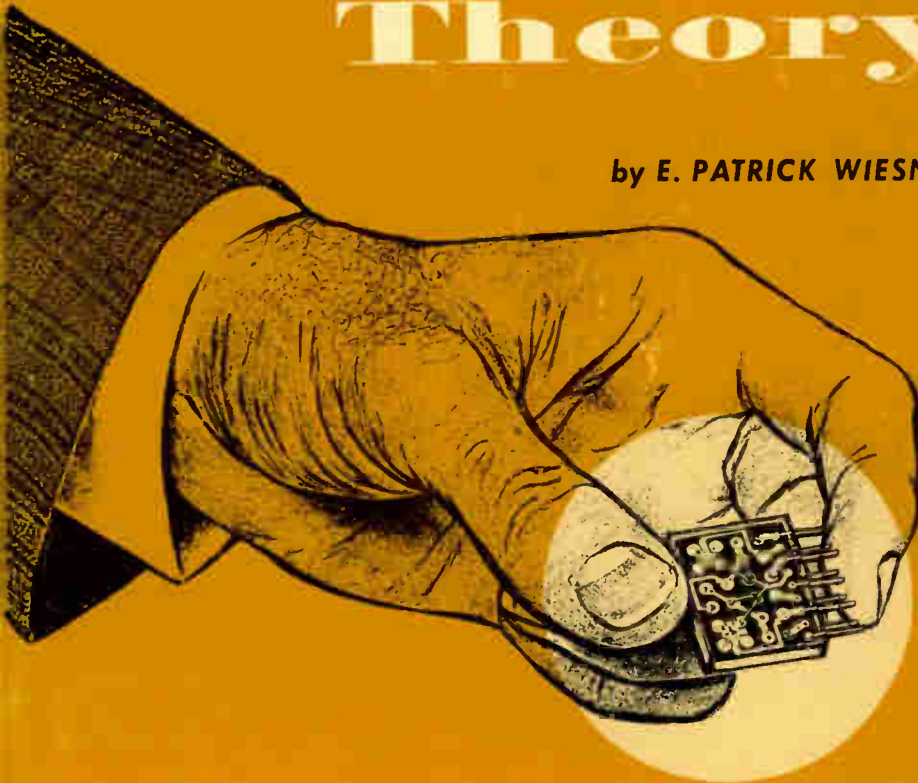
A Howard W. Sams PHOTOFACT PUBLICATION • PTW-1

Practical Transistor Theory

by E. PATRICK WIESNER



An easily understood coverage of transistor devices, how they're used, and how they work.

$2.50
Cat. No. PTW-1

Practical Transistor Theory

by E. PATRICK WIESNER

HOWARD W. SAMS & CO., INC.
THE BOBBS-MERRILL COMPANY, INC.
Indianapolis • New York

FIRST EDITION
FIRST PRINTING—JANUARY, 1964

PRACTICAL TRANSISTOR THEORY

Library of Congress Catalog Card Number: 64-16363

COVER CREDIT: The unit shown on the front cover is a pellet microcircuit supplied by P. R. Mallory & Co., Inc. Circuit elements consist of small pellets 1/10 inch in diameter by 1/16 inch thick. The pellets are inserted into a prepunched phenolic board and soldered to the printed circuit. The two units in the center are transistors.

Preface

The applications for electronics are expanding at an unprecedented rate. Pacing this explosive growth is the phenomenal development of transistor and semiconductor devices. One cannot be exposed to the field of modern electronics, even as a nontechnical participant, without being confronted with transistors and semiconductors.

This book is unique in that it attempts to cover a highly technical subject at a level which can be understood by both nontechnical persons and beginners in electronics, yet in a manner which will hold the interest of engineers and technicians.

The secret to understanding semiconductor devices is contained in the distinctions between active and passive, and linear and nonlinear types. The first chapter clearly makes these distinctions and sets the stage for discussion of the individual devices. This division is so natural that not only do difficult topics become easy to understand, but the flow of ideas moves smoothly from the discussion of linear amplifiers and radio circuits to nonlinear switching circuits and digital computer logic. This approach enables the presentation of topics not normally treated at this level or in a book of this size.

The chapter on basic semiconductor theory will give the layman a clear insight into the operation of all semiconductors, even the most complicated. The emphasis is on characteristics, and simple physics is introduced only to clarify certain points.

A detailed description of basic transistor amplifiers leads into discussion of practical radio and hi-fi circuits. Chapter 5 then deals with switching transistors, multivibrators, flip-flops, and gates. It also includes a discussion of digital computer logic design. As the various devices are introduced, their fundamental function, construction, and relationship to the others in the semiconductor family are explained. A number of applications are suggested, and several are illustrated and explained.

The chapter on integrated circuits provides information that will be useful and interesting to beginners, technicians, and engineers alike. It includes the latest information on the fundamentals of processing methods and the application of integrated circuits.

Each chapter in this book has been designed to stand alone; however, each complements the other. It is hoped that after reading this book you will know why semiconductors belong to the first family in the fastest growing industry in the world and that you will have made a lasting acquaintance with each of its members.

E. PATRICK WIESNER

December, 1963

Contents

CHAPTER 1

Basic Semiconductor Action and Materials

In order to truly grasp the operation of an automobile, one must eventually take one apart and put it back together again. Likewise, to attain a genuine insight into the workings of semiconductor devices you must take them apart and put them back together again using the "tools" of basic solid-state physics. We will disassemble the semiconductor down to its fundamental molecular structure so that you can observe and understand its action.

Electric current flows whenever electrons or other current carriers (it will be pointed out that there is another kind) move. Current carriers can be moved by applying a pressure (voltage).

Transistors, diodes, and other semiconductors are nothing more than small "chunks" of solid material in which the flow of electric current can be controlled. This ability to control current is what makes the devices so very useful. In vacuum tubes, current flows through a vacuum, while in semiconductors, current flows through a solid. Thus semiconductors are often referred to as "solid-state" devices, and the key to understanding their operation lies in basic solid-state physics.

One of the basic problems is to become familiar with new terminology. However, new words and new ideas can add to the excitement of discovery if each one is fully understood before going on to the next. When you complete this chapter, terms like majority carriers, reverse bias, junction, N-type or P-type semiconductor, and doping will all be familiar to you. A summation composed of a series of short and simple statements completes this chapter. Once these fundamentals are firmly established, the peculiarities of different devices can be easily understood.

CONDUCTORS, SEMICONDUCTORS, AND INSULATORS

The first step is to define a semiconductor. A good start would be to say that a semiconductor is a material whose characteristics lie somewhere between those of a conductor and of an insulator. This puts us in the right "ballpark," but it is not sufficient for a definition. By our being more specific about conductors and insulators the definition becomes clearer. The difference between a conductor and an insulator is in the atomic structure. A conducting material has a large number of electrons (current carriers) ready to participate in current flow when the slightest pressure (voltage) is applied. The reason that copper is such a good conducting medium is that there are numbers of free electrons ready and willing to act as current carriers. This is another way of saying that the resistance (to current) of copper is low. Later you will see that this is because the electrons are loosely bound to the nucleus of the copper atom.

A good insulator, on the other hand, is a material in which the electrons are held very tightly by the nucleus. They are so tightly held that a voltage is required to tear them away from the nucleus to make current carriers available for conduction. The resistance of an insulator is very high. When enough voltage is applied to an insulator to cause it to conduct (that is, when electrons are literally ripped away from the nucleus), the reaction is often so violent that the material is scorched or even destroyed.

Now if we agree that the characteristics of a semiconductor lie somewhere between those of a conductor and those of an insulator, you would then probably guess that the electrons are held firmly by the nucleus in a semiconductor—not too strong, not too loose. In other words, electrons don't immediately fall into conduction at the application of the slightest voltage as in a conductor; nor do they hold on so tenaciously to the point that you have to ruin the material in order to cause conduction, as is the case in the insulator.

A conducting material has many electrons available for current conduction. An insulator has no or few current carriers available for conduction. In a semiconductor material the number of electrons, or current carriers, available for conduction is few, but it can be easily increased by the application of a nominal voltage. Current carriers in a semiconductor material have not yet fallen into the flow, but by the application of a little pressure (voltage) they can be made to do so. In a conductor there are many electrons that are immediately available. In an insulator the current carriers are so tightly bound to the nucleus of the atom that they are not available for conduction.

The previous paragraph is very important and holds the key to understanding semiconductor theory. The electrical resistance of a device is determined by how easily current will flow. In a semiconductor device the resistance can be changed at will by merely varying the number of available current carriers. A semiconductor device is an easily controlled variable resistance.

With this basic idea in mind, take a closer look at how the number of current carriers available for conduction and thus the resistance of a semiconductor can be varied. Since this subject is probably quite new to you, it is necessary to go back to ground zero, make some basic definitions, and become familiar with the simple atomic structure of

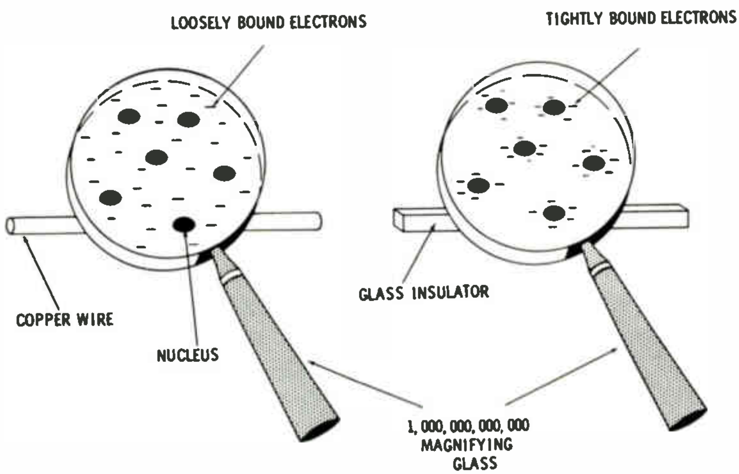


Fig. 1-1. A magnified conductor and insulator.

a semiconductor crystal. We will do this by first noting how a semiconductor differs from a conductor and an insulator.

Good electrical conductors are copper, silver, gold, and aluminum. The atoms of these elements have a large number of electrons that are loosely bound to the nucleus and easily become a part of the electric current. This situation is pictorially represented by Fig. 1-1A. Each black circle represents the nucleus of a copper atom, and each dash represents an electron tied to the nucleus. A magnification of 1,000,000,000,000 would be necessary to actually see this. It is easy to make current flow in a conductor because the electrons or current carriers are readily available and loosely bound to the nucleus. In some materials they are more loosely bound than in others. Thus some materials are better conductors than others.

