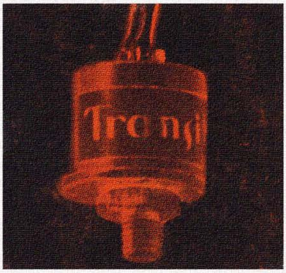
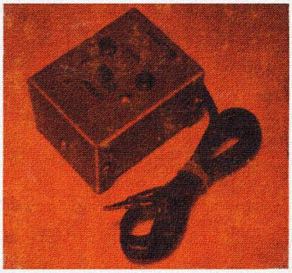
TAB BOOKS No. 116 $3.95

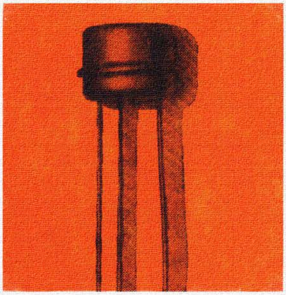
Getting Started With Transistors

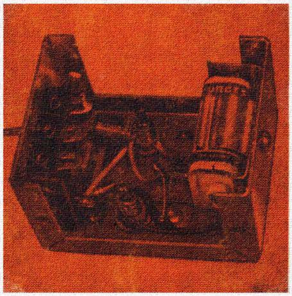
Louis E. Garner, Jr.

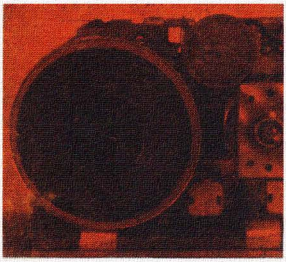
How transistors began ● learning to read electronic diagrams ● how transistors work ● amplifiers ● oscillators ● transistor types ● diodes, phototransistors, rectifiers ● transistor ratings ● testing transistors

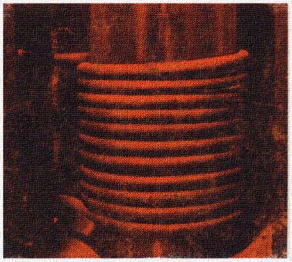












GETTING STARTED WITH transistors

LOUIS E. GARNER, Jr.

TAB BOOKS
BLUE RIDGE SUMMIT, PA. 17214

First Printing — September, 1963
Second Printing — February, 1965
Third Printing — January, 1968
Fourth Printing — June, 1968

Library of Congress Catalog Card No. 63-19542

Contents

Introduction

You don't need to know anything about chemistry to bake a cake. Just follow instructions. Electronics is different, though. You can't experiment intelligently by making connections at random, just to see "what will happen". Knowing what will happen in a circuit is an important prior condition to building that circuit.

Electronic components are numerous and complex. There are basic component categories such as transistors, tubes, capacitors, resistors and inductors. But each of the categories is composed of many variations of the basic device. There are power transistors, general-purpose transistors, high-frequency transistors, switching transistors, silicon and germanium transistors. All are transistors. All are made by a certain basic process. All obey certain basic electronic laws. But they are not interchangeable with one another because of their special abilities.

Learning how components work, how they are used is learning electronics. The components are the building blocks of electronic circuits. Many are passive units—tubes and transistors are active components; they can amplify.

Transistors are ideal building blocks for the beginner and experimenter. The voltages and currents are small and can be supplied by a small battery or dry cell instead of the bulky, expensive power supplies needed for vacuum tubes. There is no need for any connection to a potentially dangerous ac power line.

This is ideal for the experimenter who wants to make his own modifications—but they must be done intelligently, in a manner that will produce the minimum smoke. Even with a battery supply, components can be damaged when connected wrong. To make new devices you must know *what to do*—to conserve components you must know *what NOT to do*. Here's your chance to start at the bottom.

LOUIS E. GARNER, JR.

CHAPTER 1

How transistors were born

There is a rather ancient story about the three stages in a man's education. Unlike ancient jokes, it still remains interesting . . . that is, if you haven't heard it!!

It seems that man starts as a babe with no knowledge, no experience and no training. The world and everything in it is fresh and new. Everything that happens carries with it a grain of knowledge. At the beginning, however, *he knows nothing*.

Then he starts to learn. Knowledge grows by leaps and bounds. He's exposed to information from every angle—his playmates, school chums, parents, and neighbors . . . in fact, he acquires data from all his acquaintances and experiences. And he absorbs much of that knowledge, just like a sponge soaking up water. Sooner or later he acquires a vast sum of knowledge. It may be during high school, just after college, or at some similar point in his career. Exactly when he reaches the second stage is immaterial, for soon, *he knows everything*.

To put it another way—if he doesn't know it, it isn't worth learning.

Time moves on in its relentless path. Our growing man learns even more. He acquires new information, new knowledge, and, perhaps, a smattering of wisdom. Conservatively speaking, he reaches a point where he knows a hundred, a thousand, or perhaps ten thousand times as much as when he knew "everything." At this point, however, he gets a tiny glimpse of what remains to be known. He has turned full circle and is once more a babe . . . and he'll freely admit that, like a babe, *he knows nothing*.

In discussing the transistor, then, we'll start at the beginning, but we'll assume that you, the reader, are in one of those two vast intermediate stages between the points where you *know nothing and know everything.*

In the early days of radio communication, a standard receiver consisted of an antenna–ground system, a tuned (radio-frequency) circuit, a pair of electromagnetic headphones and a small "crystal" of lead sulphide (galena) or some mineral with similar electrical characteristics, such as iron pyrites, carborundum, silicon or molybdenite. The "crystal" was a two-electrode device, and hence was called a *diode.* One electrical contact was made through an adjustable sharp-pointed wire called a *cat's-whisker,* while the other contact was made through the body of the material, often by molding the individual crystal in a cup of metal alloy.

The crystal diode has a nonlinear characteristic. Unlike a linear device, such as a piece of copper wire, it offers considerably more resistance to the flow of electrons (electric current) through it in one direction than in the other. As a result, it rectifies or detects any alternating-current signal applied to it, converting the ac into pulsating direct current.

The early receiver's operation was fairly simple. Radio signals were picked up by the antenna-ground system and selected by the tuned circuit, then applied to the crystal diode. Here, the selected rf signal was converted into a pulsating dc signal which retained the original rf *modulation envelope,* containing the voice or music which the radio signal carried. The detected signal, in turn, was applied to the headphones and changed into audible sound.

In a sense, the galena crystal detector, whose use predated that of the vacuum tube by many years, is an ancestor of modern transistors. It bears approximately the same relationship to the transistor that the early two-electrode Fleming valve (circa late 1890's) bears to our modern vacuum tubes.

There were other similarities between this forerunner of the transistor—the galena diode, and the forerunner of the vacuum tube—the Fleming valve. Both were two-electrode devices (diodes). Both were used extensively during the early days of radio (circa 1890 to 1920). Both devices had nonlinear characteristics, permitting current flow easily in one direction and either preventing or severely limiting it in the reverse direction. Both devices were used as rf detectors and as ac rectifiers.

But these are similarities in application, *not in operation.* As a simple analogy, think of a canoe. While a canoe might be considered as "similar" to a bicycle in that both are used for the transportation of human beings, there are tremendous differences in the construction and in the principles and methods of operation of these two transportation "devices." In a like fashion, there are important differences between the crystal detector and diode vacuum tube, even though both are diodes and may be used in similar applications.

The crystal detector is a *solid-state* device. It depends upon the interaction of electrical bonds in solid molecular structures (hence the name) for its operation. The Fleming valve, a *thermionic* device, depends on the movement of free electrons in a vacuum for its operation. The electrons were released from their molecular and atomic bonds by the application of heat to a filament or *cathode*. The modern descendants of these two devices maintain the same basic differences. The transistor is a multi-electrode solid-state device, while the present-day vacuum (and gas-filled) tube is a multi-electrode thermionic device.

Both vacuum-tube and transistor development followed roughly parallel paths. The first practical device in both "families" were two-electrode units (diodes). Later three-electrode devices *(triodes)* were developed. These, in turn, were followed by myriads of multi-electrode and special-purpose devices. The thermionic tube family developed in logical fashion, but there is a large historical gap in the development of solid-state devices—much as if the "ancestor" galena diode was in a state of suspended animation for decades.

To understand this, let's go back a few years to the beginnings of the vacuum tube. In the early 1880's, Thomas Alva Edison discovered an intriguing phenomenon while experimenting with his newly developed incandescent lamp. If he placed a platelike electrode in the lamp bulb near the filament and connected it to the positive side of the filament's dc power source, he found that a small electric current flowed between the *plate* and its connection to the filament. This action, called the Edison effect after its discoverer, was simply a scientific phenomenon with no practical application until the English scientist Fleming developed his two-electrode "valve" in the late 1890's. The Fleming valve was the first true vacuum tube.

About 1906, a relatively few years after the invention of the Fleming valve, Lee De Forest added a third electrode to this early vacuum tube and in so doing invented the *triode* tube, which he dubbed the Audion. This third electrode—a gridlike structure—

was located between the filament (cathode) and plate electrodes of the Fleming valve. Since a small voltage applied to this *grid* could control the stream of electrons flowing from cathode to plate (and hence output current), the triode could be used to amplify as well as to detect electrical signals. This important invention laid the foundation for almost all later developments in electronic technology.

The early Audions were bulky, not too reliable and extremely expensive. But in the decades following its invention, one development followed another with almost clocklike precision. Mathematical techniques were developed for designing vacuum tubes with special electrical characteristics. Production methods were modified and refined; the hand assembly of individual vacuum tubes was replaced by efficient machine production techniques. Vacuum-tube types multiplied rapidly. Before long, tubes were available in many sizes, styles and varieties from subminiature units about the size of a cigarette to giant power tubes as tall as a man. Prices dropped, too, as production methods improved. It became possible to buy good-quality tubes for just a fraction of a dollar, a real saving compared to the many-dollar price tag of the original Audions.

By the 1920's, the vacuum tube had made possible a giant new industry—radio broadcasting. And by the 1930's, tubes were being used by the millions in radio receivers for the common man. The vacuum tube, originally an expensive experimenter's plaything, had become a household item.

The invention of the vacuum tube laid the foundation for scientific wonders—radio, television, radar, sonar, automatic controls, electronic "brains" and computers—in fact, all the electronic miracles so essential to modern living. Whole new industries were (and are) based on the vacuum tube and its application. It brought immense wealth to hundreds, new employment and opportunities to hundreds of thousands, and enjoyment, entertainment, and better living to hundreds of millions.

But for all its marvelous qualities, the vacuum tube has a number of disadvantages. It is comparatively inefficient. The power required to heat its filament to red-hot temperatures contributes little, otherwise, to the operation of the device. And the heat developed by its filament contributes to the rapid deterioration not only of the vacuum tube itself but to that of nearby electrical components. Yet the high heat is essential if the electrons are to be "freed" of the cathode. Tubes burn out, short, or become defective in so many different ways, and so often, that engineers and technicians frequently remark, *"There's nothing wrong with electronics that getting rid of the vacuum tube won't cure."*

While this statement is more wishful thinking than hard reality, it does reflect the fact that from 75% to 90% of the "troubles" which develop in receivers and other electronic equipment are caused by or the result of defective vacuum tubes. With this incentive, then, research scientists started to search for a device which could do the same job as the vacuum tube but without its disadvantages.

In contrast to the logical and fairly rapid development of the vacuum tube, the crystal detector was known—and in use—long *before* the invention of the Fleming valve. But this first of the solid-state devices lay almost dormant for nearly half a century. It was not until 1948 that Messrs. Shockley, Bardeen, and Brattain of the Bell Telephone Laboratories invented a three-electrode solid-state device capable of *amplifying* as well as detecting an electrical signal. This device, like the earlier galena detector, consisted of a small piece of crystalline metal to which body contact was made, with additional electrical contacts made through fine wire cat's-whiskers. In this case, however, two closely spaced cat's-whiskers were used.

Physically, this solid-state amplifier was a small brass tube with a metal plug near one end and an insulating plug about midway along its length. Fixed to the metal plug was a small rectangular block of crystalline germanium alloy about 1/20 inch square by about 1/50 inch thick. The two cat's-whiskers resting on the block were only a few thousandths of an inch apart, and were attached to heavier metal pins fixed in the insulating plug.

The name for this new solid-state triode was coined from two words which seemed to describe its action—*trans*fer and re*sistor,* hence, *transistor.* In 1951, with the development of the *junction transistor,* a new type which didn't require the cat's-whisker contacts, the term *point-contact transistor* was adopted to distinguish the earlier unit. Today, the original point-contact transistor is largely obsolete and no major commercial manufacturer is producing it. However, the basic schematic symbol for the point-contact transistor (Fig. 101), originally designed to show the *base* crystal and the two point-contact cat's-whiskers (*emitter* and *collector* electrodes), was retained. Today, it is used to represent all triode transistors.

The first transistors, much like the first vacuum tubes several decades earlier, had a number of undesirable features which severely limited their application to radio-electronic circuits. These included poor stability, excessive internal noise, high cost of manufacture, sensitivity to changing environmental conditions (such as heat and humidity) and difficult-to-reproduce characteristics. Often, three or four transistors of the same type had entirely different electrical characteristics.

But, for all its faults, the transistor seemed to be an answer to an engineer's prayer. With no filament to burn out, the unit operated at room temperature and, theoretically at least, seemed to have an infinite service life. Eliminating the filament in an amplifying device brought with it an even greater advantage—it eliminated a wasteful use of electric power, with the result that the transistor's efficiency was many, many times greater than that of vacuum tubes.

Following its invention, the transistor "evolved" in much the same way that the vacuum tube had several decades earlier, but

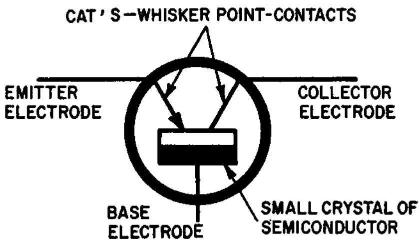


Fig. 101. Schematic symbol used to represent the transistor. The small cube of semiconductor crystal was called, appropriately, the base. *The two cat's-whisker electrodes were named the* emitter *and* collector, *respectively.*

with several important differences. When the triode tube was invented, there was no other amplifying device, and a whole new technology had to be developed. Design methods had to be perfected, components had to be designed and manufactured, and so on. But the transistor was introduced at a time when there were hundreds of applications for amplifying devices—in radio receivers, in television sets, in phonographs, in hearing aids and in all types of control equipment.

As a result, the evolution of the transistor from an experimental device to a full-fledged commercially manufactured component was compressed into just a few years. Compare this to the several decades required by the vacuum tube to reach widespread industrial use. But chronologically, the transistor is still a "baby" as far as other electronic components are concerned.

Going back to 1948–49, the first transistors were hand-made devices with many faults and undesirable features. Only point-contact types were available, and these were assembled in crude tubes, exposed to the atmosphere, as shown at 1 in Fig. 102. When it was found that moisture and chemicals in the atmosphere could cause changes in a transistor's electrical characteristics, manufacturers revised their production techniques and developed hermetically sealed transistors (2 in Fig. 102). This was the second major

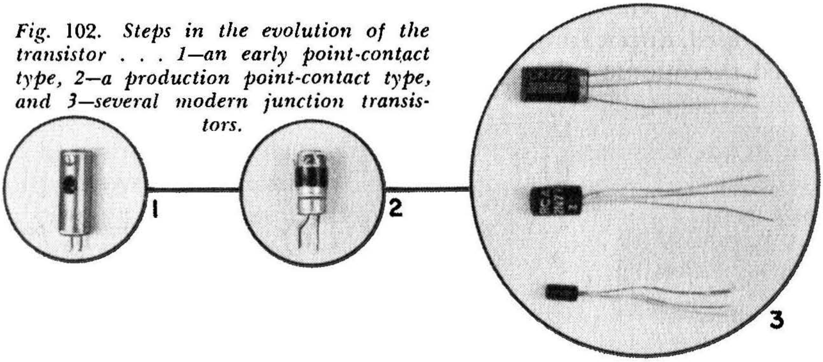


Fig. 102. Steps in the evolution of the transistor . . . 1—an early point-contact type, 2—a production point-contact type, and 3—several modern junction transistors.

step in the transistor's evolution. Finally, with the invention of the junction transistor in 1951, a variety of new designs and types became possible, and manufacturers were able to introduce production techniques which eliminated most of the transistor's early faults. This last step is illustrated at 3, Fig. 102.

Application-wise, the transistor had to fight its way against the vacuum tube, (its chief competitor) as a successful amplifier. Except for a few types of specialized military equipment, the transistor's first large-scale commercial application was in compact hearing aids. These audio amplifiers must be small, compact and extremely efficient, for necessity dictates that they be battery-operated. The transistor was a natural choice and its advantages in this application far outweighed its extremely high price (compared to vacuum tubes) and the minor faults in its operating characteristics. In a few short years, the transistor completely displaced the vacuum tube in hearing aids.

Next, the transistor was used in other types of equipment where light weight, small size and high efficiency were more important than initial cost. A few short years after, the Regency Division of I.D.E.A., Inc., Indianapolis, Ind., presented their now famous TR-1 pocket radio, the first commercially manufactured all-transistor radio (Fig. 103). This set served as a pace setter, and compact receivers manufactured years later are still similar to it.

As prices dropped and production techniques improved, the transistor was used in other receivers, in car radios, in TV sets and in all types of instruments and controls. Simultaneously with the increasing use of transistors in commercial "consumer type" products, there has been a great upswing in the use of transistors in military equipment, in computers and in all types of industrial gear.

Today, a whole class of solid-state electronic devices has been developed, duplicating almost all the functions of vacuum and gas-filled thermionic tubes. For example, crystal diodes are available

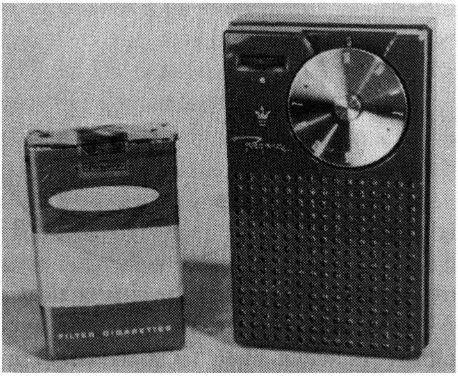


Fig. 103. A transistor pocket radio. The cigarette package is included to show relative size. Some models are even smaller than the receiver shown.

in various sizes with electrical ratings to handle such diverse jobs as detecting an rf signal, mixing extremely high-frequency signals or rectifying ac to supply substantial amounts of current to power other electronic circuits or electrical machinery. Tetrode (four-electrode) transistors are produced, as well as many special-purpose types such as subminiature low-noise units for low-signal, high-gain audio amplifiers, high-frequency types for operation up in the thousands of megacycles; high-speed switching types capable of operating in less than a millimicrosecond (a thousandth of a millionth of a second), and high-power units capable of handling electrical power in excess of 1,000 watts. (See Figs. 104 and 105.)

Along with specialized transistors and diodes, scientists and engineers have invented other types of solid-state devices. Some of these, like the bistable diode, Unijunction transistor, Trigistor, thyrode and controlled rectifier, have electrical characteristics analogous to those of a gas-filled thermionic tube (thyratron), and may be used in similar circuit applications. Others are sensitive to light and may be used as replacements for phototubes.

Still others have characteristics for which there is no thermionic-tube equivalent. An example is the *tunnel diode,* invented by a Japanese scientist, Esaki. This is a two-electrode device which has a "negative-resistance" characteristic, permitting its use as an oscillator or amplifier. In contrast, there is no diode type thermionic tube which can serve as an amplifier.

In many cases, engineers modify the characteristics of common solid-state devices to permit their use in specialized application. A familiar diode becomes a *Zener* diode, and can be used in voltage-regulation and control circuits. Or an ordinary photosensitive cell becomes a solar battery capable of converting light into electrical power. Sun batteries are used to power the radio transmitters and instruments in many of our deep-space probes and artificial satellites.

The transistor and its related solid-state cousins offer many advantages over thermionic tubes in most applications. Physically, for example, a transistor may be from 1/100th to 1/1,000th the size and weight of a vacuum tube that can do an equivalent job. Transistors are much more resistant to shock and vibration than are tubes. Some manufacturers have demonstrated this by loading

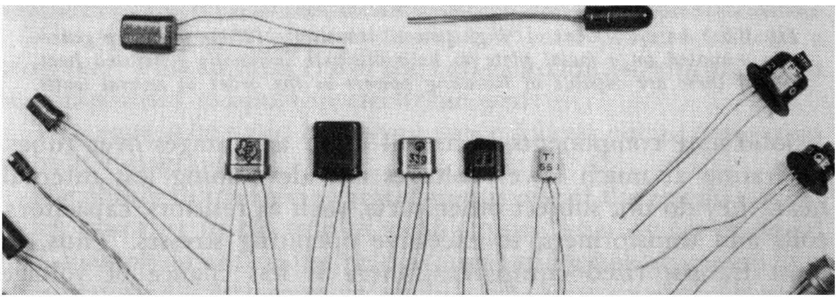


Fig. 104. Typical transistors. These are all low and medium power units, and demonstrate the variety of sizes and styles that are available. About 2,000 different types are being produced.

their transistors into a shotgun shell and firing them into a thick telephone book without damage. And with no filament to burn out, the transistor has an almost infinite service life.

Since transistors and related components are highly efficient, their power requirements, as compared to thermionic tubes, are almost negligible. An idea of the relative efficiencies of vacuum tubes and transistors may be obtained by comparing the power requirements of, say, a tube-operated automobile receiver with those of a fully transistorized set having similar performance specifications. A typical vacuum-tube receiver may require 12 volts at from 5 to 10 amperes, that is, from 60 to 120 watts. A transistor receiver, on the other hand, may require only 0.25 to 1.0 ampere at the same voltage, or from 3 to 12 watts.

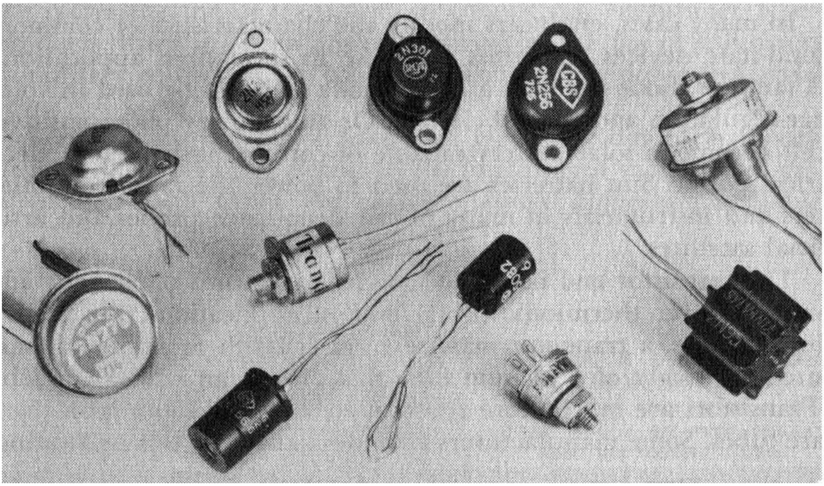


Fig. 105. An assortment of "high power" transistors. These units are generally mounted on a metal plate to help dissipate internally generated heat. All of these are capable of handling powers in the order of several watts.

Solid-state components offer still other advantages over tubes. Operating at much lower voltages and developing less internal heat, they do not subject other parts, such as resistors, capacitors, coils and transformers, to excessive operating stresses. Thus, in most transistorized equipment, there is less chance of voltage breakdown of insulation or component failure due to excessive temperatures as compared to corresponding tube-operated units.

The development of solid-state devices is not yet tapering off. New cousins for the transistor are being discovered on an almost day-to-day basis. One of the latest developments is the commercial production of *Peltier* junctions. These are devices made up of junctions similar to those found in junction transistors which are *heated or cooled* by an electric current. Current passing one way through the junction causes heat to be developed; reversing the current results in a cooling effect. Peltier junction devices show great promise for use in all-electronic air conditioning and refrigeration.

The future is hard to predict. Changes are occurring too fast to follow easily but, if past developments are any criteria, the transistor and its related devices are but a promise of things to come.

Anyone for crystal ball gazing?

CHAPTER 2

How to read electronic maps

"Get your program here—you can't tell the players without a program." This familiar cry of the ball-game huckster easily might be paraphrased to apply to electronic gear:

"Get your schematic here—you can't follow circuit operation without a diagram."

The ability to read a circuit diagram is one of the essential skills which a worker in electronics must acquire whether his eventual goal is simply to pursue the field as a rewarding hobby, or whether he is planning a long-term career in equipment repair and maintenance, as a factory technician, as a broadcast engineer, as a radio or radar operator or as a design engineer. Too often, however, this basic skill is bypassed—and not just by novices or beginners. Many experienced hobbyists and even some technicians shudder at the thought of following a diagram. They look at one as if it were an item of medieval torture.

On the other hand, a skilled individual who has taken the time to become familiar with circuit diagrams and their use looks on the diagram as a close friend. It is a useful tool which helps him to understand circuit operation and to work faster and easier. Ask any top-notch technician a question about a piece of equipment and chances are he'll come back with, "Let me see the schematic." And a design engineer would be lost without a scrap of paper and a pencil with which to sketch circuits.

While we assume that you, the reader, are a "beginner" only in your work with transistors, and not a rank newcomer to the field of electronics and that you have some, however slight, familiarity with circuit diagrams, our own experience indicates that a short review of reading techniques will prove helpful to all but the most experienced hobbyists and technicians. But if diagrams are already your "meat," and you prefer to wire a circuit from a schematic rather than a pictorial, by all means skip on to the next chapter—the time you save will be your own.

Of the familiar objects found in everyday life, the auto road map is perhaps the closest thing to an electronic circuit diagram as far as concept, layout and use are concerned. If you can read a road map, circuit diagrams will be a snap. The ability to refold a road map is not necessary, moreover!

Let's consider a typical road map. Basically, it consists of a variety of wiggly lines printed on a piece of paper. Some lines are dashed, some solid and some in color. Spotted at irregular intervals along the lines are assorted dots, stars or oddly shaped symbols representing cities, towns and villages. Rivers, bays, lakes and ocean shore lines are shown, often with pretty pictures of small boats or leaping fish. Printed at appropriate places are the names of cities and towns, lakes, rivers and so on, along with the names of prominent places such as mountains, landmarks, parks and monuments. Also printed on the map are the actual distances (in miles) between cities or towns and often between crossroads or landmarks. In addition, there may be bits and dabs of historical information —dates when famous battles were fought, data on toll bridges or tunnels, and so forth.

Most road maps are to linear scale. That is, a given length of wiggly line representing a road is directly proportional to the actual length of the road it represents. If the map is drawn to a scale of, say, 1/4 inch equals 1 mile, then a road 10 miles long would be shown as a line 2 1/2 inches long. But this scale is seldom extended to the line *width* nor to objects on the map, except for rivers, bays and larger geographical features. A village may be shown as a dot, for example, when the houses of the village are actually spread out over the area of a mile or two. A statue or other landmark may be drawn to a size which—if scale were maintained —would make it several miles high, although the statue itself is only 15 or 20 feet high.

No one pretends that a road map is an accurate picture of the countryside such as may be seen from an airplane, helicopter or man-made satellite. Depending on the firm issuing the map (usually an oil company), roads may be shown as black, red or blue lines. The actual roads themselves may be black asphalt, white concrete, gray gravel or dirty brown mud. The background may be white or lightly tinted yellow or green, whereas the actual countyside is often a crazy patchwork of odd-size plots in green, brown, tan, blue or, in winter, white. However, the map does show —and with reasonable accuracy—the *location* of important cities and towns and the length and direction of roads.

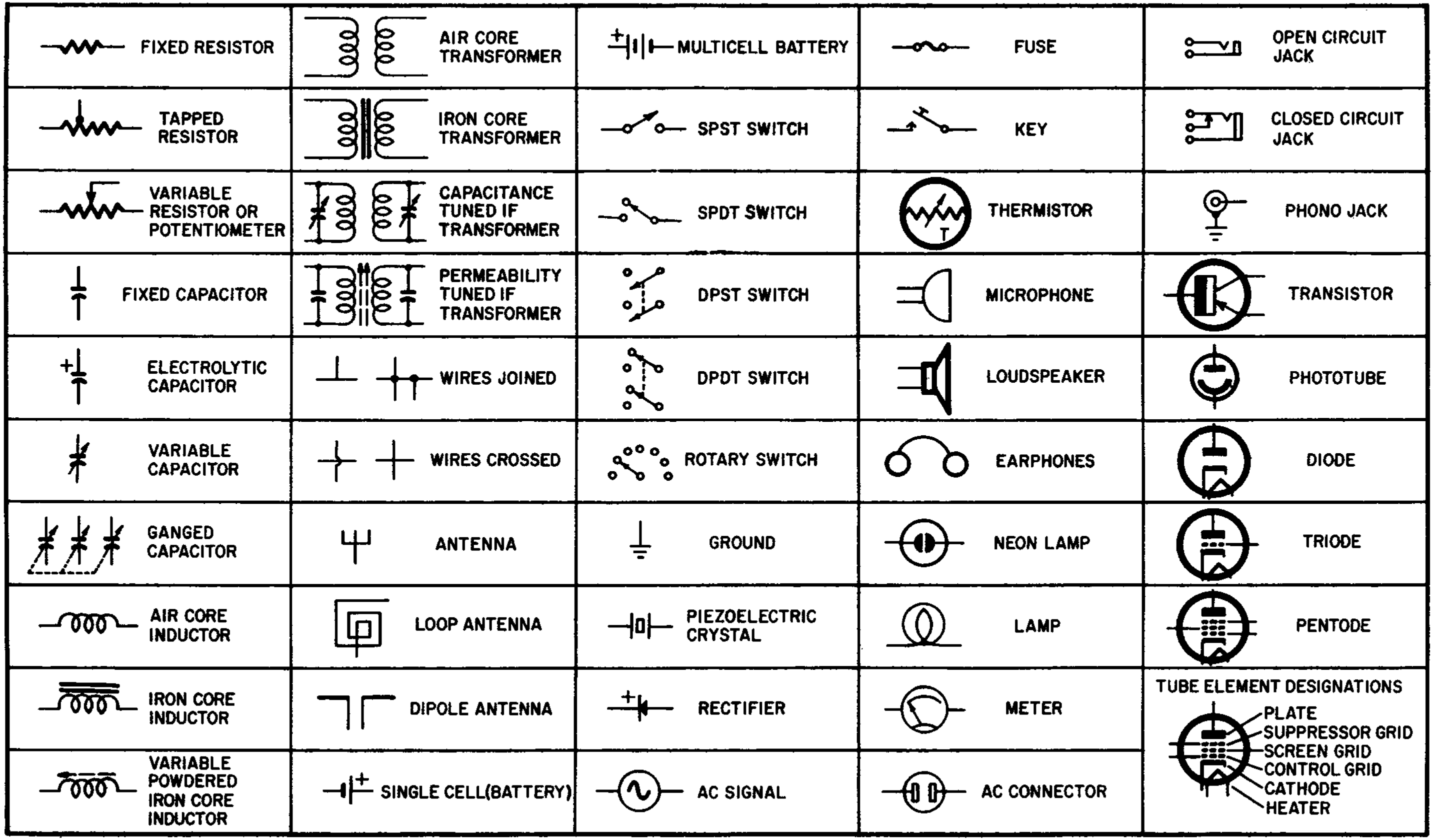
In a sense, the map shows the *scheme* of the network of roads criss-crossing the country, with *symbols* used to represent various types of roads and different-sized cities and towns. A secondary road may be shown as a dashed blue line; such a road may be paved with common gravel. A primary hard-paved road might be shown as a solid blue line. A main highway may be represented by a red line. Finally, a divided-lane expressway may be shown as a solid red line bordered with fine blue lines. Cities and towns are represented in a similar fashion. A small village may be a mere dot on the highway, a town a larger dot, while a small city may be shown as an open circle. State capitals might be shown as stars.

By analogy, the most important type of electronic equipment diagram shows the *scheme* of the circuit wiring and is called, appropriately enough, the *schematic* diagram. In practice, the word "diagram" is often omitted, and technical writers, technicians, and engineers will refer simply to an equipment's "schematic."

Like the road map, the schematic doesn't attempt to show an actual *picture* of the equipment (more about this later). Rather, it shows the electrical connections between the components making up the device. The key relationship is the use of symbols for, just as dots, circles and stars are used as symbols for towns and cities in a road map, so are special symbols used in a schematic diagram to represent the various components. Unlike the map, however, no attempt is made to scale a schematic diagram. A schematic for a miniature receiver the size of a package of cigarettes might require a full printed page, while a schematic for a powerful transmitter measuring 6 feet high by 2 feet square may require less space, even when drawn with the same-size symbols.

Schematic symbols

The common symbols used in schematic wiring diagrams are illustrated in Table 2-1. Several variations of these may be found in practice, depending on the editorial requirements of individual publishers and manufacturers. As a general rule, a specific symbol



FIXED RESISTOR	AIR CORE TRANSFORMER	MULTICELL BATTERY	FUSE	OPEN CIRCUIT JACK
TAPPED RESISTOR	IRON CORE TRANSFORMER	SPST SWITCH	KEY	CLOSED CIRCUIT JACK
VARIABLE RESISTOR OR POTENTIOMETER	CAPACITANCE TUNED IF TRANSFORMER	SPDT SWITCH	THERMISTOR	PHONO JACK
FIXED CAPACITOR	PERMEABILITY TUNED IF TRANSFORMER	DPST SWITCH	MICROPHONE	TRANSISTOR
ELECTROLYTIC CAPACITOR	WIRES JOINED	DPDT SWITCH	LOUDSPEAKER	PHOTOTUBE
VARIABLE CAPACITOR	WIRES CROSSED	ROTARY SWITCH	EARPHONES	DIODE
GANGED CAPACITOR	ANTENNA	GROUND	NEON LAMP	TRIODE
AIR CORE INDUCTOR	LOOP ANTENNA	PIEZOELECTRIC CRYSTAL	LAMP	PENTODE
IRON CORE INDUCTOR	DIPOLE ANTENNA	RECTIFIER	METER	TUBE ELEMENT DESIGNATIONS: PLATE, SUPPRESSOR GRID, SCREEN GRID, CONTROL GRID, CATHODE, HEATER
VARIABLE POWDERED IRON CORE INDUCTOR	SINGLE CELL(BATTERY)	AC SIGNAL	AC CONNECTOR	

Table 2–1. Here are some of the basic symbols commonly used in electronic circuit diagrams.

will be used for each generic class of components, even though the physical appearance and internal construction of the component may vary greatly from one type to another. For example, the multicell battery symbol is the same whether the battery represented is a small dime-size unit used in a hearing aid or a giant storage battery used to power a submarine or an electric fork-lift truck. Similarly, the symbol for a fixed capacitor is the same whether used to represent a tiny ceramic capacitor or a large transmitter unit.

In many cases, the original symbol for a component was derived from its construction in its most elementary form. As we saw in Chapter 1, the schematic symbol for the transistor is based on the representation of the base and cat's-whisker contacts of the original point contact transistor. By the same token, a capacitor's symbol (see Table 2-1) indicates, symbolically, that the capacitor is essentially two conducting plates separated by an insulating medium of some type—paper, air, oil, an electrolytic film, a sheet of ceramic or mica, or even a vacuum. An inductance in its simplest form is just a coil of wire, and the picture symbol used to represent such a coil looks very much like this. If we wind the coil on a piece of iron to increase its inductance, we indicate this fact by drawing straight lines alongside the basic coil symbol. And if the core consists of powdered iron held in a glue or plastic binder, this is shown by a series of dotted lines.

Often, a component has an adjustable value. The fact that its value can be varied—that it is a *variable* component—is shown by drawing an arrow through or alongside the basic part or component symbol. Refer to the symbols shown for a variable inductance and a variable capacitor in Table 2-1. A variable resistor (rheostat) is sometimes indicated in this fashion but, more often, is shown by an arrow at right angles to the fixed resistor symbol. This arrow represents the movable slider or "arm" of the resistor.

In addition to the symbols given in Table 2-1, a number of special symbols are used from time to time. For example, a triangle often represents a "floating ground," a common electrical connection point or piece of heavy wire that is *not* connected to the equipment's metal case or chassis. Occasionally, special symbols will be made up by combining the symbols for several basic components. A prime example here is the symbol for an if transformer. This is made up by combining the symbols for two inductance coils and those for adjustable (trimmer) capacitors. A special symbol often can be recognized from its location in the circuit or by a word identification included alongside it.

A schematic diagram is much more than a collection of symbols connected by lines representing conductors. As in the case of a road map, a symbol alone is not enough to identify an object positively. Consider a map with no place names, no route or highway numbers, no distances. It would be pretty confusing and, for practical purposes, almost useless. Thus, a schematic will include a variety of additional bits of information. Controls and switches may be identified as to function (volume control, bandswitch, tone control, power, treble, bass, etc.). Important frequencies or adjustments may be specified. Battery, tube and transistor type numbers will be shown, and, often, major components or stages may be named. In the last case, common practice is to use easily recognized abbreviations. For example, SPKR for loudspeaker, RF AMPL for radio-frequency amplifier, XTAL for crystal, AF AMPL for audio-frequency amplifier, ANT for antenna, PWR for power, GND for ground, and so forth.

Then there is the matter of the electrical components themselves. Except for a few special cases, the symbol alone is not enough to completely identify the component used. A certain symbol may indicate that a component is a capacitor—but what kind? what size? what rating? and how is it referred to when discussing the circuit? Let's review the basic components used in electronic circuits to see how they are identified *and specified.* The last item is particularly important to the hobbyist assembling a project or to the service technician replacing a defective part.

Components

It is customary to use *letter* symbols or abbreviations as well as picture symbols to identify electrical components. The more common symbols are R (resistor), C (capacitor), L (inductance coil), T (transformer), V (vacuum tube), Q or V (transistor), CH (choke), RFC (rf choke), V or E (voltage), I (current), J (jack), P or PL (plug), S or SW (switch), BATT (battery), K or RY (relay), and PHONO (phonograph). Other variations of these may be encountered from time to time. For example, TR may be used in place of Q for a transistor, SP for a loudspeaker instead of SPKR, and RL for a relay. Special letter symbols will be used in some circuits; PC (photocell), SR (selenium rectifier), D (diode) and TS (terminal strip). Where special letter symbols are used, the picture symbol generally serves as a clue to identify the component.

A few symbols have more than one meaning, depending on how they are used. For example, B may stand for the base (electrode of a transistor), for a class of amplifier operation or for a battery. Similarly, Q may refer to a transistor or to a characteristic of an inductance coil. Again, a study of the schematic diagram or any descriptive text will positively identify the meaning of the letter symbol.

In most cases, however, the letter and pictorial symbols alone are not enough to identify a *specific* component in a circuit. There may be 10 or 12 (or more) resistors in a receiver circuit, all identified by the same picture and letter (R). In practice, then, a *number* or *letter* will be added to the basic letter symbol for positive identification. Thus, we may have R1, R2, R3, R4, C1, C2, C3, T1, T2, and so on, referring to individual components in a single circuit. If two or more components are assembled in a single container or with a common connection lead, an additional letter designation may be added. In the case of a multi-section electrolytic capacitor, for example, the entire capacitor may be identified as, say, C3, with individual sections specified as C3-a, C3-b, C3-c, etc. Finally, the letter symbol may omit an identifying number, substituting in its place a second letter. This is done where there is only one component of a specific type, for example, C_T as "tuning capacitor" or J_O as "output jack."

Transistors and diodes are also identified by specific *type numbers*. Typical transistor types are 2N35, 2N169, 2N384, GT-229, 2N107, 2N170 and so on. Typical diode types could be 1N34A, 1N69A, 1N70, CK705 and SD500. As a general but far from universal rule, most semiconductor devices are identified with a type number consisting of three parts, a number, the letter N and a second multi-digit number. The number preceding the N indicates the number of active junctions and is usually *one less* than the number of electrodes. Thus, a diode will have a "1N—" designation; a triode (transistor), a "2N—" type number, and a tetrode a "3N—" type designation. Typical tetrode transistors are types 3N36, 3N37.

