stability factor S for the circuit of Fig. 12·11.

12·10 Explain, without resorting to S, why the circuit of Fig. 12·11 is more sensitive to temperature changes than the circuit of Fig. 12·12.

12·11 Derive Eq. (12·12).

12·12 When would the equivalent circuit of Fig. 12·18 be valid in representing a triode and when would the circuit of Fig. 12·19 be required?

12·13 What equations govern the circuit of Fig. 12·19? Explain each term put down.

12·14 Explain the significance of each item in Fig. 12·20.

12·15 What advantage does the equivalent circuit of Fig. 12·21 possess over the equivalent circuit of Fig. 12·20 in so far as they both represent the same transistor?

12·16 Explain how the equations governing the circuit of Fig. 12·24 are set up.

12·17 Derive the equation for the input impedance of the circuit in Fig. 12·24.

12·18 What assumption is made in deriving Eq. (12·46)? Using a commercial transistor, show from its characteristic (as given by the manufacturer) that this assumption is valid.

12·19 Explain the procedure to follow when deriving the output impedance of a transistor by using its equivalent circuit.

12·20 In what ways are hybrid parameters more useful than the other parameters discussed?

12·21 Draw the equivalent circuit of a transistor using hybrid parameters. Explain the significance of each item in this circuit.

12·22 Outline briefly how the relationship between the hybrid and resistance parameters of the same transistor circuit is obtained.

12·23 Draw a high-frequency equivalent circuit of a transistor.

12·24 Explain each component appearing in the circuit drawn for Question 12·23.

12·25 What can be done, in the fabrication of a transistor, to improve its high-frequency response? What compromises must be made?

Index