Fig. 1-1B shows a pictorial representation of an insulator such as glass, in which the electrons are tightly bound to the nucleus and a large voltage must be applied in order to produce electrical conduction. This voltage usually is so large that when the electrons are finally ripped away, the material is physically damaged.

GERMANIUM AND SILICON

There are two common semiconductor materials—germanium and silicon. These materials are crystalline, as found in natural form. This is just another way of saying that when either germanium or silicon atoms unite to form a block of material, they do so by joining hands, so to speak, and share electrons to form a simple, symmetrical crystal. The makeup of the germanium or silicon crystal is not nearly as complicated as the structure of a snowflake. As a matter of fact, the crystalline form of germanium or silicon is one of the simplest crystals. The formation is termed a crystal lattice.

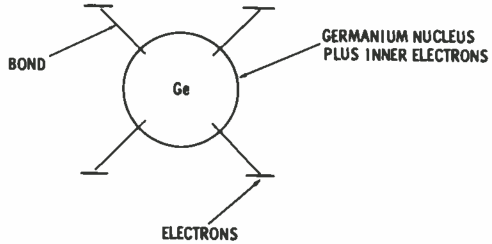


Fig. 1-2. Germanium atom.

First, look at a germanium atom as shown in Fig. 1-2. This atom has four electrons in the outer orbit that are loosely bound to the nucleus. For convenience' sake, the nucleus and the electrons that are tightly bound to it are contained in the circle marked Ge. A silicon atom could be represented in exactly the same way, except the Ge would be replaced with Si.

A number of these atoms combined form a crystal lattice, as shown in Fig. 1-3. Each atom contributes four electrons to the crystal structure and also shares those from the four adjacent atoms. Notice how this sharing results in a neat crystal structure. A silicon semiconductor material would look exactly like this except that the nucleus would be silicon rather than germanium.

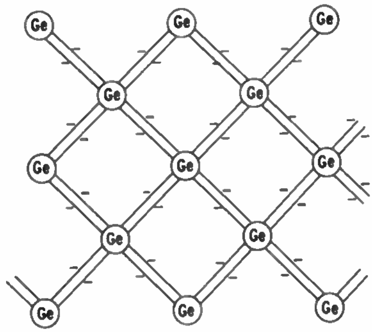


Fig. 1-3. A crystal lattice.

It is possible to produce conduction in this semiconductor material by applying a voltage that causes electrons to break away from the atoms and enter into conduction. This effect does not require high voltage as in the case of an insulator, but neither does the material act like a good conductor. It would be somewhere between an insulator and a conductor—a semiconductor. However, remember that the desired result is to be able to control the amount of conduction, and continue to investigate the properties of the crystal.

N-TYPE SEMICONDUCTOR

One of the principles that gives real importance to semiconductors is to introduce impurity atoms into the crystal lattice to obtain additional current carriers. This process of introducing impurity atoms into the basic crystal is called *doping.*

There are a number of doping materials, such as arsenic, phosphorus, and bismuth, that can be mixed with germanium. The atoms join into the crystal structure as though they were germanium. In other words, the germanium atoms have a "stranger" in their midst. The only difference is that the stranger has five electrons to contribute to the crystal, whereas germanium has only four. Since only four electrons are required to complete the crystal lattice there is one electron that is not bound to the crystal structure. The material having five electrons

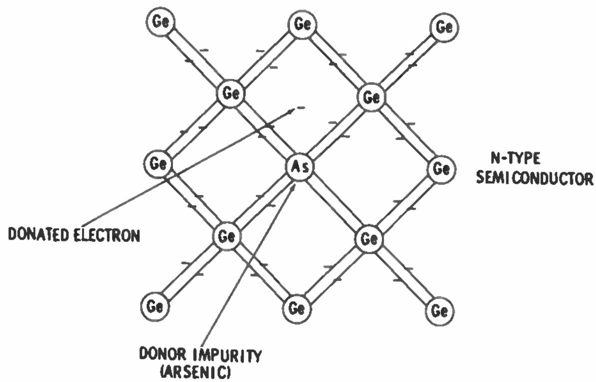


Fig. 1-4. Donor impurity.

is called an impurity. In this case, the impurity atom has donated an electron to the crystal structure and is therefore called a doNor atom (Fig. 1-4).

Semiconductor materials that contain doNor impurities are called N-type semiconductors. You can remember this by the fact that doNor has an N in it or that the electrons that were contributed have a negative (N) polarity. The number of extra electrons or current carriers controls the resistance of a semiconductor material. Obviously, the more heavily doped materials contain more donor atoms, more extra electrons for conduction, and therefore have a lower electrical resistance.

P-TYPE SEMICONDUCTOR

You're way ahead if you're wondering about the possibility of atoms having three electrons that can be added to the semiconductor crystal.

Atoms with such properties are indium, gallium, and tellurium. Indium joins with germanium as shown in Fig. 1-5. Note that insertion of an impurity atom containing only three electrons leaves a hole in the crystal structure at a point where an electron is normally located. That

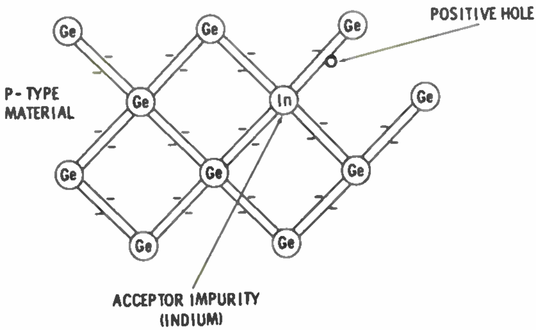


Fig. 1-5. Acceptor impurity.

is exactly what this lack of electron is called, a hole. An impurity atom that has only three electrons to contribute to the crystal structure is called an accePtor impurity because it is ready to accept an electron. The semiconductor material that is produced is called a P-type semiconductor. You can remember this fact by using accePtor, or remember that the impurity produces positive (P) holes.

In other words, N-type semiconductor material has extra electrons donated by the impurity atom, while P-type semiconductor material has excess holes contributed by acceptor impurities. The number of excess holes or electrons determines the electrical resistance of the material.

Conduction in Semiconductors

Electrons can be made to move through a crystal of N-type semiconductor by applying a voltage that will exert a force on the electron. Fig. 1-6 shows that electrons move through the crystal from the negative terminal to the positive terminal of the voltage source.

Holes can also travel through a P-type semiconductor crystal (Fig. 1-7). When an electron in a crystal bond has enough force exerted on it to break the bond, it will jump into a hole. The hole has, in effect, moved from where it was originally to the place where the electron

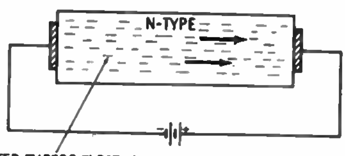


Fig. 1-6. Direction of electron flow.

jumped out of the bond. Therefore it is easy to see that holes and electrons travel in opposite electrical directions through their respective materials. By applying a voltage to N-type material, we have electron flow, and by applying voltage to P-type material, we have hole flow. At least it can be thought of as hole flow. You may be saying to yourself that it is actually electron movement, but it will be easier to understand semiconductor theory if you accept the proposition that the current carriers in N-type material are electrons and the current carriers in P-type material are holes.

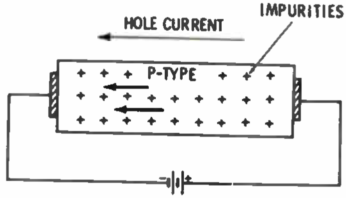


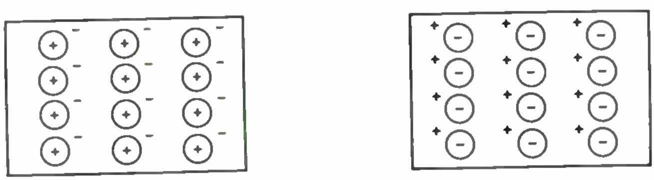
Fig. 1-7. Direction of hole flow.

At this point we should review the fact that this entire situation can be duplicated by using silicon or possibly other materials as basic semiconductor elements. Also, it is impossible to have a perfectly pure N- or P-type semiconductor. An N- or P-type semiconductor contains doNor or accePtor impurities purposely placed there. We must make a distinction between majority carriers and minority carriers. The majority current carriers are the electrons or holes that are free to move in the crystal structure, and the minority current carriers are the electrons or holes that are bound by the crystal structure and contribute to electric flow only under certain conditions.

In N-type material, electrons are the majority carriers and holes are the minority cariers. In P-type material, holes are the majority carriers and electrons are the minority carriers.

Let's recap our position by reviewing the material up to this point. Pure semiconductor materials are doped with impurities to provide a given number of current carriers. The doping produces two kinds of crystals (N and P) and two kinds of current carriers (electrons and holes). Electrons are the majority carriers in N-type semiconductors, and holes are the majority carriers for the P-type semiconductors.

Now that we understand how the semiconductor crystal is constructed and how the impurities provide electrons and holes as current carriers, let us change Figs. 1-4 and 1-5 and thus simplify the representation of a semiconductor crystal. Since the impurity atom contributes an extra electron to the crystal, we show this in conjunction with a positive charge to make up the rest of the structure. Together, these two are electrically neutral. If, in the case of a donor impurity, the extra electron were removed from the atom, that remaining would have a positive charge. In the P-type material the impurity atom has a



N-TYPE CRYSTAL

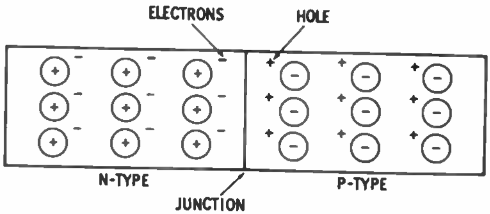
P-TYPE CRYSTAL

Fig. 1-8. P- and N-type material.

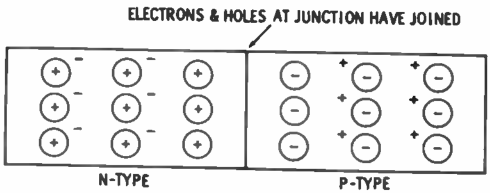
hole associated with it that has a positive charge, and therefore the rest of the atom can be considered negative. The entire impurity atom has a charge of zero—the positive hole balancing out the negative atom. Fig. 1-8 shows a representation of the P- and N-type materials. The N-type material has a number of free donor electrons associated with this large crystal structure (the two taken together are neutral). The P-type material has a number of excess holes, each associated with an acceptor impurity atom.