Other small electrical components, such as resistors, coils and capacitors, are identified by four major specifications—*type, value, tolerance* and *rating*. Of these, the component's type depends on its construction. As we shall see later, capacitors, resistors and coils may be assembled using a variety of materials or construction techniques. A component's value is given in terms of electrical units applicable to its important characteristic. The tolerance rating indicates the possible variation in a component's *actual* from its specified value. Finally, the rating is a general indication of the component's maximum capabilities, for example, the maximum voltage that can be applied to it or the maximum power it can dissipate as heat.

Tolerance and *rating* specifications are the two most likely to be misunderstood by newcomers and beginners. Let's examine these before discussing individual components in detail.

When we speak of a component's *tolerance,* we generally refer to a percentage value. Typical tolerances are 1%, 2%, 5%, 10% and 20%. If the tolerance is given as a single value—10%, for example—it means that the component's actual value may vary as

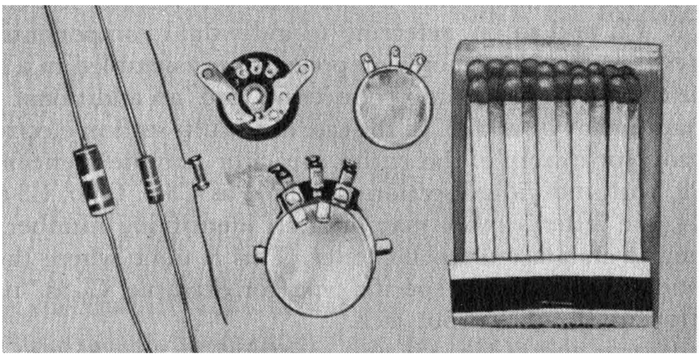


Fig. 201. Fixed and variable resistors. Three fixed resistors and two views of a small volume control are shown. The volume control is a subminiature type with built-in knob and switch. Compare this with the "standard" sized control. Bookmatches are included for size comparison.

much as 10% *on either side* of its nominal specified value or $\pm$ 10%. Thus, a resistor with a nominal value of 1,500 ohms, 10% tolerance, can have an actual value of from 1,350 to as high as 1,650 ohms. Occasionally, a dual tolerance value will be specified; an electrolytic capacitor, for example, may have a tolerance of $-$ 20%, + 40%. Such a specification means that the actual value may be from 20% below its specified value to as high as 40% above this value. Let's assume such a tolerance were applied to a 50-μf capacitor. Its actual value would then fall between 40 and 70 μf.

From the viewpoint of the project-building hobbyist or the practical service technician, a component's tolerance value can be taken as 10%, unless specified otherwise in a parts list. In any case, a component with a closer tolerance than that specified may be used. For example, if a parts list calls for a component with a 20% tolerance, you can always use parts with the same electrical value but with 10% or 5% tolerances.

A component's *rating* is almost always in terms of a *maximum* value. Thus, a capacitor with a rated *working voltage* of 150 volts dc should not be used in circuits where more than 150 volts are applied to it. Similarly, a resistor with a wattage rating of, say, 2 watts, can dissipate this much power under ideal conditions without changing value or burning out. In general, a component with a higher rating than that specified may be used without danger if there is adequate space for it. As an example, a resistor rated at

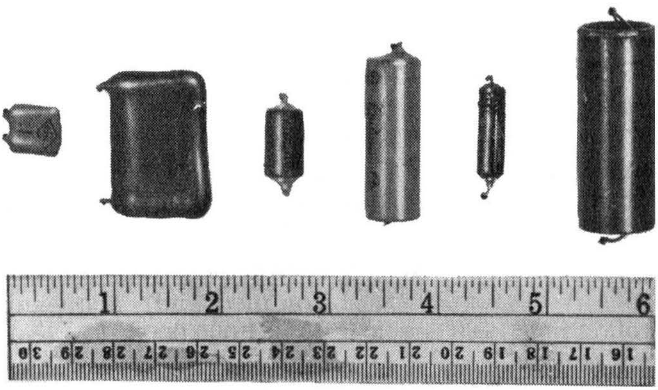


Fig. 202. Subminiature capacitors shown with "standard" sized units. From left to right . . . a pair of 0.1 μf ceramic capacitors, a pair 0.05 μf paper tubulars, and a pair of 50 μf electrolytics. Rule scale indicates sizes.

1 watt may be used as a substitute for a $\frac{1}{2}$-watt resistor *of the same resistance value* or a 150-volt capacitor may be substituted for a 100-volt unit. *But it is not safe to substitute components with ratings lower than those specified in a diagram or parts list.*

Resistors

Both fixed and variable resistors are used in transistor circuits (Fig. 201). A resistor *type* may be specified as composition (carbon) or wirewound, depending on its construction. Its *value* is given in ohms or megohms (millions of ohms), the basic units of electrical resistance. Its *rating* is given in watts. Typical resistors as used in transistor circuits are $\frac{1}{10}$, $\frac{1}{4}$, $\frac{1}{2}$, 1 and 2 watts, although larger units (5, 10, 25 watts) may be used occasionally in power circuits. Typical *tolerance* values are 10% and 20%, except for critical circuits where 1% or 5% units may be specified.

There is a fifth specification which may be applied to variable resistors (volume and tone controls, etc.)—*taper*. This refers to the way the control's resistance varies as the center shaft is turned. If the resistance variation is directly proportional to rotation (10% of its resistance with 10% rotation, 20% with 20% rotation, and so on), the unit has a *linear taper*. Special tapers giving a different percentage variation are used for special purposes. The most common of these is the *audio taper*, used primarily in audio volume and tone controls.

Capacitors

Typical fixed and variable capacitors used in transistor circuits are illustrated in Figs. 202 and 203, respectively. In fixed capacitors, the kind of dielectric used as insulation between the metal or foil plates indicates *type*. Thus, we have ceramic, paper, electrolytic and mica capacitors. Most variable capacitors use air as a dielectric, but a few subminiature types have a thin plastic film between plates to prevent accidental shorting with the close plate spacing used.

A capacitor's *value* is given in microfarads (μf) or micromicrofarads ($\mu\mu$f). A micromicrofarad (or picofarad) is one-millionth of a microfarad, or, conversely, it takes 1,000,000 $\mu\mu$f to equal 1 μf. The *rating* of a capacitor is generally expressed as a maximum dc working voltage. A voltage in excess of this value may arc through the dielectric and short the plates. As a rule, capacitor *tolerances* are not important except in extremely critical circuits. Most fixed capacitors, except electrolytics, have tolerances of from 10% to 20%. Electrolytics and some high-value ceramics may have tolerances as high as +250%.

An electrolytic capacitor's dielectric is a thin gas film formed by electrochemical action on one of its plates. Because of this, the capacitor must be used in circuits employing dc voltages and, in addition, must be connected properly to prevent the applied voltage from destroying its dielectric film. Except for a few special types, then, electrolytic capacitors are *polarized;* one lead or connection is identified as the positive lead, the other as negative. When connected in a circuit, the positive lead goes to the positive side of the dc source.

Beginners are sometimes confused by what appears to be a "polarity" marking on tubular paper capacitors—one lead may be identified by a ring or by the word *ground.* Actually, paper capacitors are not polarized. The identified lead connects to the foil plate falling on the outer perimeter of the capacitor during construction. Where a capacitor is used for bypass purposes, it is often helpful to connect the outside foil to circuit ground to obtain maximum shielding.

In highly critical circuits, fixed capacitors may be used which have a specified *temperature coefficient.* This indicates the variation in the capacitor's actual value with changes in temperature. Such capacitors are used in accurate tuning circuits to compensate for changes in other circuit components with changes in temperature.

Variable capacitors are used chiefly in tuned circuits. Their value is given in terms of *minimum* (with plates "open") and *maximum* (plates "closed") capacitance. Occasionally, their taper

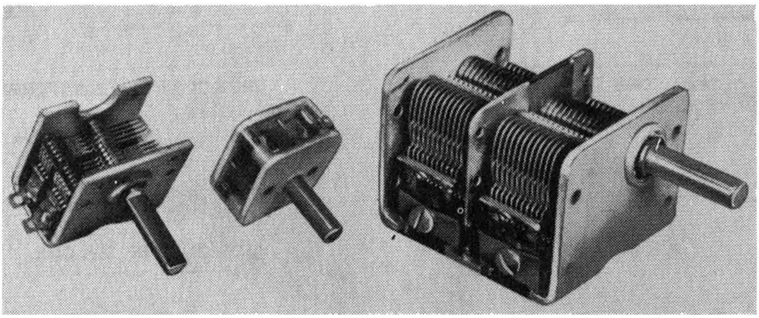


Fig. 203. Typical tuning capacitors. Miniature variable units used in transistor circuits are shown to the left, while the larger "standard" unit shows relative sizes. The very small capacitor uses a plastic film dielectric . . . the other two capacitors use air as a dielectric.

(variation in capacity with rotation) may be specified. Often, two or more variable capacitors may be "ganged" together mechanically so that rotating a single shaft changes the value of all of them. The capacitors may not be shown together on a schematic diagram, but the fact that they are "ganged" is generally indicated by a dashed line between them. The individual sections of a ganged capacitor may have similar or different characteristics, depending on circuit requirements.

Rf coils and transformers

High-frequency air-core coils may be self-supporting if wound of heavy wire or tubing or, if made of fine wire, supported on a cardboard, plastic or glass form. If a *powdered-iron* or *ferrite* core is used to increase the coil's inductance, the coil may be wound directly on the core or on a plastic or cardboard form into which the core fits. The inductance value of high-frequency coils is generally given in millihenries (mh.) or microhenries (μh.). A millihenry is a thousandth of a henry and is equal to 1,000 μh.

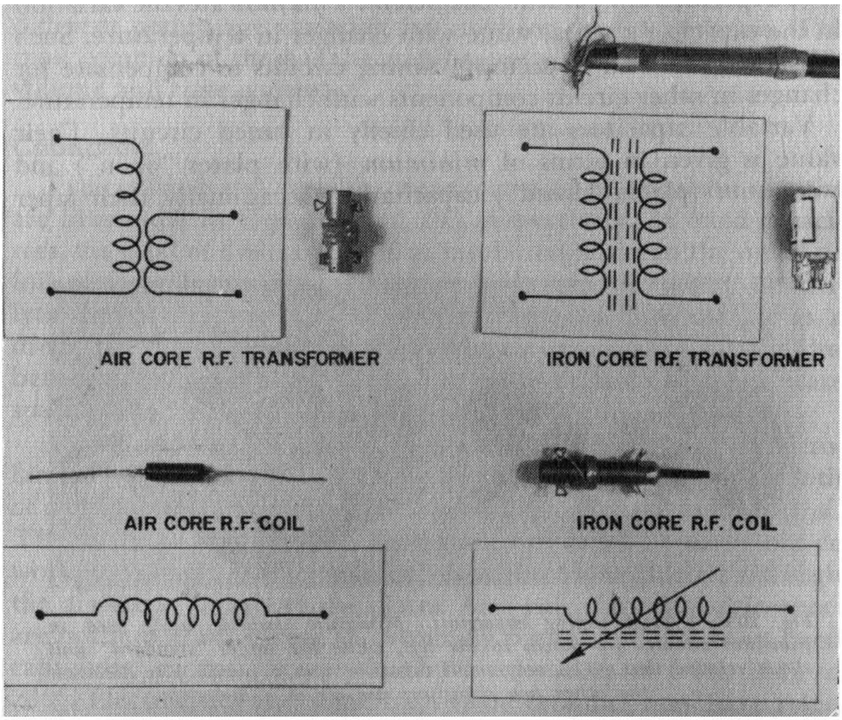


Fig. 204. Radio and intermediate frequency (rf and if) coils and transformers. Some of these use an air *core, others employ a* ferrite *core to obtain greater inductance with fewer turns of wire. An adjustable core, indicated schematically by an arrow, permits the coil's inductance value to be changed.*

Adjustable coils will have both minimum and maximum values given. Small coils such as are used in receivers seldom have a power rating, but transmitter coils may have maximum voltages and currents given as well as nominal power-handling capacity. Coil tolerances are given by some manufacturers, not by others.

A *choke* is a coil with a relatively high inductance value at the frequency at which it is used.

In addition to type (air or iron core), value and rating, two additional specifications may be given—the coil's *distributed* capacitance and its Q. Distributed capacitance refers to the inherent electrical capacity which appears to be shunting the coil; this factor is important in determining the resonant frequency of coil and capacitor tuned circuits. The frequency at which the coil itself acts as a tuned circuit is its *natural* resonant frequency; at frequencies above this value, the coil acts as if it were a capacitor.

The Q of a coil is simply the ratio of its inductive reactance to its ac resistance at a specific frequency. Q varies with frequency. Anything which causes losses in a coil will reduce its Q. As a general rule, then, the higher a coil's Q, the better its quality and the more selective will be a tuned circuit in which it is used.

If two or more coils are coupled together, the resulting combination is a *transformer*. Radio-frequency (rf) and intermediate-

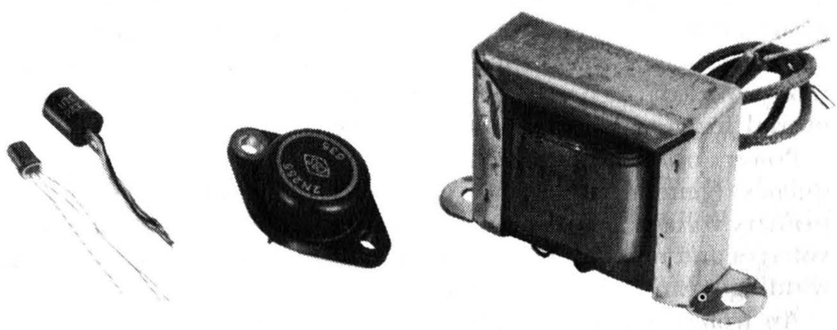


Fig. 205. Iron core audio transformers. A subminiature type, together with a typical subminiature hearing aid transistor, is shown at the left . . . it is about the size of a pencil's eraser. A power transistor is shown to the right along with its corresponding output transformer; the latter unit can handle up to 5 watts of audio power.

frequency (if) transformers are used extensively in receivers and transmitters. Transistor, rf transformers are specified in terms of ratio between primary and secondary windings and their resonant frequency with a given capacitor. Primary-to-secondary ratio may be given either in terms of impedances or turns ratio between windings. Generally, transformers are specified according to the frequency for which they are designed. Typical if values are 262 and 455 (or 465) kilocycles (kc).

Rf and if coils or transformers may be assembled in small metal cans to provide electrostatic shielding. Typical units of the types used in transistor circuits are illustrated in Fig 204 alongside their corresponding schematic symbols.

Audio and power chokes and transformers

Typical iron-core transistor transformers are illustrated in Fig. 205. Used at audio and power-line (25–60 cycles) frequencies, these are made up of copper-wire coils wound on a closed core of laminated iron. Physically, a transistor transformer may be smaller than a pencil's eraser or as large as the common transformers used for such everyday tasks as powering a doorbell or running a model electric train. The larger units look much like the transformers found in tube-operated radio and television sets and audio amplifiers.

An audio transformer's electrical specifications generally include the impedance ratio between its primary and secondary windings (in ohms), its power rating (in watts or milliwatts), the maximum dc which its primary winding can handle (in milliamperes) and the dc resistance of its windings (in ohms) as well as data on any taps provided (such as a center tap, etc.). In some cases, the unit's overall frequency response will be given.

Power transformer specifications include the operating frequency (generally 25 or 60 cycles, but higher in some cases), its primary voltage, power rating (in watts or volt-amperes), secondary voltages and maximum currents, and often the dc resistance of its windings. Any taps are specified.

An iron-core audio or filter choke is specified by indicating its nominal inductance value in henries, its dc resistance in ohms and its maximum ratings in terms of applied voltage and current. The last value is generally in milliamperes or amperes and indicates the maximum direct current at which the choke will have its specified inductance. If less dc is passed through the choke coil, its inductance will be higher; more current through the coil will reduce its inductance.

A transformer or choke is generally specified by referring to a specific manufacturer's type number.

Switches

Switch types are identified by their method of operation, number of contacts (*poles*) and number of positions since they are electromechanical rather than electronic components. Switches have ratings in terms of the maximum voltages and currents their contacts can handle safely. If the maximum ratings are exceeded, the contacts may either weld together or burn open.

The basic switch types are toggle, slide, lever, pushbutton and rotary, plus several special types (such as lock switches). Any of these may be either *positive-action* or *spring-return*. The former holds the position in which placed, while the latter returns automatically to its "normal" position as soon as the external actuating

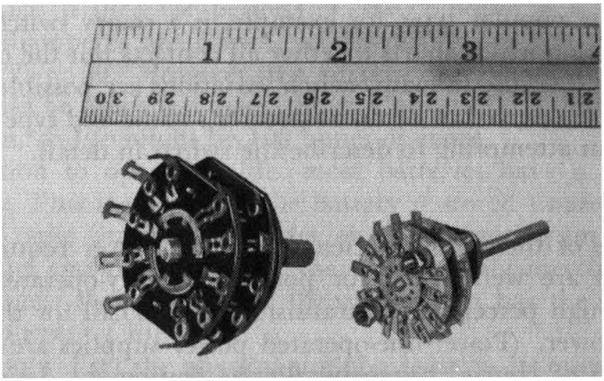


Fig. 206. Standard and miniature multiposition rotary switches. Rule scale indicates size.

force is removed. In the average household, a wall light switch is a typical positive-action type, while a doorbell pushbutton switch is of the spring-return variety. Most switches have only two or three possible positions; a rotary switch may have as many as 20. Typical standard and miniature rotary switches are illustrated in Fig. 206.

By far the simplest of switches is a single-contact type with two positions—open and closed (or on and off). Such a switch is called a *single-pole single-throw* (spst) switch. Other types include *double-pole single-throw* (dpst), *single-pole double-throw* (spdt) and *double-pole double-throw* (dpdt). The dpdt has two sets of contacts or poles and two positions into which it can be thrown. Finally, we have the multi-position rotary switches used as band-switches in short-wave sets and for similar applications. Two or more rotary switch wafers may be "ganged" on a single control shaft to form multi-pole multi-position switches.

The beginner often runs into a problem when first dealing with multi-position switches—a type designation of *"shorting"* or *"non-shorting."* The difference is a simple one. In a shorting type switch, each succeeding contact is made *before* the preceding contact is broken as the switch is turned from one position to the next. For this reason, a shorting switch is also called "make-before-break." In a nonshorting switch, on the other hand, each contact position is broken (or opened) before the next contact is made.

A number of special switch variations are possible. Most of these are made by combining the functions of the basic types discussed above. One popular type, for example, is a rotary switch with a special section which shorts together all contacts but the one (s) in use. Again, because so many switch variations are possible, special switches are referred to by manufacturer's name and type number rather than attempting to describe the switch in detail.

Batteries

Because of their high efficiency and low power requirements, transistors are well suited for portable battery-operated equipment. A high percentage of transistor circuits call for the use of battery power. (Power-line-operated power supplies are used occasionally, of course, but generally in equipment designed for large, permanent installations or with unusually high power consumption.) All types of batteries are used for transistors, from small "button" cells about the size of a dime to heavy-duty storage batteries similar to those found in autos and trucks. Single and multicell (battery) schematic symbols are given in Table 2-1.

Transistor power supply batteries are of two general classes: (a) "dry" batteries such as the familiar zinc–carbon units used in common flashlights, the more efficient akaline cell and the long-lived mercury batteries, and (b) rechargeable "storage" batteries, including both common lead–acid types as used in trucks and autos and the nickle–cadmium varieties. The majority of transistor circuits are designed to use either a single-cell 1.5-volt or multicell 3.0-, 6.-, 9.0- and 12.0-volt batteries. As a general rule, the supply voltage is not too critical and, except for a few special circuits, satisfactory operation may be obtained with both lower and higher than normal voltages, provided the battery used has low internal resistance. A transistor radio designed for 9-volt operation, for example, may still work on sources as low as 6. or as high as 12.0 volts. But *it is always best to use the supply voltage specified by the designer or manufacturer.*

Batteries may be specified by manufacturer and type number or by nominal output voltage and rated life in milliampere-hours (MAH or ma/hr) or ampere-hours (AH or amp/hr). The latter figure indicates the product of the nominal current drain the battery is intended to supply multiplied by its typical life under intermittent operating conditions. This figure varies widely with individual batteries and class of service, however. For example, a battery may be rated at 6.3 volts, 50 ma/hr. This means that the battery can supply a current of 5 ma for a total of 10 hours ($5 \times 10 = 50$) if used for periods of, say, 2 hours at a time. If only 2 milliamperes are drawn, its operating life would be 25 hours ($2 \times 25 = 50$). However, if a current of 50 ma were drawn, its life would be much less than 1 hour. Conversely, if under 1 ma were drawn, its life might be 100 hours or more.

In addition to operating life, most batteries have a definite "shelf" life. This is the life of the battery if stored unused, as in a portable radio during the winter or simply setting on a store shelf. Of the common dry cells, the rather costly mercury type has the longest shelf life, while the zinc-carbon has the shortest shelf and operating life.

Two factors affect the physical size of a battery—its output voltage and its operating capabilities in terms of ma/hr or amp/hr. Of these, the latter is the more important. A cell's output voltage is increased by adding additional cells, (two or more cells are a battery), its capacity by increasing the size of individual cells or by connecting cells in parallel. Thus, it is possible to have a single heavy-duty 1.5-volt cell that is many, many times larger than, say, 15-, 22½- or even 30 volt batteries with limited capacity. As a general rule, the larger a battery (physically), the greater its capacity and life, provided other factors (voltage and general type) are equal.

Battery polarity is extremely important. In some circuits reversing a battery's connection will simply prevent operation; in others such a move may damage the transistors and other circuit components. In batteries, polarity is generally identified by the type of position of clips or sockets used or by markings on the case. In zinc-carbon, akaline and nickel-cadmium cells, the outer shell is the negative terminal and the central "button" electrode the positive terminal. In mercury and some special zinc-carbon cells, on the other hand, the outer shell is positive, the central terminal negative. This difference is extremely important when substituting mercury-type cells with zinc-carbon dry-cell units or vice versa.

Other components

In addition to resistors, capacitors, coils, transformers and similar components we have discussed, a number of other items are used in transistor circuits. The schematic symbols for many of these are given in Table 2-1. Some are strictly "mechanical" components, such as sockets, clips, jacks, plugs, and connectors. Others are electromechanical—microphones, loudspeakers, earphones, relays, phonograph cartridges, crystals, meters and so on. A general knowledge of these items is helpful in reading circuit diagrams, in selecting replacements when servicing and in assembling hobby projects. Many of these components are illustrated in Fig. 207.

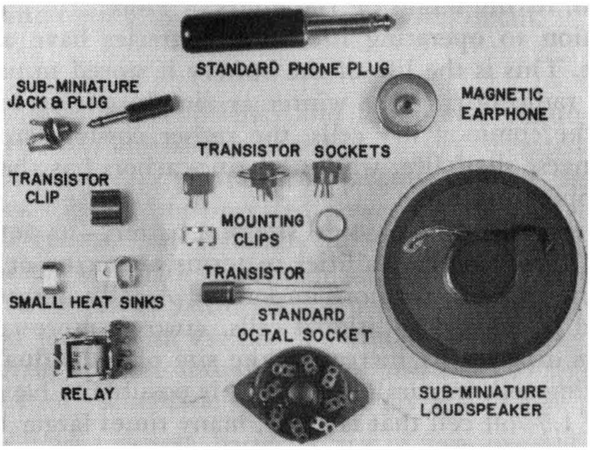


Fig. 207. Common transistor hardware. A small transistor, standard 'phone plug, and familiar octal tube socket are included for size comparison.

Microphones, phonograph cartridges, loudspeakers and earphones (or headphones) are basically similar devices. All are *transducers*—devices for changing one form of energy to another. Microphones and phonograph cartridges are used to change mechanical energy into electrical signals while speakers and earphones do just the opposite, converting electrical signals into sound vibrations by moving a small diaphragm. But the roles are interchangeable to some extent. A speaker may be used as a microphone under some conditions and, conversely, some types of microphones can serve as miniature speakers.

Three basic types of microphones find use in transistor circuits: (a) *carbon*, (b) *crystal* and (c) *magnetic* usually called dynamic, ribbon or controlled reluctance type. Of these, the carbon microphone has the highest output but requires an external source of power and is somewhat noisy, developing a slight internal "hiss." A crystal microphone uses a piezoelectric element and has an extremely high output impedance. In most cases a step-down matching transformer is needed when a crystal microphone is used in a transistor circuit. Dynamic microphones have a fairly low output but deliver good quality and are available in output impedances ranging from low to high. Of the three types of microphones, dynamic units offer the best match to transistor amplifiers. Microphones are generally specified by manufacturer and type number, although the basic type (carbon, crystal, etc.) may be given as well as output impedance.

Phonograph cartridges are basically similar to microphones in that they convert mechanical vibrations into electrical signals. Three types are in wide use: (a) *crystal,* (b) *ceramic* and (c) several forms that use *magnets.* The crystal and ceramic types are basically similar. Both use piezoelectric elements and, as a general rule, have high output levels and high output impedances. Cartridges using magnets have low to moderate output impedances and generally deliver fairly low output signal levels. Like microphones, these units are specified by manufacturer's name and type number—and may be called dynamic, variable or controlled reluctance types.

Most of the speakers used in transistor circuits are *permanent magnet* (PM) dynamic types. Basic specifications are size, given in terms of cone diameter, and voice-coil impedance, given in ohms. Typically, transistor loudspeakers run from 1½ to 12 inches in diameter, depending on application, with voice-coil impedances ranging from 3 to 45 ohms. The most popular impedances are 3–4 and 6–8 ohms, although 10-ohm subminiature speakers are used rather extensively in pocket radios.

Three basic types of earphones (often called either headphones or headsets) are used in transistor work: *magnetic, dynamic* and *crystal.* The first uses a magnet and an electromagnetic element to drive its reproducing diaphragm. The dynamic type has a moving coil to drive the diaphragm. These types have impedances ranging from 6 to as high as 50,000 ohms. Crystal earphones have a piezoelectric element and very high overall impedances (50,000 ohms or more). Common practice is to speak of earphones in terms of its overall impedance; a high impedance phone is considered the more sensitive. Physically, earphones range from tiny button units such as are used with hearing aids to fairly large assemblies which may be equipped with rubber ear covers that make them appear even larger.

Meters used in transistor instruments—as voltmeters, frequency meters and so on are specified in terms of overall *size* (diameter of face), *sensitivity* (in microamperes, milliamperes or amperes for full-scale deflection), and, often, dc *resistance.* Typically, a circuit may call for a 3-inch, 0–1 ma meter.

The relay is basically an electromagnetically operated switch. Like a switch, one of its important specifications is its number of poles and operating positions. Thus, we have spst, spdt, dpst and dpdt relays as well as rotary, contact and stepping units. In addition to contact arrangement, a relay's specification will include *coil type* (ac or dc), *resistance* (in ohms) and, often, *sensitivity* (in milliwatts or watts). In lieu of power sensitivity, the coil operating voltage, current or both may be given.

Of the remaining components, frequency-determining quartz crystals are generally specified by manufacturer's name, type number and actual frequency, while plugs, jacks and connectors are specified in general terms. If the component is an electrical part of the functioning circuit (a switch, plug, jack, and so on), it is generally shown on the schematic diagram with an appropriate symbol and identified in the parts list. A "nonfunctioning" part of the circuit (a tie-point terminal strip, socket or clip, for example), may be listed in the parts list, but as a rule, is not shown on the schematic diagram.

Color codes

Many resistors, capacitors, diodes and other components are so small physically that it is virtually impossible to mark or print their electrical value on them. Such parts are often identified by means of a *color code,* basically a series of colors to which number values have been assigned. These colors are spotted on the components as bands or dots in a specific order which indicates its electrical value. In addition, a special color code is often used to identify the terminals or leads of transformers.

Semiconductor lead connections

As mentioned earlier, diodes, transistors and related semiconductor devices are identified and specified by type number. When working with practical circuits, however, it is just as important to know the electrode lead connections as it is to know the type of component used. Unlike vacuum tubes, semiconductor lead or pin connections are not assigned numbers. Rather, their leads are arranged in a predetermined pattern with, occasionally, one lead identified by a color mark, dot or symbol. A knowledge of the various lead arrangements used is essential to anyone tracing through an electronic circuit or assembling a project from a wiring diagram.

Among the various semiconductor devices in common use, diodes are the easiest to work with as far as identifying leads is con-

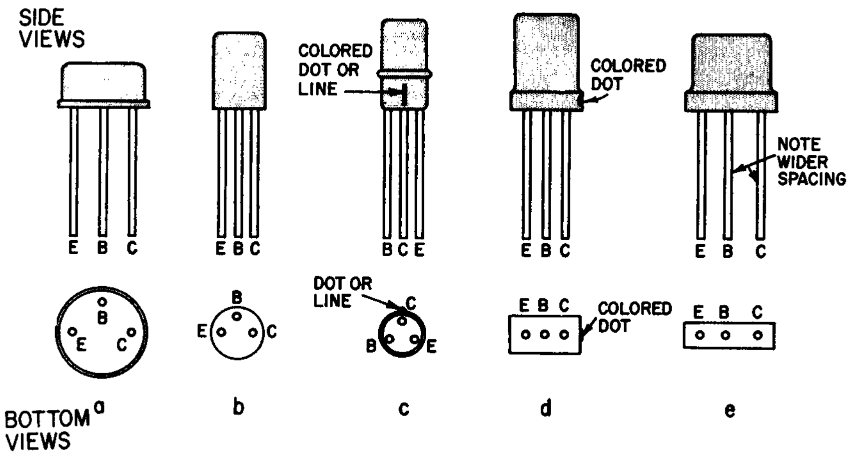


Fig. 208. Side view outlines and basing diagrams for typical low power transistors. This sketch is not to scale.

cerned. In the first place, there are only two leads—anode and cathode. Second, almost all manufacturers identify the cathode lead by a band, dot or other symbol. But transistors, with three or more leads, pose special problems, complicated by the fact that there is no convenient way of numbering the leads in the manner that tube socket pins are numbered.

Lead arrangements for most low-power transistors are illustrated in Fig. 208. Emitter (E), base (B) and collector (C) leads (or pins) are identified. A triangular pattern is used in cases (a), (b) and (c) while an "in-line" lead arrangement is used for (d) and (e). Where the leads are arranged in a symmetrical pattern or with equal spacing, as (c) and (d), a colored dot or line is used to identify the collector (C) electrode, with the other electrodes identified from here. As a general rule, the base (B) lead is between the emitter (E) and collector (C) leads. Note that it is the

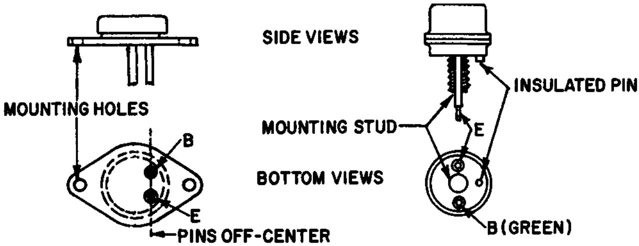


Fig. 209. Basing diagrams and outline sketches of power transistors.

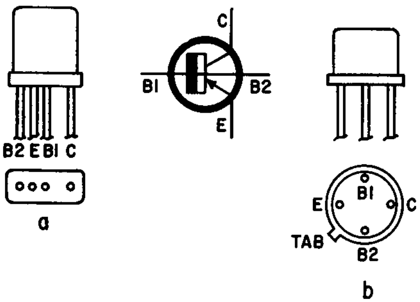
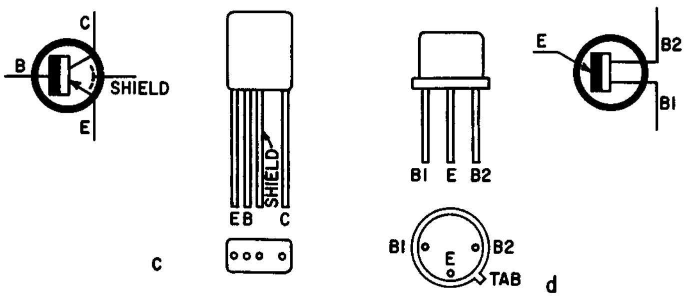


Fig. 210. Side views and basing diagrams for tetrode transistors (a, b), triode transistors with internal shield (c), and the Unijunction transistor (d). Like Figs. 208 and 209 this sketch is not to scale.



pattern or arrangement which serves to identify the leads, *not actual dimensions,* for these transistors are manufactured in a variety of physical sizes.

Power-transistor lead connections are identified in Fig. 209 With but few exceptions, the collector electrode is *always* connected internally to the metal shell or case of a power transistor,

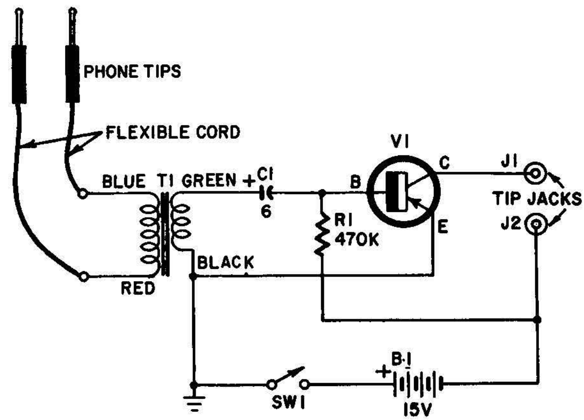


Fig. 211. Schematic wiring diagram of a single-stage headphone amplifier.

with the actual external electrical connection made by a mechanical contact (mounting screw, for example).

The lead arrangements used with other popular semiconductor devices are shown in Fig. 210. Junction tetrode connections are shown at (a) and (b), a triode having an internal shield at (c)

Fig. 212. Exterior view of the headphone amplifier.

and a Unijunction at (d). Note the wider spacing between the collector and nearby leads where an "in-line" arrangement is employed, as (a) and (c). Care must be taken not to confuse a triode with an internal shield with tetrode transistors, for similar lead arrangements are used in both cases. Use the component's type number as a guide: 2N—— numbers refer to triodes, 3N—— numbers to tetrodes.

Circuit diagrams

Three basic types of diagrams are used in electronics: (a) the *schematic,* (b) the *pictorial* and (c) the *block* diagram. Of these, the schematic is the most important and, as we discussed earlier, is made up of symbols connected with lines representing circuit

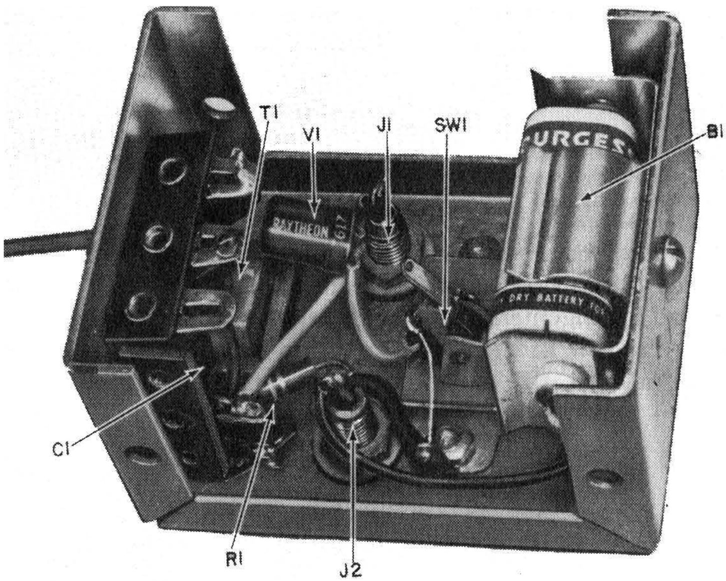


Fig. 213. Interior of the headphone amplifier, with major components identified.

wiring to show the scheme of the circuit. An experienced technician, advanced hobbyist or an engineer can tell more from the schematic diagram than from either of the other two, or from the other two types put together, for that matter. Not only that, he can actually wire a circuit from the schematic alone, and often prefers to work with it.

The pictorial diagram, as the name implies, is an actual "picture" of the equipment. Such a diagram gives a physical picture of the equipment, but is difficult to interpret in terms of circuit performance. While handy for wiring a circuit or for locating a specific part in a piece of equipment, it is of little value otherwise.

Finally, the block diagram is a simplified form of schematic. It shows a more general scheme of the unit. Here entire operational stages are replaced by blocklike symbols, with component details omitted. Block diagrams are also used to show the general interconnections between subassemblies in complex equipment.

You can obtain a better understanding of the relationship between schematic and pictorial diagrams by making a point-to-point comparison between these two types of diagrams for a single piece

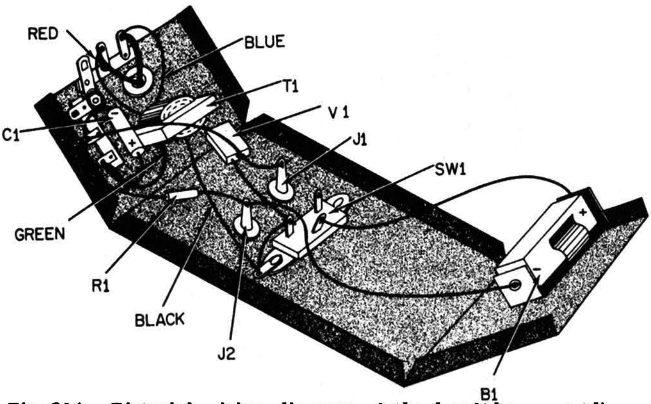


Fig. 214. Pictorial wiring diagram of the headphone amplifier. Compare this with the schematic given in Fig. 211 and the actual photograph of a wired model given in Fig. 213.

of electron equipment. The schematic diagram for a single-transistor headphone amplifier is given in Fig. 211. An exterior view of the unit is shown in Fig. 212 while an interior view, with major components identified, is shown in Fig. 213. For comparison purposes, a pictorial diagram is given in Fig. 214.

This useful piece of equipment is easily duplicated in the home workshop. Serving to increase the overall sensitivity and effective input impedance of a pair of low-cost magnetic headphones, the instrument may be used with receivers, amplifiers, signal tracers and similar devices requiring headphones. All the parts are standard and readily available through both local and mail-order supply houses.

Referring to the schematic, T1 is an Argonne type AR-104 transformer, C1 a 6-μf 15-volt electrolytic capacitor, R1 a 470,000-ohm ½-watt resistor, SW1 a spst slide switch, V1 a p-n-p transistor (more about this later) and B1 a Burgess type Y10 15-volt battery. J1 and J2 are standard phone tip jacks. The unit is assembled in a miniature aluminum box measuring 2¾ x

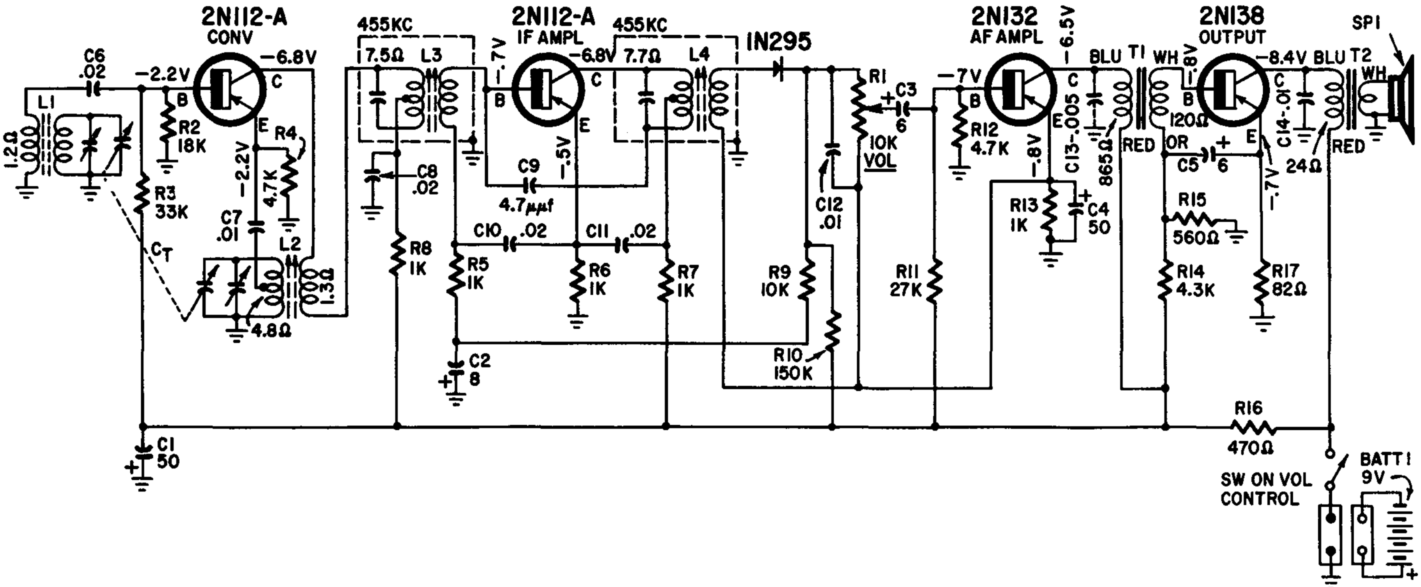


Fig. 215. Schematic diagram of a four-transistor receiver. Compare this to the interior photographs of the set given in Figs. 216, 217.

$2\frac{1}{8}$ x $1\frac{5}{8}$ inches overall. A small clip is provided for the battery. Type CK722, 2N222 or GT-222 transistors may be used.

In use, the phone tips attached to the end of the flexible cord are plugged into the equipment headphone jacks (with which the headphone amplifier is used). Standard 1,000- or 2,000-ohm magnetic headphones are connected to J1 and J2. With SW1 closed,

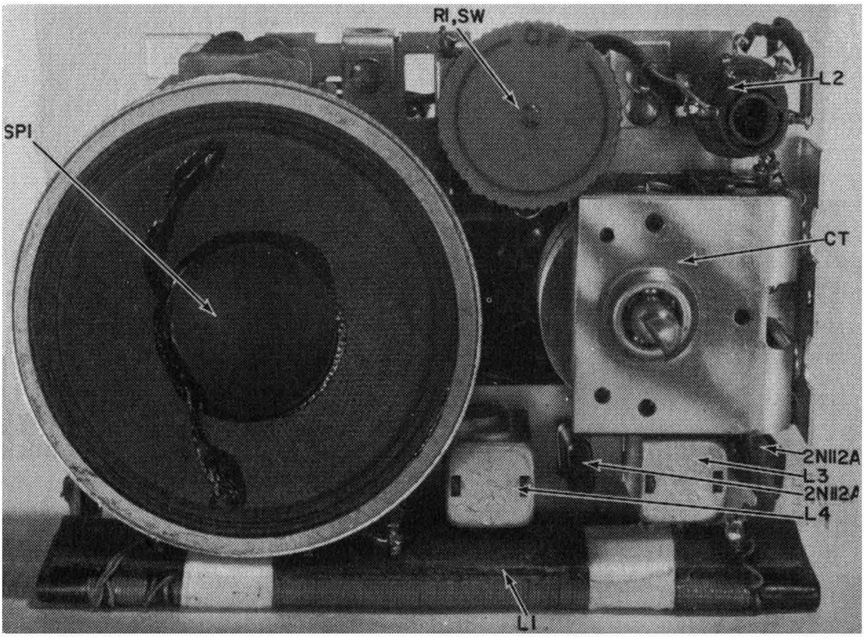


Fig. 216. Interior photograph of four-transistor receiver with leather case removed. Visible major components are identified.

the transistor amplifies audio input signals, driving the magnetic headphones to greater output volume. At the same time, T1 acts to increase the circuit's effective input impedance to approximately 20,000 ohms, or 10 to 20 times that of the magnetic headphones used alone.

Compare the location and wiring of the various components as shown in the schematic (Fig. 211), in the photograph of the wired equipment (Fig. 213) and in the pictorial diagram (Fig. 214). Note particularly the connections to the transformer, switch and battery. Note, too, that the transformer leads are *color-coded,* as identified in the schematic. As a check, *try drawing your own schematic diagram* from the pictorial and without referring to

Fig. 211. Don't worry if you don't know *how* the circuit works, we'll discuss transistor operation in later chapters.

The headphone amplifier is a relatively simple device, but the technique used to "read" the schematic diagram in terms of equipment wiring is the same whether 1 or 30 transistors are used.

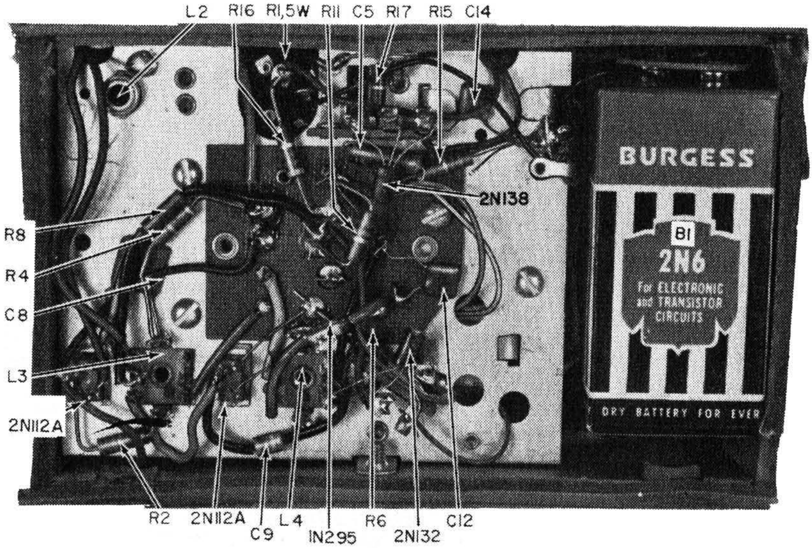


Fig. 217. "Below chassis" (back) view of the four-transistor receiver. Smaller electrical components are found here.

As a further example, the receiver schematic in Fig. 215 can be compared with the front and back interior views of the receiver in Figs. 216 and 217 respectively. Major component parts are identified in the photographs.

The schematic is of a portable receiver using four p-n-p transistors in a superhetrodyne circuit. It tunes the AM broadcast band from 500 to 1600 kc. Power is supplied by a single 9-volt battery. The if is 455 kc.

Having compared pictorial and schematic diagrams, let's see the relationship between a block diagram and a schematic. The block diagram for the receiver is given in Fig. 218. Note that each operating stage is shown by a simple block. Single lines connect the blocks, with arrows showing the direction in which signals are transferred from one stage to another. Thus, six blocks serve to illustrate the overall circuit of a receiver using 17 resistors, 15

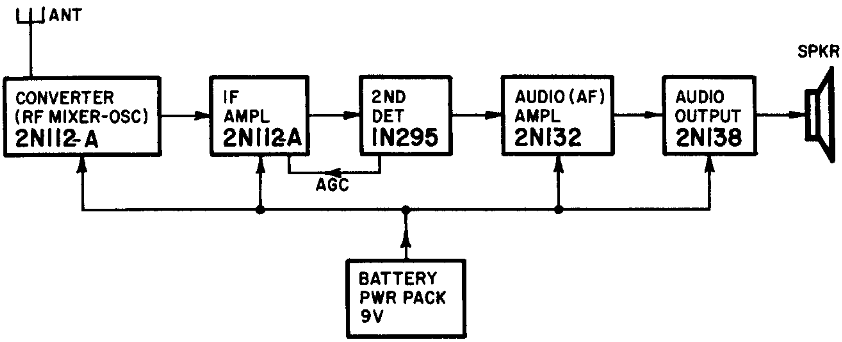


Fig. 218. Block diagram of the four-transistor receiver.

capacitors, 2 audio transformers, 4 rf and if coils or transformers, 4 transistors, a diode, a PM loudspeaker, an spst switch and a power supply battery.

Once you've fixed the basic idea of a block diagram in your mind, practice using this type of presentation by drawing your own block diagrams of receiver and amplifier circuits given in popular radio and electronic magazines. As you acquire skill through practice, you'll soon find it easy to draw a schematic from a pictorial or even from an actual piece of equipment and to prepare block diagrams from schematics. You'll find such practice a lot of fun. More important, you'll increase your knowledge of electronics and sharpen your technical know-how.

CHAPTER 3

Transistors . . . and how they work

Not too many years ago, yours truly asked a learned and somewhat stuffy college professor how a certain device worked. The prof was a smart man—he had a string of degrees and a stack of honors to prove it. And he was most cooperative.

"Why, it's very simple," he said, turning to the blackboard. He wrote a mathematical expression that nearly filled the board, mentioning in passing, "This is obvious."

"From this," he continued, "we can conclude the following." Erasing the first expression, he wrote another that completely filled the board with a little crowding around the edges.

Eight blackboards and 2 hours later, the good old prof summarized his final expression with "and that explains how the device works."

"But, prof," protested one of the other class members, "isn't there a simpler way of striking a match."

The operation of a transistor, like that of a match, can be explained with several pages of mathematical expressions and derivations. But there is a simpler way. . .

Transistors and related electronic devices depend for their operation on the electrical properties of a class of substances called *semiconductors.* This isn't a new word—we used it in earlier chapters, but we didn't define it at that time. Of course, it may be that you've guessed its meaning by now, or perhaps you are still wondering about it. Let's pause for a closer look.

Take the last part of the expression—*conductor.* Just about everyone knows the definition of an electrical conductor: a material which offers little resistance to the flow of electric current. The free movement of electrons, if you please. Copper, silver, gold and most other metals are pretty good conductors.

Materials which oppose current flow are termed *insulators.* Glass, many ceramics, rubber, most plastics, dry wood and similar materials are considered good insulators.

Unfortunately, this simple "black-and-white" classification of conductors and insulators is not strictly true. A material considered a good insulator *will conduct* if enough electrical pressure (voltage) is applied to it, just as a solid block of wood will allow water to seep through if enough water pressure is used. The chief difference between conductors and insulators lies in the ease with which one will permit current flow as compared to the other.

In a conductor, the electrons making up the outer shells of individual atoms may be detached quite easily by the application of electrical pressure or magnetic forces. These "free" electrons moving through the material constitute a flow of electric current. In an insulator, on the other hand, the electrons are bound tightly to their atomic orbits and can be dislodged only with the application of relatively strong forces.

Semi, the first part of our expression, means "half." A semiconductor, then, is a material "half like" a conductor, or, by the same token, "half like" an insulator. In general, they are crystalline materials with electrical properties intermediate between those of a good conductor, such as silver, and a good insulator, such as porcelain. For the most part, the semiconductors are elements in Column IV of the periodic table of chemical elements, including carbon, silicon, germanium, tin and lead. In addition to these, a number of compounds have electrical properties similar to those displayed by the semiconductor elements. Some of these are various sulphides, oxides and other compounds of cadmium, selenium and copper. Others are mineral-like substances such as carborundum.

In a semiconductor, the internal electrical forces bonding electrons to their atoms *(valence bonds)* may be affected easily by external physical forces, such as light, heat or moderate electrical voltages. Thus, a semiconductor may act somewhat like an insulator (or a poor conductor, which is essentially the same thing) or, with the application of suitable external forces, the same material may act as a fairly good conductor.

Holes and electrons

As we have seen, electric current flow is basically a movement of electrical charges through a material and, in a conductor, consists of a movement of negatively charged free electrons. In semiconductors, on the other hand, the current flow may include *both* negative *and* positive charges. The negative charges are electrons, just as in a conductor, while the positive charges are electrical "holes" in the material's atomic-molecular structure.

To understand what we mean by a "hole," consider a fairly stable and electrically neutral molecule. Suppose, now, that we are able to remove an electron from it. The molecule will acquire an electrical charge equal in magnitude to the charge of the missing electron, but with opposite polarity. Since the electron has a negative charge, the molecule will acquire a *positive* charge. The position which the missing electron formerly occupied in the molecular structure has become an electrical "hole" which can, under the proper circumstances, be "filled" by another electron.

Our slightly altered molecule will readily accept an electron and will become electrically neutral once again. If this electron is obtained from a similar molecule nearby, the second molecule is left with a hole and a net positive charge. The second molecule, in turn, might "rob" still a third molecule to replace its missing electron. In this way, the hole can readily *migrate* through the material, effectively jumping from molecule to molecule and behaving as *if it were an actual particle with the mass and size of an electron, but with a positive charge.* When discussing semiconductor action, it is often convenient to treat holes as if they, like electrons, were actual subatomic particles.

Holes, having a positive charge, move through a semiconductor in a general path opposite to that taken by electrons, but at about half the speed of negative particles. A movement or drift of either electrons or holes (or both) through a semiconductor constitutes a flow of electric current. As far as external circuits are concerned, it is impossible to differentiate between electron current and hole current. Since both electrons and holes "carry" an electric current, they are called, appropriately enough, *carriers.*

In a given semiconductor material, either of three conditions may exist: (1) there may be equal numbers of holes and free electrons, (2) there may be more holes than electrons and (3) there may be more electrons than holes. If there are equal numbers of both holes and electrons, we have a "pure" or *intrinsic* semiconductor. If the positive holes predominate, we have a positive or *p-type* semiconductor. Finally, if there are more electrons than holes, we have a negative or *n-type* semiconductor.

Since holes are the predominant current carriers in p-type materials, any electrons present are termed *minority carriers.* In a similar fashion, holes are minority carriers in n-type semiconductors.

A pure (intrinsic) semiconductor (Column IV of the periodic table) such as germanium may be given p-type properties by alloying it with a small amount of an "impurity" element from the

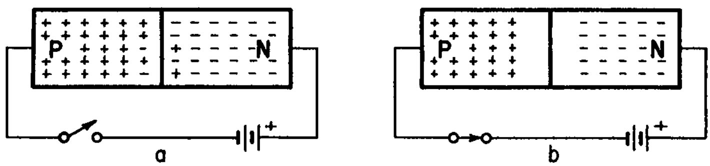


Fig. 301 a, b. Electron and hole conduction in a simple p-n junction. Holes are represented by plus signs (+), electrons by minus signs (—). In the junctions shown at (a), (b) the two semiconductors are doped equally.

Column III, such as indium, boron, aluminum or gallium. The germanium is said to be *doped* by the impurity. Because the net effect is to increase the number of holes present in the semiconductor and hence the material's ability to "accept" free electrons, the impurity element is termed an *acceptor.*

Conversely, if an intrinsic semiconductor such as germanium is doped with an element from Column V of the periodic table, such as phosphorus, arsenic, or antimony, the resulting alloy has n-type characteristics. In this case, since the impurity element contributes or "donates" free electrons to the material, it is called a *donor.*

The diode junction

Before tackling the operation of the transistor as a solid-state amplifying device, let's examine the operation of its somewhat less complex "cousin," the semiconductor diode. Referring to Fig. 301-a, consider a series circuit made up of a dc power source (a battery, for example), a spst switch and a single junction of p-type and n-type semiconductor materials. The negative terminal of the power source is connected through the switch to the p-type material, with the positive terminal connected to n-type semiconductor. The p-type material, while it has a few free electrons, has an overwhelming preponderance of positive holes, represented in the diagram by "plus" signs. The n-type material, on the other hand, has a preponderance of negatively charged free electrons, represented by "minus" signs, plus a few holes.

With the switch closed (Fig. 301-b), the electrical pressure (voltage) applied by the battery causes the positive charges to migrate toward the negative terminal. Remember that *unlike charges attract.* At the same time, the negative electrons rush toward the battery's positive terminal, leaving the central junction virtually depleted of current carriers. Except for the relatively few electrons present in the p-type material and the few holes in the n-type semiconductor, there can be no exchange of current carriers at the junction and hence little or no flow of current through it. Since little or no current flows through the circuit, the p-n junction acts as a *high resistance,* virtually an open circuit.

Suppose, now, that the battery polarity is reversed (Fig. 301-c), with the negative terminal connected to the n-type material and the positive to the p-type. Here, the current carriers move toward the junction and are neutralized. In effect, the positive holes, moving away from the positive side of the battery *(like charges repel),* migrate to and across the junction into the n-type material, where they are filled by the abundance of free electrons. Similarly, many of the free electrons in the n-type semiconductor migrate toward the junction and away from the negative terminal. Here, they cross the junction into the p-type material, filling and neutralizing the positive holes. As the action continues, new holes are formed constantly in the p-type material and new electrons are released in the n-type semiconductor by the electrical pressure applied by the battery.

This constant exchange of holes and electrons at the junction, plus the formation of new current carriers and their continual movement toward the junction, constitutes a flow of electric current which is maintained by the battery. A sizable current can result, and the p-n junction now behaves as a low resistance.

From this we see that the two-electrode (diode) p-n junction is a unilateral device. It permits an appreciable current flow in one direction and little or no current flow in the opposite direction. When the battery voltage is applied to make the junction act like a low resistance, the diode is said to be *biased* in its *forward* or conducting direction. If the voltage is applied so that the junction acts like a high resistance, the diode is said to be biased in its *inverse* or nonconducting direction. The resistances (and currents) obtained under these conditions are called the *forward resistance* (current) and *back* or *inverse resistance* (current), respectively.

If the semiconductors making up the junction are doped so that the p-type material has just about as many holes as the n-type material has free electrons, the electric current *at the junction* will be made up of holes and electrons in equal quantities. Just as many holes will be carried over into the n-type material as electrons are transferred into the p-type semiconductor.

On the other hand, if one semiconductor is doped more heavily than the other, one or the other current carriers (holes or electrons) may dominate the junction current, and there may be a greater carryover in one direction than the other. As an example, if the p-type material is doped more heavily than the n-type, more

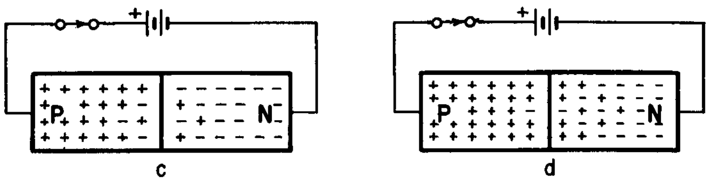


Fig. 301 c, d. In (c) the two semiconductors are doped equally. In (d), the p-type semiconductor is doped more heavily than the n-type and hence has more holes than the n-type has free electrons.

holes will carry past the junction into the n-type semiconductor than electrons are carried into the p-type, and holes will dominate the junction current. This condition is illustrated in Fig. 301-d.

Although the current *within* the semiconductor diode is made made up of both holes and electrons, the current in the *external* circuit (battery and leads) consists of electrons only. It is impossible to determine by examination of the external circuit whether holes or electrons predominate within the diode. As far as the external circuit is concerned, electrons leave the negative battery terminal and move through the connecting wire to the n-type semiconductor. At the same time, other electrons move from the p-type material (as holes are formed within the semiconductor) and back to the positive battery terminal. Diode current may be determined by using Ohm's law and dividing the battery voltage (in volts) by the diode's resistance (in ohms).

Conversely, a diode's resistance may be determined by measuring circuit current (using a suitable series ammeter or milliammeter) and dividing this into battery voltage. Either back or forward resistance may be calculated in this way simply by reversing battery polarity and obtaining a new current reading. Both resistance values, of course, are measured in ohms.

A diode's actual resistance in a practical circuit will depend, not only upon the electrical characteristics of the individual diode, but upon the applied voltage, upon ambient temperatures and so on. If *too much* voltage is applied in the nonconducting direction, the junction may break down, with its resistance dropping to a low value. The voltage at which this breakdown occurs is called the diode's *Zener voltage.* A number of special-purpose diodes are designed for operation within their breakdown region and are called Zener diodes.

Basic transistor action

A triode transistor is made of alternate regions of p- and n-type semiconductors. In one respect, a p-n-p transistor might be considered as p-n and n-p diodes joined in one assembly in such a way that there is a single, relatively thin, n-type layer. Thus, the triode transistor is a *two-junction* device as contrasted to the diode, which has a single junction. Depending on how the transistor is used, each junction may be biased in either its forward (low-resistance) or reverse (high-resistance) direction. In operation, transistor action depends on the control of current through *both* junctions by varying the current through one.

A typical circuit arrangement is illustrated in Fig. 302-a. As in the diode, holes are represented by "plus" signs, electrons by "minus" signs. Battery B2 is used to apply a fixed bias voltage between the outer p-type electrodes and hence *across both junctions.* Battery B1 is connected to apply a bias between the central n-type electrode and the lower p-type electrode when the spst switch (SW) is closed. A current-limiting resistor (R) is connected in series with B1 and SW to keep circuit current within a safe value when the switch is closed.

With the switch (SW) open, the positive carriers in the upper and lower electrodes tend to migrate away from B2's positive and toward its negative terminal. Thus, the holes in the upper electrode will move *away* from the top junction and toward the negative battery terminal, leaving the junction area more or less depleted of charges. At the same time, the holes in the lower electrode will move away from B2's positive terminal and toward the lower junction. Since no external voltage is applied directly to the central electrode (SW open), the negative electrons in this region will remain more or less evenly distributed, although a few may move toward the lower junction.

Under these conditions, the upper junction has a high resistance. There are relatively few carriers near it to effect the interchange and carryover necessary to conduction. The lower junction has somewhat lower resistance, what with a fair accumulation of holes on the p-side and a few electrons available on the n-side. However, with no voltage applied to the central electrode, there is no pressure for additional electrons to accumulate near the lower junction and its resistance is much higher than that of a "conducting" junction. Since the upper junction has a high and the lower a moderate resistance, little or no current can flow between the two p-type electrodes.

Now, if a bias voltage is applied to the central electrode by closing switch SW (Fig. 302-b), the situation changes radically. The electrons in this region immediately start moving toward the lower junction and it becomes biased in its forward (low-resistance) direction. Conduction takes place and electrons move from the n-type material into the lower p-type electrode. At the same

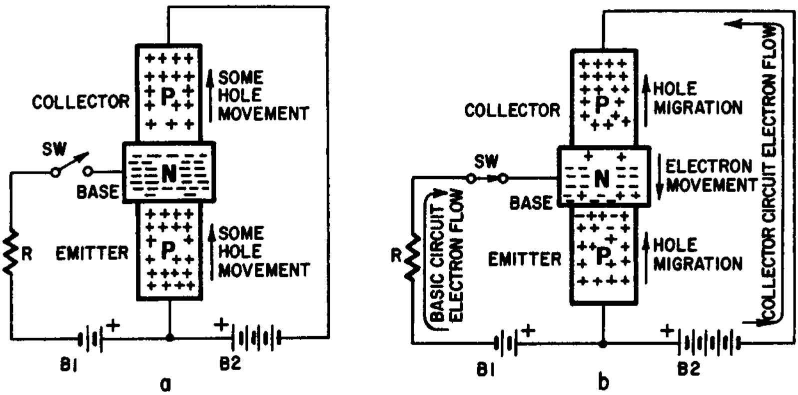


Fig. 302. How a p-n-p transistor can be used as a current amplifier. As in Fig. 301, holes are represented by plus signs, electrons by minus signs. Drawing (a) shows transistor operation without base bias, (b) with base bias current.

time, of course, holes move from the lower p-type electrode into the n-type region. Some of these holes are not neutralized immediately and move across the thin n-type area to the upper junction and into the *upper* p-type material. Since there are now suitable carriers available on the n-side of the upper junction, its resistance drops from a high to a moderate value.

With bias applied to the central region, then, the overall resistance between the two outer p-type electrodes drops from a very high to a moderate or relatively low value made up of the moderate-resistance upper junction in series with the low-resistance lower junction. With this drop in effective resistance, the circuit current supplied by B2 increases by a substantial amount.

Since the lower p-type electrode injects or "emits" holes into the n-type region, it is called the *emitter*. Similarly, since the upper p-type electrode receives or "collects" charges from the central region, it is called the *collector*. The central electrode is called the *base*, not only because its central location makes it "basic" to transistor action, but because it is the electrode corresponding to the small cube or "base" of semiconductor material used in the original *point-contact transistor* (see Chapter 1).

With no series resistor (R) to limit current in the emitter–collector circuit and with, in this case, a higher voltage applied between emitter and collector electrodes, the collector current is much greater than base current. In addition, a relatively small change in base current will cause a corresponding but much larger, change in collector current. This, then, is a basic condition for amplification. *By varying a relatively small current through one junction* (base–emitter), *we control a larger total current through both transistor junctions* (upper collector–base and lower base–emitter junctions).

To utilize the control of one current by another in a practical amplifier, we need simply apply our input "control" signal to the base–emitter electrodes and to insert a useful "load" or output device of some sort in series with the collector lead. A typical load might be a pair of earphones, a resistor, a relay or the primary winding of a transformer.

At this point, we can examine the impedance relationships in the transistor's "input" (base–emitter) and "output" (collector–emitter) circuits quite easily. The input impedance depends primarily upon the resistance of the base–emitter junction. Since this is biased in its forward (conducting) direction, its impedance is fairly low. The output impedance depends on the combined impedances of the collector–base and base–emitter junctions. As we have seen, the collector–base junction is *not* biased heavily in its forward direction and, therefore, has a moderate resistance. This, coupled with the low base-emitter junction resistance, gives a moderately high output impedance.

In summary, we find that the basic transistor amplifier stage shown in Fig. 302 has a moderately low input and a fairly high output impedance. This difference between input and output impedances, in itself, could supply power gain (amplification), *even* if the input and output currents were on the same order of magnitude. After all, an input current of, say, 100 μa into a 1,000-ohm impedance represents less power than a similar output current through a 50,000-ohm load.

Transistor characteristics

To identify any item positively, whether it is an electrical device such as a vacuum tube, relay or transformer or a piece of hardware such as a machine screw, nail, rivet or hinge, we must describe it in general terms common to all similar items. For example, to identify a machine screw, we must specify (a) material of construction, (b) finish, (c) diameter or physical size, (d) length, (e) type of thread, (f) number of threads per inch, (g) type of head and so on. While the general characteristics may be common to all items in a class, individual units may differ considerably according to the *exact* values assigned to each specification. Thus, a 6-32 ½-inch brass screw with a flat head is *not* interchangeable with a 10-24 ¼-inch cadmium-plated steel screw with a round head and Phillips slot. But *both* are machine screws and both are similar in that the same terms are used to describe them and both serve as fasteners.

As we have seen, several kinds of information must be specified when dealing with electronic components. These include the type of device (transistor, tube, relay, etc.), physical size and shape, electrode or terminal connections, maximum operating conditions and basic electrical characteristics. All of this information, taken together, constitutes the component's *specifications.*

Regardless of individual characteristics, all transistors are multi-electrode semiconductor devices in which the current (voltage) through one pair of terminals can control the current through another pair. We can classify a transistor in greater detail by indicating its number of electrodes, the semiconductor material used in its manufacture (germanium, silicon, etc.), its type of construction (point-contact, junction—more about this in a later chapter) and the internal arrangement of its semiconductor layers (for example, p-n-p). This much information is adequate for identifying the type of device and providing a general area of classification, but to identify a specific unit accurately, we need detailed data.

Characteristic curves

Since an individual transistor's electrode currents and other electrical properties are not fixed constants but variable which depend upon the transistor's inherent qualities as well as upon applied voltages and exact operating conditions, we cannot describe its electrical properties fully by using a single set of fixed values as we might for a resistor or capacitor, for example. Instead, we must use a graphical description which consists of a set or "family" of curves. Each curve in the family shows the variation of a dependent variable when plotted against an independent variable, with all other electrical quantities held constant.

A set of transistor characteristic curves is analogous to the families of curves used to specify the characteristics of vacuum tubes. It is used in much the same way by practical engineers to design circuits and to predetermine the operating performance of a particular transistor under a given set of operating conditions.

Transistor characteristic curves may be prepared with the transistor connected in any of several ways and with any of a number

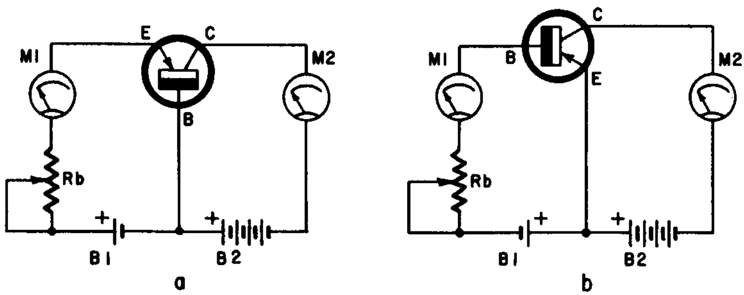


Fig. 303. Basic circuits used to determine a transistor's d c characteristics. The base serves as a common electrode at (a), emitter at (b.)

of circuit currents held fixed as the "constant" quantity. For example, using the circuit arrangement shown in Fig. 303-a, we could measure collector voltage (B2), collector current (indicated by M2), emitter voltage (B1) and emitter current (indicated by M1 and controlled by Rb).

Suppose we used this circuit and held the collector current constant, measuring the emitter current for various emitter voltages. This would give us a set of values which could be plotted on a graph. Then we could choose a *new* value of collector current, holding it constant while preparing another set of values showing emitter current for various emitter voltages. After repeating these measurements a number of times, we could plot our results on a piece of graph paper and we thus have a set (family) of curves showing emitter voltage vs emitter current for various values of constant collector current. These would be the transistor's *input characteristics* for the particular circuit used.

In a similar fashion, we could prepare families of curves showing collector voltage vs collector current for a constant emitter current (*output characteristics*), emitter current vs collector voltage for a fixed collector current (*forward characteristics*) or collector current vs emitter voltage for a constant emitter current (*feedback characteristics*).

But all of these families of curves would show the transistor's operating characteristics only when used in the circuit arrangement given in Fig. 303-a. Similar curves could be prepared by using the circuit shown in Fig. 303-b. Here, for example, collector current could be plotted against different values of collector voltage for a fixed value of base current. Additional curves would be

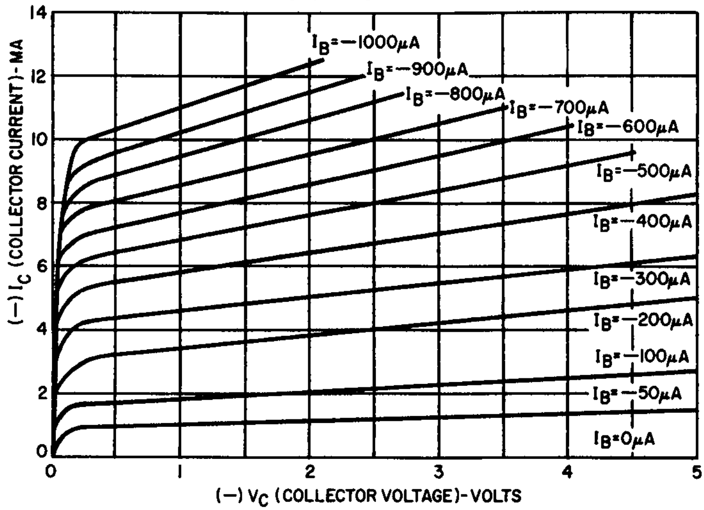


Fig. 304. A family of curves showing the d c operating characteristics of a typical p-n-p transistor. A graph such as this one may be prepared by using a circuit like that shown in Fig. 303. (b).

plotted on the same graph for other values of base current to make up the entire family. Such a family of curves are the *collector* or *output characteristics* for the transistor in the second circuit arrangement. A typical family of collector characteristic curves for a p-n-p surface-barrier transistor are shown in Fig. 304.

Maximum ratings

Like any other electrical component, the transistor may be seriously damaged if excessive voltage is applied to its electrodes or if too much current is drawn through it. An excessive voltage applied to the emitter and collector electrodes, for example, may "punch through" the thin base layer, destroying both junctions and rendering the transistor useless. Unusually high currents will overheat the transistor and, in some cases, may even melt its electrodes.

Of course, heat damage may be caused by externally applied as well as by internally generated heat. External heat sources include hot soldering irons and electrical components which run "hot," such as vacuum tubes and power resistors, plus common heating devices such as radiators, hot-air registers and sun lamps. *A transistor need not be in use to be damaged by excessive heat.* Serious damage may occur if a transistor is stored in an overheated room or if transistorized equipment is placed close to a hot stove, furnace or similar heat source.

Internally generated heat is the result of a power loss. The transistor, like any other electrical or mechanical device, is not 100% efficient. Some part of the power applied to it will be lost during operation and this will show up as a temperature rise. If the heat is dissipated rapidly enough through the emitter and collector electrodes, the junction temperatures may stay within safe limits. But inefficient operation, overload and similar conditions might cause permanent damage.

To help circuit designers, technicians, hobbyists and other users of their products to work within safe limits, transistor manufacturers specify maximum ratings for their units. As a minimum, most manufacturers specify the following values:

(a) *Maximum collector-to-emitter voltage.* Referring to the basic amplifier circuit shown in Fig. 305, this is the value measured by voltmeter V3. It may range from as low as 4.5 volts for small high-frequency units to well over 200 volts for special transistors designed for switching or high-power applications. The measurement is generally made with the base connection "open." In addition, some firms also will specify *maximum collector-to-base voltage* (emitter open) and *maximum emitter-to-base voltage* (collector open). In Fig. 305, these would be the voltages measured by voltmeters V2 and V1, respectively, with the appropriate electrodes disconnected. Since a transistor's characteristics vary with temperature, maximum voltages are generally listed at a specific ambient temperature.

(b) *Maximum collector current.* Specified in milliamperes or amperes, this value may range from less than 10 ma for subminiature transistors to 30 amperes or more for heavy-duty units. Referring to Fig. 305, this would be the current measured by meter M3. A few manufacturers will give an additional listing of *maximum emitter current* (measured by M2, Fig. 305). As a rule, this second figure will be about the same as the first.

(c) *Maximum power dissipation.* Varying considerably from one type of transistor to another, this value depends on the transistor's construction, its physical size and the ambient temperature conditions under which it is used. It indicates the maximum

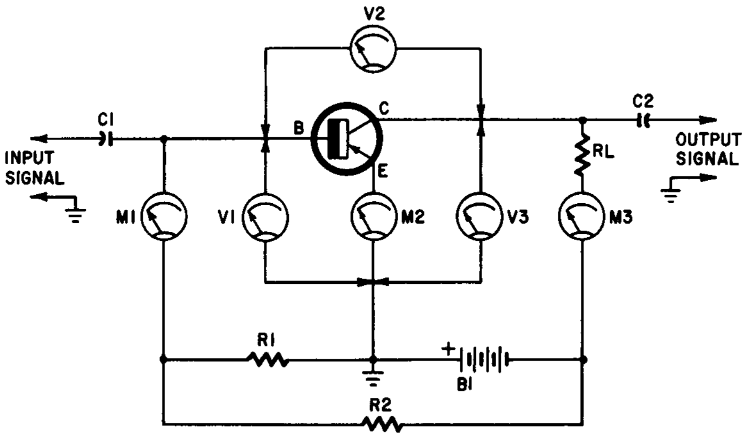


Fig. 305. *Important voltages and currents in a typical transistor amplifier stage. A p-n-p transistor is used.*

electrical power the device can safely dissipate as heat and is generally specified in milliwatts or watts. Typical values may range from less than 10 mw for subminiatures to over 200 watts for power transistors. In a few instances, a manufacturer may list *maximum collector dissipation* instead of total device dissipation.

(d) *Maximum storage temperature.* Generally given in degrees Centigrade, this is the maximum temperature to which the transistor can be subjected without permanent damage. The rating may be a fixed value, such as 85°C, or may be given as a temperature range as, for example, −55° to 85°C. A few manufacturers will specify a *maximum operating temperature* as well as storage temperatures.

While vacuum tubes, resistors and other electronic components generally have *maximum ratings,* these values are often conservative and little or no damage will result if the components are operated slightly past their ratings for short periods. This is *not* always the case with transistors, however. The ratings specified by the manufacturer should be considered as *absolute.* Operation of a transistor at values in excess of its ratings, even momentarily, may cause permanent damage.

Transistor parameters

If connected in similar circuits with identical operating conditions (bias voltages, ambient temperatures and so on), all transistors of a given manufacturer's type should have about the same electrical characteristics as far as such factors as input and output impedances, power output and gain are concerned. Since these characteristics are constant for a given transistor under specified operating conditions, they are called the transistor's *parameters.*

The expression "parameter" is borrowed from mathematics and is a term referring to *arbitrary constants,* values which are constant only under a specified set of conditions, as distinguished from true or absolute constants (such as *pi,* 3.14159), which always have the same value.

Since transistor parameters are arbitrary constants, the individual units specified as well as their numerical value will vary considerably from one manufacturer to another, with various transistor types, with the test circuit arrangement used, with bias currents and with actual operating conditions. In most cases, too, a "specified" value represents a typical or average for a particular transistor type (number) under stipulated conditions. For example, a manufacturer may indicate that a specific unit has a gain, say, of 65. In practice, production units of that type may have gains ranging from 55 to 75, depending on the manufacturer's normal "spread" of values. The more important parameters which manufacturers specify include (a) input and output impedances (or admittances), (b) amplification factor, (c) frequency cutoff, (d) collector cutoff current, (e) power gain, (f) power output, (g) interelectrode capacitances and (h) noise figure. Let's examine each of these in turn.

(a) *Input and output impedances (admittances).* Both of these values are specified in ohms. The input impedance is the effective load which the transistor stage will present to a signal source; the output impedance is its internal impedance when used to drive an external load. Some manufacturers will specify admittances rather than impedances. These are essentially the reciprocals of the impedances and are expressed in *mhos.* These two parameters are useful when coupling transistors to each other or to other devices. A maximum transfer of energy between a source and its load occurs when the output impedance of the source equals the input impedance of the load, or, in more familiar terms, when their impedances are *matched.*

(b) *Amplification factor.* This is a basic measure of the maximum gain possible when the transistor is used as an amplifier. It is expressed as a "pure" number rather than in specific units and indicates the ratio between a transistor's input and output signal currents. Analogous to the *mu* (μ) of a vacuum tube, it is termed *alpha* (α) when the transistor is connected with its base grounded and *beta* (β) when the emitter is grounded. Except for the virtually obsolete point-contact transistor, alpha is always less than 1, but beta values may run into the hundreds. Beta is referred to as *base current gain.*

(c) *Frequency cutoff.* When a transistor is used as an amplifier, its effective gain varies with the frequency of the signal handled. This variation is due to a number of factors, including the nature and value of internal impedances, the time required for holes and electrons to migate through the semiconductor material, and so on. The frequency at which the transistor's gain is approximately 70% of its low-frequency value is called the frequency cutoff (or *cutoff frequency*). This value depends on each transistor's type and construction. For example, a typical high power transistor may have a frequency cutoff of only 8 or 9 kc, while a carefully manufactured high-frequency type may have one in the thousands of megacycles.

Most transistors can furnish considerable gain at signal frequencies well past their nominal "cutoff" values. Because of this, some prefer to specify the high-frequency characteristics of their transistors by a *figure of merit.* This is essentially the highest frequency at which the transistor has unity (1) gain, and, therefore, approximates the highest frequency at which the unit will serve as an oscillator.

(d) *Collector cutoff current.* Sometimes called *leakage* current, this value is expressed in microamperes or milliamperes. Basically, it is the current flowing when a specified reverse bias voltage is applied between the base and collector electrodes, emitter open. In a sense, it indicates the transistor's "quality" as far as internal leakage currents are concerned. A good *rule of thumb* . . . the lower the collector cutoff current, the better the transistor.

Many low-cost transistor testers measure a transistor's *leakage* by checking collector current with a voltage applied between emitter and collector, base open. This is approximately equal to the transistor's collector cutoff current multiplied by its beta.

(e) *Power gain.* In its simplest form, gain is the ratio of the output to input signal levels in an amplifier. Since a signal level may be expressed in terms of voltage, current or power, any of these terms may be used when specifying gain. However, power levels are the most significant for many types of work. A transistor's power gain is generally given in terms of decibels (db), a logarithmic unit. In practice, the power gain is equal to 10 times the logarithm (to the base 10) of the ratio in output to input power levels, with both powers expressed in the same terms (watts or milliwatts).

(f) *Power output.* Expressed either in milliwatts or watts, a transistor's signal power output is significant when working with circuits in which electrical power is needed to drive an electrical or electromechanical device, such as a relay, loudspeaker, earphones, transmitting antenna or lamp bulb. Where the "quality" of the signal handled is important, as in record players and radio receivers, the power output figure is useful only where distortion levels are indicated. Some transistor amplifiers, for example, are rated in terms of "undistorted power output." Such a figure is considerably less than the *maximum* power output.