JUNCTIONS

A semiconductor junction is formed by taking a small block of N-type and a small block of P-type material and fusing them together to form a single crystal. This is called a P-N junction (Fig. 1-9A).



(A) PN junction before carriers have moved.



(B) Final position of carriers at PN junction.

Fig. 1-9. Action of current carriers at a PN junction.

Bear in mind that the atoms represented by the large circles are bound into the crystal and therefore are not easily moved, whereas the electrons and the holes are more or less free to move if an attraction or repulsion force exists.

The most logical question to ask at this point is, why don't all the electrons in the N-type material drift over into the P-type material and fill the holes. It does happen to the electrons and holes right at the junction. The electrons located in the N-type material immediately adjacent to the junction cross over the junction and fill the holes immediately adjacent to the junction in the P-type material. This results in the configuration shown in Fig. 1-9B. This action at the junction results in the existence of a wall which keeps the other electrons and holes from moving across the junction. Note that on the P side of the junction, a large negative wall is set up by the P-type material. On the other hand, a large positive wall is set up by the N-type material. Remember that the atoms that make up the wall on either side are fixed into the crystal so they cannot move. In order for an electron to cross from N to P it must be given a big enough push to overcome the negative wall on the P side. Conversely, in order for a hole to go from P to N, it must be given a big enough push to overcome the positive wall on the N side of the junction.

It is rather simple to give these current carriers the necessary push by applying a voltage of the proper polarity to overcome the barrier set up at the junction. Fig. 1-10 shows the application of a voltage of the proper polarity to cause current to flow through the junction. Note

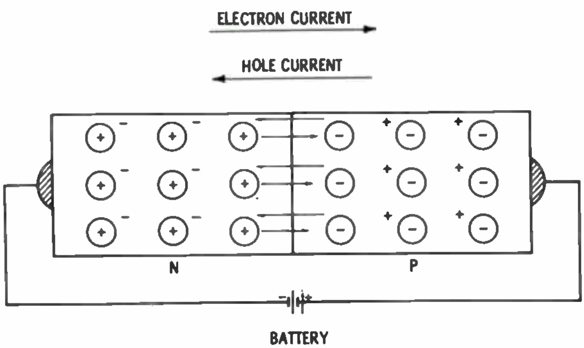


Fig. 1-10. Electron-hole flow produced by applied voltage.

that electrons flow from N to P and holes from P to N. The number of carriers taking part in the conduction can be controlled by adjusting the amount of applied voltage. In other words, there will be more current flowing if a higher voltage is applied. This is not so unusual; conductors behave in somewhat the same way. But let us look at an interesting effect which is not true for conductors.

If the voltage applied to this diode junction is in the opposite direction, no current will flow. The applied voltage has the effect of making the wall at the junction higher. In this case, as is shown in Fig. 1-11, for an electron to travel from N to P, not only would it have to overcome the natural wall at the junction, but it also would have to overcome the voltage of the battery. A PN junction turns out to be a device in which current is allowed to flow in only one direction. When the applied voltage is of one polarity, the current carriers are given a sufficient push to overcome the wall at the junction. When the applied voltage is of the opposite polarity, the wall at the junction is reinforced and the current carriers find it even harder to move across.

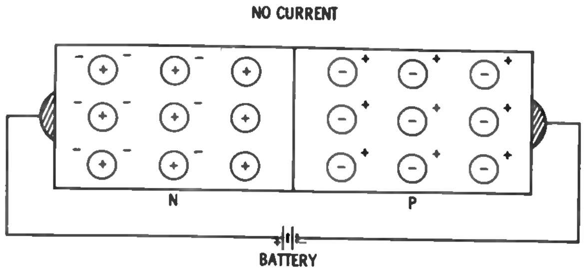


Fig. 1-11. Zero current flow with voltage applied.

When the applied voltage is of such a polarity as to cause current to flow across the junction, the junction is said to be "forward-biased"; when the applied voltage is of such a polarity as to prohibit current from flowing across the junction, the junction is said to be "reverse-biased." The word "bias" indicates that a battery or voltage is involved, while "forward" and "reverse" indicate the polarity or direction of the voltage.

We have just described a semiconductor diode. The diode is a device which, under normal conditions, allows current to flow in one direction only. Using brute force, it is possible to cause current flow in the reverse direction. However, this requires the application of a voltage large enough to rip electrons out of the crystal structure. Even this action is useful in the construction of voltage regulation devices, such as zener diodes.

When a junction is reverse-biased, there will be a very small amount of leakage current. This is due to the presence of the minority carriers that are in the material by virtue of the fact that it's impossible to completely exclude them. It is important to note that if somehow we were able to inject minority carriers into a reverse-biased junction, they would immediately go across the junction. For example, the junction in Fig. 1-11 is reverse-biased, and the majority carriers (electrons in the N material and the holes in the P material) are forced to stay on their own side of the junction. However, if there are any free electrons present in the P material, they would immediately be conducted over to the N material. The potential wall at the junction would be downhill and the biasing battery would give them an extra push. Actually, this gives us our first insight as to how a transistor works.

PN Junction Characteristic Curve

A curve that shows the relationship between the current flow through a junction and the voltage applied to it gives valuable insight into the operation of semiconductor junctions. This curve is shown in Fig. 1-12. Notice that there is no mention of N, P, electrons, or holes on the curve. It deals in terms of voltage and current. Also notice

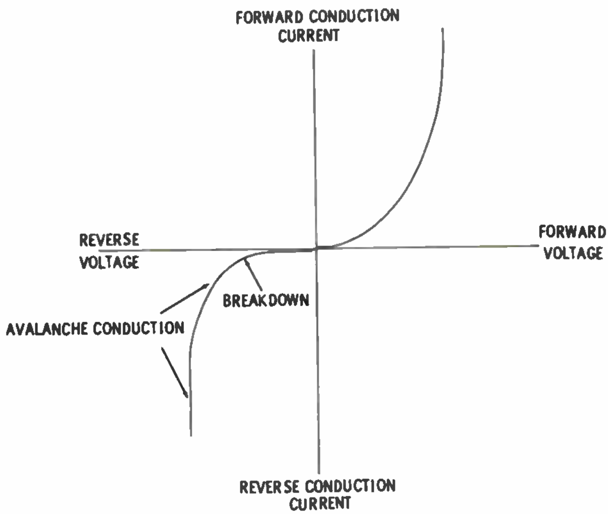
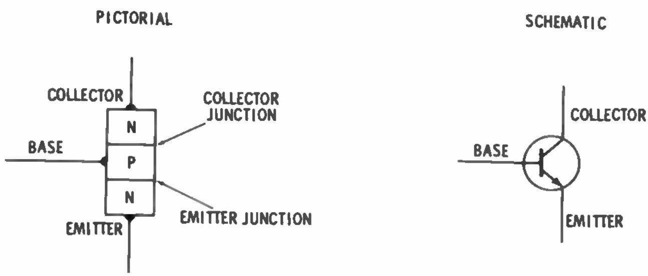


Fig. 1-12. Characteristic curve for a semiconductor diode.

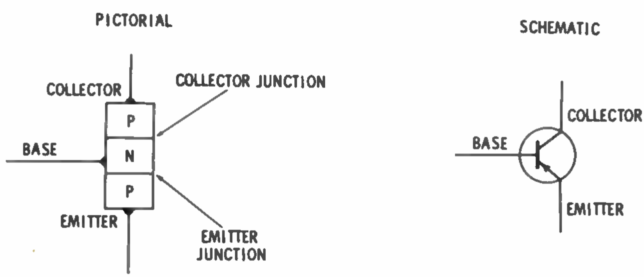
that as the forward voltage increases, the forward current increases. A point is reached when a further increase in the forward voltage causes the current to rise to a point where the junction may be destroyed. When a reverse-bias is applied to the junction, a very small amount of current flows. This small amount of current is called leakage current. A further increase in the reverse-bias voltage finally causes the junction to break down in the reverse direction, and there is an avalanche conduction in the reverse direction. For some crystals this is not destructive and can be used advantageously. As you will learn later, this is the portion of the curve that describes the action of zener diodes.

The Junction Transistor

The transistor is a three-element semiconductor consisting of two junctions. The sandwich is made up of the two types of material (N and P). It can be either an NPN transistor or a PNP transistor. Fig. 1-13 shows the pictorial and schematic representation of an NPN and



(A) NPN transistor.



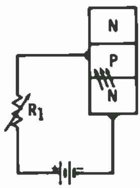
(B) PNP transistor.

Fig. 1-13. NPN and PNP transistors.

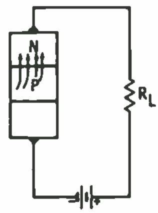
PNP transistor. The direction of the arrow on the schematic differentiates between PNP and NPN. The three elements of a transistor are the base, emitter, and collector. The names originated with the idea that the emitter emits the carriers and the collector collects them, while the base is common to both the emitter and collector.

The junction between the emitter and base is called the emitter junction. The junction between the collector and base is called the collector junction.

Controlling the resistance of the transistor is accomplished by controlling the number of current carriers available to the base region. Fig. 1-14A shows how the emitter junction is forward-biased. By forward biasing the emitter junction, majority carriers flow from the emitter (N region) to the base (P region). After these electrons enter



(A) Emitter junction (forward bias).



(B) Collector junction.

(C) Complete schematic.

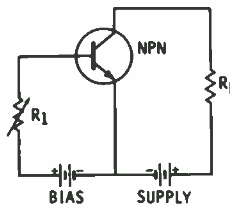


Fig. 1-14. Transistor bias.

the base region, they are no longer majority carriers as far as the collector junction is concerned, they are minority carriers and are conducted immediately through to the collector. Fig. 1-14B shows how the collector junction is reverse-biased. Therefore, we can control the number of current carriers available to the base by merely adjusting the variable resistor in series with the supply battery. The emitter emits the majority carriers into the base region where they become minority carriers and flow across the reverse-biased collector junction. Thus, the current applied to the base and emitter controls the resistance of the transistor. This is exactly analogous to the way in which the voltage applied from grid to cathode of a vacuum tube controls its resistance. Fig. 1-14C shows the equivalent transistor schematic for Fig. 1-14.

ACTIVE, PASSIVE, LINEAR, AND NONLINEAR

Every component in the entire field of electronics can be classified as either active or passive, and linear or nonlinear. This gives rise to the comforting thought that there are only four groups of components to consider: passive and linear, passive and non-linear, active and linear, and active and nonlinear. Categorizing semiconductor components in groups of families with similar characteristics will simplify the work of becoming familiar with them. You will be able to associate various devices as having generally the same properties and behaving basically the same way.