(g) *Interelectrode capacitances.* Expressed in terms of micromicrofarads ($\mu\mu$f), these specifications refer to the effective capacitance between various pairs of electrodes, between the emitter and collector or between base and emitter, for example. They are of importance when the transistor is used in high-frequency tuned circuits, such as rf and if amplifiers, or where the transistor handles pulselike signals.

(h) *Noise figure.* Any electronic device may introduce a certain amount of random signal power or noise in signals which it handles. This noise signal is generally the result of thermal (heat) agitation of current carriers (holes and electrons). A transistor's noise factor (or *noise figure,* NF) is a quantity which compares its performance with a theoretically ideal noiseless amplifier having the same input and output impedances, gain, power output and other characteristics. This theoretical amplifier is assumed to have a noise factor of 0 db, with its output serving as the reference level. The transistor's noise factor, then, is the ratio between its noise power output and that of the reference amplifier. It may be expressed as a simple numerical ratio or in terms of db (in the same way that power gain is specified). From this, we can see that it also may be defined as the ratio of the signal-to-noise ratio of its output signal to that of its input signal. Good quality transistors have noise factors lower than those of comparable vacuum tubes.

P-n-p and n-p-n types

In our discussion of basic transistor action earlier in this Chapter, we assumed that the transistor used a p-type material in its emitter and collector electrodes and had an n-type base (Fig. 302). It is called, appropriately enough, a p-n-p transistor. An

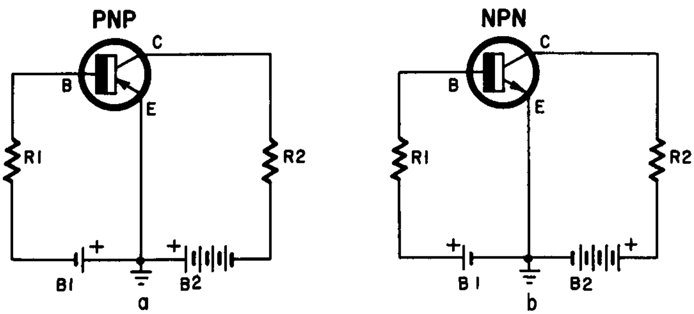


Fig. 306. Bias connections for p-n-p (a) and n-p-n (b) transistors. The d c polarities are generally referred to the electrode which is common to both the input and output connections . . . the emitter in this case.

equally good transistor may be assembled with n-type emitter and collector electrodes and a p-type base. This is identified as an n-p-n transistor and is shown, schematically, by reversing the direction of the arrowhead on the emitter electrode. Identical stages using p-n-p and n-p-n transistors are illustrated in Figs. 306-a and -b, respectively.

In practice, the n-p-n transistor operates in much the same fashion as the p-n-p unit, except that the roles of electrons and holes are interchanged and battery polarities are reversed. Where a p-n-p unit is used, for example, the collector and base electrodes are generally biased *negatively* with respect to the emitter. If an n-p-n unit is used, on the other hand, these electrodes are biased *positively* with respect to the emitter. Commercial circuits may use all p-n-p types, all n-p-n types or a combination of the two, depending on the personal preferences of the individual designer, economic considerations, availability and the particular circuit.

Transistor circuit configurations

Any electrical amplifying device, whether a vacuum tube, transistor or magnetic amplifier, has, basically, four signal terminals. The control signal is applied to a pair of *input* terminals, while the amplified signal is obtained from a pair of *output* terminals. Since the triode transistor has only three electrodes (emitter, base and collector), it follows that one of these must be common to both the input and output circuits. In fact, *any* of the three may be the common one, depending on circuit arrangement.

In practice, then, the input signal is applied between one electrode and the common one, while the output is obtained between the remaining and common electrodes. From this, it is easy to see that there are three basic transistor circuit configurations: (1) *common-emitter,* (2) *common-base* and (3) *common-collector.* The common electrode is often connected to circuit ground. As a result, the word "grounded" may be substituted for "common" when describing a particular configuration.

Since the vacuum tube predates the transistor as a practical amplifying device, it is reasonable to compare these two units. In typical applications, the transistor's *emitter* is roughly analogous to a vacuum tube's *cathode,* the *base* to a tube's *grid* and, finally the *collector* to the *plate.* Carrying this analogy a step further, we find that the common-emitter circuit corresponds, roughly, to the familiar common-cathode tube circuit; the common-base arrangement is similar to the popular grounded-grid amplifier and the common-collector is roughly equivalent to a tube's grounded-plate circuit (known more popularly as a *cathode follower*).

Regardless of these similarities in applications and basic circuit configurations, however, we must remember that *the transistor is not equivalent to a vacuum tube.* Not only do the two devices differ in physical appearance and in principles of operation, but there are significant differences in electrical characteristics, power requirements and interelectrode impedances. Unlike the vacuum tube, there is a direct resistive connection between all three electrodes in a transistor.

The three basic transistor circuit configurations, together with the corresponding vacuum-tube circuits, are shown schematically in Figs. 307, 308 and 309. Let's examine these in detail.

Common-emitter amplifier

Providing the highest voltage gain with a given transistor type, the common-emitter configuration (Fig. 307) is the most popular of the three basic circuits. It is the one encountered most often in commercial and experimental equipment and, in this respect, is like its analogous vacuum-tube circuit, the grounded-cathode amplifier.

Referring to the schematic diagram, base-bias current is supplied by battery B1 through series current-limiting resistor R1. Collector current is supplied by B2. Depending on the circuit application, the load may be a resistor, choke coil, relay, head-

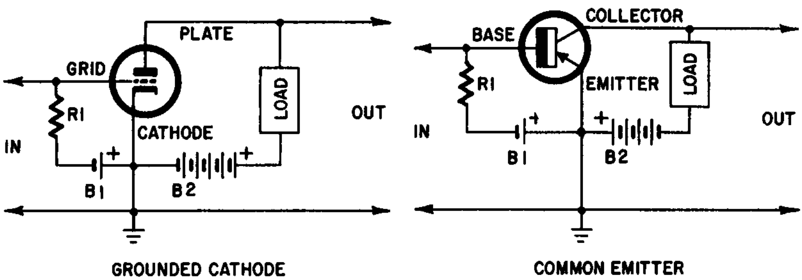


Fig. 307. The basic common-emitter transistor circuit configuration together with the analogous vacuum tube circuit arrangement. Often, the words "common" and "grounded" are used interchangeably . . . thus, a common-emitter stage might be specified also as grounded-emitter.

phone or the primary winding of an af, if or rf transformer. If desirable, a single battery may be used to supply *both* base-bias and collector currents. With the circuit shown, this can be done by removing B1 and returning R1's lower lead to B2's negative terminal, readjusting R1's value as necessary to maintain proper bias current.

In operation, the input signal is applied to the base–emitter circuit across R1 and B1. A dc blocking capacitor may be used in series with one of the input signal leads to avoid changes in base bias current. The amplified output signal is developed in the collector–emitter circuit, appearing across the collector load. Here, the output signal may be coupled through a matching transformer to, say, a loudspeaker's voice coil or it may be transferred to another stage for additional amplification or, if the collector load is a relay, headphone or similar electromechanical unit, may serve to actuate the device.

Assuming a p-n-p transistor to be used, a swing of the input signal toward a negative polarity will cause an increase in collector current and hence an increase in the voltage drop across the collector load. This *reduces* the negative voltage available at the collector (supplied by B2) and hence causes the collector voltage to swing in a *positive* (less negative) direction. The opposite action occurs in the case of an n-p-n transistor, but since the battery voltage is reversed, the net result is the same. In other words, the common-emitter circuit causes a reversal between input and output signal polarities. When the *input* signal swings in a *negative* direction, the *output* signal swings in a *positive* direction, and vice versa. When dealing with symmetrical sine waves, this is equivalent to a 180° phase shift. The transistor circuit shares this polarity-reversing characteristic with its analogous vacuum-tube circuit.

The general characteristics of the common-emitter circuit, then, are the ability to supply good voltage, current and power gains, a low to moderate input impedance, and a moderate to high output impedance. It is suitable for use at audio and radio frequencies and is fairly stable when junction transistors are used. The gain and power output obtained from a common-emitter stage depend on the type of transistor, operating currents and the nature of the collector load.

Common-base amplifier

Roughly equivalent to the grounded-grid vacuum-tube circuit, the common-base amplifier (Fig. 308) has high inherent stability. However, it *does not* offer the high degree of isolation between input and output circuits that characterizes the grounded-grid arrangement.

Referring to the schematic diagram, emitter bias current is supplied by battery B1 through emitter resistor R1. Collector current is furnished by B2. Again, the collector load may be a resistor, choke, transformer primary winding or an electromechanical device. As in the case of the common-emitter configuration, a single battery may be used. In this case, B2 is removed and the lower lead of the collector load returned to B1's negative terminal. In addition, B1's voltage would be raised and R1 dropped in value, with a suitable current-limiting resistor connected between the base electrode and B1's negative terminal. This new resistor would be bypassed with a large capacitor.

In operation, the input signal is applied to the emitter–base circuit across R1 and B1. As before, a dc blocking capacitor may be used in series with one of the input leads. The amplified output signal is developed in the collector–base circuit, appearing across the collector load. Although a small *current* gain may be obtained with some point-contact transistors, the current gain is always *less* than unity (1) with more popular modern types. The collector (output) current is always slightly less than the emitter's (input) current. A voltage or power gain is still possible, however, due to a

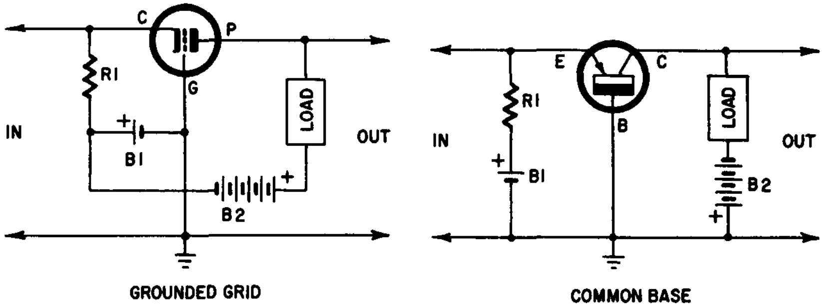


Fig. 308. The grounded-grid circuit is often used for circuit isolation. The common-emitter stage does not provide the isolation of its vacuum-tube counterpart.

large difference between the transistor's input and output impedances.

Assuming a p-n-p transistor to be used, a swing of the input signal toward a negative polarity will cause a decrease in the emitter–base bias current. This, in turn, will reduce collector current and decrease the voltage drop across the collector load, thus increasing the negative voltage available at the collector electrode (supplied by B2) and causing the collector voltage to swing in a negative direction. Since the output signal swings in a negative direction when the input signal swings toward a negative polarity, there is *no signal phase reversal.*

The common-base circuit has a low input impedance and a high to very high output impedance, and can supply good voltage and power gains, although these are somewhat less than those obtained with the common-emitter configuration. It's current gain is very low, however, actually less than 1 where popular transistor types are used. The basic circuit is extremely stable and may be used at audio or radio frequencies. Actual gain and power output levels depend on the type of transistor used, operating currents and the nature of the collector load.

Common-collector amplifier

Unlike the other two basic configurations which feature low to moderate input and high to very high output impedances, the common-collector configuration (Fig. 309) has a *high input* and *low to moderate output impedances.* Because of this characteristic, the circuit, like its analogous vacuum-tube arrangement (the familiar cathode follower), is used primarily for impedance-matching applications.

Referring to the schematic diagram, a single battery, B1, supplies dc operating power for the entire circuit. As in the first circuits discussed, however, separate power sources could be used. Base bias current is supplied through current limiting resistor R1, while emitter current is supplied through the emitter load. As before, the load might be a resistor, choke or similar component. A large bypass capacitor, C1 (shown dotted), may be connected across the power source (B1) as an effective short for ac signal currents.

In operation, the input signal is applied to the base–collector circuit across R1. A dc blocking capacitor may be used in series with one of the input leads. The output signal is developed in the emitter-collector circuit, appearing across the emitter load. The polarity of this signal is such as to partially cancel and to effectively reduce the base-emitter voltage, and thus to limit the *voltage gain* which the circuit can deliver. Since the input signal voltage must be greater than the emitter-to-ground signal (output) for the transistor to operate, *this circuit delivers a voltage gain approaching, but slightly less than, unity (1).* In other words, the output signal voltage is always less than the applied (input) signal.

Another way of explaining this action is to consider the emitter load as an unbypassed impedance introducing a degenerative voltage (or inverse feedback signal) which, in turn, effectively cancels much of the stage's voltage gain. In this respect, the common-collector circuit is quite similar to the analogous cathode follower, which also has a gain of less than unity.

While the inverse feedback signal developed across the emitter load limits stage voltage gain to less than 1, it has a quite desirable secondary effect. The emitter load is between the low-impedance base-emitter junction and circuit ground, thus raising the stage's input impedance. For practical purposes, the circuit's input impedance is a value approximating the emitter load impedance multiplied by the transistor's gain (beta). There is a limiting factor on the magnitude of the input impedance, regardless of how large the emitter load is made. This is the impedance of the high-resistance base–collector junction. With commonly available transistors, the input impedance of this circuit is limited to values under 10 megohms.

Assuming a p-n-p transistor to be used, a swing of the input signal toward a negative polarity will cause an increase in base bias current and hence an increase in emitter current. This, in turn, causes an increase in the voltage drop across the emitter load, reducing the positive voltage available at the emitter electrode (with respect to circuit ground). Since a swing of the input signal in a negative direction causes the output signal voltage to swing in a negative (less positive) direction, the input and output signals are *in phase.*

In summary, the common-collector circuit has a high input and low output impedance. Although its voltage gain is less than 1, it can supply reasonable current and power gains. It does not reverse signal phase. Fairly stable, it is suitable for use at both audio and

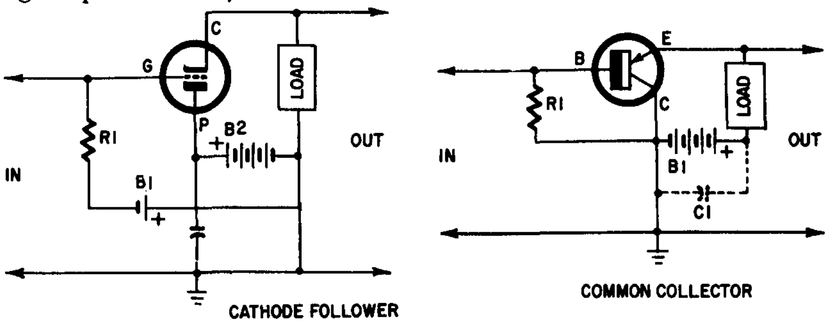


Fig. 309. Both the common collector and its vacuum-tube counterpart, the cathode follower, are used to provide minimum circuit loading.

radio frequencies. As with the analogous vacuum-tube cathode follower, one of its principal applications is as an impedance-matching device, matching the high impedance of one circuit or device to the low impedance of another. Finally, since it offers a low output impedance coupled with reasonable power gain, it is often used as power output stage in receivers and amplifiers.

Phase

Advanced workers as well as beginners often have difficulty in understanding the phase relationships between the input and output signals in an amplifier stage. As we have seen, phase reversal occurs in the common-emitter circuit, but not in the common-base or common-collector amplifiers.

A simple analogy will help in understanding this important relationship. Consider two runners starting a race. Let us suppose that both start on their *left* feet, both take an equal number of equal length strides and both end the race in a dead heat, crossing the finish line with their *right* feet extended forward. The two runners would be *in phase.*

Suppose, now, that the conditions are the same, but that one runner takes slightly shorter strides than the other. To keep up with his opponent, he takes more strides during the race. Again, we'll suppose both end in a dead heat, but that one runner crosses the finish line with his *right* foot extended, the other with his *left* foot forward. In the second case, the two runners are *out of phase.*

Returning, now, to the case of an amplifier, we can consider the relationship between signal waveforms. The two basic situations are illustrated in Fig. 310 in which an amplifier stage is represented by a block diagram and typical input and output signal waveforms are shown. In a true amplifier, the output signal would

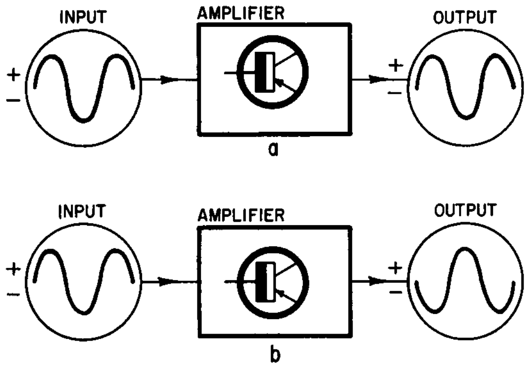


Fig. 310. What we mean by phase when speaking of an amplifier stage. At (a), the input and output signal waveforms are in phase. At (b), the two signals are out-of-phase. In the latter case, we say that phase reversal has taken place.

have greater amplitude than the input signal, but we have assumed the same amplitude for purposes of illustration.

In Fig. 310-a, the input and output signals are *in phase.* When the input signal swings in a positive direction, the output signal also swings in a positive direction. Similarly, when the input signal swings in a negative direction, the output signal changes in a negative direction.

In Fig. 310-b, the input and output signals are *out of phase.* When the input signal swings in a positive direction, the output signal changes in a negative direction, and vice versa. In effect, the output signal is just like the input, but turned "upside down."

When dealing with sine waves, the positive-going and negative-going half-cycles are identical. If we consider a full cycle (consisting of one positive and one negative half-cycle) to represent 360

Table 3–1.

Basic Transistor Amplifier Circuit Characteristics
(Junction Transistors)

Basic Circuit	Input Impedance+	Output Impedance+	Power Gain	Voltage Gain	Current Gain	Phase Change
Common-emitter	Low to Moderate	Moderate to High	Best	Best	Very Good	Reversal
Common-base	Very Low to Moderate	High to Very High	Very Good	Good	Less Than* 1.0	None
Common-collector	Moderate to High	Low to Moderate	Moderate	Less Than 1.0	Very Good	None

* Point-Contact Transistors—Current gain may exceed 1.0; not stable in common-emitter circuit.

+ Hi Power Junction Transistors—Both input and output impedances may be under 100 ohms.

electrical degrees, then the effect of phase reversal is as if the entire signal has been shifted forward or back along its axis by a single half-cycle, that is, 180°. Thus, signal-polarity reversal in an amplifier is often treated as an electrical *phase shift* of 180°. In Fig. 310, one and one-half cycles are shown in each case.

Summary

In Table 3-1, the relative characteristics of the three basic amplifier configurations are itemized for comparison. Typical values are not shown, for these will vary considerably with different types of transistors, with operating bias currents and with the nature and magnitude of the output loads used. For example, a typical low-power transistor in the common-emitter configuration might have an input impedance of 1,000 ohms, an output impedance of 20,000 ohms and a power gain of 35 db. On the other hand, a power transistor in a similar circuit configuration, and with similar applied voltages, could provide a similar power gain, but might have an input impedance of only 10 ohms and an output impedance of 40 ohms. A medium power transistor could have values between the two.

CHAPTER 4

How to talk transistor language or, Don't let the big words scare you

Radio receivers and transmitters, record players, TV sets, test instruments, radar and sonar gear, industrial controls . . . in fact, almost any type of electronic equipment you can name . . . are all *signal handling* devices. This is true whether the equipment uses magnetic amplifiers, vacuum tubes, or such solid-state semiconductor devices as diodes and transistors. Depending on its intended function, each individual piece of equipment may handle its signal (s) in a variety of ways; it may amplify, rectify, measure, interpret, convert, detect, translate, modulate, or attenuate its signal (s), using the signal to operate a loudspeaker, light a cathode ray tube, deflect a meter pointer, energize a lamp bulb, or actuate a relay or solenoid.

"Fine," you ask, "but *what's* a signal?"

As the clever politician would say—"That's a *good* question!"

And the politician, after congratulating you on your astuteness in asking such a question, would then ramble on with several thousand words having absolutely no relationship to the question . . . or, in many cases, very little relationship to anything in particular. We'll try to do better than the politician.

Let's start with a simple example. Take a length of two-conductor cable—ordinary lamp-cord will do. At one end station a friend with a dc voltmeter. Place yourself at the other end with several flashlight batteries and a few pieces of hook-up wire.

Your friend connects the voltmeter to his end of the cable and measures zero voltage, which is not too surprising. While he's making this measurement, you connect one of your flashlight batteries to your end of the cable. Immediately he measures a dc voltage. Whether an up-scale or down-scale (needle moves to *left* of zero on meter scale) reading is obtained depends on how he connects his meter leads and how you connect your battery. If he obtains a down-scale reading, he may simply reverse his meter leads to obtain the proper meter polarity.

Your act of connecting the battery to the cable has transmitted an electrical signal to your friend. This signal is called a *step function*, for the voltage on the line has "stepped" from one fixed level (0) to another (battery voltage). If you connect another battery in series with the one you have, doubling the voltage applied to the line, you'll develop another step function . . . and will send another signal to your friend. Here, the line voltage has shifted from one value (battery voltage) to another (twice the battery voltage).

Suppose, instead, you reversed your battery connection. The voltage checked at the other end of the line would drop to zero, then jump back to its original value, *but with opposite polarity.* Your friend would have to reverse his meter connections to obtain an up-scale reading. Let's go a little further . . . instead of reversing the battery once, you continue to do so at the rate of, two reversals per second. And, to simplify our discussion, we'll call one of the conductors the *ground* lead. Your friend, then, will measure a positive voltage with respect to ground for a half-second, then a negative voltage with respect to ground for another half-second . . . for as long as you continue to reverse the battery.

If the line voltage were plotted in form of a graph, using *time* as a base line, the resulting curve would consist of a series of short dashes above the 0-voltage level to represent the periods during which a positive voltage was measured, with a corresponding series of alternating dashes below the zero voltage line representing the periods when a negative voltage is present. The alternating dashes would be connected with short vertical lines at each end to show the change from a positive voltage, through zero to a negative voltage. The resulting pattern would resemble the top wall of an ancient castle and would represent the *waveform* of our signal.

An electrical signal, therefore, consists of a *change* in voltage or current through a circuit. It may have a definite polarity . . . essentially a pulsating dc voltage . . . or may reverse polarity, as in our last experiment. If the polarity reverses, it is, of course, an ac signal. By definition, a signal is any physical phenomena that can be converted into electrical voltages or currents, such as sound waves through air or water, electromagnetic vibrations, variations in light levels or even the slow movement of a piece of industrial machinery.

Any signal has certain easily-defined characteristics. It has *amplitude* (the value of voltage), it has a *waveform,* and, if ac, it may have a specified *frequency* if it is not a random signal—like speech. In our example, the last signal developed had a frequency of 1 cycle per second (cps), for it went through its full cycle of positive and negative variations during this period. If the battery had been reversed twice as fast, developing quarter-second positive and negative impulses, the frequency would have been two cps.

A signal's waveform may be determined by plotting instantaneous-voltage or current values on a graph, just as in our example. This procedure is not practical, however, unless we deal with signals of very low frequency. More generally, the waveform is observed on the screen of an oscilloscope, an instrument which electrically traces the waveform pattern. Typical-signal waveforms encountered in transistor circuits are illustrated in Fig. 401.

A *sine-wave* pattern of an ac signal, with its voltage varying from a peak-positive value to a peak-negative value is shown in Fig. 401-a. The variation is not sudden, however, but follows a smooth, gradual curve. The waveform of the alternating current delivered by a power line is a sine-wave, generally with a frequency of 60 cps (50 cps and 25 cps ac is furnished in some areas).

If a sine-wave signal is passed through a rectifier, such as a semiconductor diode, "half" of its cycles are "stripped" away. As you'll recall, a semiconductor-diode junction offers high resistance to the flow of current in one direction. The resulting waveform, a *half-wave rectified sine-wave,* is illustrated in Fig. 401-b.

Often, two rectifier diodes are used, each working on alternate half-cycles, and having their outputs combined in such a way as to reinsert the missing half-cycle, thereby giving a signal waveform of single polarity (all positive half-cycles, for example), as seen in Fig. 401-c.

A *modulated sine-wave* signal is illustrated in Fig. 401-d. This is the type of signal developed by a common radiotelephone (am) transmitter. In effect, it is a high-frequency sine-wave whose instantaneous amplitude is varied in accordance with the waveform of a low-frequency signal. The low-frequency signal is the high-frequency signal's *modulation envelope*. If the instantaneous frequency of the high-frequency signal is varied, instead of its amplitude, the signal is frequency-modulated (fm).

The waveform seen in Fig. 401-e results when a sine-wave signal of relatively high amplitude gradually loses strength and finally drops to zero. It is called a *damped wave-train*. Two successive wave-trains are shown.

The *square-wave* signal in Fig. 401-f is the pattern which would result in our last battery experiment. The signal has a fixed positive value for a short time interval, drops through zero, to a fixed negative voltage, and so on. 1½ cycles (3 half-cycles) are shown.

Let us return to our battery experiment. If the battery had been connected to the line for a very short interval, disconnected for a

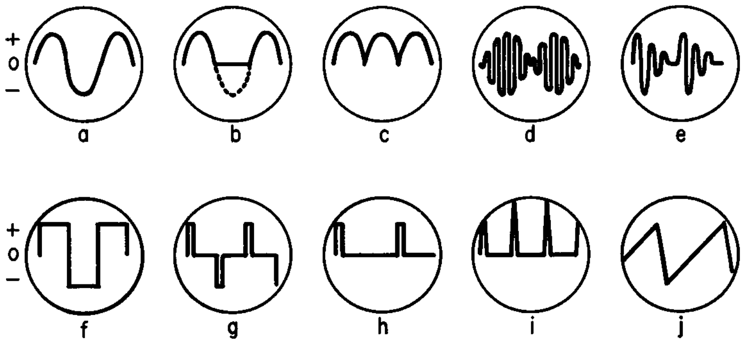


Fig. 401. *Basic signal waveforms: (a) sine wave; (b) half-wave rectified signal; (c) full-wave rectified signal; (d) modulated high-frequency signal; (e) damped wave train; (f) square wave; (g) rectangular bidirectional pulses; (h) unidirectional pulses; (i) sharp pulses or spikes; (j) saw-tooth signal.*

fairly long interval, reconnected with reversed polarity for a short interval, disconnected . . . and so on . . . the result would be a series of positive and negative rectangular pulses, as shown in Fig. 401-g. These are called *bidirectional* pulses, since they have alternate-polarity values. Had the battery not been reversed, *unidirectional* pulses, as shown in Fig 401-h, would have been developed. Extremely sharp and narrow pulses of high amplitude are generally called *spikes*—see Fig 401-i.

Pulses, spikes and damped wave-trains (oscillations), since they may occur erratically in some types of circuits, are called *transients*.

A smooth, linear (straight-line) variation in voltage or current from one level to another, followed by an almost instant return to its original value, with the entire cycle repeated, develops a waveform which bears a close resemblance to a familiar household tool. Fig. 401-j. Such a pattern is called a *saw-tooth* waveform.

Learning to recognize and to identify various signal patterns is an interesting past-time, and can be invaluable both in the study of practical circuits and in the troubleshooting and repair of commercial equipment or experimental projects. A variation in the waveform of a signal from its expected pattern often gives an important clue as to circuit behavior.

AMPLIFIERS

As we have seen, the transistor is essentially a current-amplifying device which may be used in any of three basic configurations, depending on which electrode is chosen as common to both the input and output circuits. In the common-emitter and common-collector

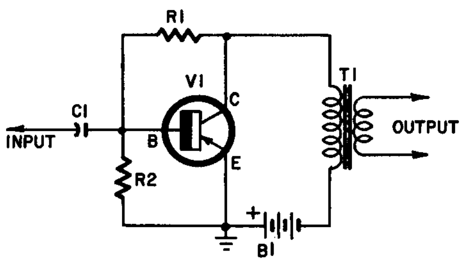


Fig. 402. The single-ended amplifier used for audio amplification is always operated "Class A". Distortion occurs when the bias is of a value that permits any portion of the signal to move the instantaneous bias point to a nonlinear portion of the operating curve. This distortion can be put to a practical use in special circuits not intended for the reproduction of sound.

configurations, for example, a relatively small current applied to the input terminals can control a much larger current in the output circuit. Although no actual current gain is achieved in the common-base amplifier (except with certain point-contact transistors), a gain in signal amplitude (voltage or power) is possible due to the large difference between input and output impedances.

While the three configurations are basic to all transistor applications, there are two ways in which *any* of the three may be used . . . *single-ended* and *push-pull.* These two arrangements are fundamental to *all* types of amplifying devices, whether vacuum or gas-filled tubes, transistors or other solid-state units. Typical single-ended and push-pull circuits, using p-n-p transistors in the common-emitter configuration, are illustrated in Figs. 402 and 403, respectively.

The single-ended stage, as we can see by referring to Fig. 402 is one of the three standard configurations in its basic form. That is, the input signal is applied between one electrode (base) and a common one (emitter), with the amplified output signal obtained between the remaining (collector) and common electrodes. In the circuit shown, the transformer's (T1) *primary winding* serves as V1's collector load, with base bias supplied by a voltage divider made up of R1 and R2. C1 serves as an input dc-blocking capacitor, preventing a change in base bias when a signal source is connected between the stage's input terminal and circuit ground. Operating power is supplied by a single battery, B1. In other applications, V1's load might be a tuned circuit, resistor, chcke coil, relay or some other component; and one of several resistor-connections could be used to supply base-bias current.

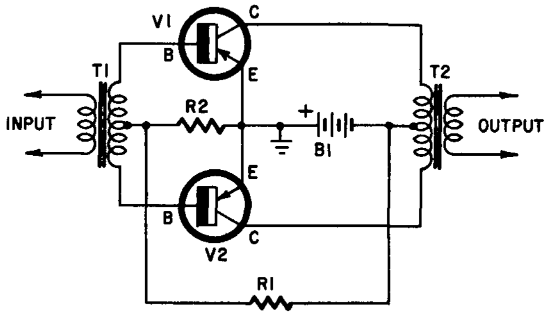


Fig. 403. The push-pull amplifier can be biased "Class A" for the least distortion or "Class B" for the greatest power output. Compromise amplifiers (like AB) are used for more power and have more distortion than those in Class-A operation. Class-C bias is used only with L-C tuned circuits.

Employing two transistors in a single stage, the push-pull amplifier can deliver considerably more output power and can handle signals of much greater amplitude than a single-ended stage using the same type transistor. In a broad sense, however, the push-pull circuit actually consists of two single-ended stages connected "back-to-back."

To demonstrate this concept, refer to Fig. 403. Use a small piece of paper or your hand to cover the top "half" of the circuit. Study the connections to V2. Then move the paper to cover the lower part of the circuit, studying the connections to V1. In each case you'll find that the exposed circuit is basically a single-ended common-emitter stage with a transformer (T1) input and a transformer (T2) collector-load. In both cases, operating power is supplied by B1, with base bias provided by resistive-voltage divider R1-R2.

In operation, two identical, but inverted, input signals are required to drive the stage. These are obtained from center-tapped-input transformer T1. During ½-cycle, a positive-going signal is applied to one transistor and an identical negative-going signal to the other. Signal polarities are reversed during the succeeding half-cycle, reversed again during the next half-cycle, and so on.

Let's assume that the base bias applied by voltage-divider R1-R2 permits each transistor (V1 and V2) to draw a small amount of collector current. These equal currents, flowing in *opposite* directions through T2's center-tapped primary winding, develop magnetic fields in T2's core which cancel each other. Thus, there are no flux linkages to couple with T2's secondary, nor is there any tendency for T2's core to reach magnetic saturation.

When a positive-going signal is applied, for example to V1, its collector current is reduced. At the same time, the negative-going signal applied to the base-emitter circuit of V2 increases its collector current. The currents through T2's primary winding, although still flowing in opposite directions, are no longer equal and hence do not produce magnetic fields which cancel. Instead, magnetic-flux linkages are developed which are proportional to the *difference* between V1's and V2's collector currents. These flux linkages, coupling with T2's secondary, develop an amplified-output signal. On the next half-cycle, a negative-going signal is applied to V1's base-emitter circuit and a positive-going signal to V2. A similar situation occurs, but with the V1's current increasing and V2's collector current decreasing.

Therefore, one transistor serves to "push" the amplified-output signal while the other serves to "pull," hence the circuit's name.

While we have used a common-emitter configuration in our example, push-pull operation need not be confined to just this one type of circuit. Push-pull common-collector or common-base arrangements are just as practical. The essential ingredients for push-pull circuitry are: (a) an input signal or signals, which affect two transistors (or other amplifying devices) in an equal but opposite fashion, and (b) an output device capable of combining the two amplified signals delivered by each "half" of the stage. For best operation, the two transistors should be *balanced* . . . that is, should have identical ac and dc characteristics and should provide equal gain.

Distortion

When a transistor is operated as a *linear* amplifier, its output current is *directly proportional* to the variations in the input signal. Thus, the output-signal waveform is a faithful reproduction of the input signal. If a transistor is operated as a *non-linear* amplifier, the output-signal waveform becomes *distorted.* This condition occurs when an amplifier stage is operated improperly and may be caused by a defective transistor, by a change in operating voltages (or currents) due to defective circuit components, by improper bias currents or by applying an input signal greater in amplitude than the stage is capable of handling. In the last case the amplifier is said to be *overdriven.*

Referring back to Fig. 402, let us assume that a sine-wave signal is applied to the stage's base-emitter input circuit—Fig. 401-a. Let us suppose, furthermore, an incorrect value has been chosen for R1, so that excessive base bias is applied to the transistor. Collector current, under these conditions, is so high that the available battery voltage (B1) is dropped across T1's primary winding. On positive half-cycles, V1's collector current can be reduced and will develop an amplified-signal waveform which duplicates the input signal. On negative half-cycles, however, V1's collector current cannot increase, since this current is already at a maximum. The stage is essentially *saturated.* Since collector current cannot increase to follow the variations of negative-going half-cycles of the input signal, this "half" of the input signal is not amplified. It is "stripped" away, instead, and the output signal waveform would appear essentially like the pattern shown in Fig. 401-b, although inverted (remember that phase inversion takes place in a common-emitter stage). The negative-going half cycles of the applied signal have been *clipped,* and severe signal distortion results.

If the base bias had been of some smaller value— (not enough to saturate the stage, but still enough to permit saturation on the *peaks* of the negative-going half-cycles) then only the peaks would be clipped. The stage would still distort the signal, but the distortion would not be as severe as before.

Let us now, consider the opposite situation. Let's assume that R1 is open due to either a defective resistor or broken (or unsoldered) lead. The stage is operated without base bias and is at *cut-off* or there is no collector current flow.

When the sine-wave input signal is applied, the negative half-cycles will cause appropriate increases in collector current, and will develop an amplified reproduction of themselves in the output circuit. The positive-going half-cycles will not affect collector current, however, for the current cannot be reduced below zero. As before, the output signal waveform will appear somewhat like the pattern shown in Fig. 401-b; the positive-going half-cycles of the applied signal have been clipped.

If a little, but not adequate, base bias had been applied to the stage, only the extreme peaks of the positive-going half-cycles would be clipped. Again, distortion would occur, but would be less severe than if the stage were operated without bias.

Had adequate bias been applied to the stage, and the input signal been of such amplitude as to drive the transistor to saturation on its negative peaks and to cut-off on its positive peaks, then *both peaks* of the output signal would be clipped. In extreme cases, the output-signal waveform might take on the appearance of a square-wave, as shown in Fig. 401-f.

By using mathematical-analysis techniques, it can be shown that any repetitive-signal waveform (including all of those shown in Fig. 401), regardless of its complexity, is made up of a fundamental *sine-wave* signal plus *even-* and *odd-harmonic* signals of varying amplitudes. The mathematical expression derived by such an analysis is known as the *Fourier Series.* When this series is used to analyze the distorted-signal waveform obtained from a non-linear amplifier, it is found that the output signal contains the original (fundamental) signal *plus* higher order harmonics. For example, if a 120 cps sine-wave signal is passed through an improperly operating stage, the distorted output signal still has a basic frequency of 120 cps. However, it also contains *additional*-signal components at 240 cps (2nd harmonic), 360 cps (3rd harmonic), and so on.

Distortion which results in the addition of harmonics to the original signal is termed *harmonic distortion* and, can be determined by filtering out the original (fundamental) signal and comparing its amplitude to those of the harmonics. This result is generally expressed as a percentage figure.

In addition to introducing harmonic distortion, a non-linear amplifier can cause another type of difficulty. When several *different* signals are amplified simultaneously by an ideal (linear) amplifier. the composite-output signal will contain amplified versions of *the* original signal is termed *harmonic distortion* and can be detive separation that existed in the input. However, when several different signals are handled by a non-linear stage, they tend to modulate one another, so that the composite-output signal contains not only the original signals but also their *sum*-and-*difference* frequency signals.

As a practical example, suppose that 70 and 7,000 cps signals were passed through a linear amplifier. The output signal would include amplified versions of these two signals and nothing more. However, if these two signals were handled by a non-linear stage, the output would contain both original signals (70 and 7,000 cps), *plus two new signals.* One would represent their difference-frequency (7,000 − 70, or 6,930 cps), and the other their sum-frequency (7,000 + 70, or 7,070 cps), *in addition to any harmonic signals resulting from harmonic distortion.* The latter could include harmonics of both original signals as well as harmonics of the sum- and difference-frequency signals. Distortion which results when two or more signals cross-modulate each other is called *intermodulation distortion,* and like harmonic distortion is expressed as a percentage figure.

Summing up, non-linear operation may occur if a transistor is operated with incorrect bias, if it is overdriven or if the transistor is defective. If a p-n-p stage is operated with too little bias, the positive peaks of the input signal may be clipped, and the negative peaks clipped only if the stage is operated with excessive bias. In an n-p-n stage, the opposite situation holds. In both, however, both peaks of an amplified signal will be flattened (or clipped) if the stage is overdriven. Non-linear operation may introduce either harmonic or intermodulation distortion, or both.

Amplifier classes

The operating parameters of a given transistor with a specific load in any of the three basic configurations depend primarily on its bias currents (and voltages). Of these, the input-bias current is probably the most important single factor in determining such values as stage gain, collector current, power output and mode of operation. Transistor-amplifier stages may be operated in any of three basic *Classes,* plus one sub-class, depending on the relationship between stage-bias and input-signal amplitude. These stages are identified as *Class A, Class B, Class C* and the sub-class *Class AB.* The designations are equivalent to a similar classification system used for identifying various types of vacuum-tube-amplifier stages.

In the common-emitter and common-collector configurations, the input bias is the stage's base current. In the common-base circuit, the input bias is emitter current. When an external ac-signal is applied to a stage, the instantaneous bias current varies up and down in accordance with the variations in the input signal. These variations in turn, cause corresponding changes in the instantaneous value of the stage's output current, developing an amplified signal across the transistor's output load.

Taking the common-emitter circuit as an example, one will note the steady-state or "zero-signal"-collector current approximates the transistor's *beta* times the fixed base-bias current. If a p-n-p transistor is used, the collector current increases when the input signal swings in a negative direction, reaching its maximum value when the input signal approaches a negative peak. When the input signal swings in a positive direction, the collector current decreases, reaching its minimum value when the input signal is at its positive peak. A similar, but opposite, action occurs when an n-p-n transistor is used.

The operation of the three basic amplifier Classes is illustrated in Fig. 404. In each case, a single-ended circuit is shown, using a p-n-p transistor in the common-emitter configuration. N-p-n transistors may be used in similar circuits by reversing all dc (battery) polarities.

In a Class-A amplifier, a moderate base-bias current is employed, setting the zero-signal-collector current at a relatively high value. With an input signal applied, the collector current varies to either side (above or below) of its steady-state value, *but never reaches zero*, even on the positive peaks of the applied signal. Thus, the full cycle of the input signal controls the instantaneous amplitude of collector current, and the collector-current and output-signal waveforms accurately duplicate those of the input signal. With a typical signal, such as a sine-wave, the collector-current excursions above and below its steady-state value are equal, and thus cancel when averaged. As a result, the average collector current of a Class-A amplifier remains the same whether or not a signal is being amplified.

Class-B operation is obtained when our typical common-emitter stage is worked with *zero-base-bias current.* With this mode of operation, the steady-state collector current would be zero if an ideal transistor were used. Where a p-n-p-type is employed, as in our example, the collector current remains cut off during the entire positive half-cycles of the applied signal, flowing only during the negative half-cycles and reaching a maximum when the input signal reaches a negative peak. Thus, only alternate half-cycles of the applied signal are effective in controlling the amplifier's output, and the output-signal waveform becomes a series of half-cycle pulses, resembling the input signal as it would appear if passed through, for example, a half-wave rectifier. In effect, alternate-half-cycles of the applied signal are "clipped off," with the remaining half-cycles greatly amplified. The average value of col-

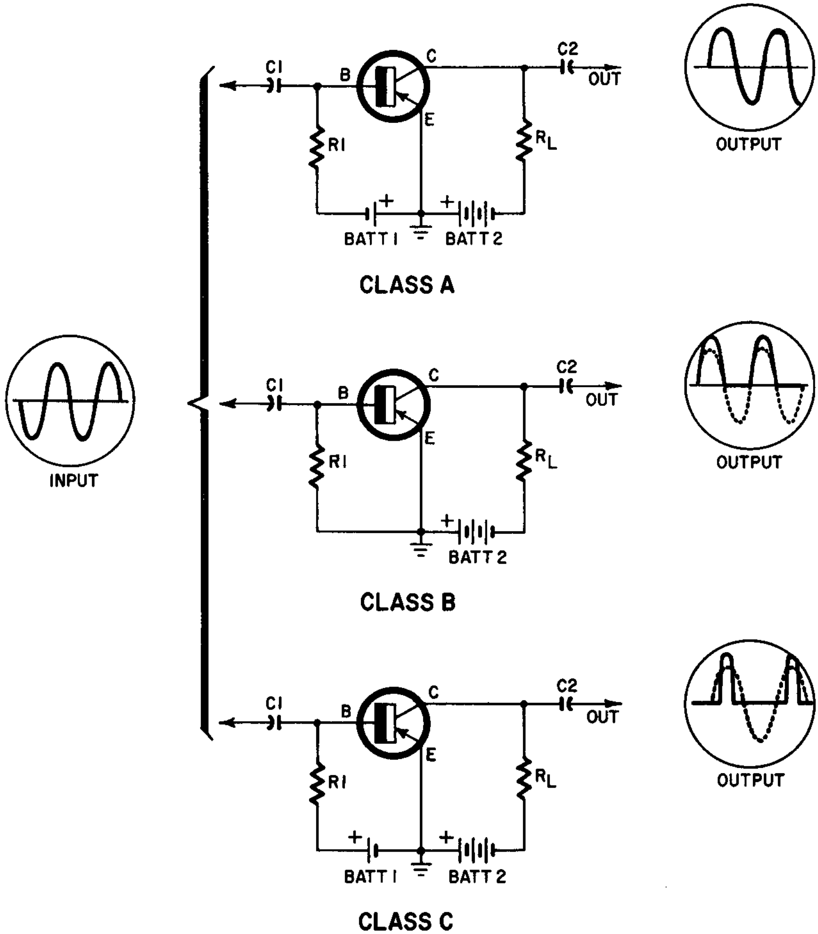


Fig. 404. The three basic arrangements of amplifier operation—Class A, Class B and Class C. In addition there is a fourth, intermediate between A and B—called, appropriately, Class AB.

lector current rises from zero to a large value when an input signal is applied to a Class-B stage, with its actual value directly proportional to the amplitude of the applied signal. The greater the input signal, the greater the increase in average collector current.

In a Class C amplifier, a *reverse*-base bias is employed, and collector current remains not only under zero-signal conditions, but doesn't flow unless the alternate peaks of the applied signal are sufficient to overcome this initial bias. Once the bias is overcome, collector current can flow, and will increase and decrease in amplitude to follow the variations in signal peaks. Thus, only the alternate *peaks* of the applied signal have control over collector-current amplitude, and the output-current waveform becomes a series of narrow pulses. As with a Class-B stage, the average value of collector current rises from zero to a large value when an input signal is applied, with its actual value depending on the amplitude of the applied signal.

The Class-AB amplifier is intermediate between Class A and Class B. A small input (base) bias is used. There is some collector current flow under zero-signal conditions, but not as much as with Class-A operation. When an input signal is applied, the collector current is reduced to zero for a large portion of positive half-cycles, but there is a net-average increase in collector-current flow due to the high current peaks on negative half-cycles. In practice, true Class-B operation is employed infrequently in commercial-transistor circuits, and textbook and magazine references to "Class-B amplifiers" generally refer to Class-AB stages.

Where low-level signals are involved, it is common practice to employ Class-A amplifiers to provide signal gain, and Class-AB, -B, or -C stages as modulators, mixers or detectors. On the other hand, where large amplitude signals are handled, as in power-output stages, Class-AB, -B, and -C stages are more popular. Class-AB and -B amplifiers are used in audio work, while Class-C circuits are encountered most often in radio-frequency equipment, in special types of control equipment and, occasionally, in test instruments. For a given transistor, maximum-stage gain is generally obtained when the unit is operated as a Class-A or Class-AB amplifier, while maximum power output is obtained with Class-B or -C operation.

The *efficiency* of an amplifier stage is proportional to the *ratio* of the *output-signal* (ac) power obtained to the *input (dc)* power required for operation. Since collector current flows continuously in a Class-A stage, even when no signal is being amplified, it requires dc power at all times and, therefore, is the *least* efficient of the three amplifier classes. Circuit-efficiency increases as we go from Class-A to Class-C operation, hence the greater popularity of Class-AB, -B, and -C stages in circuits handling substantial amounts of power. Percentagewise, a perfect Class-A amplifier may approach 50% efficiency, while an ideal Class-C stage can have an efficiency from 95% to 99%.

Unfortunately, single-ended stages operated as Class-AB, -B, or -C amplifiers introduce considerable distortion. Compare the input (sine-wave) and output-signal waveforms given in Fig. 404. As a result you will find that practical audio circuits use Class-AB or -B operation only in a push-pull arrangement. Here, one "half" of the stage amplifies the input signal's positive-going half-cycles, the other "half" the negative half-cycles. The two amplified signals are recombined in the output load to form a perfect (distortion-free) output signal. In rf circuits or special amplifiers employing a tuned load, any amplifier class may be used, for the resonant action ("flywheel" effect) of the tuned circuit tends to restore missing parts of the amplified signal.

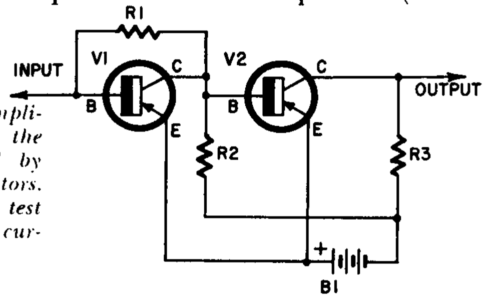


Fig. 405. Direct-coupled (DC) amplifier stages are needed to amplify the very low frequencies not passed by transformers or coupling capacitors. This circuit form is often used in test equipment—especially for meter current amplification.

Interstage coupling

Individual transistor stages are seldom used alone except in the simplest types of electronic equipment. In a Public-Address amplifier, for example, two or three stages may be connected in series, or "cascaded," to provide signal gain. In turn, these are coupled to the power output stage which drives the loudspeaker(s) or other output devices. Several methods may be used for coupling the output signal of one transistor stage to the "input" of another, depending on the intended application of the circuit, types of transistors used and similar factors. The four most popular interstage-coupling techniques are given in Figs. 405, 406, 407 and 408. Circuit polarities are for p-n-p transistors in the common-emitter configuration.

Perhaps the simplest way to couple one transistor stage to another is to provide a *direct*-electrical connection from the "output" electrode of one to the "input" electrode of the succeeding stage. This type of *direct-coupling* is shown in Fig. 405. In operation, V1's base-bias current is supplied through R1, acting in conjunction with the collector load R2. Resistor R2, then, serves both as V1's collector load and as part of the base-bias network for V2, forming a voltage-divider with V2's base-emitter circuit. Finally, R3 serves as V2's output load.

Capable of handling dc as well as ac signals, the direct-coupled circuit is simple, economical of parts and useful in many types of special-purpose equipment. It suffers from one serious disadvantage, however. Since the collector current for the first stage has a

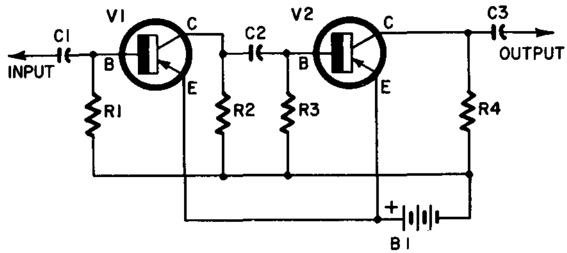


Fig. 406. Resistance-capacitance (R-C) coupling is used where space and economy are a consideration. It may be necessary to use a 3- or 4-transistor amplifier to get the same gain possible when using a transformer as a coupling element.

great influence on the base current of the second stage, all load and bias resistors have rather critical values. Often, the actual values of resistors used must be determined experimentally to match the *specific* transistors used in the circuit. Changing to other transistors, even of the same manufacturer's type, may cause undesirable changes in circuit operation and performance. In addition, this dependence of one stage on the next for its dc operating parameters makes such circuits extremely sensitive to temperature variations unless special compensation techniques are used.

If we are interested in amplifying only ac signals, we can avoid the difficulties encountered in direct-coupling by isolating one transistor stage from another *as far as dc is concerned.* This situation can be accomplished easily by providing each transistor with its own base-bias and output-load resistors as well as dc-blocking (or ac coupling) capacitors between stages.

A typical two-stage *resistance-capacitance* (R-C) coupled amplifier is illustrated in Fig. 406. Here, capacitor C1 functions as the input and C3 as the output-coupling capacitors. V1's base-bias is provided through R1, while R3 serves a similar function for V2. First and second-stage-collector loads are R2 and R4, respectively. Finally, C2 couples the amplified signal appearing across R2 to V2's base electrode while, at the same time, blocking the dc voltage present on V1's collector and preventing its application to the second stage. Thus, only the ac component of the amplified signal is transferred from one stage to the next.

While both direct- and R-C-coupled circuits can give satisfactory performance in a variety of applications, neither permits the full potential gain of the transistors to be utilized. As you'll recall, the common-emitter circuit has a low to moderate input impedance and a high output impedance (Chapter 3). In order to achieve a maximum transfer of signal power from a source to a load, the source impedance must *match* that of the load. Where several stages are connected in cascade, one serves as a "source" for the next, with the stage following becoming the "load." Since there is no provision for matching the high output impedance of

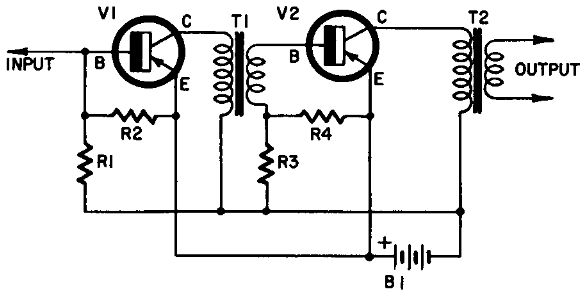


Fig. 407. Transformer-coupling gives greater gain. Higher cost of the transformer is offset by eliminating a transistor, coupling capacitor and bias resistors.

one stage to the much lower impedance of the next, we will obtain a loss of efficiency in signal transfer, thereby reducing overall gain considerably. In addition, fairly large coupling capacitors must be used between stages in R-C circuits to prevent a serious loss of low-frequency signals. This condition occurs due to voltage division between the capacitor's increasing impedance (at lower frequencies) and the effective-input impedance of the following stage.

For maximum efficiency, then, we must use a coupling method which serves to match interstage impedances as well as to transfer the amplified signal.

There is a readily available component suitable for matching circuit impedances . . . the transformer. With *transformer-coupling* used between stages, a nearly perfect impedance match may be achieved. Such an arrangement is illustrated in Fig. 407. Here, the turns ratio between transformer (T1 and T2) primary and secondary windings is chosen to match the high-output impedance of V1 to V2's moderate input impedance, and V2's output impedance, in turn, to the input impedance of a third stage or other load. Except for the type of collector load used, the circuit's basic operation is much like that of other common-emitter circuits. V1's base-bias current is furnished by voltage divider R1-R2, V2's by R3-R4. The transformers, of course, automatically provide dc isolation between stages.

When matching a high impedance to a lower one, as in our example, a *step-down* turns ratio is used between primary and secondary windings. In matching a common-emitter stage to an extremely high-impedance load, such as a pair of crystal headphones, a *step-up* turns ratio would be used.

Transformer-coupled amplifiers are used at both audio and radio frequencies. At audio frequencies (20 to 20,000 cps), small iron-core units are employed. At rf and if values (200 kc to 100 mc, or more), air-core or powdered-iron (ferrite) core transformers may be used. With low-power transistor types, typical primary impedances range from 2,500 to 20,000 ohms, while secondary impedances are from 200 to 2,000 ohms, depending on intended application.

A modified version of the transformer-coupled amplifier is shown in Fig. 408. Here, the two-winding transformers have been replaced by tapped inductive *impedances* . . . coils L1 and L2. Acting as autotransformers, the coils serve both as collector loads and as impedance-matching devices. In practice, the ratio of the coil's overall impedance to the impedance between its output tap and "ground" (lower) terminals approximates the ratio between the primary and secondary impedances of a conventional interstage transformer. Since there is no provision for dc isolation with impedance-coupling, blocking capacitors C1 and C2 are added to the circuit. As before, voltage dividers R1-R2 and R3-R4 provide base bias for the first and second stages, respectively.

In commercial electronic equipment, direct-coupled circuits are found in medical, industrial and military equipment. R-C circuits are used at low signal levels in many types of subminiature equipment (such as Hearing Aids). Transformer and impedance-coupled circuits are found in audio amplifiers, radio receivers, transmitters, instruments and in similar types of equipment, as well as in circuits where high power levels are handled. The impedance *ratios* found in circuits using multi-watt power transistors are comparable to those used in low power circuits, but the *actual impedance values may be much lower.*

Complementary circuits

In a vacuum tube, electron flow is from the cathode to plate and from the plate through the external circuit (plate load). If a negative-going signal is applied to its grid, plate current decreases, and similarly a positive-going signal increases plate current. In transistors, on the other hand, the *direction* of electron flow as well as the relative effects of positive and negative-going signals depend on the *type* of transistor.

The operation of an n-p-n transistor in the common-emitter configuration, for example, is analogous to that of a vacuum tube.

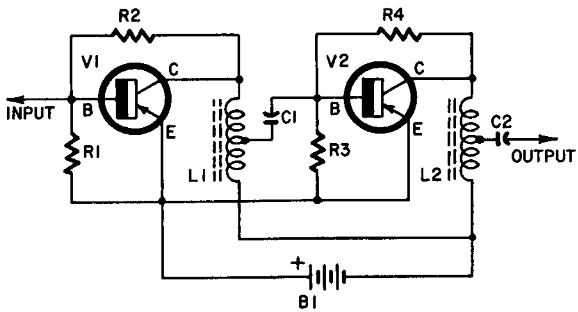


Fig. 408. Impedance coupling uses an inductor (choke) in place of R2 in Fig. 406. Here the choke is tapped to provide a high impedance to the collector of V1 and a lower value to the base of V2. The tapped choke also has an autotransformer effect and could be considered a form of transformer coupling.

As far as the external circuit is concerned, electron flow is from emitter to collector, and from collector through the external load. As in a vacuum tube, a negative-going signal applied to the base electrode will cause a decrease in collector current, while a positive-going signal will cause an increase in collector current.

When the operation of a p-n-p transistor is considered, however, the situation is reversed. That is, electron flow, as far as the external circuit is concerned, is from collector to emitter. A negative-going input signal causes an increase and a positive-going signal a decrease in collector current. Thus, the operation of a p-n-p transistor, for practical purposes, is just the opposite . . . or the *complement* . . . of that of n-p-n types. This complementary correspondence between two basic types of transistors is unique and has no analogy in vacuum-tube circuitry. It may be used as a foundation for a whole class of practical transistor circuits for which there is no parallel in vacuum-tube work. Typical complementary circuits are illustrated in Figs. 409 and 410.

Employing common-emitter configurations, the two-stage cascaded amplifier shown in Fig. 409 utilizes the complementary properties of n-p-n and p-n-p types to achieve direct-coupling between stages. Here, C1 serves as an input-coupling capacitor and

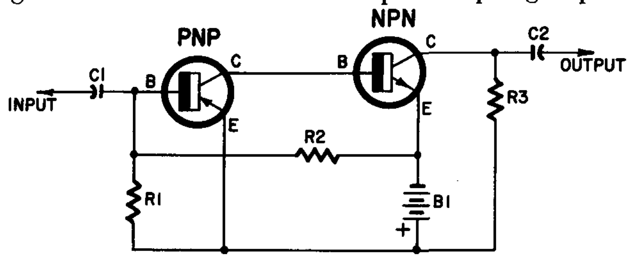


Fig. 409. Complementary circuits have no direct vacuum-tube circuit equivalent. (This would require a tube which emitted electrons from its plate and attracted them with its cathode, using a positive value of bias for control.) Complementary transistor circuits are an easy method of direct coupling.

first stage bias is furnished by voltage divider R1-R2. The base-emitter circuit of the second stage serves as the first stage's collector load, with the p-n-p transistor's collector-current becoming the base-bias current for the n-p-n unit. Finally, R3 is the second stage's collector load and C2 the output coupling capacitor.

In operation, a negative-going signal applied through C1 will cause an increase in the p-n-p's base current. This, in turn, brings about a corresponding, but much greater, increase in the p-n-p's collector current and in the base bias applied to the n-p-n stage. An increase in the second stage's base current results in an amplified increase in its collector current, developing a negative-going output signal across its load, R3. In each case, the ratio of collector current increase to input base-current increase is proportional to stage gain.

Similarly, a positive-going signal applied to the input, causes a decrease in p-n-p base current, a decrease in its collector current and hence in the n-p-n's base current, and a decrease in collector current through R3, thus developing a positive-going output signal.

This amplifier, then, uses a minimum of components, has high gain, excellent low frequency response (since there is no interstage-coupling capacitor), and provides an output signal that is *in phase* with its input signal.

Another interesting circuit which utilizes the complementary properties of p-n-p and n-p-n types is shown in Fig. 410. Sometimes called a "single-ended push-pull" amplifier, this circuit requires only a single input signal, does *not* need a center-tapped output load (as in a conventional push-pull stage), will provide true push-pull amplification and is capable of comparatively high-power output and high efficiencies.

Referring to the diagram, p-n-p-base bias is furnished by R1 in conjunction with current-limiting resistor R5 and emitter resistor R3. Similarly, n-p-n-base bias is supplied by R2 in conjunction with R4 and R6. The two emitter resistors tend to stabilize and balance circuit operation. The output *load*, which may be a resistor, transformery primary, or loudspeaker-voice coil, is con-

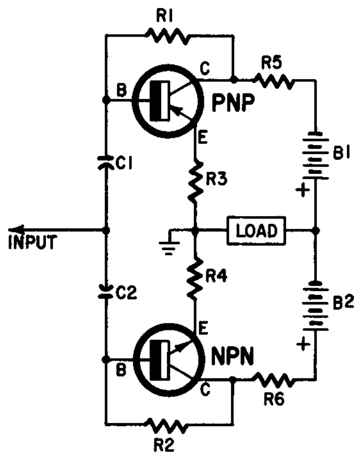


Fig. 410. This complementary symmetry circuit is a push-pull arrangement. It needs no transformers. This circuit is seldom used for audio since it is not usually economically practical to get a low-enough output impedance or high-enough load (speaker) impedance. This circuit does have uses in other devices.

nected between the junction of the emitter resistors (R3 and R4) and the split-power supply, B1 and B2. The input signal is applied through dc-blocking capacitors C1 and C2 to both transistors simultaneously.

Under zero-signal conditions, the collector currents of the n-p-n and p-n-p transistors are equal and thus cancel. Electron flow is from the minus terminal of B1, through R5, from collector (C) to emitter (E) of the p-n-p transistor, through R3, through R4, from E to C of the n-p-n unit, through R6, and back to the positive terminal of B2. B2's negative terminal is connected directly to B1's positive terminal, completing the circuit. There is no dc through the load.

If a negative-going signal is applied through C1 and C2, the p-n-p's collector current increases and the n-p-n's current decreases. Since the two currents are no longer equal, they cannot cancel, and the net *difference* current flows through the load. Electron flow in the difference-current path is from B1's negative terminal, through R5, from C to E of the p-n-p transistor, through R3, through the *load*, and back to B1's positive terminal. Electron flow, in this case, is from left to right through the load.

Suppose, now, that a positive-going signal is applied to the stage. The p-n-p's collector current decreases and the n-p-n's current increases. Again, the two currents are no longer equal and cannot cancel, leaving a net difference current to flow through the load. In this case, electron flow is from B2's negative terminal, through the *load*, through R4, from E to C of the n-p-n unit, through R6, and back to B2's positive terminal. Electron flow is from right to left through the load.

In conclusion, there is no current through the load under zero-signal (steady-state) conditions. When the p-n-p transistor's current is greater than that of the n-p-n unit, the difference current flows in one direction through the load. Conversely, when the n-p-n's collector current is greater than that of the p-n-p unit, the difference current flows in the opposite direction through the load. Since the electron flow through the load is first in one direction and then in the other on alternate half-cycles of the applied signal, a true ac output signal is developed.

Obviously, proper operation of the complementary single-ended push-pull amplifier requires that the two transistors (p-n-p and n-p-n) have virtually identical ac and dc characteristics, with the exception of polarity. In other words, the two units must be perfectly symmetrical. Because of this, push-pull circuits employing the complementary principle are often termed *complementary-symmetry* arrangements.

OSCILLATORS

Used to develop an alternating-current (ac) signal when powered by a dc (battery, for example) source, oscillators are encountered in transistorized-radio transmitters, receivers, TV sets, signal generators, alarm circuits, power converters, measuring instruments and in other electronic equipment. The signals delivered by practical oscillators may have a variety of wave-forms, ranging from sine-wave to pulse and square-wave shapes, depending on individual applications. Operating frequencies may be from a few cycles per second to hundreds of megacycles.

An oscillator is an amplifier with its output circuit coupled back to its input in such a way that any signal reinforces itself. Thus, any change in circuit currents (or voltages) is reamplified time and again, generating an ac signal. The basic prerequisites for an oscillator are: (1) an amplifier capable of supplying sufficient gain to overcome circuit and feedback losses, (2) a circuit for coupling the amplifier's "output" signal back to its input in such a way that the input signal is *in phase* with the output signal, and (3) some means of controlling the periodic rate (frequency) at which the circuit operates. There are many ways to accomplish these three jobs and, hence, many different types of oscillators. In most cases, however, the differences between oscillators fall into two categories . . . (a) the exact method used to obtain a feedback signal, and (b) the means used to control frequency of operation.

Although the majority of transistor oscillators are simply transistorized versions of older vacuum-tube arrangements, there are several important differences between the transistor and tube circuits. Most vacuum-tube oscillators, for example, operate as *Class C* amplifiers and develop their own grid-bias voltage. Common transistor oscillators, on the other hand, are generally *Class A* or *Class AB* amplifiers operating with a fixed base-bias current. In addition, typical vacuum-tube circuits have high input and output impedances. Since there is usually a significant difference between the input and output impedances of a transistor amplifier, the feedback arrangement chosen must take these into account to prevent excessive losses.

Any of the three basic amplifier configurations may be used as an oscillator, provided suitable means are employed for obtaining the feedback signal, for overcoming circuit losses, and for controlling the frequency of oscillation. In the common-collector circuit, for example, the feedback circuit must convert a current gain into voltage gain and must match a low output to a high input impedance. Where the common-base circuit is used, impedance matching becomes a real problem, due to the wide differences between this circuit's input and output impedances. Still another problem is encountered where the common-emitter configuration is employed . . . that of converting the out-of-phase "output" signal to an in-phase feedback signal. As you will recall (Chapter 3), the input and output signals are in-phase in the common-base and common-collector arrangements, but out-of-phase in the common-emitter configuration. However, the common-emitter's high gain makes it most attractive for this type of application and it is, perhaps, the most popular of the three configurations in practical use.

The "tickler feedback" oscillator

Using a p-n-p transistor in the common-emitter configuration, the oscillator circuit shown in Fig. 411 employs a two-winding transformer, T1, to provide the feedback signal needed to start and sustain oscillation. Base bias is supplied by voltage divider R1-R2, by-passed by C1. The transformer's primary winding, L1, serves as a collector load and its secondary, L2, as a feedback coil. Since L2 is used to "tickle" the circuit into operation, it is appropriately called the *tickler-feedback* coil. In practice, the L1/L2-turns ratio is chosen to match the high collector impedance to the moderate base impedance, while the coil connections are chosen to invert the feedback signal, giving it the proper in-phase relationship. The circuit's operating frequency is determined by a resonant circuit made up of primary winding L1 and its distributed capacities, Cd (shown dotted). In some applications an external capacitor may be connected across the transformer to tune the oscillator to a specific frequency.

In operation, base-bias current can flow when power is first applied to the circuit, with the amount determined by the ratio of R1 and R2. This, in turn, permits a flow of collector current through L1, setting up magnetic-flux linkages which when coupling with L2, induce a feedback-signal voltage that is applied to the transistor's base-emitter circuit. If we consider that the by-pass capacitor C1 is virtually a short-circuit as far as ac is concerned, we can see that L2 is essentially between the base and emitter electrodes.

Thus, changes in the transistor's collector current develop a signal voltage in resonant circuit L1-Cd. This signal, in turn, is coupled to L2 and fed back to the transistor's "input" (base) circuit in such a way as to reinforce the signal developed across L1-Cd. This action continues as long as power is supplied to the circuit, developing an ac signal at L1-Cd's resonant frequency.

Tickler-feedback oscillators can be used at audio and radio frequencies, depending on the choice of transistors and on the values of L1, L2, and other components. For proper operation, however, it is essential that the signal furnished by L2 to the base-emitter circuit be *in phase* with the signal developed across the collector load (L1-Cd). In practical circuits, the proper connections to L2 can be determined experimentally—if the circuit fails to oscillate, for example, the connections to the feedback coil (L2) are reversed.

Other oscillators

As discussed earlier, practical oscillators may be developed using any of a variety of feedback techniques. For example, the transformer arrangement used in the tickler-feedback oscillator may be replaced with a tapped-inductance or coil, much in the same way that a tapped coil can be used in place of a transformer for interstage coupling. One version of such a circuit is often termed a *Hartley* oscillator after the inventor of a vacuum tube oscillator using a tapped coil for feedback. In a similar fashion, feedback may be provided by capacitor networks or combinations of capacitors, coils or resistors. Where precise frequency control is needed,

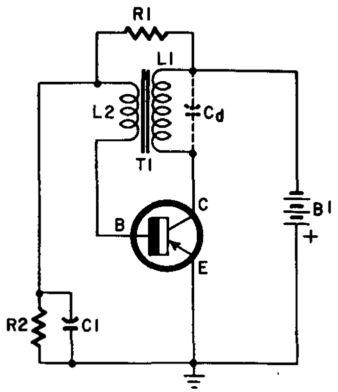


Fig. 411. A basic transistor oscillator circuit using a transformer (T1) for impedance matching and feedback.

a piezo-electric quartz crystal may be incorporated into the feedback circuit. Such a crystal is electrically equivalent to a high "Q"-tuned circuit (see discussion of coils in Chapter 2).

Almost any of the basic-oscillator circuits may be operated as a *blocking oscillator* if the feedback signal amplitude is increased sufficiently and if the base-bias current is reduced to the point where the stage approaches Class AB or Class B operation. For example, the circuit shown in Fig. 411 could be converted to blocking-oscillator operation by increasing the number of turns on L2's winding, removing R2, and increasing the values of R1 and C1.

With these changes, the negative-going signal delivered by L2 to the base electrode on alternate half-cycles may be of much greater amplitude than the transistor's normal base bias, charging C1 to a value exceeding B1's supply voltage. Once charged in this way, the voltage applied by C1 will place a reverse bias on the base-emitter junction, thus biasing it in its high resistance (or non-conducting) direction. With such a bias on the transistor, C1 can discharge only through R1 or the high resistance base-emitter junction. At the same time, the transistor's collector current is cut off and the transistor stops operating as an oscillator. The circuit remains in its non-oscillatory condition until C1 discharges sufficiently for normal base bias to reassume control. Thus, the oscillator operates only in short bursts (until C1 is charged) and is "blocked" the rest of the time, hence its name. The basic frequency of a blocking oscillator is determined by the L-C resonant circuit (L1-Cd), while its *blocking rate* is determined by the circuit's R-C *time constant* (R1-C1 in our example).

Depending on the type of transistor and circuit parameters used, blocking oscillators can supply pulses, saw-tooth signals, and other special waveforms, at audio, ultrasonic or radio frequencies. Such circuits are used extensively in TV and Radar equipment, in computers, instruments and in many types of military and industrial control devices.

The multivibrator

From our discussion of the common-emitter configuration in Chapter 3, we found that the output signal of such a stage is 180° out-of-phase with respect to its input. If two common-emitter stages are connected in cascade, there is an additional 180° phase-shift in the second stage, or a total shift of 360° from the input of the first to the output of the second stage. With a 360° (full cycle) phase shift, the amplified signal is *in-phase* with the original signal. From this, we can conclude that a two-stage R-C-coupled amplifier will oscillate if its "output" is coupled back to its "input." Such is indeed the case, and an oscillator employing this principle is illustrated in Fig. 412.

Referring to the schematic diagram, two p-n-p common-emitter stages are connected in cascade, with the output of the second stage (V2) coupled back through C1 to the "input" of the first (V1). In effect, the two stages are *cross-coupled,* for V1's collector is coupled to V2's base through C2, and V2's collector is coupled to V1's base through C1. V1's base bias is furnished through R1, and V2's through R3. R2 and R4 serve as collector-load resistors for the first and second stages, respectively. Operating power for the entire circuit is furnished by B1.

Since there are no L-C-tuned circuits involved in this type of oscillator, the circuit's repetition rate (frequency) is determined primarily by the R-C time constants. To understand how this can happen, let us consider circuit operation. . .

When power is first applied, bias current is supplied to both transistors simultaneously. Since there are always minor differences in the conduction characteristics of transistors, even of the same type, one or the other of the two will start conducting first. For purposes of discussion, let us say it is V2. When V2 starts conducting, its collector current, through R4, will cause a voltage drop, thereby developing a *positive-going* signal. This signal, transferred through C1 to V1's base, tends to reduce V1's collector current. With the current through R2 (V1's collector load) reduced,

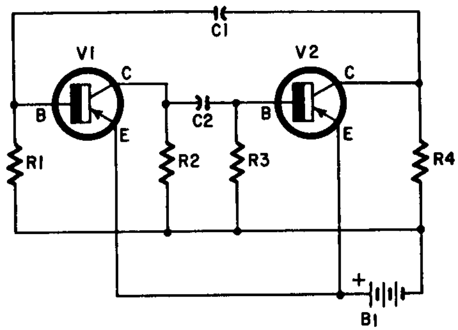


Fig. 412. Circuit of a transistor multivibrator. The schematic is essentially that of a two-stage amplifier with capacitive feedback between its input and output circuits.

there is less voltage drop across this resistor, and a *negative-going* signal is developed. V1's output signal, then, transferred through C2 to V2's base, increases V2's collector current still more and further increases the positive-going signal developed across R4 and delivered back to V1's base.

As this action takes place, C2 is charged through R2 and V2's base-emitter resistance. When C2 is fully charged, its voltage is essentially the same as that of R1 and it can no longer deliver a negative-going signal to V2's base. With no applied signal to continue the increase in V2's collector current, there is no longer a positive-going signal developed across R4 and hence no signal delivered to V1's base. When this occurs, V1's collector current starts to increase, developing a positive-going signal across R2. This signal is transferred to the base of V2 as C2 discharges.

Of course, when a positive-going signal is delivered to V2's base, its collector current (now at a maximum) is reduced, thereby developing a negative-going signal across R4. This new signal, coupled through C1 back to V1's base, causes V1's collector current to increase further.

Thus, the conduction characteristics of V1 and V2 have been reversed and C1 can now charge through R4 and V1's base-emitter resistance. While C1 is charged through this path, C2 is being discharged through V1's collector-emitter resistance and through R3. This action continues until C1 is charged and C2 discharged, at which point the action once again reverses and the original conditions are reestablished.

In summary, C2 charges through R2 and V2's base-emitter resistance. It is, in turn, discharged through V1's collector-emitter circuit and through R3. Feedback capacitor C1 is charged on alternate half-cycles through R4 and V1's base-emitter resistance, and is discharged through V2's collector-emitter circuit and through R1.

The waveform of the signal developed is shaped by the charge and discharge curves of the R-C networks involved and, hence, does not have the sine-wave characteristics of signals developed by L-C oscillators. It is, instead, a complex pulse-like waveform which is very rich in harmonics. For this reason, this general type of circuit is called a *multivibrator.* Depending on the choice of transistor types and circuit parameters, multivibrators can deliver a variety of signal waveforms over a broad range of frequencies. Commercially, such circuits are used in computers, TV and radar equipment and in many types of electronic instruments and control devices.

TIME TERMS

In Rodgers and Hammerstein's famous musical, *The King and I,* the King of Siam sings a philosophical little song entitled "A Puzzlement." The King wonders about the ways of the world, and particularly some of the customs of the British, who are, to him at least, strange and illogical.

When dealing with electronic circuits, we encounter expressions involving *time* quite frequently. These are often "a puzzlement"—not just to newcomers, but to more advanced workers as well. We speak of the *transit time* of current carriers (holes move slower than electrons), the *time-constant* of R-C circuits, the *timing* of pulse generators, the *time-base* of graphs or waveforms and so on.

Now time itself, although intangible, is certainly neither strange nor unusual. We all use time measuring devices . . . clocks, watches, and calendars . . . in our everyday activity, and frequently schedule our affairs according to time intervals . . . "see you at eight," . . . "we'll eat at six," . . . "it's a date for Thursday." Our timing periods are well known . . . centuries, decades, years, months, weeks, days, hours, minutes, and seconds. Where accurate measurements are needed, as when timing athletic events, we may deal with tenths or even hundredths of a second.

If we consider time, as Einstein did, but another dimension, then the chief difference between time as we normally think of it and "time" in the electronics sense lies *in the intervals used for measurement.* An hour is a rather large unit if we need measurements accurate to a tenth of a second. By the same token, a tenth . . . or even a hundredth . . . of a second may be too large a unit for measuring certain types of electronic phenomena. For example, suppose we consider a signal of 1,000 cps (1 kilocycle or 1 kc). This is not a very high frequency, even as audio signals go . . . yet the time required for one full cycle is only a *thousandth* of a second. If we want to determine the instantaneous amplitude of the signal at some point during its cycle, we must deal with even smaller units. Going a step further, the time required for one cycle of a 1,000,000 cps signal (1 megacycle or 1 mc) is but a *millionth* of a second, and for a 1,000 mc signal, but a *thousandth-millionth* of a second! By comparison, a wink of the eye would take ages.

The more common terms used to express time intervals in electronic circuits are . . .

Millisecond—thousandth of a second.

Microsecond—millionth of a second.

Millimicrosecond—thousandth-millionth of a second.

Time constant

In our discussion of blocking oscillators and of multivibrators, we mentioned the term "R-C time constant." This is a familiar expression and is used frequently when dealing with electronic circuits. It indicates the product of the resistance of the circuit in ohms and its capacity in farads (or ohms in megohms and capacity in microfarads) and is the time required, in seconds, for the capacitor to charge to 63% of its final voltage when a dc voltage is applied to the circuit . . . or conversely, to discharge to 37% of its original voltage when the capacitor is discharged through the resistance.

Consider the simple series circuit shown in Fig. 413-a. It consists of a source of dc voltage (B1), a SPST switch (SW), a fixed resistor (R), and a capacitor (C). A voltmeter is connected to measure the dc voltage across the capacitor.

With the switch open, we'll assume that C is discharged and V measures zero volts. When the switch is closed, the capacitor starts to charge through the resistance. It doesn't reach full charge immediately, however, for it takes a certain length of time for the electrons to travel through the circuit and accumulate upon the capacitor's lower plate. Checking our voltmeter constantly, we can plot the capacitor's voltage (representing its charge)

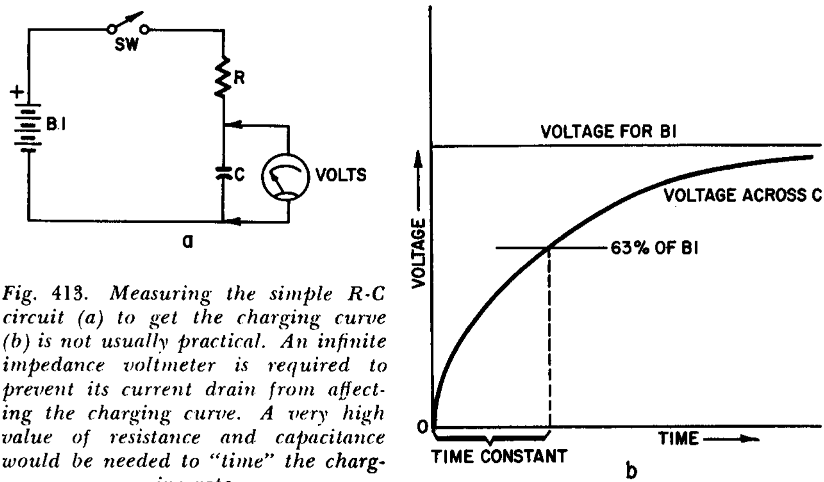


Fig. 413. Measuring the simple R-C circuit (a) to get the charging curve (b) is not usually practical. An infinite impedance voltmeter is required to prevent its current drain from affecting the charging curve. A very high value of resistance and capacitance would be needed to "time" the charging rate.

against time intervals. The resulting graph would look similar to that given in Fig. 413-b. The *time* required for the capacitor's voltage to reach 63% of the battery's voltage is the circuit's R-C *time constant,* and equals the product of R's value in megohms by C's value in μf. In practical circuits, the time constant is often a tiny fraction of a second.

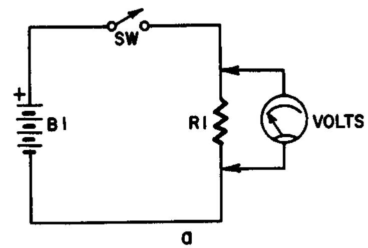
If we waited until the capacitor reached its full charge, then opened SW and connected R across C to discharge it, C's voltage would drop gradually, following a curve almost identical to its *charge* curve, but inverted. The time required for the capacitor to discharge to 37% of its original voltage under these conditions also equals the circuit's time constant.

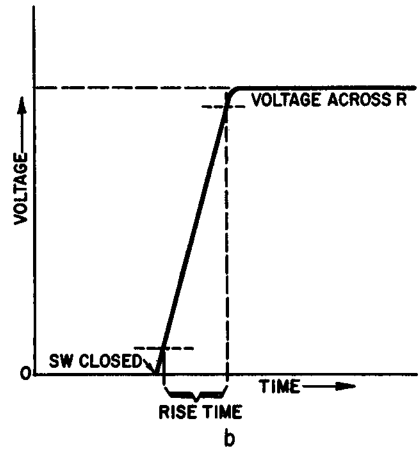
The charge and discharge characteristics of R-C circuits are utilized quite frequently in special purpose oscillators (such as multivibrators), in filters and for modifying waveforms. Oscillators which depend on the discharge of an R-C circuit for their operation are sometimes called *relaxation oscillators.* A very simple form of relaxation oscillator may be assembled by connecting an appropriate Zener diode (see Chapter 3) in place of the voltmeter in the circuit shown.