Active and passive mean approximately the same thing in electronics as they do in everyday conversation. An active component is one where the output power is greater than the input. In other words, there is a gain connected with the device. For instance, a vacuum tube or a transistor amplifier is considered an active device. An amplifier having an input voltage of 0.2V and an output voltage of 20V would have a voltage gain of 100 associated with it. This would be an active device.

A passive device is one that produces no power increase. Examples of passive devices include resistors, capacitors, inductors, and diodes. When these devices are used in simple circuits, the output never exceeds the input. For instance, a tapped bleeder across the output of a power supply is nothing more than a common voltage divider, and it serves the purpose of reducing the output to a usable value. This is one example of a passive device.

A linear device or circuit is one at which the output is proportional to the input. In other words, if the input is doubled, the output is doubled; if the input is halved, the output is halved. If the input to a voltage divider is multiplied by 2, the output voltage is multiplied by 2. If the input voltage is halved, the output voltage is halved. Resistors are a linear circuit component. Other linear circuit components—that is, devices that have an output that is proportional to the input—are linearly operated vacuum tubes, transistors, field-effect transistors, and (under certain conditions) tunnel diodes and other amplifying devices. The reason for saying under certain conditions is because transistors and tunnel diodes and most active devices can be either linear or nonlinear, depending on how they are operated. For instance, a linear amplifier is an amplifier in which the output is proportional to the input.

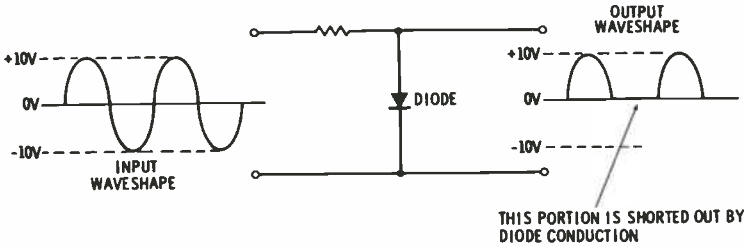


Fig. 1-15. AC voltage applied to a diode.

A nonlinear device is one in which the output is not proportional to the input. A good example of a nonlinear circuit element is the diode (which conducts current in only one direction). As shown in Fig. 1-15, when the voltage across the diode is in the forward direction, the output is nearly proportional to the input. But when the voltage across the diode is in the reverse direction, the output is nonexistent. It may help to say that a nonlinear device always distorts the input waveshape. In the case of a diode, if the input is a sine wave, the output is a half sine wave. Other examples of nonlinear circuit components are capacitors, zener diodes, vacuum-tube diodes, switching transistors, silicon-controlled rectifiers, etc. In every case the output is a distorted representation of the input.

To say that a device is nonlinear does not mean that it is bad or not useful. As we shall see, nonlinear devices play very important parts in many areas of electronics. They are used extensively in television receivers, transmitters, computers, etc.

SUMMARY

Semiconductors are solid-state devices.

Germanium and silicon are semiconductor materials that have a crystalline structure.

The process of adding impurity atoms to semiconductor crystals to provide current carriers is called doping.

There are two types of current carriers—electrons and holes.

Electrons are majority carriers in N-type material and minority carriers in P-type material.

Holes are majority carriers in P-type material and minority carriers in N-type material.

Forward bias causes majority-carrier conduction.

A diode junction conducts in only one direction.

For proper operation of a transistor, forward-bias the emitter junction and reverse-bias the collector junction.

A transistor is a variable resistor whose resistance depends on the number of current carriers that are emitted into the base region.

An active component produces gain; a passive component does not.

A linear component produces an output that is proportional to the input.

A nonlinear device distorts the input.

CHAPTER 2

Passive Semiconductor Devices

This chapter is concerned entirely with passive semiconductor devices. Passive means that the output from the device is less than the input to it. Both linear and nonlinear devices will be discussed. Some of the more important passive semiconductor circuit elements are resistors, thermistors, diodes, breakover diodes, photodiodes, and thermoelectric junctions.

THERMISTORS

Most ordinary resistors are made of carbon or metals. These materials have a positive temperature coefficient; this means that as the temperature increases, the resistance also increases. For example, a resistor having a positive temperature coefficient and a value of 1,000 ohms at 70°F might become 1,050 ohms at 150°F.

A thermistor is a resistor with a negative temperature coefficient. When the temperature increases, the resistance decreases. A thermistor is a piece of semiconductor material that has been doped to provide a given resistance. When the thermistor is heated, the electrons in the crystal gain energy, leave the crystal lattice, and become free current carriers. This means that applying heat to a semiconductor material is equivalent to applying a voltage to it. A big disadvantage of semiconductor circuits is their sensitivity to temperature change. Designers go to great lengths to overcome this temperature problem. In many cases, thermistors are used to compensate transistor circuits for temperature changes.

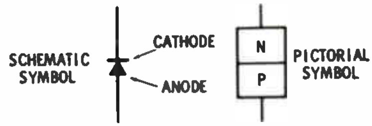
Thermistors are used to stabilize the effects of temperature variations in various types of circuitry. We have all experienced the annoyance of having to readjust a radio or TV receiver after it warms up. This need for readjustment is due to the change in characteristics of some of the devices in the set after they warm up. Thermistors are used in modern circuitry to compensate for these changes so that this type of adjustment will not be necessary. An example of how this compensation works is that of a series-connected, 1,000-ohm carbon resistor and a 1,000-ohm thermistor. At 70°F, the resistor and the thermistor each has a value of 1,000 ohms; the total resistance is 2,000 ohms. At 150°F, the resistance of the carbon resistor increases from 1,000 ohms to 1,050 ohms. This is a 5% change in the resistance. At the same time, the resistance value of the thermistor changes from 1,000 ohms to 950 ohms. Even though the resistance of each device has changed, the total resistance remains at 2,000 ohms regardless of the temperature. This basic approach to temperature compensation is used in many types of electronic circuits. Thermistors are particularly useful for stabilizing excessive current flow in power transistors, thus preventing over-current damage.

DIODES

The diode is a single-junction semiconductor device. The characteristic that makes the diode a valuable circuit element is that it allows current to flow in only one direction. The P-N junction offers a low resistance to current in one direction and a high resistance to current in the opposite direction. Current flow through a diode that encounters a low resistance is said to move in a forward direction; the direction of the high-resistance encounter is called the reverse direction. This gives rise to the terms *forward current* and *reverse current*. Reverse current is usually referred to as leakage current because it is an undesirable characteristic.

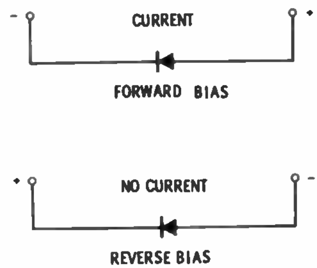
Fig. 2-1 shows the schematic symbol for the diode. The bar represents the cathode, and the arrow represents the anode. Fig. 2-2 shows how to bias a diode. It is forward-biased (for maximum current

Fig. 2-1. Schematic and pictorial symbol for a diode.



flow) when the anode is positive and the cathode is negative. It is reverse-biased (for minimum current flow) when the anode is negative and the cathode is positive.

Fig. 2-2. Polarity of voltage for forward and reverse bias.

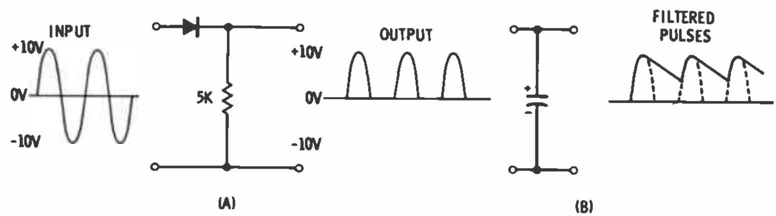


As the forward bias applied to diode is increased, the current will increase to the point where excessive heat will damage the junction. When a diode is reverse-biased, current is blocked from flowing through the junction. If the reverse voltage is sufficiently increased, the diode will break down and in most cases the excessive current will destroy the junction.

The two most important characteristics of a diode are the maximum forward current and the maximum reverse voltage that the diode can withstand. The maximum forward current is an indication of how much current the diode can carry before it is damaged. The maximum reverse voltage indicates how much reverse voltage the unit can block before it breaks down. Exceeding either of these ratings will usually destroy the diode.

Rectifiers

The most common use for diodes is to convert alternating current into direct current. The diode is a nonlinear device that fits this task perfectly (Fig. 2-3A). Since it passes current in only one direction, it clips off the negative half of the sine wave, leaving only the positive half. This produces a choppy DC, but since all the pulses go in the positive direction, it can be considered as pulsating direct current. These pulses can be smoothed, as shown by Fig. 2-3B. This figure shows the addition of a simple capacitor that makes up a filter. The purpose of the resistor is merely to complete the circuit and provide a load for the diode. As the input signal goes positive, the diode is forward-biased and is nearly a short circuit. Thus, the input voltage appears across the load resistor. When the input voltage goes negative, the diode is reverse-biased and is nearly an open circuit; therefore, none of the input voltage is passed to the load resistor. The resulting waveshape is shown in Fig. 2-3A at the output of the half-wave rectifier.



(A) Rectifier section. *(B) Filter section.*

Fig. 2-3. Diode rectifier and filter.

A capacitor is placed across the load resistor to filter the ripple in the DC output. The capacitor stores energy by charging when a voltage is applied and then releasing this energy into the circuit when the charging pulse is removed. Thus, during the negative half cycle when the diode is blocking the input voltage, the capacitor is discharging into the load resistor, giving the effect of smoothing the output DC waveshape (Fig. 2-3B).

It is possible, by using two diodes and a transformer as shown in Fig. 2-4, to obtain twice as many output pulses, thereby improving the regulation of the DC output voltage. This is called full-wave rectification. When the input voltage goes positive, the transformer causes diode D1 to be forward-biased and the positive half cycle appears across the load resistor. When the input signal goes negative, diode D2 is forward-biased, causing the negative half cycle to appear across the

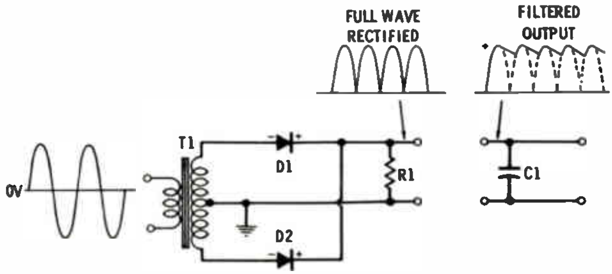


Fig. 2-4. Full-wave rectifier.

load resistor. The circuit is built so that both diodes conduct in a manner that produces the same polarity of voltage across resistor R1 in Fig. 2-4. This has the effect of converting the negative half cycle to a positive one. The center-tapped secondary of the transformer puts each diode into a separate circuit. When one diode is forward-biased, the other diode is reverse-biased. Thus, there is conduction on both halves of the input cycle, and full-wave rectification results.

Fig. 2-5 shows another method of constructing a full-wave rectifier. This is called a bridge rectifier because the diodes are connected in the form of a square and the input and output terminals are at opposite diagonals of the square. When the applied AC input voltage is positive,

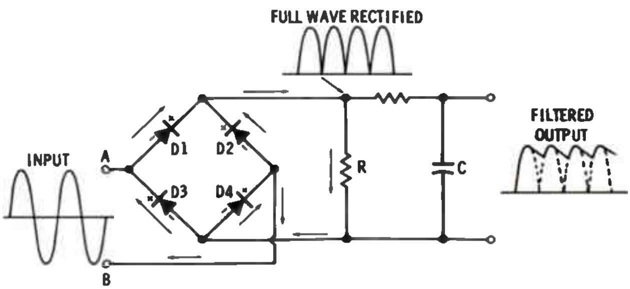


Fig. 2-5. Bridge rectifier.

D1 and D4 are forward-biased and D2 and D3 are reverse-biased. Current flows from A (+ with respect to B) through D1, through the output circuit, through D4, and back to the B terminal at the input. When the input voltage is negative, D2 and D3 are forward-biased and D1 and D4 are reverse-biased. Current flows from B through D2, through the output load resistor, and through D3 to input terminal A. It can be seen from this diagram that the current of both positive and negative alternations flows in the same direction through the load resistor; this provides full-wave rectification. If a filter network is placed across the load resistor of a bridge rectifier, the output will be filtered to produce a nearly smooth DC.

At first glance you might think that a bridge rectifier has a definite disadvantage in that it requires more diodes for full-wave rectification than the previous power-supply example. The truth of the matter is that the bridge rectifier is probably less expensive than the other, because the bridge rectifier does not require a transformer to accomplish full-wave rectification. In addition, the diodes can have a lower voltage rating, since in each conduction path there are two diodes rather than just one, as in the case of the transformer circuit. This means that under conditions of reverse bias, two diodes will divide the reverse voltage that they must block. If the voltage to the diodes is 200V peak-to-peak, the diodes of the bridge rectifier need only be capable of blocking 100V each; whereas in the transformer-type rectifier system, the single diode must be capable of blocking the entire 200V.

Diodes are widely used in power-supply circuits to rectify AC signals. Semiconductor diode circuits are designed in the same manner as vacuum-tube diodes; however, they have advantages in that they require no heater voltages and take up very little space compared to vacuum tubes. They can be designed to carry very high forward currents and to withstand large reverse voltages.

Breakdown Diodes

The breakdown diode is more commonly known as the zener diode. The word breakdown, or avalanche, is more descriptive because it actually indicates what happens inside the zener diode, which is a passive, nonlinear semiconductor. Fig. 1-12 in Chapter 1 shows that when a diode is sufficiently reverse-biased, a point of breakdown is reached. This action damages most diodes, but it was found that silicon diodes could be used to great advantage at this point. Fig. 2-6 is an expanded view of the reverse-bias portion of a silicon-diode characteristic curve. Notice that for a small change in reverse voltage a large change occurs in the reverse current. At this point the voltage across

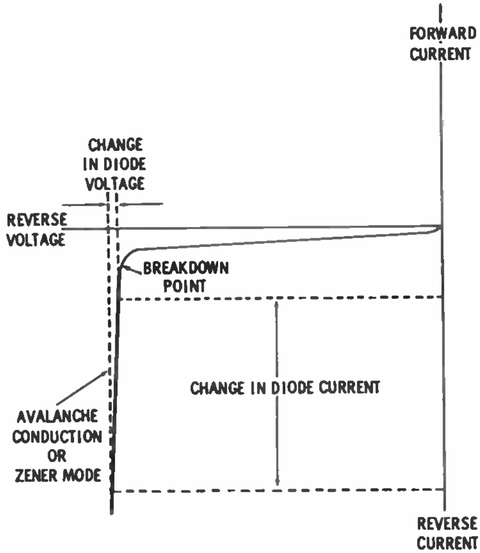


Fig. 2-6. Breakdown diode characteristic.

the breakdown diode is nearly independent of the zener current. The breakdown diode can therefore be used to regulate voltages. That is, it can be used to hold voltages constant regardless of the amount of current drain.

Fig. 2-7 shows the circuit of a breakdown-diode voltage regulator. When the zener diode is forward-biased, it behaves as a normal diode, but when it is caused to break down in the reverse direction, the voltage across it remains constant even when there is a change in the current or applied voltage. To take advantage of its voltage-regulating ability,

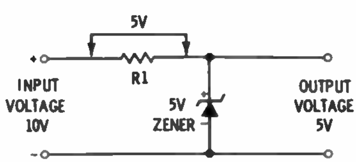


Fig. 2-7. Zener voltage regulator.

the zener diode must be biased in a manner opposite that of a normal diode. Zener diodes are specified in terms of their breakdown voltage.

In other words, a 5-volt zener would maintain a voltage of 5 volts when broken down in the reverse direction. If the input voltage to the regulator circuit were 10 volts, there would be 5 volts across the breakdown, or zener, diode and 5 volts across the resistor. If the input voltage were to rise to 15 volts, there would still be 5 volts across the zener diode and 10 volts across resistor R1. This zener-diode circuit maintains the output voltage at a constant 5 volts, regardless of the input voltage changes.

The zener diode can be used for many purposes, one of which is regulating the output voltage of DC power supplies for transistor circuits. For instance, consider the 30-volt transistor power supply

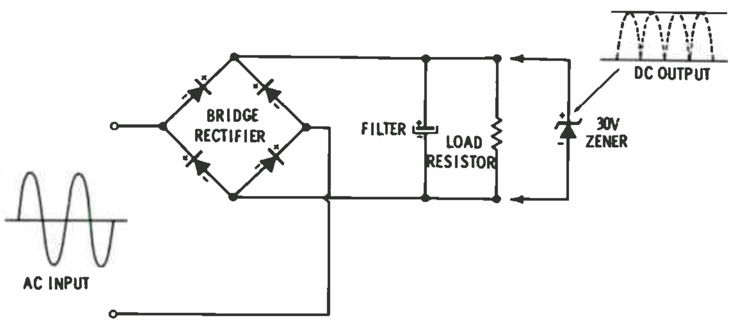


Fig. 2-8. Regulated power supply.

shown in Fig. 2-8. Without the zener control, the output voltage of this power supply would change with input voltage fluctuations and load changes. It would be considered a poorly regulated power supply. If the nature of the transistor circuit it is supplying is such that it cannot tolerate these fluctuations, the regulation may be improved by the addition of a proper breakdown diode. In addition to the regulating action, the addition of the breakover diode decreases the ripple in the final DC output.

Another example of a circuit where zener diodes are used is in the generation of square waves. The zener circuit for producing square waves is shown in Fig. 2-9. The function of the normal diode (D1) is to remove the negative half of the sine wave. The positive half cycles have a peak of 15 volts. The zener diode breaks down at 5 volts, thus shorting out the peak of the input signal. The resultant output waveshape is nearly a square wave. Square-wave pulses of this type are used in many electronic circuits, such as those in computers, radar systems, and television receivers.

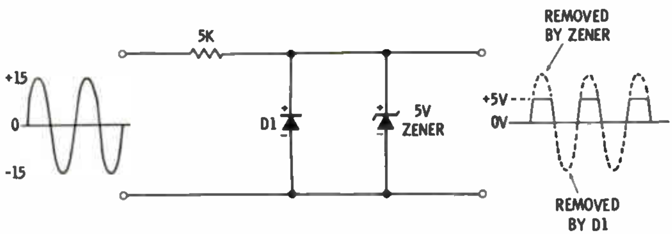


Fig. 2-9. Square-wave generator.

Diode Gates

Thousands of diodes are used in the electronic gating circuits of modern digital computers. A gate is a device that has an output only if certain input conditions are satisfied. It can be compared to an ordinary fence gate that opens if certain input conditions are fulfilled. When the gate is unlocked and nothing is standing in the way, the gate opens if someone pushes on it. Suppose, in a certain computer, we want to turn on a front-panel indicator light when: the motor is on, the computer is not performing a specific operation, and the operator pushes a certain button. This function would be performed by a gate. Voltages at the input of this gate indicate when all three of these prerequisites are fulfilled. When all three voltages are present, indicating that everything is ready, the gate opens and there is an output which turns on the lamp.

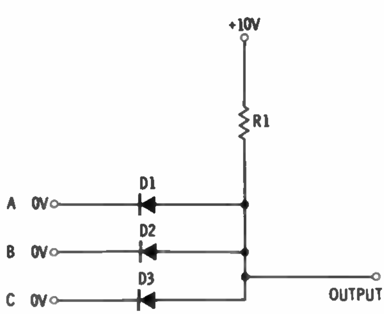


Fig. 2-10. AND gate.

The lighting of the lamp mentioned in the previous paragraph can be accomplished by using an AND gate. If the three conditions are considered as A, B, and C, we can say that when the voltages are present for A and B and C, then there will be an output to turn on the lamp—thus the name AND gate. Fig. 2-10 shows a diode AND gate. Notice that when the input voltages to the diodes are at zero, current flows from +10, through resistor R1, and through the diodes. The forward-biased diodes can be considered as short circuits. Therefore, when all zeros are present at the input of the diodes, the output will be zero because it is shorted through the conducting diodes to the input "zeros." In this case, no voltage indicates that conditions A and B and C are not present, and the gate is closed. If one of the inputs is raised to +5 volts, the corresponding diode has less forward bias and does not conduct as heavily. However, each of the other two diodes is still strongly forward-biased and will continue to short the output to zero volts. This is merely saying that fulfillment of one input condition is not enough to cause the gate to open. Should two of the +5-volt inputs be present, the output would still be shorted out by the third conducting diode, and the output would still be zero. The conditions for opening the gate have not yet been fulfilled. If all three input voltages are raised to +5 volts, all three diodes have less forward bias and the output will be connected to the three +5-volt inputs through the diodes. This voltage will then be used to turn on the lamp. The output voltage will rise only when all three input conditions are fulfilled. Only when A and B and C are present will there be an output voltage.