Rise time

There is another "time" expression which occasionally causes a scratching of heads . . . *rise time.* It is encountered most often

Fig. 414. Circuits like this and that of Fig. 413 are usually measured with more sophisticated devices than voltmeters. A typical method is with an oscilloscope. A camera can be used to record the rapid action and measurements made on the photograph.





when dealing with circuits handling square-waves, pulses, and other step-like functions. This term can be explained quite easily if we refer to Fig. 414.

Another simple-series circuit is shown in Fig. 414-a. It is quite similar to the one we used in discussing time constant, but with the capacitor omitted and the voltmeter (V) connected across our resistor (R1).

Suppose we had a stop-watch accurate to a millimicrosecond and could read the voltmeter at such short intervals. We close the switch and measure the *instantaneous* voltage across R1. We would find that a definite period of time is required for the voltage across R1 to equal battery voltage. This period will depend on such things as circuit-distributed capacities, lead inductances and so on. If we plotted R1's voltage against time, our curve might look something like that given in Fig. 414-b. Such a curve, of course, is a step function.

By definition, then, *rise time* is the time required for a step function (such as a square-wave or pulse) to go from 10% to 90% of its peak value. In practice, rise time is generally given in milliseconds, microseconds or, in high frequency circuits, in millimicroseconds.

CHAPTER 5

How Transistors are Made or , Growing Crystals Without a Garden

An oft-quoted and very ancient recipe for rabbit stew admonishes the reader to, *"First, catch a rabbit."*

Good advice for its day, this basic instruction may be a little hard to follow in our modern world of sprawling suburbs, paved superhighways, canned and frozen food and pre-cooked dinners. Not only are rabbits somewhat scarce—at least in their natural state—there's a good chance that 99 persons out of 100 wouldn't know *how* to catch a rabbit.

Although the march of technology has made obsolete the first part of our time-honored recipe, its basic philosophy is as true today as it was when Daniel Boone blazed a trail through the wilderness. To make *anything*—be it rabbit stew, an office building, an automobile, a dish, a telephone, a pistol, a radio receiver or a toothpick—you must start with proper basic ingredients. The ingredients needed will vary with your intended product and with the stage at which you start your manufacturing or assembly process. To make steel, for example, you start with pig iron. The pig iron itself, however, is a "manufactured" product which, in turn, is extracted from a basic raw material mined in mineral form—iron ore. In like fashion, a radio receiever is assembled from such basic ingredients (or components) as capacitors, coils, resistors, tubes or transistors, copper wire, and so on, with each of these products manufactured from more basic materials.

The "heart" of the transistor, the crystal diode, and related devices, as we have in Chapter 3, is a small multi-layered crystal of semiconductor material and, like the rabbit in our rabbit stew, is the one *essential* ingredient in device operation. This small crystal is mounted on a metal, ceramic or glass base called the *header* and covered with an outer case of copper, steel, plastic or glass, depending on intended application and upon the preferences of individual manufacturers. Electrode leads or pin connections generally are made of copper-alloy wire, plated with tin, solder, silver or gold. Often, a plastic "potting" compound is used to fill empty spaces within the case to cushion against shock and seal against possible contaminants. However, the case, header, leads, potting and other parts of a practical commercial unit are important only in a supplementary way, for the *semiconductor crystal is actually the heart of the transistor* (diode, etc.).

Basic transistor construction

Knowing the basic ingredient needed to make a transistor, we can follow the sage advice given in our ancient recipe and, *first, catch a semiconductor.* Semiconductor elements are materials which lie between the metals and the insulators in their ability to conduct electricity. They are found in column IV of the periodic table of chemical elements. Included in this group are the elements carbon (C), silicon (Si), titanium (Ti), germanium (Ge), zirconium (Zr), tin (Sn), hafnium (Hf) and lead (Pb). All these elements have four electrons in the outer shell of their atoms and hence have a *valence* of 4.

In crystalline form, pure germanium and silicon are electrically neutral and relatively poor conductors. Each atom shares its four valence electrons with four neighboring atoms, and there are no free electrons left over to "carry" an electric current. If sufficient energy is added to the crystal (by heating, for example), the atoms will start to vibrate. Under these conditions, an electron occasionally can acquire enough energy to break its valence *bonds* and to move freely through the crystal. When this happens, of course, the broken bond is left unsatisfied and will readily accept nearby free electrons. In effect, the resulting crystal acts as if it has a particle the size and mass of an electron, but with a positive charge.

As you will recall from our discussion in Chaper 3, the *absence* of an electron in the molecular structure is called, for convenience, a *hole.* This hole can, in itself, serve as a current carrier as it migrates from one atom to another through the material. Thus, adding energy to a crystalline semiconductor can change it from a relatively poor conductor (insulator) to a fairly good conductor, for the excess energy creates *pairs* of holes and electrons which can serve as current carriers.

Our discussion in Chapter 3 also covered, rather briefly, the formation of p-type and n-type semiconductors by adding various "impurity" elements to the basic semiconductor material. The p-type semiconductors were formed by adding *acceptor* elements from column III of the periodic table, and n-type semiconductors by adding *donor* elements from column V to our intrinsic (column IV) or "pure" semiconductors. A p-type semiconductor has an excess or surplus of holes and these are the predominant current carriers. In similar fashion, free electrons are the dominant current carriers in n-type semiconductors.

Atoms of elements found in column III of the periodic table have three electrons in their outer shell and hence a valence of 3. Included in this group are boron (B), aluminum (Al), scandium (Sc), gallium (Ga), yttrium (Y), indium (In), lanthanum (La), thallium (Tl) and actinium (Ac). Of these, boron, aluminum, gallium and indium generally are used as acceptor elements in commercial transistor production. In contrast, atoms of elements in column V of the periodic table have five electrons in their outer shell and a valence of 5. Included here are nitrogen (N), phosphorus (P), vanadium (V), arsenic (As), nobium (Nb), antimony (Sb), tantalum (Ta) and bismuth (Bi). Of these, phosphorus, arsenic and antimony are used commercially as donor elements.

Let us consider a crystalline structure made up of atoms with a valence of 4 (germanium or silicon). As we have seen, such a crystal is neutral electrically, for all the electrons are held in covalent bonds. If we replace one of the atoms in the crystal with an atom having a valence of 3 (such as aluminum), the resulting crystal is not "complete"—in that one valence bond is left unsatisfied. The structure has a positive hole and will readily accept nearby negative electrons. It is in this fashion that our p-type semiconductor is formed.

On the other hand, if our valence 4 atom is replaced with a valence 5 atom (such as arsenic), only four of the electrons are needed to complete the crystalline structure, and thus the fifth electron becomes "free" to serve as a current carrier. Our crystal then becomes an n-type semiconductor.

Basically, all semiconductor devices are multilayer units of alternating p- and n-type materials, forming one or more p-n junctions. In a triode transistor, there are two essentially parallel junctions. All transistor construction techniques, then (as we shall see later, there are many) have as their final goal the formation of these two parallel p-n junctions as close together as is practicable and with accurately controlled characteristics.

Transistor manufacturers must rely on the skills and knowledge of many branches of science and engineering. For example, skilled mechanical engineers and machine designers are called on to create the intricate machinery needed to handle and process the microminiature semiconductor components. Transistor design itself is the responsibility of the solid-state physicist. Using the mathematics of atomic physics to develop his basic designs, the physicist calls on the metallurgist and the physical chemist to supply necessary data concerning semiconductor alloys and crystal characteristics. Often, chemists are called on to help work out production details and to organize and supervise any processes involving chemical reactions. Electronic engineers get into the act, too, for they work out circuit applications; as a result, they are expected to set basic electrical specifications for the types of transistors and other semiconductor products needed in their equipment designs.

The basic steps in manufacturing a transistor include: (a) obtaining a suitable semiconductor material, (b) changing the material into crystalline form, (c) cutting and shaping the crystal into final size, (d) forming a pair of parallel, closely spaced p-n junctions within the crystal, (e) attaching electrode leads, (f) testing the partially assembled unit, (g) assembling the transistor into its case, potting (or encapsulating) and attaching external leads or pins, (h) sorting, branding and final tests. Depending on exact manufacturing techniques, some of these steps may be combined or additional substeps added. In general, however, the chief differences between techniques is found in step (d), the exact method used to form the two p-n junctions which form the heart of the transistor. Let us follow through a typical manufacturing process step by step.

Growing crystals

Our first step, of course, is to obtain a suitable semiconductor material. As we have discovered, the most popular materials in current use are germanium and silicon. To be useful in transistor manufacture, the semiconductor must be *extremely* pure. This is essential if we are to maintain an accurate control over characteristics by "doping" with our own impurity (acceptor and donor) elements. Typically, the undesired impurity concentration in transistor-grade semiconductor material may range from 1 part in 1,000 to as little as 1 part in 100,000,000,000.

Germanium, one of our basic raw materials, may be obtained from a number of sources, including flue dusts from zinc smelters and ashes from certain types of coal. Silicon is more plentiful, for

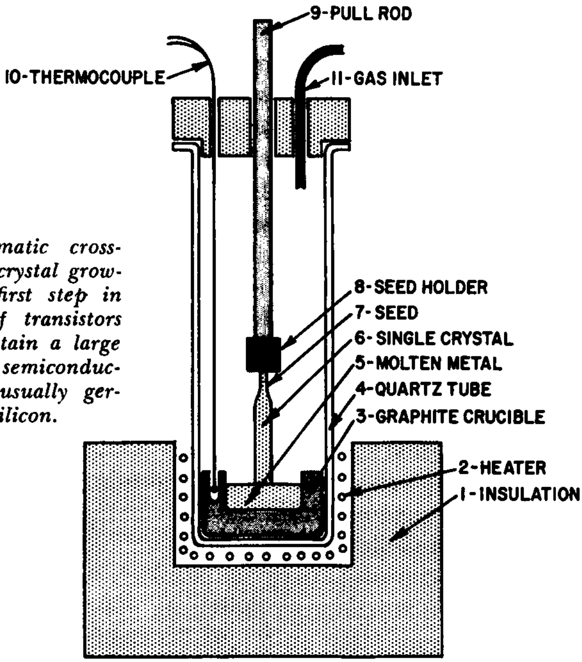


Fig. 501. Diagrammatic cross-sectional view of a crystal growing furnace. The first step in the manufacture of transistors and diodes is to obtain a large ingot-like crystal of semiconductor material . . . usually germanium or silicon.

almost all rock and sand have silicon as one of their primary elements. Regardless of the original source of the raw material the semiconductor element is taken through a multistage refining and purification process which involves both chemical reactions and metallurgical techniques, resulting in ingots or bars of relatively "pure" metal.

The last stage in the purification of semiconductor materials is generally a technique called *zone refining*. This method utilizes an interesting characteristic of the metallic elements: most impurities are more soluble in the metal when in its liquid state. In practice, then, a narrow molten zone is made to travel along the length of a semiconductor bar. Higher-purity metal "freezes" behind the molten area, with the impurities tending to concentrate in the molten zone until the end of the bar is reached. This process can be repeated over and over until the desired degree of purity is obtained, with the "impure tip finally cut off. The heating used to produce a narrow molten zone in the metal is generally accomplished by induction coils carrying rf energy. These circle the metal bar and, by induction, generate circulating eddy currents in the metal, causing it to heat until the metal melts. Depending on technique, the narrow coils may be moved

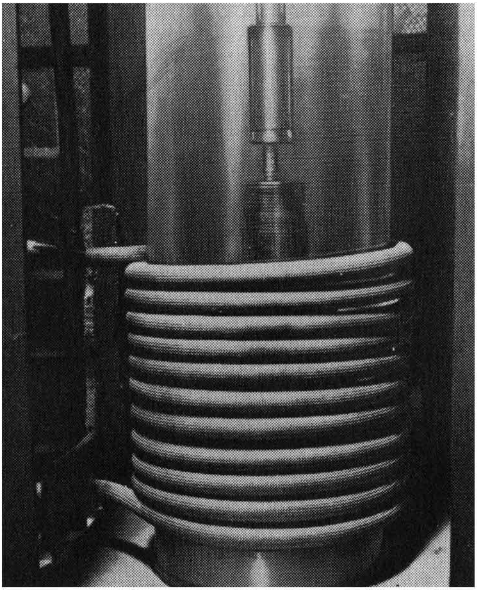


Fig. 502. Close-up view of the furnace shown in Fig. 501. The induction coil and growing crystal of germanium are clearly visible.

the length of the bar, or the bar itself can be moved through fixed-position coils.

Once the refining and purification process is completed, the next step is to form or "grow" a single large crystal of the material, doping it with precise amounts of the acceptor or donor elements to give it the desired p- or n-type characteristics. One of the basic methods used for this process is shown schematically in Fig. 501, while a detailed view of the actual heating coils is given in Fig. 502.

Referring to Fig. 501, the equipment used includes a quartz tube (4) to exclude atmospheric gases, a graphite crucible (3), heating coils (2), a motor-operated pull rod (9) to which is attached a clamplike holder (8) for a "seed" crystal (7). Temperature is checked accurately by a thermocouple (10). Special provision is made for filling the quartz tube with an inert gas through gas inlet (11), thus preventing possible contamination by airborne impurities.

In operation, the semiconductor metal (5) is melted in its crucible by rf induction. Coils (2) surrounding the crucible are fed with high-frequency electrical energy; the resulting electro-

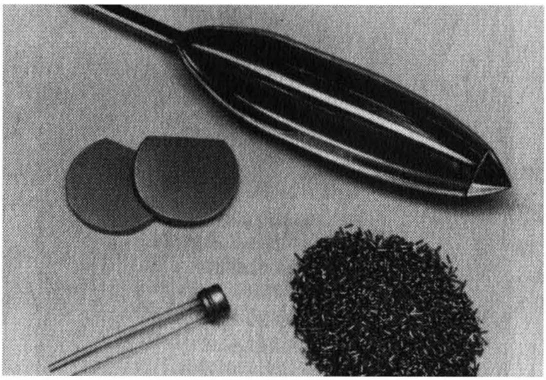


Fig. 503. A large single crystal of semiconductor material is shown in the background. "Discs" sliced from such a crystal are shown to the left, and an assembled transistor is in the left foreground. The pile of tiny rods in the right foreground are unassembled transistors . . . short pieces of semiconductor!

magnetic field develops eddy currents in the semiconductor which, in turn, heat it to the melting point. The molten metal is maintained within a few degrees of its melting point. At the proper time, the pull rod (9) is lowered into the pool of molten metal and gradually withdrawn. A small piece of crystal is clamped on the end of the pull rod, and this serves as a "seed" on which a larger crystal is grown. The large single crystal has the same molecular pattern (orientation) as the seed crystal used. The molten semiconductor may be doped with acceptor or donor impurities as desired to develop either a p- or n-type crystal. The crystal growing process continues until all the molten metal is used, with the resulting crystal assuming a long torpedo-shaped form (Fig. 503).

Cutting and shaping

In the next manufacturing step, the ingotlike crystal is sliced into thin, waferlike discs, using diamond-tipped power-driven saws. These discs may be from a few thousandths to a few tenths of an inch thick, depending on the type of transistors being produced. In any case, the disc thickness must be kept to accurate tolerances. Typical discs are shown in Fig. 503, while a diamond "crystal slicing" machine is illustrated in Fig. 504.

The sliced discs, in turn, are cut into still smaller sections, again using diamond-tipped saws. Each disc may be diced into dozens, or even hundreds, of rectagular wafers or bars, each one not much larger than the tip of a pencil. These extremely small bits of semiconductor crystal, with further processing, will become the "hearts" of our final transistors.

Alloying and assembly

At this point, we must identify the general type of transistor we are producing, for production techniques differ widely after the

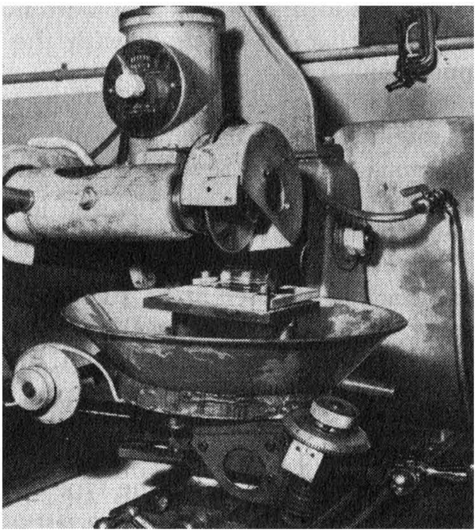


Fig. 504. Slicing discs from a crystal of germanium. A special diamond-tipped saw blade is used. Disc thickness must be accurate to a minute fraction of an inch.

basic crystal is grown, cut into waferlike discs and finally diced into small wafers or bars. For the moment, we will assume that our goal is an *alloyed-junction p-n-p transistor,* and that our basic crystal, during its growing process, was doped with donor impurities to produce an n-type semiconductor.

The next step, then, is to form the two p-n junctions. This is accomplished by plating or pressing small metal dots on each side of the thin rectangular wafers, then heating the resulting assembly in a special oven. The metal dots contain an acceptor type impurity, such as indium. When the small assembly is heated, the metal dots melt and dissolve part of the wafer, producing an alloy solution. When the heat is removed, the dots recrystallize, forming p-n junctions on each side of the n-type wafter. Generally, one metal dot is made larger than the other and serves as the collector electrode. The wafer itself is the transistor's base, and the smaller dot is its emitter.

There are several variations of this basic technique. In some cases, for example, the tiny wafer is first etched electrochemically, reducing its thickness in the center, and thus the thickness of the base layer. By reducing the carrier transit time through the base region, this method improves the resulting transistor's high-frequency characteristics. If an n-p-n transistor were needed, a wafer of p-type semiconductor would be used, with the metallic dots containing a donor type impurity.

With the alloying step completed, we have the "heart" of our transistor, and the next step is to attach suitable leads to the emitter, base and collector electrodes. Several techniques may be used here, including alloying, soldering, welding and pressure bonding under heat. The electrode leads may be aluminum, gold, indium, nickel, or other metals. As a general rule, a resistive contact is used for attaching leads to exposed areas, such as the emitter and collector regions, or to the base in alloyed-junction types. Where the exposed electrode region is extremely thin, a semiconductor contact might be employed. As an example, suppose that the base region is so thin that it is impossible to attach a lead without overlapping the emitter and collector electrodes. Here, the base lead could be doped heavily with the same type of impurity used in the base (a donor element in the case of a p-n-p transistor, for example). In this way, a small overlap into the emitter and collector areas will have relatively little effect on transistor operation, for these two electrodes "see" the base lead as simply an extension of the base region.

The assembled transistor is now attached to the connection leads or pins on its *mounting base* or *header* using spot-welding techniques. Afterward, the unit may be tested, and then mounted in its case, encapsulated (potted), retested and branded with manufacturer's name and type number.

Other techniques

The alloying technique which we have examined in detail, while quite popular, is only one of several used in the commercial production of transistors and related semiconductor devices. The original (and now obsolete) *point-contact* transistor, for example, consisted of a small cube of n-type germanium alloy on which rested two closely spaced leads or cat's-whiskers. Small p-type areas were formed by impurity diffusion under each lead, thus providing the two p-n junctions needed for transistor operation. Point-contact construction techniques are still used in the manufacture of single junction diodes (Fig. 505). Compare this with the alloy-junction transistor assembled using the techniques described above (Fig. 506).

Today, two methods are generally used in the manufacture of transistor junctions. Each has a number of refinements and modi-

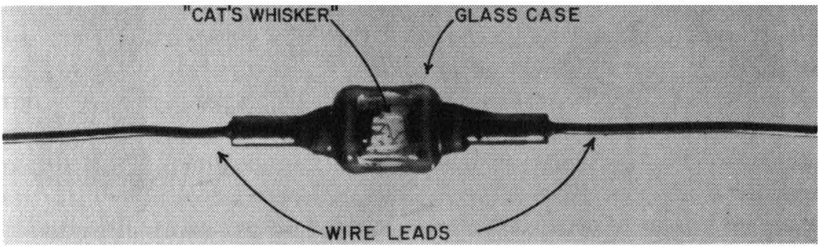


Fig. 505. Detailed view of a point-contact diode, showing the wire cat's whisker. *The original point-contact transistor used two closely spaced cat's whiskers as its emitter and collector electrodes.*

fications, depending on the transistor characteristics needed and the production methods preferred by individual producers. The alloying technique is an example of the *impurity-contact* method, as are the *point-contact, surface-barrier* and *vapor-diffusion* techniques. All these methods rely on the external contact of impurity elements with the basic semiconductor crystal. With the alloy technique, for example, impurity pellets are applied to the crystal, then alloyed with it through heat to form p-n junctions. In point-contact assemblies, external cat's-whiskers (wire leads) of a donor or acceptor element rest against the crystal surface. In the surface-barrier transistor, opposite sides of the crystal wafer are electro-etched with fine streams of a suitable electrolyte solution carrying an impurity element. When the etching has progressed sufficiently, electric currents are reversed, plating small dots of the impurity element on opposite surfaces. In some cases, a *microalloy* is formed by heating the completed surface-barrier transistor, causing the dots to melt, alloy with the semiconductor base and recrystallize, much in the same way that conventional alloyed-junction transistors are made.

The vapor-diffusion technique is one of the latest of the various *impurity-contact* methods. A wafer of the basic semiconductor crystal is placed in a closed crucible containing one or more impurity elements. The original crystal has been doped to have either n- or p-type characteristics, depending on the type of collector needed in the final transistor. When heat is applied to the crucible, the impurity element is vaporized and diffused into the semiconductor crystal, developing, say, a base region. Later, an

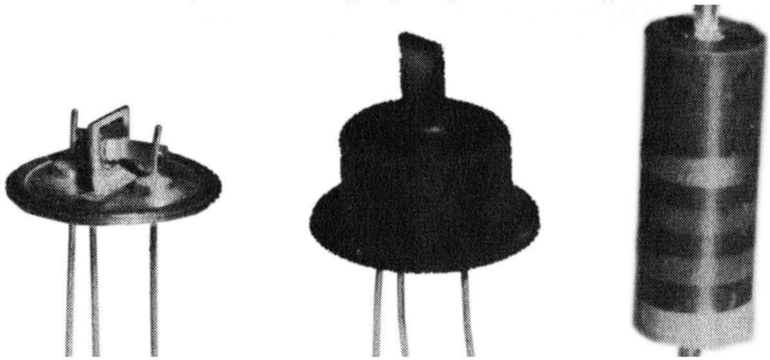


Fig. 506. Uncased (left) and cased views of an alloy-junction transistor . . . the small carbon resistor (right) is included to show relative sizes.

emitter region may be added. If both acceptor and donor impurities are used at the same time, one will generally diffuse into the crystal faster than the other, creating three alternate regions of p-type and n-type materials. The next result is something like a candy bar, with the outer surface of the crystal of one type of material (like the chocolate coating on a bar), a subsurface layer of another type of semiconductor and, finally, a central core (similar to the filling in candy). This is converted into a three-layer sandwichlike transistor by cutting and etching techniques which expose the inner layers.

If desired, alternate transistor junctions may be formed as part of the original crystal growing process. The *grown-junction* method includes three general techniques: (a) rate growing, (b) meltback and (c) grown diffused. All of these techniques rely on the formation of p-n junctions either during the initial crystal growth or by using impurity elements included within the original crystal, in contrast to the *impurity-contact* methods which rely on the external application of acceptor and donor elements *after* the crystal is formed.

The *rate-growing* process takes advantage of a previously noted characteristic of semiconductor alloys, the tendency for some types of impurities to concentrate in a liquid (melted) semiconductor. In practice, the molten metal in the crucible (Fig. 501) has both donor and acceptor elements added to it. Of these, the donor element is more sensitive to crystal growth rate, with the amount of this impurity taken up by the crystal varying as growing conditions change. As the large single crystal is being grown, then, the power is turned off, allowing the crystal to grow much more rapidly. At the proper point, a greater than normal amount of power is applied to the heating coils, causing the growth to stop and part of the crystal to remelt. The power is then turned off again, allowing the metal to cool and the crystal to restart growth. In the narrow region where the crystal growth was stopped and restarted, the acceptor impurity tends to concentrate, changing the normally n-type semiconductor into a thin p-type region. Repeating this process results in the formation of alternate p- and n-type areas throughout the length of the crystal. Later, the crystal can be cut into sandwichlike discs containing n- p- and n-layers. Finally, the discs are diced into tiny n-p-n bars, leads are attached and the resulting transistors mounted in their cases.

As in the rate-grown technique, the molten semiconductor is doped with both acceptor and donor impurities in the production of *meltback* transistors. Here, however, a constant growing rate is used, so that the resulting crystal has n-type properties. After slicing into disclike wafers and dicing into thin bars, the crystal is subjected to one more process step before leads are attached. Heat is applied to one tip of the bar, melting a small drop of the semiconductor. When the heat is removed, the drop recrystallizes, with the acceptor elements concentrating at the drop's point of contact with the unmelted bar, thus changing this part of the semiconductor into a thin p-type region and producing an n-p-n transistor. A typical meltback transistor, uncased but attached to its *header*, is shown diagrammatically in Fig. 507; the juncture between the semiconductor bar and the melted and recrystallized tip is the transistor's base electrode.

Somewhat similar to the rate-growing technique, the last of the *grown-junction* methods, the *grown-diffused* process, takes advantage of the different diffusion rates of donor and acceptor elements in a growing crystal. Growth is started with the molten semiconductor doped to one characteristic. Additional donor and acceptor impurities are added to the molten metal as crystal growth proceeds. Of these, one impurity tends to concentrate more rapidly at the boundary of the growing crystal and the molten pool, changing a thin area here into the opposite type semiconductor. If growth were started with an n-type semiconductor, for example, an acceptor impurity would be chosen which diffuses into the crystal more rapidly than the donor impurity, thus producing a p-type material at the crystal boundary. As crystal growth continues, the other impurity eventually diffuses into the crystal, re-

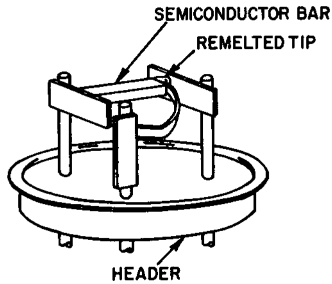


Fig. 507. Construction details of a meltback-*junction transistor.*

establishing a dominance of donor elements and forming a second n-type layer. Later, of course, the grown crystal is sliced into a three-layered "sandwich" of n, p and n semiconductors, then diced into n-p-n transistors.

Power transistors

All of the transistor construction techniques discussed have been used in the manufacture of low-power or *small-signal* transistors. In general, these transistors have maximum current ratings of less than 100 ma and maximum power dissipations of, at the most, a few hundred milliwatts. Except for physical size and maximum ratings, power transistors are basically similar to small-signal units. Like their smaller "cousins," they are two-junction solid-state devices made up of alternate p- and n-type semiconductor regions, with beta values which approximate those of low-power types. Both n-p-n and p-n-p types have been manufactured. Their electrical specifications, compared to small-signal types, are on a higher order of magnitude, however. Maximum current ratings may range up to 25 or 50 *amperes,* while power dissipations of 50 to 100 *watts,* or more, is not uncommon.

Physically much larger than common small-signal transistors, power types generally are manufactured using one of the *impurity-contact* methods, with perhaps the majority made using the *alloyed-junction* technique. Recently, some manufacturers have adapted the *vapor-diffusion* technique to the production of high-power high-frequency units. In the latter technique, the double diffused semiconductor crystal is etched away until a miniature tablelike projection, or *mesa,* is left projecting from the collector electrode, exposing the base and emitter regions. But regardless of the technique used to form the final junctions, the semiconductor crystals used in power transistor construction are grown using the same technique employed in growing crystals for small-signal units (Fig. 501).

"Exploded" and cross-sectional views of one type of power transistor are given in Figs. 508-a and 508-b, respectively. This is a

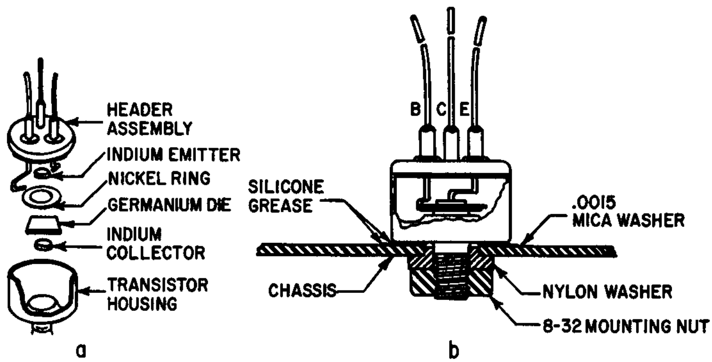


Fig. 508. An "exploded" view of a p-n-p alloy-junction power (multiwatt) transistor is shown at (a) . . . mounting details for this unit are given at (b). For maximum heat dissipation, most power transistors are mounted on a metal chassis or "heat sink."

p-n-p junction alloy type in which indium metal dots are pressed into each side of an n-type germanium die, then alloyed to it by heating in a suitable oven. One indium dot becomes the transistor's collector electrode, the other serving as an emitter. The small piece of germanium is the transistor's base, with a connection made to it through a nickel ring which surrounds, but doesn't touch, the indium emitter. The collector electrode is connected directly to the transistor's outer case to facilitate heat dissipation. In use, the transistor generally is mounted on a metal chassis, although insulated from it with a thin mica washer and a nylon shoulder washer, as shown in Fig. 508-b. This permits maximum heat transfer from the transistor to the chassis, but maintains an "open" circuit electrically.

The basic assembly stages for different types of power transistors are really quite similar with closeup interior and exterior views of a completed unit given in Fig. 509. The differences between this power transistor and the one shown in Fig. 508 are more physical than electrical, for both are manufactured using the alloyed-junction technique and both are p-n-p units. As with most power transistors, the collector electrode is connected, both physically and electrically, to the unit's outer metal case.

Summary

As we have seen, there are two general methods for producing the closely spaced p-n junctions needed for transistor operation:

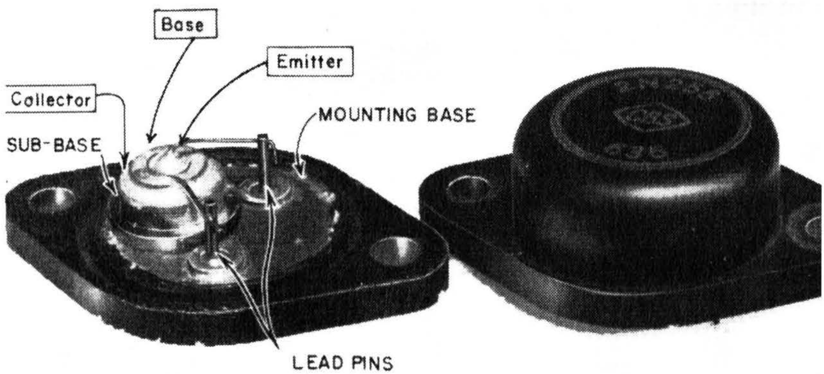


Fig. 509. Detail view of an uncapped (left) and capped multiwatt p-n-p transistor.

(a) *impurity contact* and (b) *grown-junction.* The several techniques available under each method are summarized in Fig. 510. The differentiating feature between them is the manner in which the donor and acceptor impurity elements are added to the basic semiconductor crystal. With the impurity-contact method, the impurity elements are added *after* the crystal is grown, either by physical contact (as in the point-contact transistor), by electroplating (as in the surface-barrier type), by alloying, by a combination of plating and alloying (as in microalloy transistors) or by vapor diffusion (as in *mesa* types). Where the grown-junction method is used, the impurity elements are added to the molten semiconductor *before* the crystal is grown, with the final junctions formed either by varying the rate of growth, by remelting

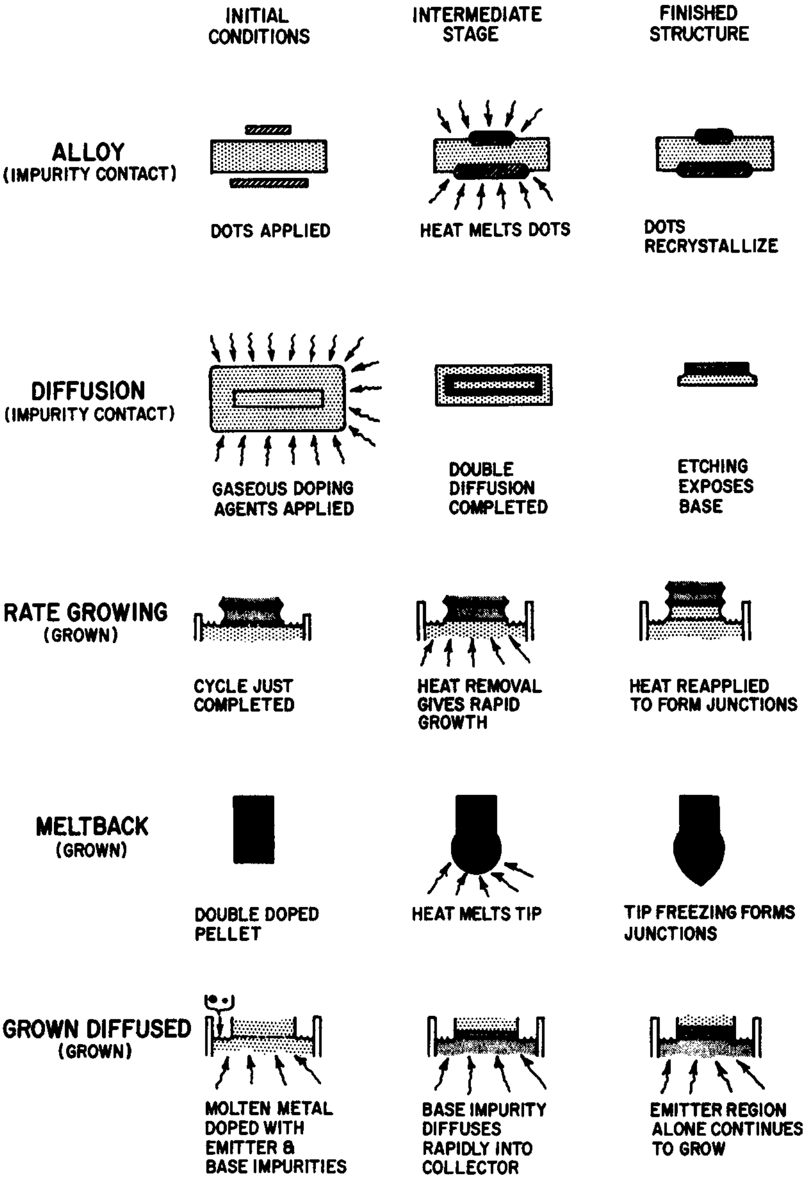


Fig. 510. Summary of important transistor production techniques. These methods, used to produce various types of junction transistors, represent the different techniques that may be employed. The important thing is that the production methods used permit the formation of two junctions having accurately controlled physical and electrical characteristics.

part of the shaped transistor (as in the meltback type) or by adding additional donor and acceptor elements as the crystal is grown.

All of the techniques may be used for producing both p-n-p and n-p-n transistors, all are suitable for mass production, and all may be used to manufacture a variety of types, including audio, rf, low-noise, low-power and high-power units. As a general rule, however, power transistors are manufactured using impurity-contact methods.

Type selection

Unfortunately, transistor production techniques have not yet been refined to the point where a manufacturer can produce a large batch of transistors with closely controlled characteristics. Instead, a given production run will include units falling within a relatively broad range of specifications. One of the last production steps, then, is a final test of important characteristics and the assignment of type numbers. In a given production batch, for example, a certain percentage of the transistors may have a very high beta, another percentage will have a lower beta, and still another group may have a medium beta. Typically, those falling within predetermined high beta ranges would receive one type number, those with a medium beta range a second type number, and those with low beta values still a third number, *even though the entire batch was manufactured at the same time under virtually identical conditions.*

As a general rule, the published *maximum ratings* for various transistor types should be taken at their face value. Electrical characteristics, such as beta, cutoff frequency, distributed capacitances and so on, are "typical" or average, rather than absolute, values. A transistor with a given type number, for example, may have a beta higher (or lower) than another unit with the same type number, even when produced by the same manufacturer. The tolerances to which the transistors are selected depends, to some extent, upon their type and price. For example, a premium-priced "high-quality" transistor may have a beta range of only 3 to 1. That is, the maximum beta may be only three times as much as the minimum beta for transistors of that type number. Typically, a good-quality transistor of a specific type number may have a listed beta of, say, 30, but typical production units of that type may have individual betas from as low as 20 to as high as 60. Lower-priced "experimenter's" transistors, on the other hand, may have a beta range as high as 5 or 6 to 1.

CHAPTER 6

How to identify transistor types

Differentiating between similar objects can be a simple chore or an exacting task, depending on one's training, experience and knowledge, and the degree to which details must be identified. Caucasians, for example, complain that "all Orientals look alike." And in TV Westerns, both heros and heavies will refer to the Indian without further description, as if all Indians looked, thought, and acted alike. Chances are that the reverse situations are just as true—that many Orientals have difficulty distinguishing between various Europeans, and that the early Indians tended to group all light-skinned people together as "white men."

That reminds one of the story about the little boy who peeked through a board fence into a nudist camp. When asked later whether he saw men or women in the camp, he shrugged his shoulders and replied matter-of-factly: "I couldn't tell—they didn't have their clothes on."

There are, of course, many different ways of classifying human beings. If we stick to basics, there are only two types, male and female. Going a step further, we could classify according to color of hair, skin, or eyes. On the other hand, if we ignore color as an identifying trait, we could classify according to facial features or physical attributes—tall, short, fat, thin, and so on. Or, we could ignore physical characteristics completely and classify on the basis of type of dress, costume, or uniform, according to nationality or according to speech or voice accents. The possibilities are endless, and if *all* classifications and sub-classifications were utilized at the same time, it is likely we would soon find that there is only *one* person in each of several billion sub-classes.

In one sense, transistors are a lot like people. If we stick to basics, there are only two types—p-n-p and n-p-n. But we can go further, classifying according to physical size, type of semiconductor material (germanium or silicon), style of case, electrical specifications (such as beta, cutoff frequency, or power handling capacity), method of construction, or even according to lead or pin arrangement. Again, if we utilize *all* the possibilities, we are likely to find that there is just *one* transistor in each of several million sub-classes.

While people, as our little boy found, may be easier to classify with their clothes *on,* transistors can be just the opposite. If we wish to classify according to type of construction, for example, they are easier to identify with their clothes (cases) *off,* for many different transistors may be assembled in the same style and size case; externally, then, they would all look alike. For that matter, many types of component can be (and are) assembled in transistor style cases, including such units as subminiature relays, diodes, resistors, special purpose capacitors, and even microminiature iron core transformers. Unfortunately, removing a transistor's case is likely to damage it beyond repair unless special tools are used.

There is one saving grace, however. As we learned in Chapter 2, and as many of you probably knew beforehand, transistors are stamped with an identifying *type number* by the manufacturer. By using this as a guide and referring to the manufacturer's published technical literature, we can learn such things as the transistor's lead connections, its maximum ratings, and its technical specifications, as well as its basic electrical type (n-p-n or p-n-p) and material of construction.

When referring to transistor data charts, you'll find quite often that one or more special identifying symbols are given in addition to the usual electrical specifications and maximum ratings. These are generally in the form of various letter combinations, such as PC, SB, MADT, and so on. Contrary to what you might expect at first glance, these are not the initials of the manufacturer nor of his Chief Engineer. Nor do the symbols represent company slogans, a la the famous one used by a major cigarette manufacturer . . . LS/MFT . . . even though clever copywriters from Madison Avenue could easily dream up appropriate mottos—Pretty Cute for PC, Some Boy for SB, or Many A Dreamy Transistor for MADT. Instead, these symbols refer to the basic transistor type as far as manufacturing technique is concerned, and hence indicate the transistor's internal construction.

TYPES OF TRANSISTORS

As we found in Chapter 5, there are two basic methods of manufacturing transistors, (a) *impurity contact* and (b) *grown junction*. These general methods, in turn, include a variety of specific techniques, each with its own advantages and limitations. Some of these techniques are less costly than others, but are better suited to the manufacture of certain types as far as electrical specifications are concerned. Others require complex production equipment or precise controls, but permit the manufacture of units with special characteristics or have a high yield.* For example, alloy junction transistors are relatively inexpensive to produce but, as a general rule, are low-frequency types. Surface barrier transistors, on the other hand, manufactured using a more complex technique, have excellent high frequency characteristics but limited power handling capabilities.