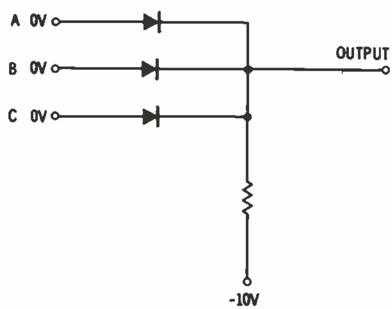


Fig. 2-11. OR gate.

Suppose you want to turn on the output lamp of the computer when any one of the input voltages is present. You would want the presence of either A or B or C voltage to turn on the output lamp. This would require an OR gate. An OR gate is one where any input being present will turn on the output. A diode OR gate is shown in Fig. 2-11. In this case, all diodes are again forward-biased. When zero voltage is present at all inputs, the output voltage is zero because it is connected to the zero-volt input through the conducting diodes. If the A input voltage rises to +5 volts, the corresponding diode becomes more forward-biased and the output is shorted to the +5 volts through the more heavily conducting diode. The input voltage is conducted through the diode to the output, even though the B and C inputs remain at the zero level. The same action will be repeated if any of the other inputs are turned on. If A or B or C is present at the input, there will be an output.

Fig. 2-12 shows a combination of two AND gates and one OR gate

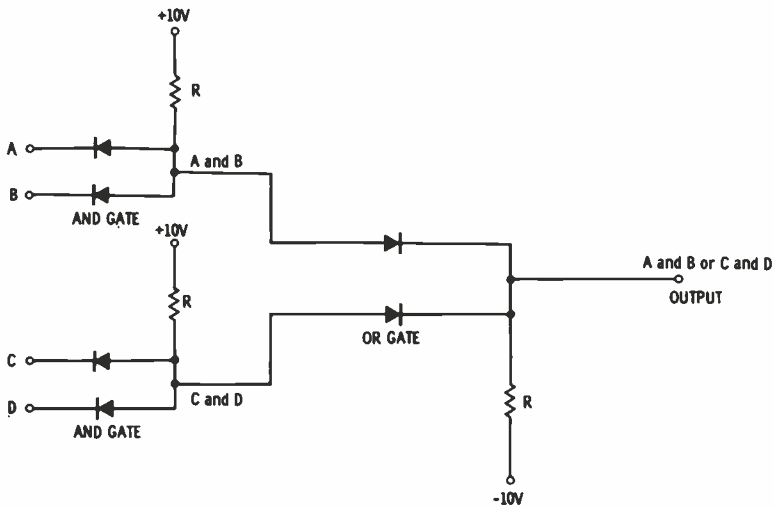


Fig. 2-12. Logic circuit.

used to accomplish a single logic function. Diodes are popular because they are relatively inexpensive, high switching speeds can be attained, and the circuitry simplifies the problem of designing, maintaining, and servicing the equipment.

Photodiodes

Current carriers in semiconductors are sensitive to many different types of energy. In the case of thermistors, the semiconductor is sensitive to heat energy. Light energy may also be used to excite electrons so that they will move out of the crystal structure and enter into conduction as current carriers. A photodiode is one of a group of such devices. It is a straightforward PN junction, having a plastic lens in the top of the container through which light can reach the junction.

When used in a circuit, this type of diode is biased in the reverse direction. When the semiconductor junction is exposed to light, energy

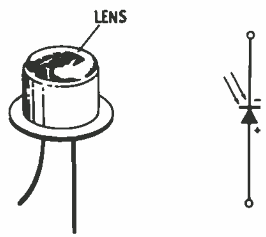


Fig. 2-13. Photodiode.

is absorbed by the electrons; the energy they gain permits them to overcome the junction potential, and the diode conducts in the forward direction. One might say that the energy of the light is used to overcome the barrier produced by the PN junction. Fig. 2-13 shows a sketch of a typical photodiode. It is contained in a case similar to that in which a transistor is packaged. The top contains a lens through which the light passes. Fig. 2-13 also shows the schematic symbol for a photodiode. The small arrows indicate impinging light.

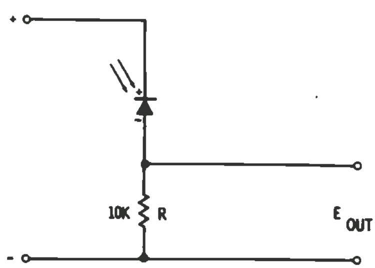


Fig. 2-14. Circuit for indicating light intensity.

The diagram in Fig. 2-14 shows a photodiode circuit used for measuring light intensity. When no light illuminates the diode, no current flows in the circuit and thus the output voltage is zero. The photodiode conducts when light hits it. Its resistance decreases by an amount that depends on the intensity of the light. Therefore, the current through R and the output voltage are controlled by the intensity of the light.

THERMOELECTRIC COOLING

Semiconductor materials are also beginning to find wide application as heat pumps. This means that they can accomplish the same tasks that a refrigerator performs. Refrigerators are already employing semiconductor materials as heat pumps instead of the standard compressor and motor. Fig. 2-15 shows a typical semiconductor thermocouple. The principle is based on the fact that when an electron goes from a

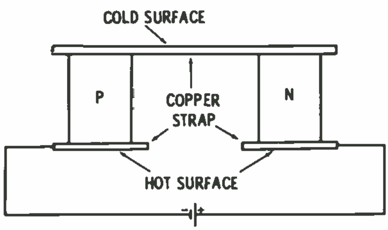


Fig. 2-15. Semiconductor thermocouple.

low-energy state to a high-energy state, it must absorb energy. The electrons in N-type material are plentiful and therefore are in a high-energy state, whereas in the P-type material the electrons are in a relatively low-energy state. Thus, when an electron travels from P-type material to N-type material, it absorbs heat energy from the surroundings, which are, as a result, cooled. On the other hand, when electrons go from a P-type material to N-type, they give up energy and thus dissipate heat. This is what happens when the electrons leave the material and go to the hot surface copper strap through the battery back to the N-type. The net result of the entire system is that the top copper strap is cooled while the lower one is heated. The heat is pumped from the cold surface to the hot. This does not violate any of the laws of thermodynamics. Heat that is pumped from the cold side of the junction and all of the energy must be dissipated by the hot junction. The most common means of dissipating this heat is by using a heat sink. In actual application the thermoelectric cooler is sandwiched between the cold plate and the heat sink, as shown in Fig. 2-16, which illustrates the sandwiching of multiple thermocouples

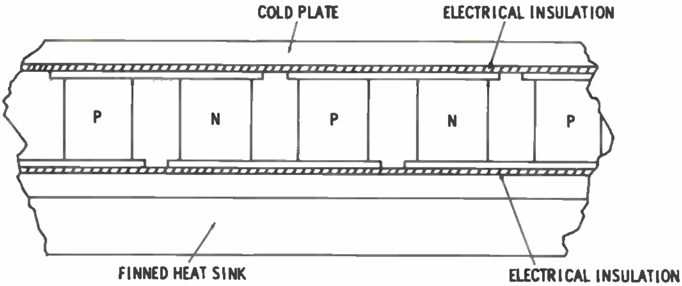


Fig. 2-16. Multiple thermocouples.

to form a refrigeration system. It is also possible to reverse the flow of current and produce a reversal of the hot and cold surfaces.

One of the applications of thermoelectric coolers at the present time is as an improved heat sink for power transistors. The cooling effect is used to maintain a low case temperature for the transistor and thereby increase the power-handling capabilities of the transistor.

CHAPTER 3

Transistors

Undoubtedly the best-known semiconductor device is the transistor. It was discovered in 1945 at the Bell Telephone Laboratories by John Bardeen and W. H. Brattain; since then it has revolutionized the electronics industry. There are many different types of transistors, each with characteristics that make it best suited for a particular application. The basic difference among transistor types is the method of construction. The two main types of transistors are germanium and silicon. The basic semiconductor material in one type is germanium, and in the other type it is silicon. The economics of the manufacturing process usually determines which type of semiconductor material is used. Some manufacturing methods work best with germanium, and others work best with silicon. A few methods are suited to both materials.

There are two configurations for transistors—the NPN and the PNP. These two configurations complement each other. The existence of complementary transistor types makes possible many circuits that were not feasible with vacuum tubes. A PNP transistor can be replaced with a similar NPN; and the circuit can be made to function exactly the same, merely by reversing the polarity of the power supply. The only difference between complementary NPN and PNP transistors is the direction of current flow.

The schematic symbol for a transistor dates back to the first transistors constructed at the Bell Labs. These were known as point-contact transistors. The emitter and collector were fine wires placed very close together on the surface of a semiconductor material called the base. The emitter has the arrow on it, signifying the emitting of current carriers into the base region. The collector is so called because it is thought of as collecting the current carriers emitted by the emitter. Only a few types of point-contact transistors are still manufactured, and these are only for very special applications. A large majority of the transistors used today are of the junction type. For this reason, we will not consider point-contact transistors, since their operation is very similar to that of junction-type transistors.

CONSTRUCTION

There are a number of ways to manufacture transistors, and each method has certain advantages and disadvantages. Transistor junctions may be manufactured in one of three basic ways—rate growing,

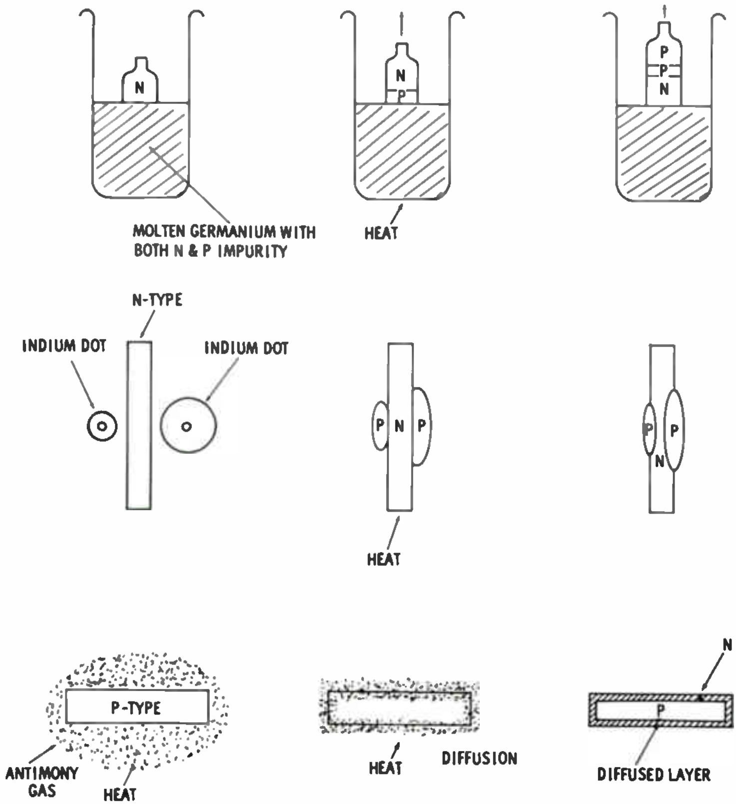


Fig. 3-1. Transistor manufacturing methods.

alloying, and diffusion. Sometimes these three ways are combined to form special transistors. The basic principles of these three methods are illustrated in Fig. 3-1. Junctions may be grown out of a melt of the molten semiconductor material. As a germanium crystal is slowly pulled out of the melt, the crystal will grow. The melt contains both N- and P-type impurities, and whether an N- and P-type crystal forms depends on the rate at which the crystal is grown. Rate growing is not used for making either PNP or NPN silicon transistors; because of the high cost of manufacture, they cannot be competitive with the germanium transistors made in the same way.

Alloying may be likened to welding, where a junction is made between two metals by the application of heat. The alloy transistor is constructed by applying two impurity dots to the opposite sides of a semiconductor wafer and then supplying enough heat to alloy the two

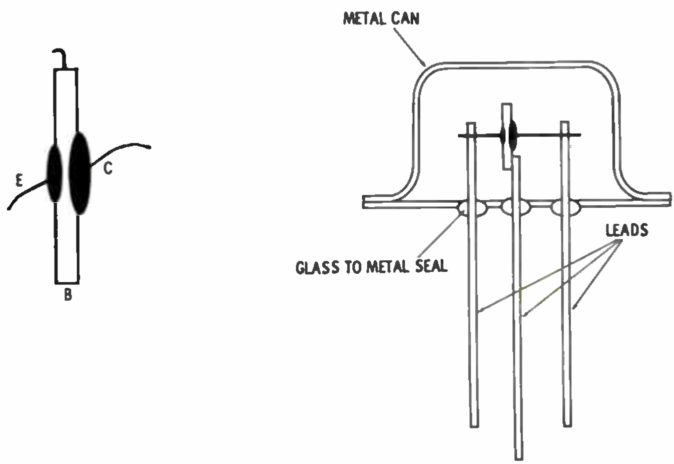
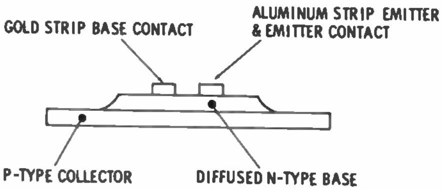


Fig. 3-2. Alloy transistor construction.

dots to the wafer. Fig. 3-2 shows the geometry and mounting of the alloy-type transistor. The collector dot is made larger than the emitter because the emitted current carriers tend to spread out. In this way, all or most of them are collected. Both PNP and NPN germanium and silicon transistors are made by the alloying method.

Diffused mesa transistors get their name from their physical construction (Fig. 3-3). The collector is a flat wafer of P-type material into which a *mesa* of N-type material is diffused. This forms the collector base junction. A gold strip is used to make contact with the base. An aluminum contact makes up the emitter and emitter contact. The aluminum strip on the N-type base forms a junction of dissimilar metals which has a diode characteristic. Therefore, it functions as both emitter and emitter contact. The physical mounting of the mesa transistor is shown in Fig. 3-3.



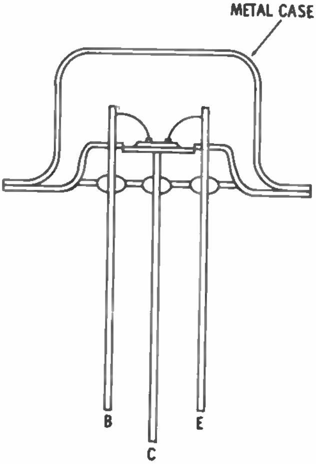


Fig. 3-3. Mesa transistor construction.

Transistor manufacturers found that they could get better electrical characteristics from some transistors if a collector with two different doping levels, or resistances, was used. To do this, films are grown on existing collectors, as shown in Fig. 3-4. These are called *epitaxial*

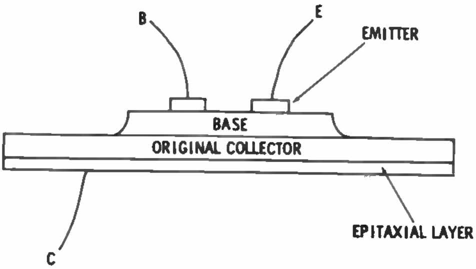


Fig. 3-4. Epitaxial-mesa transistor.

films, and transistors with this feature are called epitaxial transistors. Epitaxial means that the crystal structure of the film is lined up with the crystal structure of the collector material.

Planar transistors are made entirely by diffusion, and a silicon material is always used. The name planar is derived from the fact that all of the connections (E, B, and C) are on the same plane (Fig. 3-5). A major advantage of this transistor is that the junctions are covered with silicon dioxide (SiO_2), an inert material that protects the junc-

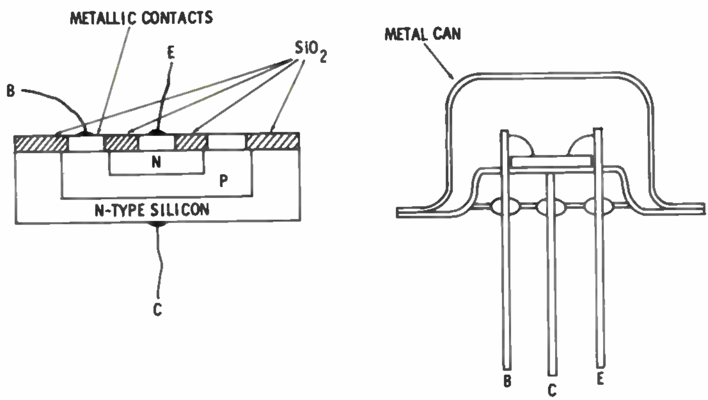
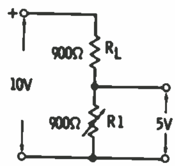


Fig. 3-5. Planar-passivated transistor.

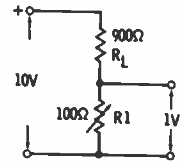
tions from the corroding effects of the atmosphere. Thus the transistor is said to be "passivated," or protected. An oxide coating forms easily on the silicon surface, passivating it, and thus accounting for the exclusive use of silicon for this type of transistor.

TRANSISTOR AMPLIFIER

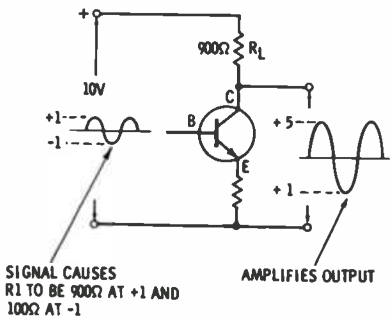
To accomplish amplification using a transistor, you need only to cause the transistor to operate as a variable resistor. Before continuing, let us investigate what is meant by amplification. Consider the voltage divider in Fig. 3-6. The top resistor is known as the load resistor (R_L), and the variable resistor (R1) represents the transistor. The variable resistor is capable of changing resistance from 100 ohms to 900 ohms. When R1 is set to 900 ohms, there will be exactly half the voltage at the output, as shown in Fig. 3-6A. Since both R_L and R1 are equal,



(A) Voltage divider set for high output.



(B) Voltage divider set for low output.



(C) Transistor substituted for the variable resistor.

Fig. 3-6. Amplifier analogy.

there will be 5 volts across each resistor, and therefore the output voltage will be 5 volts. When the variable resistor shown in Fig. 3-6B is set at 100 ohms, its value equals one tenth of the total resistance. This means that one tenth of the total voltage will appear across the variable resistor and, therefore, across the output. The output voltage in this case would be 1 volt. Consequently, by varying the resistance of the variable resistor from 100 ohms to 900 ohms, it is possible to make the output voltage change from 1 volt to 5 volts, a total change of 4 volts. Now, if you substitute a transistor for the variable resistor, as shown in Fig. 3-6C, and apply a 2-volt peak-to-peak sine wave to the base, the resistance of the transistor will vary. If the resistance in the transistor changes from 100 to 900 ohms, the output will swing from 1 to 5 volts. We have made a 2-volt peak-to-peak input result in a 4-volt peak-to-peak output. This device would have a gain of 2; the output is twice the input.

By applying a signal to the base of a transistor, we can cause the resistance of the transistor to vary. We can also design the associated circuitry so that the output voltage will change by a greater amount than the input voltage, thus resulting in gain. The bias battery shown in Fig. 3-7 is used to maintain forward bias on the emitter junction in order that the transistor can operate on the linear portion of its characteristic curve.

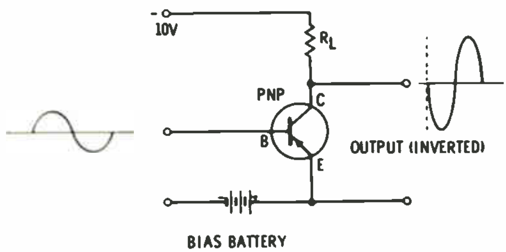


Fig. 3-7. PNP transistor amplifier.

The forward bias changes in synchronism with the input signal. Thus, current carriers in the base-emitter region will follow the input signal. In the PNP transistor amplifier of Fig. 3-7, when the input signal is at a maximum, the current carriers will be at a minimum. When the input signal is negative, the current carriers will be at a maximum. The reason is that the input signal adds to the bias voltage. When the input is positive, it is subtracting from the bias and there are fewer current carriers. This accounts for the inversion of the output (Fig. 3-7). (The output goes negative when the input goes positive.) Current carriers entering the base region are conducted to the collector; and since the number of current carriers varies in synchronism with the input signal, the resistance of the transistor varies in synchronism with the input signal.

Amplifier Biasing Methods

All transistor-biasing methods are based on the rule that the emitter junction should be forward-biased and the collector junction should be reverse-biased to provide linear transistor amplification. This is an important point to remember when analyzing, building, or servicing transistor circuits.

The amplifier in Fig. 3-7 has a disadvantage in that it employs a separate biasing battery. This is expensive and bulky, particularly since it is possible to use the supply battery to provide bias (Fig. 3-8). In this case, a resistor replaces a battery at quite a saving in cost and space. R1 is a biasing resistor; current from the battery flows through R1 and through the emitter base junction, thus providing forward bias. The amount of bias depends on the resistance of R1. The collector junction is reverse-biased by the positive voltage applied to the collector through load resistor R_L.