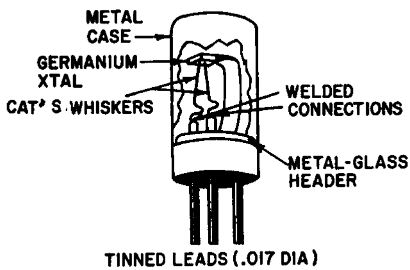


Fig. 601. Cutaway view of point-contact transistor construction. Although considered obsolete, it has characteristics not possible with junction types.

Within certain limitations, however, any of the basic manufacturing techniques may be used to produce a variety of transistors with a broad range of electrical specifications. Alloy junction types, for example, may be produced with low, medium, or high beta

* *Yield*—the number of accepted transistors resulting from a given production run. The higher the yield, the lower the number of rejects. Ideally, the yield should be 100%.

characteristics, and with power capabilities ranging from a few milliwatts to as high as a hundred watts (or more). Although not a preferred technique for producing very high frequency transistors, the alloy junction methods can be used to produce rf as well as audio types. Similarly, grown junction transistors can be produced with various beta specifications, with different cutoff frequencies, and in a variety of sizes with different power ratings. In choosing a specific construction technique, then, the transistor manufacturers must consider such factors as . . . (1) personal preferences, (2) materials used, (3) electrical characteristics

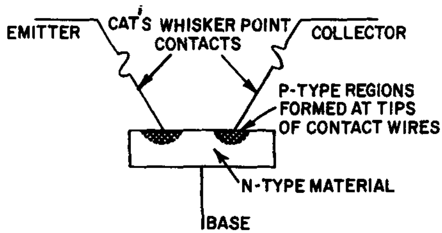


Fig. 602. This pictorial representation of the point-contact transistor shows the formation of p-type regions on the n-type base material.

needed, (4) special production equipment required, (5) quantities to be produced, and (6) expected yield.

For discussion purposes, it is convenient to group transistors into "types" according to internal construction (and hence method of manufacture). Using this as a guide, then, let's take a look at commonly available types. . . .

Point-contact transistors

Historically, the point-contact transistor could be considered the grand-daddy of all modern types, for it was the first type invented and the first type to be produced commercially. Identified by the letter symbols *PC* or *PT*, the point-contact type is no longer produced by major commercial manufacturers and hence is generally considered an obsolete design. Since PC types were the first to be produced, any units encountered are likely to have "low" type number designations, typically, numbers 2N22 through 2N26 and 2N30 through 2N33 are point-contact types.

The basic mechanical construction of a typical PC transistor is shown in Fig. 601, while its electrical construction is shown symbolically in Fig. 602. In this unit, the *emitter* and *collector* electrode connections to the *base* are made through two fine wire cat's-whiskers resting on (or welded to) the crystal's surface a few thousandths of an inch apart. The base itself is a small wafer of n-type or p-type semiconductor crystal and, typically, may measure only 0.05 inches square by, say, 0.02 inches thick. After assembly, the transistor is formed by applying electrical pulses and heat to its cat's-whisker electrodes, causing small amounts of the electrode material to diffuse into the base crystal as an impurity, and changing the conduction characteristics of the crystal in the areas immediately adjacent to the points of contact, thus developing the two p-n junctions essential to transistor operation. With an n-base transistor, as shown in Fig. 602, p-type areas are formed, and vice-versa.

Electrically, the n-base PC transistor is similar in operation to a modern p-n-p junction type, while the p-base type has general characteristics roughly equivalent to an n-p-n unit. As a general rule, point-contact transistors have exceptionally high gain—so

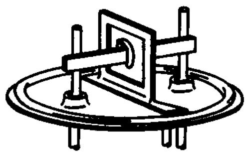


Fig. 603. Typical internal construction of an alloy-junction transistor. Connections to base, emitter and collector are welded to low-heat conduction alloy leads.

high, in fact, that they are unstable when used as amplifiers except in the common-base configuration. They make excellent oscillators and switches, and have good high frequency characteristics. It is significant that the earliest rf transistor types were invariably point-contact designs.

On the debit side of the ledger, PC transistors are difficult to manufacture with controlled specifications and hence have a low yield. In addition, they tend to be noisy and, as mentioned above, are unstable in some types of circuits. Requiring higher operating voltages than other types, they have limited power handling capabilities due to the small size of the cat's-whisker electrodes.

Alloy junction transistors

Identified by the letter symbols *AJ*, *JA*, or, often, simply *A*, the alloy junction transistor is perhaps the most popular type in current production. AJ transistors are used extensively in audio amplifiers, hearing aids, radio receivers, test instruments, computers, power supplies, power inverters, and related types of equipment. The internal construction of a typical alloy junction transistor is illustrated pictorially in Fig. 603, while its electrical configuration is shown symbolically in Fig. 604. As we learned in Chapter 5, the transistor consists of a small wafer of p or n-type semiconductor crystal, with small dots of an impurity element pressed onto opposite sides and then alloyed with the semiconductor by the application of controlled heat. If the wafer is, say, n-type germanium, and the metallic dots are indium, the resulting transistor is a p-n-p unit. The indium, when melted, alloys with the n-type germanium (or silicon) and changes the areas immediately adjacent to the dots into p-type material.

There are two basic variations of the alloying technique. The *diffused alloy* transistor is similar to the conventional alloyed junction shown, but with the addition of a diffused base region on the emitter side. The net result is to reduce current carrier transit time through the base region and thus to improve the transistor's high frequency characteristics; diffused alloy transistors are often called *drift* types. In the *alloy diffused* transistor, the initial wafer is pre-doped to the characteristics desired in the emitter electrode.

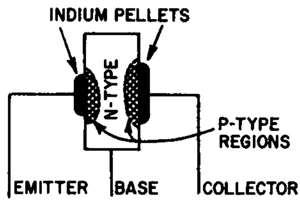


Fig. 604. Pictorial cross-section of base shows how pellets are diffused into the base wafer. Joining pellets to base material forms an alloy of the two materials.

Afterwards, the base region is formed by diffusion out of the emitter dot, which is made of an alloy containing both donor and acceptor impurities; as you will recall, different impurity elements have varying diffusion rates through a semiconductor crystal.

As a class, alloy junction transistors offer perhaps the widest range of different characteristics of currently available types. With somewhat less gain than PC types, they are stable in most circuit configurations. They can be manufactured to reasonably close specifications and, in production, have a good yield rate. Most high power transistors are alloy junction types. In general, they have much lower inherent noise than PC types, with selected types having a lower noise level than premium quality tubes. Frequency-wise, most alloy junction transistors are best-suited for operation at audio and low radio frequencies, but some drift (diffused alloy) types have cutoff frequencies in the hundreds of megacycles and hence are suitable for use in the VHF range. AJ transistor voltage ratings are moderately low, but, because of the large electrode contact areas, they are capable of handling relatively large currents.

Grown junction transistors

Differing from the various alloy types in that their p-n junctions are formed either during the initial crystal growing process or by diffusion methods utilizing impurities included in the original crystal, grown junction transistors are identified by the letter symbols *G, GJ,* or *GD* (grown diffused). In contrast to the alloy types, which are generally wafer-like in general shape, *GJ* transistors have a bar-type structure, as shown pictorially in Fig. 605. The grown junction transistor's electrical configuration is illustrated in Fig. 606. Physically, the multi-layered bar which makes up the final

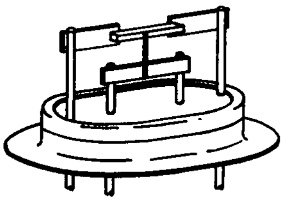


Fig. 605. Grown-junction transistors have a bar construction instead of the wafer. Tabs are welded to leads that come out from the bottom of the header. These are emitter, base and collector, from left to right.

transistor may measure about one-one hundredth of an inch on each side by about one-tenth of an inch long.

If we ignore the obsolete *double-doped* process, grown junction transistors are produced using any three basic techniques: (a) rate grown, (b) meltback, and (c) grown diffused. All three were discussed in Chapter 5. In addition to the three basic processes, a diffusion technique is sometimes employed in the formation of meltback transistor junctions, resulting in a *meltback-diffused* type. Meltback and meltback-diffused transistors are often identified by the symbols *MB* and *MBD,* respectively, rather than the various G symbols used more commonly for identifying grown junction types. Both p-n-p and n-p-n transistors may be produced using the outlined processes, depending on the donor and acceptor impurities used, on the basic semiconductor material, and on the exact technique employed. Although requiring precise controls during crystal growth and junction formation, the various grown junction techniques have a fairly good yield rate and permit the manufacture of units with reasonably consistent characteristics.

Most grown junction transistors have somewhat better high frequency specifications than corresponding alloy junction types. This is due, in part, to the lower interelectrode capacities resulting from their smaller base area. On the other hand, their smaller physical size limits their power dissipation, making their structure less suitable for the construction of high power (multi-watt) transistors. In general, however, GJ types are used in applications similar to those in which AJ types are found, but with greater emphasis on rf circuits (such as receivers) and less emphasis on power applications.

Surface-barrier transistors

For transistor action to take place, we must have two p-n junctions in close proximity to each other and separated by a layer of semiconductor material, the base region. Current carriers—electrons and holes—must travel through this intervening region in moving from the emitter to the collector electrode. It follows, then, that the transit time required for the current carriers will have a direct effect on the speed with which the transistor can respond to applied signals, and hence on its high frequency characteristics. In basic terms, the thinner the base region, the better the transistor's high frequency specification and the higher its cutoff frequency.

Unfortunately, there is a lower limit on the practical size of the base region where transistor junctions are formed using alloy

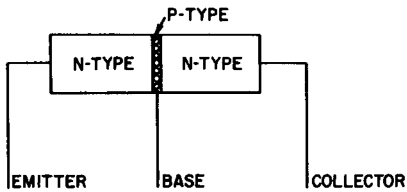


Fig. 606. Bar of semiconductor material has a grown p-type layer between two n-type sections.

or grown junction techniques. If the wafer used in alloy junction transistors is made too thin, it will be next to impossible to cut and shape using available equipment and, in addition, will be quite fragile and subject to damage during handling and processing. In grown junction transistors, on the other hand, attempts to reduce base region thickness below a minimum dimension can (and will) result in frequent shorts between the much larger emitter and collector regions, thus ruining the transistor. A determined effort to reduce the base region thickness, using an *easily controlled* production technique, resulted in the development of the surface-barrier transistor and its family of related types.

A schematic view showing the construction of a surface-barrier transistor is given in Fig. 607. It consists of a small wafer of semiconductor material with two etched out areas opposite each other. Small metallic dots are plated on either side of the etched out regions and become the emitter and collector electrodes. In the manufacturing process, two needle-fine streams or jets of a metallic salt solution are directed against opposite faces of a thin wafer of semiconductor material. A dc voltage is applied to the streams and to the wafer, with the polarity chosen so as to electrolytically etch away the sprayed area of the semiconductor. This etching process is carried on until the center of the wafer is only a few ten-thousandths of an inch thick. At this point, the voltage polarity is *reversed,* stopping the etching process and electroplating the two metallic dots on opposite sides of the wafer. If the basic wafer is, say, n-type germanium, an indium salt (such as indium sulfate) solution may be used for the etching process, with indium dots plated on the wafer in the final process. A transistor produced using these materials would have p-n-p characteristics. Once the plating process is completed and the transistor cleaned, the unit

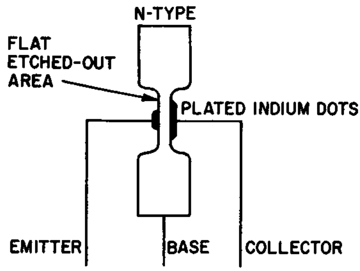


Fig. 607. Surface-barrier transistor base is similar to that in Fig. 604. Base section is made thinner with etching process.

can be tested, leads attached, and the assembly mounted on its header and cased.

In the basic surface-barrier transistor, there is no diffusion of the impurity elements into the base or semiconductor wafer, and hence its action depends on metallic junction contact alone. The unit may be subjected to additional processing steps, however, to obtain more desirable characteristics. For example, a combination of heat and applied voltage can result in surface diffusion, or sufficient heat may be applied to cause the metallic dots to melt, alloy with the semiconductor, and to recrystallize, much in the same way that alloying takes place in the alloy junction transistor, but on a much smaller scale. Finally, a combination of alloying and diffusion techniques may be employed.

By starting with the basic transistor, then, and processing in various ways, a whole family of transistor types can be produced. As with the other types discussed, these can be identified quite conveniently by using appropriate letter symbols—*SB* for surface-barrier, *SBDT* for surface-barrier diffused type, *MAT* for micro-alloy type, and *MADT* for the micro-alloy diffused type. The basic construction of a micro-alloy (MAT) type is illustrated in Fig. 608.

Transistors in the surface-barrier family share a number of characteristics. They are all high frequency types, with cutoff frequencies ranging up into the hundreds of megacycles. A refined and improved version of the micro-alloy diffused transistor, for example, has been used as an oscillator at frequencies as high as *several thousand megacycles.* The ultra-thin base layer used in these types is easily punctured by excessive collector-to-emitter voltage, however, and hence these units have a relatively low maxi-

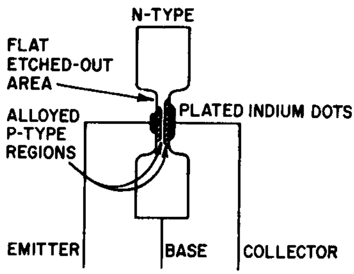


Fig. 608. This method can be considered as a combination of the surface-barrier and alloy-junction processes.

mum voltage coupled with limited power handling capabilities. As a result, transistors in this family find their widest application in high frequency receivers (such as FM and TV sets), in low power VHF and UHF transmitters, and in the high-speed switching circuits employed in some types of computers.

Other types

While the great majority of the transistors in current production are of the types just discussed, these are by no means the only types possible. New manufacturing techniques are being developed on an almost month to month basis, and these, either alone or in combination with older techniques, may result in new types of transistors, or even whole families of types. Along with the development of new techniques, there is a constant refinement in established manufacturing processes, resulting not only in higher yield rates but, often, in the development of improved types.

As we learned in Chapter 5, one of the latest of the impurity contact methods involves the formation of alternate p-n junctions by the diffusion of impurities in vapor form into the basic crystalline wafer. Afterwards, etching and cutting methods are used to expose the alternate layers and to permit lead connections to be attached to the base and emitter electrodes. Physically, the resulting transistor looks much like a flat surface with a projecting table-like formation. The larger flat area serves as the transistor's collector and has excellent power dissipation qualities, while the thin base layer resulting from the vapor diffusion permits improved high frequency characteristics. The net result is a transistor which combines power handling ability with a relatively high frequency cut-off. Because of the table-like shape of the projecting junction areas above the collector electrode, these units are called *MESA transistors* (in Spanish, mesa is the word for table and frequently is applied to the high, flat plateaus found in the Western United States).

In some cases, manufacturers will combine the features of two basic production techniques. One firm, for example, is producing transistors by using a combination of alloy and diffusion processes. Here, the transistor is built up on a base of p-type semiconductor. Two small impurity pellets are placed on the surface. One pellet (which eventually becomes the base electrode) contains only an n-type impurity. The other contains both p-type and n-type impurities, but with the p-type chosen to have a slow diffusion rate. When the resulting assembly is heated, the n-type impurities diffuse rapidly together, forming a thin n-type layer. At the same time, the pellets melt into an alloy with the semiconductor base material. When the completed unit is cooled, the alloy recrystallizes, forming p-type (emitter) and n-type (base) electrodes, with the thin n-type region between the original p-type base and the p-type emitter pellet. Transistors manufactured using this technique have thin base regions and relatively large collector electrodes, hence good high frequency characteristics combined with fair power handling capabilities. Since the technique is called the Post Alloy Diffusion Process, transistor types manufactured using it are identified as *PADT* types.

OTHER SEMICONDUCTOR DEVICES

Like the "Ruler of the Queen's Navy" in Gilbert and Sullivan's famous comic opera, The *H.M.S. Pinafore,* the transistor has "cousins . . . whom he reckons up by dozens." The transistor, then, for all its types, styles, sizes and varied circuit applications, is but one of a vast class of devices utilizing the special properties of semiconductor materials. Included in this vast group are single and multiple junction two-electrode devices (diodes), three electrode units other than the triode transistor, and devices with four or more electrodes (tetrodes, etc.). The variety of currently available and potential devices is so great that a mere listing by type and application would require several chapters, while a detailed examination of each would require many volumes, each larger than this book.

Obviously, space limitations prohibit as detailed a study of other semiconductor devices as we made of the triode transistor. However, we will find it worthwhile to take a quick look at the various units that may be encountered in commercial and experimental equipment, or that may be used in conjunction with more familiar transistor circuits. In this way, you can obtain a better understanding of semiconductor operation and, equally important, can make better use of the transistor and its cousins in your hobby or profession. When examining these other devices, keep in mind the special physical properties of semiconductor materials, for these unique properties are basic to the operation of *all* semiconductor devices.

In review, semiconductor materials are crystalline elements or compounds with conduction characteristics falling midway between those of good conductors, such as copper or silver, and good insulators, such as glass or ceramic. They have two really unique properties: (1) the ability to change their conduction characteristics when subjected to external forces, such as electrical or magnetic fields, heat, or light, and (2) the ability to conduct by means of two different types of current carriers, holes or electrons (or both).

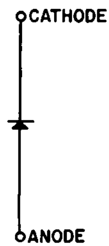


Fig. 609. The diode symbol does not indicate the abilities of the unit used. This could be a low-voltage, low-current detector or a high-voltage, heavy current rectifier.

Diodes

By definition, a diode is simply a two electrode device. Unfortunately, such a definition can cover, as the expression goes, a multitude of sins, for the very nature of semiconductor materials makes possible a fantastic array of two terminal devices, all with different electrical characteristics, divergent circuit applications, and varying physical traits. To avoid confusion, then, the expression *semiconductor diode* (or, simply, *diode*) is used in conventional practice, as a descriptive term only for single junction, two electrode devices having unilateral conduction characteristics —that is, devices which offer little or no resistance to current flow in one direction, but which act as a high resistance or open circuit when current polarity is reversed. When dealing with other types of diodes, an additional descriptive term is used, for example, *Bistable* diode, *Tunnel* diode, and so on.

As you may recall, we prefaced our discussion of transistor operation with a detailed examination of the p-n junction diode and its function (see Chapter 3, Fig. 301). The junction used in a diode may be fabricated using any of the manufacturing processes we have reviewed, while the diodes themselves, like transistors,

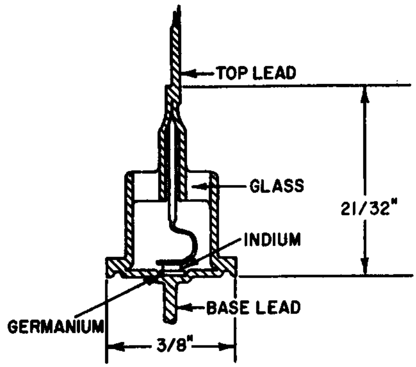


Fig. 610. Cut-a-way view of a top-hat rectifier diode. Base material is connected to outside case to carry the heat away from the junction and prevent damage to the semiconductor material.

may be made of germanium, silicon, or various semiconductor compounds. If we ignore external appearances, variations in manufacturing technique, and differences in exact electrical specifications, there are basically two classes of diodes in current use: (a) signal handling types and (b) power rectifiers. Both operate in essentially the same fashion, except for the magnitude of currents handled, and both are identified with the same schematic symbol, as shown in Fig. 609.

Most signal handling diodes are rather small physically. A popular point-contact diode of this type was illustrated in Fig. 508, Chapter 5. Typically, their maximum voltage ratings range from a fraction of a volt to as high as several hundred volts, depending on intended application. Current ratings are generally less than 20 ma., but a few types can handle currents as high as 100 ma. Application-wise, these diodes are used as detectors, modulators, mixers, clippers, and limiters in receivers, test instruments, and control equipment, and as unilateral coupling elements in computers and special amplifiers. (Fig. 610 shows a cross section of a diode.)

A typical experimenter's application for a signal handling diode is shown schematically in Fig. 611. Here, the diode (D1) is used as a detector (demodulator) in a simple radio receiver. In operation, rf signals are picked by the antenna (ANT.) and selected by tuned circuit L1-C1. The selected rf signal is detected by diode D1, with the audio portion of the detected signal coupled through C2 to a single stage common-emitter amplifier using a p-n-p transistor, V1. The amplifier stage, in turn, drives a pair of standard headphones. Receiver operating power is supplied by a three volt battery, with V1's base bias current furnished through series resistor R1.

There is some overlap in specifications between the smaller power rectifier diodes and the larger signal handling types. This results in a rather hazy line of division between types and a tendency on the part of designers to use diodes within this area interchangeably. As a general rule, however, most power rectifier diodes are designed to handle currents ranging from .1 ampere

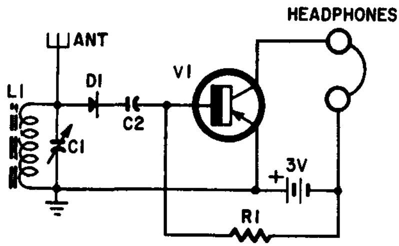


Fig. 611. Simple crystal detector and single transistor amplifier is a typical circuit for the beginning experimenter.

(100 ma) up to hundreds of *amperes*. Their voltage ratings are comparable to those of signal handling types.

The relatively heavy current handling requirements of power rectifiers necessitates a number of differences in their physical construction as compared to signal handling diodes. For example, most signal handling types are quite small physically, and are mounted in glass, ceramic, or plastic housings. A few types are assembled in metal transistor style cases. Generally, the anode or cathode leads are independent of, and insulated from, the housing, even where a metal case is used. The junction and contact areas are small, to minimize distributed capacities and to obtain a minimum current carrier transit time through the semiconductor material, thus assuring good response to high frequency signals. In contrast, the junctions used in power rectifier diodes are relatively large and have broad contact areas. These features are needed to insure low internal resistance in a forward direction and to permit the units to handle the large currents for which they are designed. In addition, most power rectifier types are assembled in metal cases, with one electrode connected to the outer case both physically and electrically to insure good heat dissipation. As you may recall, a similar technique is used in the construction of power transistors, with the collector electrode connected directly to the metal case.

Application-wise, power rectifiers are used to convert ac into pulsating dc. A typical circuit is shown in Fig. 612. Here, 117 volts ac such as may be obtained from a household power outlet, is applied through current limiting resistor R1 to diode D1. The diode conducts on alternate half-cycles, charging capacitor C1 to peak line voltage. The capacitor, in turn, discharges slowly through bleeder resistor R2 and to the *load* connected to the output terminals, maintaining a reasonably steady output dc voltage

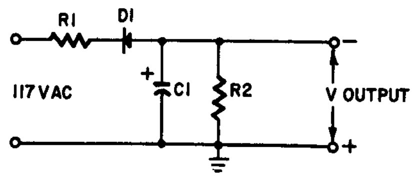


Fig. 612. Power-rectifier circuit is typical of that used in many table radios.

with the polarity shown. Since the diode conducts only on alternate half-cycles (when its anode is positive with respect to its cathode), the entire circuit is a *half-wave dc power supply*.

Power rectifiers, like signal handling diodes, can be connected in series or in parallel. A *series* connection is used where the *applied voltage* is greater than the breakdown voltage of a single junction. A *parallel* connection is used where the *current requirements* are greater than can be handled by a single junction. In practice, however, series connections are more popular than parallel arrangements, because better results are obtained with a single large junction than with several small junction diodes connected in parallel.

Zener diodes

All semiconductor diodes have a breakdown voltage. This is the voltage level at which the junction fails and allows a sudden increase in current when the voltage is applied in the diode's inverse or non-conducting direction. As mentioned in Chapter 3, this is the diode's *Zener voltage*. Some diodes are designed and selected for use at their Zener voltage and are called, appropriately enough, Zener diodes. Typical applications for these devices are in voltage regulator circuits, as voltage reference elements, as clippers and limiters, and as protective devices for voltage sensitive components, such as meters and transistors.

Photocells

Most semiconductor substances are sensitive to light energy as well as to heat and to the stress of electrical fields. Light energy impinging upon the material's surface causes the formation of hole and electron pairs within the crystalline structure. This can result in several types of phenomena, depending on the type of

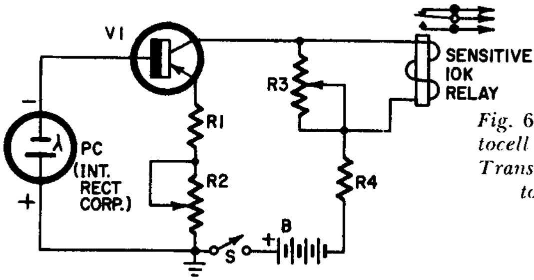


Fig. 613. Light falling on photocell changes bias on transistor. Transistor conducts and collector current closes relay.

material involved, on whether there are p-n junctions present, and on other factors. In general, however, the freeing of the electron and hole current carriers by the light energy causes a drop in the material's resistance. Where a p-n junction is involved, the change in resistance can be considerable. If the material is multilayered and thin enough to be transparent to light, the electrons may tend to accumulate on one surface, generating a small potential and, in effect, converting light energy into electrical power. It is this last phenomenon that is employed in the famous sun batteries used so extensively in artificial earth satellites and space probes.

A large variety of semiconductor substances are used in manufacturing these devices, including such semiconductor compounds as lead and cadmium sulfides, selenium, and, of course, germanium and silicon alloys. The majority of types are two-electrode units (diodes), but some manufacturers assemble certain types of their transistors in such a way that light can strike the p-n junctions, thus producing a three-terminal (triode) *phototransistor;* the resulting device can be used either as a conventional transistor or as a light-sensitive photocell, depending on the circuit requirements.

A typical semiconductor photocell application is illustrated in Fig. 613. Here, a self-generating photocell (PC) serves as a source of base bias current for a single stage direct-coupled amplifier using a p-n-p transistor, V1, in the common-emitter configuration. With the photocell dark, little or no base current is supplied to the transistor and collector current is very low. As a result, the relay remains open. When light strikes the photocell, a small voltage is developed and applied to V1's base-emitter circuit. This results in a flow of base bias current and a corresponding, but amplified, flow of collector current, closing the relay. Operating power is supplied by the battery, B, with resistors R1-R2 serving to limit the transistor's emitter current and as a control over the transistor's input impedance. R3 serves as a shunt across the relay and permits an adjustment of relay sensitivity. R4, in series with the relay, serves to limit V1's collector current. Circuits similar to this one may be used for assembling electric eye burglar alarms, doorway annunciators, counters, and related alarm and control equipment.

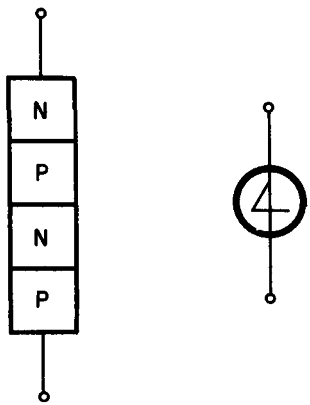


Fig. 614. The four-layer diode is a bistable device with characteristics similar to those of a gas-filled diode.

Bistable diodes

The diodes we have examined thus far have been single junction devices. In contrast, the *Bistable* diode, although having only two electrodes, has *three* semiconductor junctions. Its construction is shown pictorially in Fig. 614. Consisting of four alternate layers of p-type and n-type semiconductors, the Bistable diode can operate in either a high resistance non-conducting or in a low resistance conducting state, depending on the polarity and amplitude of applied voltages. The unit's four-layer construction is symbolized by its schematic symbol, shown in Fig. 615. The symbol resembles the numeral "4" (for 4-layer), with the slanting line serving as an arrowhead indicating the forward direction of the device when in a conducting state.

Since the Bistable diode operates as a two terminal, two-state device, its action is roughly analogous to that of a gas-filled diode, such as a neon bulb, or a Zener diode used at close to its Zener voltage, and it can be used in similar applications. Thus, its primary circuit functions are in relaxation oscillators, switches, bistable configurations, counters, and various computer logic circuits. A typical Bistable diode saw-tooth relaxation oscillator circuit is illustrated in Fig. 615. In operation, capacitor C1 is charged

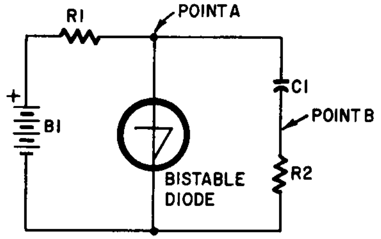


Fig. 615. Typical 4-layer diode relaxation oscillator circuit. Neon lamp or a gas-filled diode can replace the semiconductor but battery voltage must be increased considerably.

slowly by the battery (B1) through the series resistors R1 and R2. R1 is generally much larger than R2. The charging time depends on the R-C time constant of the circuit, or, essentially, upon the values of R1 and C1, since R2 is relatively small. During the charging period, the Bistable diode remains in its non-conducting state. However, as soon as the voltage across C1 reaches the breakdown or firing voltage of the diode, this unit suddenly switches to its conducting state. Its resistance may change from, say, 10 megohms or more (in its non-conducting state) to less than 20 ohms when it fires. At this point, C1 is rapidly discharged through the Bistable diode and through current limiting resistor R2 until the voltage remaining across C1 (and the current through the diode) is no longer sufficient to hold the diode in its conducting state, at which time the diode switches back to its high resistance state, allowing C1 to be recharged once again. This action is repetitive, of course, as long as power is supplied to the circuit. A saw-tooth signal voltage is developed at point A during C1's charge, with a sharp pulse developed at point B as C1 is discharged.

Tunnel diodes

Invented by a Japanese scientist and often called the *Esaki*

Diode in his honor, the Tunnel diode is a single junction, two terminal device having a *negative resistance* characteristic. Commercial units may be manufactured of germanium, silicon, or various semiconductor compounds, such as gallium arsenide, and, generally, are mounted in small cases about the same size as those used for subminiature transistors. Although the Tunnel diode

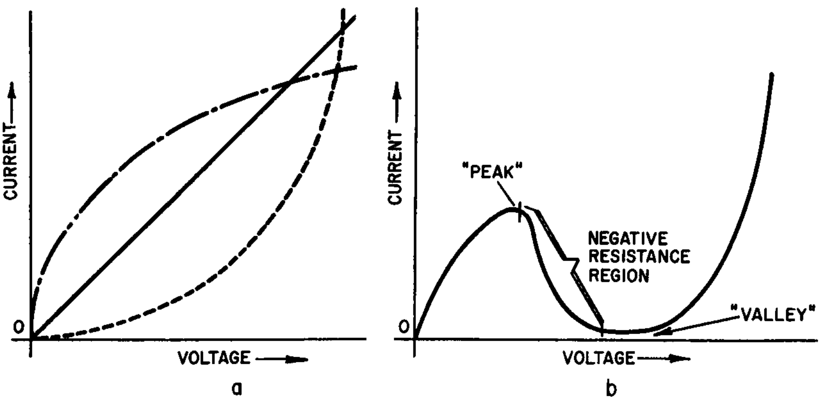


Fig. 616. The straight solid line in (a) is the linear characteristic of a resistor. Nonlinear devices may have curves like the broken and dashed lines. The Tunnel diode has a more complex curve (b) and the reverse slope (negative resistance region) allows the diode to oscillate.

is basically a two terminal device, some manufacturers provide a third lead or pin connection connected to the diode's case to serve as a shield.

In order to understand the operation of the Tunnel diode in practical circuits, we must first discuss the term *negative resistance.* This can be explained best by comparing the characteristic curve of the Tunnel diode with that of more familiar devices. Referring to Fig. 616-a, the characteristic curve of an ordinary resistor is shown by the solid line; this indicates that the current through the resistor, represented by the vertical axis, is directly proportional to the voltage applied to it, as represented by the horizontal axis. In other words, if we double the voltage, we double the current; if we triple the voltage, we triple the current, and so on. Since the "curve" is a straight line, the resistor is said to have a *linear* characteristic.

Some types of semiconductors, special resistors and other devices, have a *non-linear* characteristic curve. Typical non-linear curves are shown by the broken and dashed lines. Here, as in the simple resistor, current is still proportional to voltage; if we increase the voltage applied to the device, the current through it increases. But the relationship is no longer a simple one. In one case (broken line), circuit current increases rapidly up to a point, then increases slowly with further increases in applied voltage. In the other case (dashed line), the current increases slowly at first, then starts to increase rapidly as we approach higher voltages. In all three cases, however, *the current always increases, however little (or much), with increases in applied voltage.* The reverse is true, too, for a decrease in applied voltage will result in a corresponding decrease in current. The three devices which these curves represent are all positive resistances.

Let's take a look, now, at the characteristic curve of the Tunnel diode, as given in Fig. 616-b. As the applied voltage is increased, circuit current increases until a peak value is reached. Afterwards, further *increases in voltage bring about decreases in current,* and vice-versa. This continues until a minimum or valley point is reached, at which point circuit current once more starts to increase. From this point on, circuit current continues to increase with further increases in applied voltage. In the area between the peak and valley current values, then, circuit action is just the opposite of that in a conventional transistor, and the Tunnel diode is said to have a negative resistance in this region.

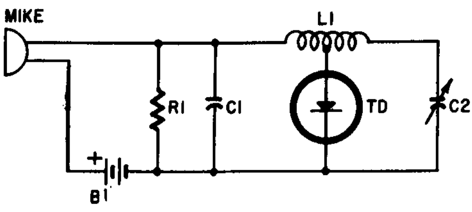
Negative resistance is a dynamic characteristic, for it shows up *only* if we consider *changes* in applied voltage. If we consider the steady diode current with a specific applied voltage, even if this voltage biases the diode in its negative resistance region, the diode still has a real resistance value, which can be determined quite easily by applying Ohm's law, and dividing the voltage at that point by the diode's current.

A Tunnel diode is made by unusually heavy doping of the semiconductors used in forming the p-n junction. Normally, excessive doping reduces the diode's reverse breakdown voltage. Theoretically, the limit is reached when the reverse breakdown voltage is zero, but this does not work out in practice. In the Tunnel diode, then, the doping is so heavy that the diode is still *in a reverse breakdown condition even with a slight forward bias.* It is this condition which accounts for the first peak in diode current as applied voltage is increased. As the forward voltage is increased still more, the reverse breakdown condition is more or less wiped out by the applied bias, and circuit current drops until the minimum or valley point is reached. Afterwards, the Tunnel diode behaves like a conventional forward-biased diode.

The Tunnel diode derives its name from the speed with which current carriers traverse the junction, almost as if there were a tunnel through this area. The action is so fast, in fact, that Tunnel diodes can be used at frequencies up into the thousands of megacycles.

The Tunnel diode's more important electrical specifications are: maximum forward and reverse currents in ma., power dissipation in milliwatts, peak current in ma., valley current in ma., peak voltage, valley voltage, series resistance, series inductance, terminal capacitance, and equivalent negative resistance. As a general rule, a Tunnel diode's quality is indicated by its *peak to valley*

Fig. 617. This simple circuit is actually a transmitter. Tunnel diode (TD) is an oscillator. Carbon microphone (MIKE) modulates the oscillator. Changes in voltage across R1–C1 cause the oscillator frequency to vary.



current ratio. Referring back to Fig. 616-b, this is the ratio of the diode's first peak current to its minimum valley current, and generally falls between 4 and 40, with most currently available units having ratios between 6 and 12.

A typical circuit application for the Tunnel diode is illustrated schematically in Fig. 617. Here, the Tunnel diode is used in a high frequency FM oscillator circuit as part of a miniature hand-held transmitter or *Wireless Microphone.* In operation, the instrument's basic frequency is determined by tuned circuit L1-C2, and by the Tunnel diode's interelectrode capacity. Dc operating power is furnished by a battery (B1) with the exact voltage needed to bias the diode in its negative resistance region determined by a voltage divider made up of the carbon microphone's dc resistance and fixed resistor R1, bypassed by C1. Acting as a negative resistance to cancel the internal resistance of the tuned circuit, the Tunnel diode maintains the L-C circuit in an oscillatory condition once it is shocked into oscillation by an external signal . . . or simply by the act of applying power. Sound vibrations striking the microphone's diaphragm cause instantaneous changes in the microphone's effective resistance. This, in turn, varies the instantaneous bias voltage applied to the Tunnel diode and its interelectrode capacity. Finally, changes in the diode's effective capacity results in corresponding changes in the resonant circuit's instantaneous frequency, developing a frequency modulated output signal.

Unijunction transistors

Originally called a double-base diode, the Unijunction transistor is a transistor by definition only. It has a single junction, like a diode—hence its name, *Uni* (one) junction—rather than the two junctions which characterize the triode transistor. The con-

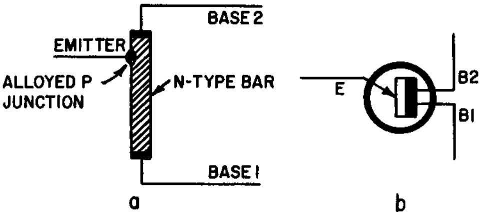


Fig. 618. Unijunction transistor has two bases, no collector.

struction of a Unijunction transistor is shown in Fig. 618-a, while its schematic symbol is given in Fig. 618-b. This unit consists of a bar of one type of semiconductor material (such as n-type silicon) with electrical connections at each end and with a wire of one of the *impurity* elements (such as aluminum) alloyed to the bar at at a point slightly off-center, where it forms a p-n junction. The semiconductor bar is the *base*, with its two electrical contacts identified as Base 1 (B1) and Base 2 (B2), while the p-n junction is the emitter.

In normal operation, a dc voltage is applied across the two base connections, setting up a potential drop inside the base bar which biases the emitter junction in its reverse, or non-conducting, direction. Thus, if a small external voltage is applied between the emitter and, say, B1, little or no current can flow and the emitter to B1 circuit acts like a high resistance or open circuit. However, if the externally applied emitter voltage is of the right polarity and is increased sufficiently to overcome the internally developed bias, current carriers (holes, for example) can be injected from the emitter into the base, permitting some current flow through the emitter-B1 circuit. At the same time, the presence of these injected current carriers in the base decreases the internal voltage drop, reducing the effective reverse bias and permitting the externally applied voltage to become more effective, thus increasing the emitter-B1 current still more. This effect is cumulative, with the result that the emitter-B1 current increases rapidly up to the maximum that the power supply can deliver, or until the transistor is destroyed; to prevent self-destruction, then, it is customary to include a current limiting device (such as a resistor) in the emitter dc supply circuit. In summary, the emitter-base 1 junction can be switched rapidly from a high resistance to a low resistance state by the application of an external voltage.

Since the Unijunction operation is essentially that of a switch from a non-conducting to a conducting state, its operation is analogous to that of a gas-filled thermionic tube, or thyratron. Like the thyratron, it is useful primarily in switching and relaxation oscillator circuits. A basic Unijunction relaxation oscillator circuit is shown in Fig. 619. A single battery B1, supplies both the interbase (B1 to B2) bias and emitter voltages. In operation, the emitter junction is biased internally in the non-conducting state. However, the emitter-B1 junction is also biased by the

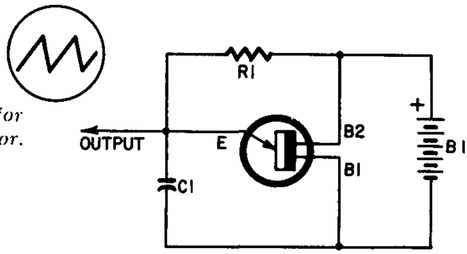


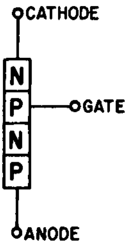
Fig. 619. Simple circuit is for an efficient sawtooth generator.

charge appearing across capacitor C1. The voltage across this capacitor is built up gradually as it is charged by the battery (B1) through the current limiting resistor R1. When the voltage across C1 is sufficient to overcome the internally developed reverse bias, the emitter-B1 junction switches rapidly from a non-conducting to a conducting state, quickly discharging the capacitor and dropping its voltage until the internally developed bias can once more assume control; afterwards, the emitter-B1 junction switches back to its non-conducting state, allowing C1 to charge once more. This action is repetitive as long as dc is supplied. The voltage across C1 gradually builds up to a maximum as it is charged through R1, then drops suddenly when discharged by the Unijunction, then builds up again, and so on, forming what is essentially a saw-tooth waveshape. For a Unijunction of fixed characteristics, and with a given supply voltage, the oscillator's repetition rate (frequency) depends on the R1-C1 time constant.

Controlled rectifiers

In a sense, the Controlled Rectifier is somewhat similar to the

Unijunction transistor, for both are three terminal semiconductor devices primarily suited to switching circuitry. The Controlled Rectifier, like the Unijunction, can be switched from a high resistance non-conducting state to a low resistance conducting state by the application of an external signal. As far as internal construction is concerned, on the other hand, the Controlled Rectifier is more like the Bistable diode we examined earlier. Referring to the symbolic drawing of the Controlled Rectifier's construction



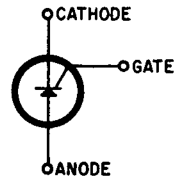


Fig. 620. Comparable to a 4-layer diode with an extra connection (gate) the semiconductor controlled rectifier has characteristics similar to a gas-triode or thyratron.

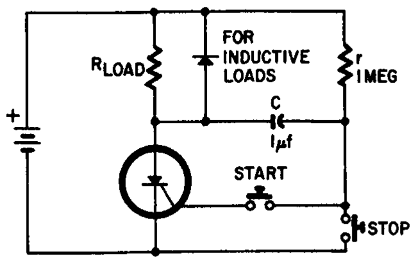
given in Fig. 620-a, we see that this unit, like the Bistable diode, is a multi-layered device made up of four alternate layers of p and n-type semiconductor materials, forming three p-n junctions. It differs from the Bistable diode in that an electrical connection is made to one of the internal layers, with this serving as the *gate* electrode. The Controlled Rectifier's schematic symbol is given in Fig. 620-b.

If we ignore the gate connection for a moment, the Controlled Rectifier's operation essentially duplicates that of the Bistable diode. It offers a high resistance to the flow of current when voltages are applied in *either* its nominal *forward* or *reverse* directions, up to a point. If the applied voltage in its forward direction exceeds a specific breakover voltage, the anode-cathode resistance drops suddenly to a very low value. From here on, the unit behaves much like a conventional rectifier, permitting a flow of current in its forward direction, but blocking it in its non-conducting direction (cathode positive). Thus, the unit behaves much like a standard rectifier, with but a small change—it conducts in its forward direction *only* when the applied voltage exceeds a specific value.

Now we can consider the function of the *gate*. A relatively small voltage applied to this electrode will cause the Controlled Rectifier to switch to its low resistance state, even when the applied forward voltage is *below* the unit's breakover voltage. In other words, the gate electrode *triggers* or *fires* the device, making it act like a conventional rectifier diode. It accomplishes this by injecting carriers into the central junction area. Once the Controlled Rectifier is switched to a conducting state, it remains in this state indefinitely after the removal of the gate signal as long as there is no interruption in its forward (anode) current. Once the anode current is interrupted, the unit switches back to its high resistance state until another triggering pulse is applied to the gate electrode (or until the applied voltage exceeds the breakover voltage rating).

The Controlled Rectifier generally is made of silicon, although other semiconductor materials can be employed. Similar devices are made by a number of manufacturers under their own trade

Fig. 621. Controlled rectifier circuit replaces relay in control applications. Start and stop push buttons can be located some distance from device being controlled.



names. It is called a *Controlled Rectifier,* or *Silicon Controlled Rectifier* (SCR), *Trigistor, Trinistor* and a *Thyrode* for instance. While these units have differences their applications are so similar they can be considered identical. Units, currently available, range in size from small types capable of handling a few hundred milliamperes to heavy duty units with ratings as high as 50 to 100 amperes, or more. Voltage ratings are as high as several hundred volts. The typical gate signal is a pulse of a few hundred milliamperes at from 5 to 10 volts.

Circuit applications for the Controlled Rectifier are many and varied. Generally, the units are used as heavy duty switches or as rectifiers in such equipment as battery charging regulators, temperature controllers, motor controllers, servo controls, and so on. A typical application is shown schematically in Fig. 621. Here, a Controlled Rectifier is used as a heavy duty switch. In operation, the current through the load device (which may be a motor, heating element, solenoid, or similar device) is limited by the Controlled Rectifier. Two control buttons are provided—a *Start* button which applies a signal to the gate electrode, triggering the Controlled Rectifier, and a *Stop* button which connects a charged capacitor (C) across the unit, impressing a reverse voltage on the Controlled Rectifier and switching it back to its high resistance state. During operation, the capacitor (C) is charged by the IR drop across the load, through a high-value resistor (R). A diode connected across the load protects the Controlled Rectifier from inductive switching spikes or pulses. A typical unit is shown in Fig. 622.

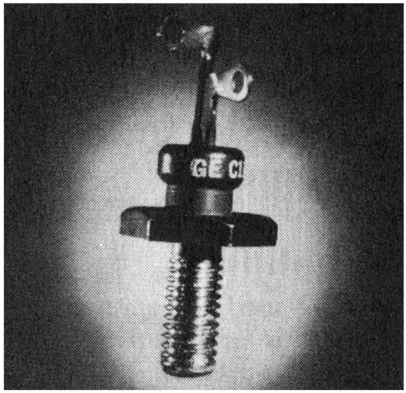
Spacistors

The Spacistor is a semiconductor device which retains many of the basic characteristics of the transistor, but which utilizes quite a different principle in its operation. The Spacistor is very small, requires relatively little power for operation, has no filament, and, theoretically, has an almost unlimited life. Its internal construction is shown pictorially in Fig. 623. Basically, the unit consists of a p-n junction, including the base (B) and collector (C) electrodes, to which two additional electrodes have been added (a pressure contact injector (I) and an alloyed contact *modulator* (M)). In practice, the injector and modulator serve as input terminals, while the base and collector are the device's output electrodes.

A basic Spacistor circuit is shown semi-schematically in Fig. 624. The p-n junction is biased in its non-conducting (high resistance) direction by a battery (B1), connected, through the output (LOAD), to the base and collector electrodes. A fairly strong electric field, or *space-charge* region (SC), is set up in the area of the junction. Still another battery (B2) biases the injector (I) negatively with respect to the space-charge region which, for practical purposes, has a voltage approaching that of B1's positive terminal. Although B2's positive terminal connects to I, B2 supplies a lower dc voltage than B1, hence the injector is less positive (or, which means the same thing, more negative) than the SC region. Thus, while the injector can emit a few current carriers (electrons) into the space-charge region, its emission, and hence its current, is limited by the space charge. The other electrode, the modulator (M), is connected to the SC region at a point intermediate between the injector and collector electrodes and, like the injector, is biased negatively with respect to the SC region by the battery B3. From this, we see that the voltage on M sets up a field which prevents the movement of holes from the P-type semiconductor to the SC region; since it (M) injects few or no current carriers into the SC region in itself, it draws little or no current, and presents a high impedance. Now, when an input signal is applied to the I and M electrodes, the net effect is to vary the injector's bias voltage up and down in accordance with the input signal variations, thus varying its current carrier emission into the SC region. This, in turn, controls the B-C current, developing an amplified signal across the output (LOAD) impedance.

As far as electrical characteristics are concerned, the Spacistor has very high input and output impedances and an exceptionally good high frequency response. The high input impedance results from the negative bias voltages applied to the I and M electrodes which keep current flow from these two electrodes into the SC region to

Fig. 622. The SCR is mounted on a flat metal surface by means of the threaded stud (thin washers may be used for insulation) to dissipate heat from the semiconductor material.



a minimum. The high output impedance results partially from the fact that the p-n base-collector junction is biased in its reverse (high resistance) direction, and partially because the modulator, located between the I and C electrodes, keeps hole current to a minimum. Finally, the unit's excellent high frequency characteristics result from the fact that the current carriers are moved rather rapidly by strong electric fields rather than allowed to diffuse slowly through the semiconductor crystal.

Tetrode transistors

By definition, a triode is an electrical device having three electrodes. Similarly, a tetrode is a device with four electrodes. The earliest tetrode transistors, like the original triodes, were point-contact units, and were made simply by adding another fine wire cat's-whisker, with this additional electrode serving as a second emitter. Thus, the point-contact tetrode had two emitters, one base, and one collector. When used in the common-base circuit configuration, the tetrode's operation was somewhat analogous to that of a vacuum tube having two control grids and a single plate, in that it had two inputs (emitters) with a common output (collector); the unit was useful in mixer and modulator circuits, but, like point-contact triodes, was never manufactured in large quantities and is no longer in production.

The low power junction tetrode is in production by at least three major manufacturers, while at least one firm has introduced a high power (multi-watt) tetrode. The construction of a low power tetrode is shown symbolically in Fig. 625-a, while the sche-

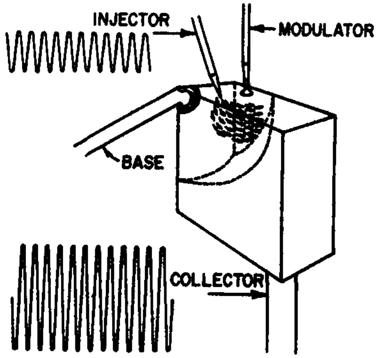


Fig. 623. A pictorial representation of the connections to the spacistor semiconductor.

matic symbol used to represent it is given in Fig. 625-b. The symbol shown is for an n-p-n tetrode, with the emitter arrowhead reversed for p-n-p types. The electrical specifications and maximum ratings of tetrode transistors are comparable to those of similar triode units, except for much improved high frequency response.

Basically, the junction tetrode is a triode transistor with two connections to the base electrode. It very easily could be called a double-base triode, for it has only the two junctions characteristic of triode types. The electrode connections, then, are the emitter (E), the conventional base (B1), the second base (B2), and the collector (C). In operation, the second base normally is biased with a current opposite to that applied to the conventional base. This bias sets up an electric field within the base area which repels the current carriers (holes or electrons) traveling from emitter to collector, and thus forces the major conduction to take place in the immediate vicinity of the B1 base lead. By keeping carrier diffusion through the base region to a minimum and, at the same time, limiting transistor action to an area immediately adjacent to B1's lead, the transistor's high frequency response is extended considerably.

Like the triode, the tetrode may be used in any of the three basic circuit configurations. A typical common-emitter tetrode amplifier is shown in Fig. 625-C. Referring to this diagram, conventional base bias is supplied base 1 by power supply battery B1 through current limiting resistor R1. The reverse bias necessary for tetrode operation is supplied to base 2 by battery B2 through its current limiting resistor R2, bypassed for ac by capacitor C2. Battery B3 supplies collector current and C1 and C3 serve as input and output dc blocking capacitors, respectively. The LOAD may be a resistor, coil, primary winding of a transformer, or similar impedance. The battery polarities shown are for an n-p-n tetrode; all dc polarities would be reversed for a p-n-p unit.

In operation, this single stage amplifier performs much like a conventional triode stage, except for its improved frequency re-

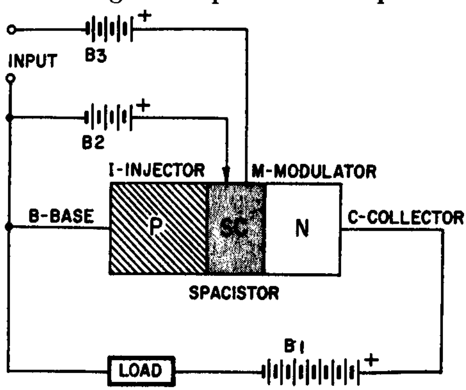
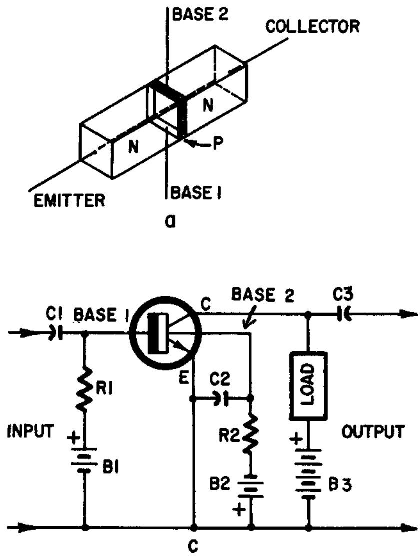


Fig. 624. The spacistor can be considered a tetrode transistor—it has four leads. The injector and modulator have different bias values applied.

sponse. Thus, the input signal is applied between the conventional base connection (1) and emitter, while the amplified output signal is developed across the LOAD and is obtained between the collector and emitter electrodes. The second base connection (2) is used for biasing only and does not enter into the signal circuits.

The tetrode need not be restricted to just one mode of operation, however. A variable bias can be applied to the second base for gain control purposes, if desired, or, if preferred, both base connections may be used as input terminals, giving a stage with two inputs and a common output. With the latter arrangement, the tetrode is an excellent mixer, converter, or modulator. Finally,



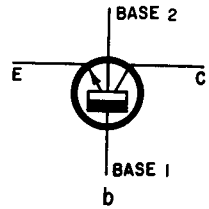


Fig. 625. A tetrode transistor semiconductor material representation (a) with its schematic symbol (b) and a basic circuit (c).

the tetrode may be used as a conventional triode simply by allowing one of the base connections to *float* (that is, not connecting it into the circuit). Different electrical characteristics are obtained, depending on which base connection (1 *or* 2) is allowed to float and which is used, of course.

CHAPTER 7

How to keep out of trouble or, the care and feeding of Mighty Mites

Transistors, like military tanks, are pretty rugged devices. Compared to a vacuum tube, for example, the average transistor can withstand from 30 to 40 times as much vibration and acceleration shock and has an operating life hundreds of times longer.

Mechanical damage can occur if transistor cases are crushed, whether in a vise, by being stepped on, or by over-squeezing with pliers. Connecting leads or pins can be broken off short or pulled out if the worker doesn't exercise reasonable care when installing or removing the units. Finally, internal damage can occur if transistors are overheated or used at voltage and current values above their maximum ratings. Unless you have a direct pipeline into a transistor manufacturing plant and can obtain all the transistors you want at no cost, you'll find it worthwhile to follow a few basic rules when working with these marvelous, but somewhat tender, devices.

AVOIDING HEAT DAMAGE

As we learned in earlier chapters, several of the junction forming techniques use controlled heat as a means for alloying semiconductor materials and for establishing impurity (donor and acceptor) diffusion through the basic crystalline structure. From this, it follows that junction temperatures which approach those used in the original manufacturing process may cause a permanent change in a transistor's electrical characteristics. In an alloyed junction unit, for example, a high temperature may cause the alloying or diffusion process to start again, with the emitter and collector areas expanding and, eventually, touching each other, thus shorting out the base region; if this happens, the device is ruined and is no longer a transistor.

Thus, transistors and transistorized equipment shouldn't be stored next to heat sources, near furnaces, radiators or ovens, for example, nor exposed to concentrated heat or light, such as spotlights or heat lamps, even for relatively short periods. Many a personal portable radio receiver has been damaged when the owner left it on the seat or dash of his closed automobile during the summer, with the hot sunshine streaming through the car windows or windshield and beating on the set's case. Most semiconductor manufacturers specify a *Maximum Storage Temperature* for their products.

As a general rule, silicon and gallium arsenide semiconductor devices can withstand higher temperatures without damage than

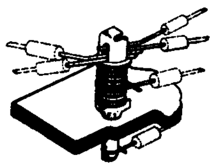


Fig. 701. Push-in, spring-loaded terminals need no soldering. Connections are easily changed. Such devices eliminate heat problems in experimental hook ups. (Vector Electronic Co.)

can germanium units. But even where heat damage is not permanent, a considerable, although temporary, change may occur in a semiconductor device's *operating* characteristics. A transistorized receiver or amplifier which gives excellent performance at average temperatures—say, 30°F. to 90°F.—may distort, lack sensitivity or power output, squeal or oscillate, or become completely inoperative at much higher (or lower) temperatures.

In experimental work, soldering and unsoldering of components (including transistors) can be avoided by using spring-loaded connectors, such as the one shown in Fig. 701.

Heat sinks

Transistors and related devices may be damaged by internally generated as well as by externally applied heat. Whenever electric current flows through a conductor, whether it be a transistor, lamp bulb, or a piece of copper wire, a portion of the electric power is converted into heat energy by molecular friction, raising the temperature of the conductor. The amount of electrical energy converted into heat depends on the resistance of the conductor and the current flowing through it. Mathematically, the power (in watts) converted into heat is equal to the square of the current (in amperes) multiplied by the resistance of the conductor (in ohms), or . . .

$$P \text{ (watts)} = I^2R.$$

The fact that electrical energy can be . . . and is . . . transformed into heat so readily is one reason we have fuses (or circuit breakers) in our buildings. Most house wiring, for example, is 12 or 14 gauge copper wire, with individual circuits fused for 20 or 15 amperes, respectively. If excessive current is drawn through the wiring, the wire temperature may rise high enough to damage its

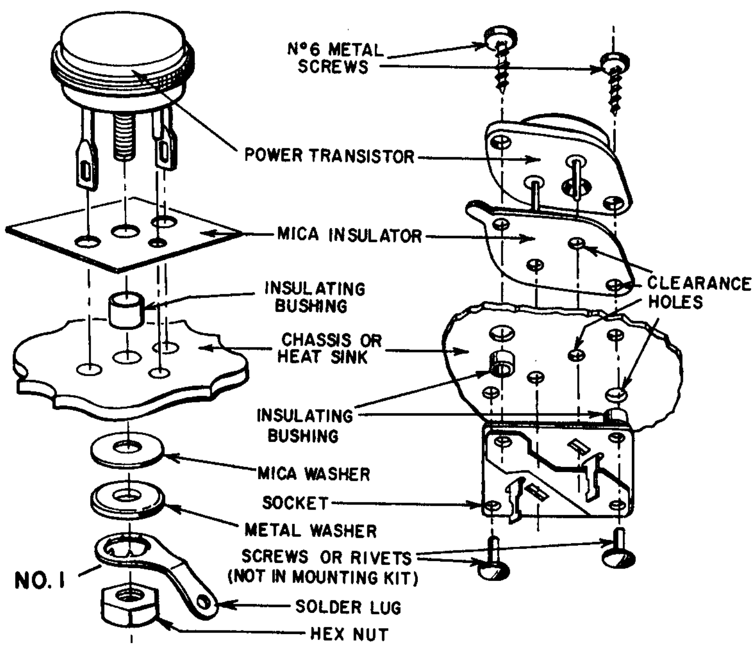


Fig. 702. Power transistor mounting must provide for heat dissipation. Mica insulators are coated with silicone grease for better conduction. (Motorola Semiconductor Prod. Inc.)

insulation, or even enough to set fire to wooden studs or other building materials. However, the fuses blow when excessive current is drawn, thus opening the circuit and preventing a fire.

When a transistor, diode or similar device handles a fair amount of electric power, the junction current is likely to be moderately large. This, in turn, can produce a good deal of heat, raising the junction temperature to dangerous levels. Generally, an *increase* in semiconductor junction temperature causes a *decrease* in resistance. Under some conditions, a self-sustaining reaction can develop which may cause the transistor to destroy itself. For example, junction temperature increases and its resistance drops, permitting greater current flow and thus a further increase in temperature, with a corresponding drop in resistance, and a still greater increase in current, and so on. Since the current and junction temperature increases quite rapidly once the action starts, such a condition is called *collector current runaway,* or, since junction temperature plays a prominent part in the action, *thermal runaway*. Unless special steps are taken to limit circuit current or junction temperature to within safe values, the device is soon destroyed.

One effective way to get rid of unwanted heat, thus preventing an excessive rise in temperatures, is to absorb the heat with a good heat conductor (such as a mass of metal) and then to dissipate it into the surrounding environment through radiation and convection. In medium and high power (multi-watt) transistors, this is accomplished by using several techniques. First, the transistor's collector electrode is connected both electrically and physically to the unit's metal case. This permits a rapid transfer of heat energy from the junction to outside the transistor. Second, the transistor is attached to a larger mass of metal (Fig. 702) which absorbs heat from the transistor's case and dissipates it into the air. Since the metal "gets rid" of the unwanted heat, keeping the transistor within safe temperature values, it is called, quite appropriately, a *heat sink*. An additional technique, used where relatively large amounts of heat must be dissipated, is to provide external cooling, either by means of a fan circulating air across the heat sink or by using solid state thermoelectric elements, connected with reverse polarity to obtain cooling.*

In practice, a transistor's heat sink may be a permanent part of the device—more or less an extension of its regular metal case—or a separate mass of metal to which the transistor is attached. Typical built-in and external heat sinks are illustrated in Figs. 703 and 704, respectively. Note that radiating fins are provided to improve heat dissipation where higher powers are handled. Even where a built-in heat sink is used, however, most equipment designers will mount the transistor on a metal chassis or plate.

Temporary heat sinks are used to prevent damage from ex-

** NOTE: Some p-n junctions, properly mounted, will convert electric power into heat when current is passed through them with one polarity, but will absorb heat when the current is reversed. These thermoelectric elements may be used for both heating and cooling, and may be assembled in groups or batteries to improve their effectiveness. All-electronic refrigerators and air-conditioners have been built with these units.*

ternally applied heat when transistors are wired permanently into circuits (or are removed for tests). Here the heat sink is clamped to the lead being soldered at a point *between* the transistor's case and the point at which the soldering iron is applied, thus serving to drain off heat before it reaches the semiconductor junction.

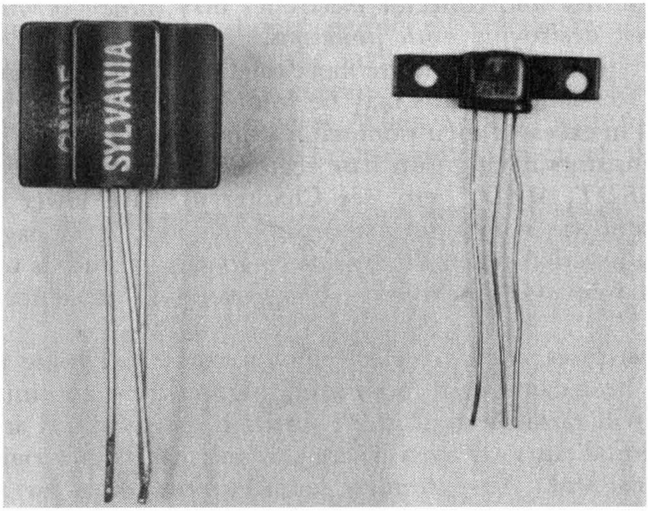


Fig. 703. For low-power purposes it is possible to make heat fins part of the transistor case (left). Sometimes mounting tabs are solidly attached to the case (right) for attaching to a heat dissipating mass.

AVOIDING ELECTRICAL DAMAGE

Most of the world's great philosophers recommend a policy of moderation in all things or, as the "wet" said to the prohibitionist, "even too much water is bad for you—you can drown." The philosophy of moderation is as important when handling transistors as it is in dealing with life's temptations. Too *much* voltage or current can ruin a transistor in short order, even if the unit's power dissipation or external temperature limits are not exceeded. The *Maximum Ratings* given by semiconductor manufacturers are for the user's, not the manufacturer's, benefit. They are not pretty numbers used simply to fill up a specification sheet nor almost meaningless figures to impress potential customers, like some auto engine horsepower ratings. They should be taken seriously.

Excessively high voltages applied to a transistor's electrodes may cause various types of damage. Internal arcing may take place, either between electrode pins or between insulated elements and the transistor's outer metal case. If the transistor is potted or encapsulated with a plastic compound, dielectric breakdown may occur, with internal leakage paths established. Under some conditions, a high voltage may completely destroy a semiconductor junction. For example, a high voltage applied between a transistor's emitter and collector electrodes may *punch through* the base layer, destroying *both* junctions.

Some types of transistors are more tolerant of high voltages than others. Occasionally, units may be found which can handle voltages well in excess of their nominal maximum ratings. As a general rule, transistors having thin base regions, such as *SB* and related types (*SBDT, MADT,* etc., see Chapter 6), have lower voltage ratings and are *much more* susceptible to voltage damage than are alloy junction units. In any case, a good basic rule is to *never exceed a transistor's maximum voltage ratings,* even if for an instant.

In most cases, excessive electrode currents cause some type of internal heat damage. A moderately high current continued for a while will raise the transistor's junction temperatures and may start thermal runaway even where the unit has an otherwise adequate heat sink. An extremely large current, even for a short period, may melt the internal electrode leads or fuse the junctions, destroying the base region and shorting the transistor. This may occur even where the high current level is of such short duration that the transistor's external (case) temperature does not rise much above normal values. Where transistors are used at relatively high environmental (ambient) temperatures, it may be necessary to *derate* the units, limiting electrode currents and power dissipation to values appreciably below their normal maximum ratings.

Electrically, the chief difference between *power* (multi-watt) transistors and common *small signal* types lies in their maximum current ratings. A typical small transistor may have a maximum collector (or emitter) current rating of from 20 to 100 ma. A power transistor, on the other hand, with similar voltage and gain (beta) specifications, may have a current rating of from 1 to 10 *amperes,* or more. But whether a transistor's maximum current rating is given in milliamperes or amps., your best bet, to avoid damage, is to stay well within the specified limits.

A standard Volt-Ohm-Milliammeter, VOM, or separate meters with suitable ranges should be available whenever you are working with transistors, whether testing experimental circuits, assembling amplifiers or receivers, adjusting control equipment, or making repairs in factory built gear. The instrument is used to check interelectrode voltages and, where necessary, circuit currents, with the measured values compared to the transistor's specifications as insurance that maximum ratings are not exceeded. Circuit polarity must be taken into account, of course, for p-n-p and n-p-n types require different dc polarities. As a general rule, the instrument chosen should permit dc voltage measurements from under one volt up to the maximum voltages encountered in your circuits, and current measurements from less than 1.0 ma. to several hundred milliamperes. When working with multi-watt transistors, an ammeter with a maximum range of up to 10 amperes is quite useful.

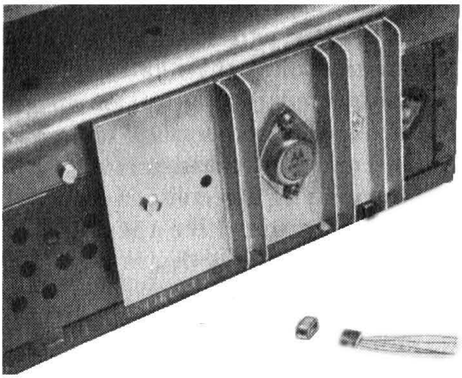


Fig. 704. Large cast or extruded aluminum heat sinks are used to help dissipate the heat by convection. If the heat sink is insulated from the chassis it is not necessary to use other insulators (Fig. 702). *The aluminum can be anodized to form a very thin insulating coating.*

Occasionally, transistors will be damaged in experimental circuits even where the dc supply voltage is well under maximum ratings, where currents are relatively small, and where temperatures are normal. In such cases, the damage may be caused by a *transient* (momentary) high voltage spike or quite possibly by a current surge when the test set up is turned on and off. A sudden change in the current through a large inductance, such as the primary of an unloaded iron-core transformer or a heavy filter choke, may develop voltages twenty times higher than the dc supply voltage. A voltage spike can punch through a transistor's base region in a flash, destroying the unit. Current surges, as may result when a discharged high value capacitor is connected to a voltage source (the capacitor acts as a momentary short circuit), can cause damage too, but are not as common a source of trouble as are voltage transients.

Voltage spikes, where encountered, may be minimized by connecting a load resistor across the offending component (choke, transformer, etc.), or, in extreme cases, by the use of a *damping* diode. Current surges can be eliminated by adding a small resistor in series with the circuit. Proper resistor values (or diode types) may be determined experimentally to suit individual circuit conditions.

TESTING TRANSISTORS

Whether or not transistors can *go bad* is a highly debatable point. A semiconductor engineer will state, with conviction, that good quality transistors, used within their ratings in properly designed circuits, and not subjected to physical or electrical abuse, will last indefinitely. An experienced service technician charged with the repair and maintenance of transistorized equipment, on the other hand, will state with equal conviction that transistors can and do become defective in use. Like the five blind men of Hindustan groping for the elephant, both are right, yet partially wrong. Let's clear up the smog.

A good quality transistor without latent defects, used well within its maximum ratings in a well engineered circuit, *will* last indefinitely, provided it is not subjected to physical or electrical abuse, either deliberately *or accidentally.*

On the other hand, top quality transistors are not used in all commercially manufactured equipment. Transistors may have latent defects, impurities introduced during manufacture or assembly which gradually poison the unit by establishing internal leakage paths or by causing slow changes in the semiconductor characteristics. Designers, in an effort at economy, may use transistors at close to (or slightly over) their maximum ratings as is often done with tubes. Electrical breakdowns in other components —a shorted capacitor, open resistor, or failure of insulation, for example—may subject the transistor to electrical overloads. And accidents are always possible. An owner may connect the wrong type of battery to the equipment, or may connect a battery with reverse polarity. Finally, electronic equipment may be subjected to physical abuse—may be dropped, left out in the rain, kept near a source of high temperature, or used in a corrosive atmosphere. As a result, *transistors do become defective in use.*

Transistor defects

By far the most common transistor defects are: (a) an increase in leakage between electrodes, with actual shorts developing in extreme cases, (b) an open between electrodes, often caused by severe overloads which melt internal connection leads, and (c) a severe drop in gain (beta) as a result of latent defects or operation in excess of ratings. Less common defects are: (1) an *increase* in gain (which may cause critical circuits to be thrown into oscillation), (2) an increase in internal noise, (3) a drop in the unit's high frequency response (rendering it unfit for use in rf circuits), and (4) a severe change in operating characteristics (the transistor may still be useable, but may require different operating bias currents).

These defects can cause various operational complaints, depending on how the defective transistor(s) is used in the circuit and the nature of the equipment. In a receiver or audio amplifier, a leaky transistor may cause distortion (garbled sound), low power output, and short battery life. A shorted transistor will usually "kill" operation (dead set) and may cause a resistor or other component to burn open. Sometimes, the same basic operational complaint may be caused by any of several transistor defects; weak operation, for example, may be caused by a leaky or open transistor, by a low gain unit, or by a severe change in a transistor's operating characteristics and bias requirements.

Transistor testers

Commercial transistor testers may be divided into three groups: (a) inexpensive checkers capable of testing for shorts and opens and giving approximations of leakage and overall gain (dc beta) with fixed bias, (b) medium priced units providing more accurate tests of leakage and gain, and permitting tests at varying bias levels, and (c) expensive laboratory instruments allowing a detailed analysis of all important transistor parameters under a variety of operating conditions.

As a general rule, the less expensive transistor checkers are quite easy to operate and provide a GOOD-?-BAD type of reading, although some units give a numerical indication of gain values. The expensive laboratory instruments, on the other hand, are more difficult to use and may require the operator to perform a series of tests, interpreting the results on a comparative basis. The extra effort required for operation pays dividends, however, for the results obtained permit a more detailed analysis of transistor characteristics and expected performance.

Aside from greater versatility, a wider range of test conditions, and improved accuracy, there is another important difference between lower cost transistor checkers and laboratory grade analyzers. The majority of inexpensive checkers test transistors under *static* conditions, that is, with fixed dc bias currents or signals. The

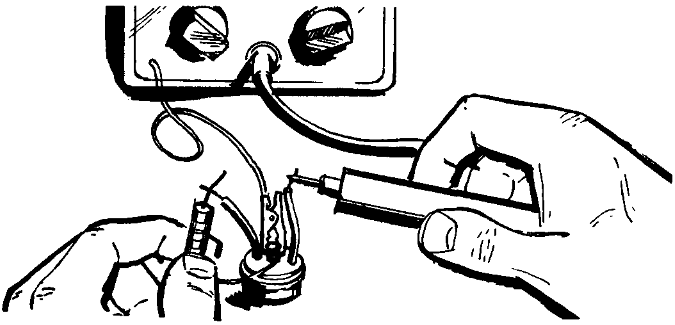


Fig. 705. Power transistor being tested with an ohmmeter. Hand-held resistor limits transistor current to safe value.

laboratory type analyzers, on the other hand, generally test transistors under *dynamic* conditions with adjustable bias currents and known ac signals. Dynamic tests more nearly duplicate the transistors' actual operating conditions in a circuit and hence are much more valuable.

In addition to general checks of overall quality and electrical characteristics, transistor testers are used quite extensively to select individual units with specific parameters. Transistors are manufactured to rather broad tolerances, and even units of the same type number may have different gain and leakage values. Where transistors are needed with specific characteristics, or where closely matched pairs are needed for critical push-pull circuits, the professional tester provides the user with means of accomplishing the job without resorting to trial and error methods.

Ohmmeter tests

If a transistor used in a piece of electronic equipment is thought to be defective, but a commercial tester is not available, a *substitution* technique may be employed. Here, the suspected unit is replaced in the circuit with a new transistor of the same type. Unfortunately, the substitution test, while an excellent one, requires a rather large (and expensive) stock of replacement transistors. As an alternative method, a transistor can be given a simple, but usually adequate, check with a standard VOM. The instrument's Ohmmeter ranges are used to measure dc resistances between the transistor's various electrodes, and these values are compared with the *expected* results based on our general knowledge of transistor characteristics.

Before discussing actual test techniques, however, let's review a few basic facts about the transistor's internal construction. In discussing transistor operation in Chapter 3, we found that a triode transistor, in one sense, may be looked on as two p-n semiconductor junctions connected internally back-to-back and sharing a common element (the base). A simple p-n junction, of course, is a diode and, if in good condition, should have a high resistance when biased in its reverse or non-conducting direction and a low resistance when biased in its forward or conducting direction. Going a step further, it is easy to conclude that the two diode junctions making up the transistor can be checked independently simply by applying our test leads to the proper pair of electrodes: the base-emitter junction is checked by measuring between the base and emitter leads. See Fig. 705.

A standard Ohmmeter has an internal source of dc voltage (its battery) which may be used to bias a diode junction in *either* its forward or inverse direction merely by reversing the test leads.

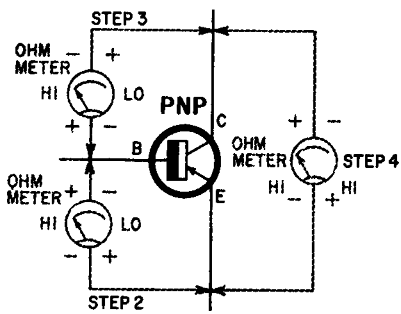


Fig. 706. Ohmmeter readings for p-n-p transistor tests.

With a suspected transistor removed from its circuit, then, the test technique is a simple one.

STEP 1. Choose a *series type* Ohmmeter with an internal battery supplying 4.5 volts or less. Select a range permitting readings up to 500,000 ohms. Short the leads together and *zero* the instrument. If in doubt about the type of meter, connect a 1,500 ohm, ½ watt resistor in series with one of the test leads and deduct its value from subsequent readings. This resistor limits the Ohmmeter current through the transistor to a safe value.

STEP 2. To check the base-emitter junction, connect the instrument's leads to the base and emitter pins. Note the resistance reading obtained and *reverse* the leads. One of the two measurements should be from 5 to 100 times higher than the other.

STEP 3. To check the base-collector junction, repeat Step 2, but connect the Ohmmeter leads to the transistor's base and collector pins. Again, be sure to reverse the Ohmmeter leads to obtain two measurements.

STEP 4. For a final check, connect the test leads to the emitter and collector leads and repeat the two measurements.

If *identical* resistance readings are obtained when the Ohmmeter leads are reversed in either Step 2 or 3, the junction checked is de-

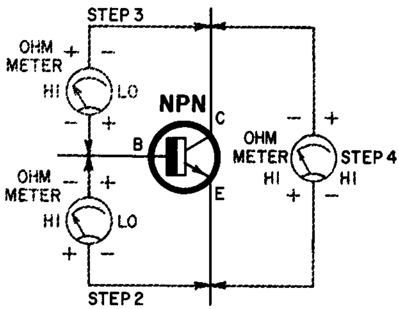


Fig. 707. Ohmmeter readings for n-p-n transistor tests.

fective. If the two readings are both low, the junction is either leaky or shorted. If both readings are very high, the junction is partially open. In either of these cases, the transistor is defective and should be discarded. As a general rule, the higher the ratio between forward and reverse resistances, the better the diode junction. A good quality junction may have a forward to back ratio as high as 50 to 100 to one.

The final test, Step 4, is used to check the two diode junctions in series. Both resistance measurements (forward and inverse) should be moderately high but, in some cases, one reading may be appreciably higher than the other. The ratio will not be nearly as high as that obtained when the junctions are checked individually, however. If the two resistance values are the same, or one is larger than the other, the transistor is probably in good condition, provided satisfactory results were obtained in Steps 2 and 3. If the resistances are quite low, however, the transistor is leaky, and there may be an internal short between the emitter and collector electrodes. In this case, the transistor is defective.

Actual resistance values measured will vary considerably from one transistor type to another. Small signal transistors generally have forward resistances of a few hundred ohms, or less, and inverse resistance running into the tens of thousands of ohms. Power (multi-watt) transistors have much lower forward and inverse resistance values, but the *ratio* between the two readings is about the same as those of smaller units.

The basic steps involved in carrying out the Ohmmeter test are summarized schematically in Fig. 706. Note that whether a high (HI) or low (LO) resistance measurement is obtained . . . assuming the transistor to be in good condition . . . depends on the dc polarity of the test leads, the junction checked, and whether a p-n-p or n-p-n transistor is tested. Relative readings for p-n-p transistors are indicated in Fig. 706, for n-p-n units in Fig. 707.

An important point before closing this Chapter—the Ohmmeter test simply indicates whether or not the transistor is an operable unit—that is, whether it is excessively leaky, shorted, or open. This test does not indicate the transistor's gain, its power handling capacity, its noise level, nor its frequency response. Therefore, it is not a complete substitute for more sophisticated tests using commercial transistor testers. But it is a good test to know and to be able to use.

FUTURE DEVICES

At this point, you may have started to wonder about the length of this book . . . about the number of transistor and semiconductor devices which are being manufactured, and about how many more devices will be described before we can sum up with . . . *and that covers all the semiconductor devices which you will ever encounter.* Fortunately (or unfortunately, depending on your viewpoint), we can not make such a statement. Even as this is being written, even as you, the reader, are scanning the printed word, new semiconductor devices are being developed, new units have passed the development stage and are being manufactured, and theoretical physicists are exploring still newer paths of semiconductor device design. The general field of semiconductors is a vibrant, growing, dynamic field. In a sense, the surface has been barely scratched as far as overall potential is concerned. The future will see new techniques, new devices, and improved versions of existing devices.

Along with improvements and refinements in manufacturing technique will come the development of special purpose semiconductor devices, units which are not new in the sense of the Tunnel diode or Spacistor, but which represent straight-forward adaptations and modifications of existing designs. For example, several firms are planning (or are now producing) junction transistors in which the emitter and collector electrodes have identical characteristics, permitting them to be used interchangeably in circuit design. As an auxiliary feature, these units can be used to design *bi-directional* amplifiers when used in the common-base configuration. In other cases, manufacturers are adding additional electrodes to simplify circuit design or to permit the readjustment of internal bias values for optimum operation. Along these lines, one firm has introduced a *double-emitter* transistor; such a unit is especially valuable for use in mixer and converter circuits.

Still another path of development involves combinations of known devices into a single unit, for example, producing complementary transistors, in which p-n-p and n-p-n types are combined and direct-coupled in a single case, resulting in a high-gain transistor. Several transistors have been combined into a single package, developing a unit analogous to multifunction vacuum tubes. A *composite transistor* (a multi-element device) has been developed that has the external characteristics of a single transistor, with emitter, base, and collector electrodes, but which, internally, is equivalent to a multistage, thermally compensated direct-coupled amplifier. The result is a transistor-like device with a beta of up to 30,000 or more, as compared to betas of, at the most, a few hundred for more familiar transistors.

As to the future? Your own imagination is probably as good a guide as any. Given a need, a technical requirement, or a goal and there's a good chance that scientists and engineers will deliver the goods!

PRINTED IN THE UNITED STATES OF AMERICA