Transistors are very sensitive to changes in temperature. In fact, as the temperature increases, the number of current carriers increases

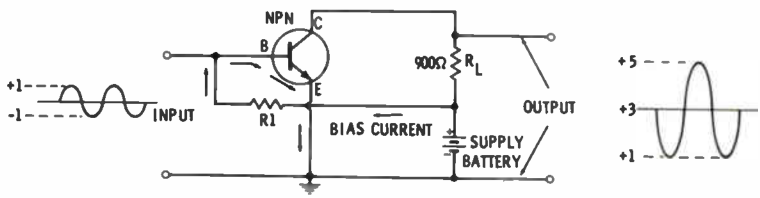


Fig. 3-8. NPN transistor amplifier.

and the resistance decreases. Consider Fig. 3-8; bias current flowing in the emitter circuit causes collector current to flow, and this collector current causes the transistor to heat. As the transistor heats, more current flows in the circuit because of the temperature dependence of the transistor. This action, which may continue to the point where the transistor destroys itself, is called *thermal runaway*. Destruction does not always occur; the transistor normally does not get too warm because the heat is radiated, or carried off, by air currents in the unit. However, thermal runaway is a serious problem.

A circuit for stabilizing a transistor amplifier against thermal run-

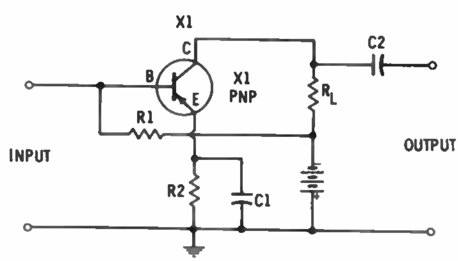


Fig. 3-9. Emitter resistor used to stabilize a transistor amplifier stage.

away is shown in Fig. 3-9. An emitter resistor (R2) is placed between the emitter and ground. If the transistor begins to heat, the increased current in the collector circuit causes an increased voltage drop across R2. As a result, voltage on the emitter rises, reducing the forward bias on the base. As a result, less current flows in the emitter junction. This, in turn, reduces the collector current and permits the transistor to cool. The capacitor across emitter resistor R2 is called an emitter bypass capacitor; it is used to maintain the emitter of the transistor at AC ground.

The circuit of Fig. 3-9 may be further stabilized against changes in operating characteristics due to changes in temperature. This is done by insertion of R3, as shown in Fig. 3-10. R3 provides more reliable

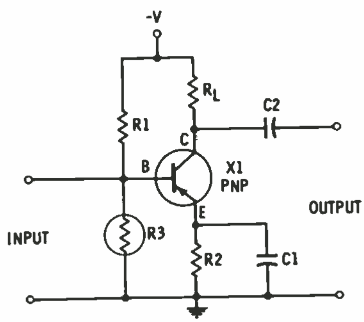


Fig. 3-10. Stabilized transistor amplifier.

and more stable operation of the transistor amplifier by holding the base voltage constant, regardless of changes in the transistor parameters due to changes in temperature. R1 and R3 may be thought of as a voltage divider that keeps the base voltage constant. R2 keeps the emitter voltage constant. The only disadvantage is that this circuit draws more current from the battery than either of the previous circuits; however, the additional stabilization is worth the extra current drain on the battery.

Coupling Methods

To provide increased gain, most transistor circuits require that two or more amplifiers be cascaded. The output of one amplifier is fed into the input of the next amplifier and amplified, until the desired output level is reached. One of the most common types of coupling is capacitive coupling. The amplifier shown in Fig. 3-11 uses capacitive cou-

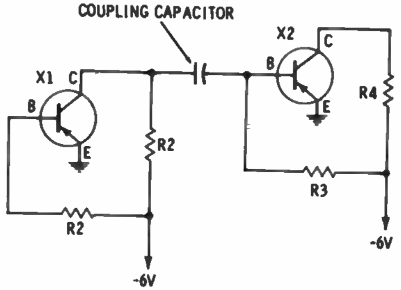


Fig. 3-11. Capacitive-coupled transistor amplifier.

pling. The purpose of capacitor C1 is to block the DC voltage at the collector of the transistor and to pass the AC signal. This DC voltage must be blocked because it would upset the biasing of the next stage. Capacitive coupling of transistor stages is the most economical but not the most efficient method. The reason is that each transistor amplifier has different input and output impedances. Usually, capacitive coupling does not afford an efficient impedance match between stages. As a result, the signal coupling is not as good as it could be.

The most favorable condition for signal transfer exists when the output impedance of one stage is matched to the input impedance of the next stage. Impedance is a specific name for AC resistance. A transformer is used as an impedance-matching device between transistor stages. Fig. 3-12 shows a transformer-coupled amplifier. The load resistors of the amplifiers are replaced by the primaries of the trans-

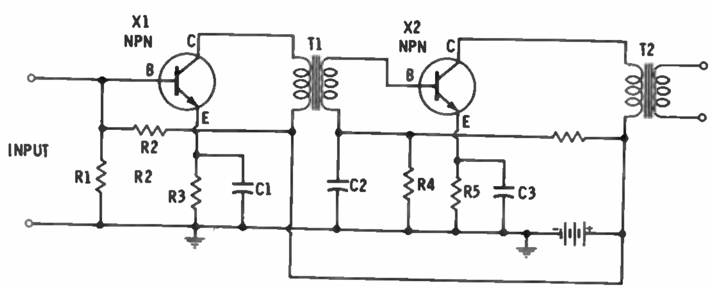


Fig. 3-12. Transformer-coupled transistor amplifier.

formers. C2 blocks the DC bias current for X2 from returning to ground, and it also provides an effective AC ground for the secondary of transformer T1. If C2 were not in the circuit, the series resistance of bias resistors R4 and R6 would reduce the level of the applied signal. The AC signal is coupled from the primary of the transformer to the secondary. The disadvantage of transformer coupling versus capacitor coupling is the higher cost of the transformer. Sometimes, however, the added expense is justified.

Unlike vacuum tubes, transistors are available in complementary types, NPN and PNP. Transistors may be direct-coupled using opposite transistor types, PNP and NPN, as shown in Fig. 3-13. The first stage is a PNP transistor amplifier. The second stage is an NPN amplifier that is direct-coupled to the previous stage. Transistor X2 receives bias

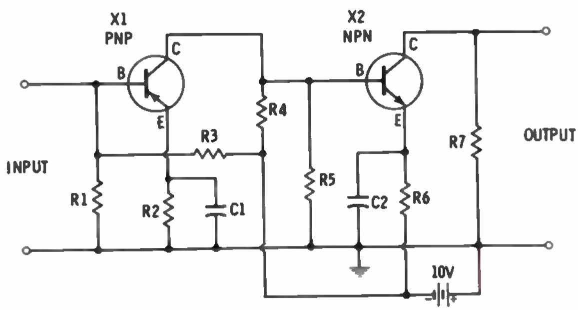


Fig. 3-13. Direct-coupled transistor amplifier.

current from the load resistor of the previous stage. The emitter resistor (R6) of transistor X2 is returned to the negative terminal. Also, load resistor R6 for the NPN transistor (X2) is returned to positive. This, in effect, turns the power supply around for use with the NPN stage. It is direct-coupled and takes advantage of the complementary symmetry of NPN and PNP transistors.

AMPLIFIER CONFIGURATIONS

Up to now we have been discussing what is known as the common-emitter amplifier configuration. There are two other ways to operate a transistor as an amplifier—the common-base and the common-collector configurations. The word "common" is sometimes used interchangeably with "grounded." It refers to the element that is common to the input and output circuits. The common-emitter circuit is probably the most popular because it provides good voltage, current, and power gain.

Common-Emitter Amplifier

Common-emitter amplifiers are shown in Figs. 3-8 and 3-9. The base of both the PNP and NPN types is referenced to ground. This base becomes one of the terminals in the input circuit and also one of the terminals of the output circuit of a common-emitter amplifier.

Common-Base Amplifier

Fig. 3-14 shows a common-base circuit. The input signal is fed into the emitter and drives the base junction. An amplified version of the driving signal flows in the collector circuit. Note that there is no inversion in the common-base amplifier as there is in the common emitter.

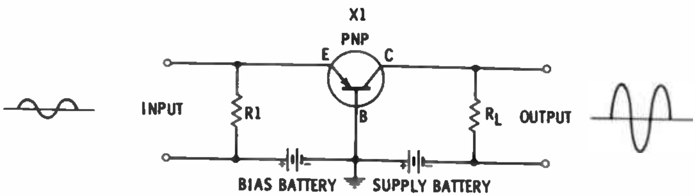


Fig. 3-14. Common-base amplifier.

The voltage gain of a common-base amplifier is lower than that of a common-emitter amplifier. The common-base amplifier does, however, afford a large degree of current gain and isolation between the input and output circuits. For this reason, it is often used for amplifying high-frequency signals where stability is desired. Note that even though this configuration differs greatly from the common emitter, the emitter junction is still forward-biased and the collector junction is still reverse-biased. The vacuum-tube counterpart of the common-emitter circuit is the grounded-grid amplifier.

Common-Collector Amplifiers

The third circuit configuration is connected so that the collector is common to both the emitter input and the base output circuit (Fig. 3-15). A common-collector circuit is often referred to as an emitter follower. The counterpart in vacuum-tube circuitry is the cathode follower. The reason for the name "emitter follower" is that the voltage across the emitter resistor tends to follow the base voltage, in the same way that the cathode voltage follows the grid voltage in a

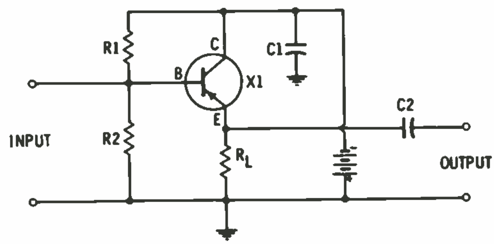


Fig. 3-15. Common-collector amplifier.

cathode follower. The load resistor (R_L) in Fig. 3-15 is in series with the emitter. Note that the collector is connected directly to the battery negative. The purpose of C1 is to ground the collector to AC signals. The voltage gain of an emitter follower is always less than 1. It is therefore not of much use as a voltage amplifier. The power gain is greater than 1, but less than that of other configurations. You might ask why the emitter follower is used at all. The reason is that it has a very high input resistance, a very low output resistance, and good current gain. Signal transfer between circuits is very sensitive to the resistance of both the circuits into which they feed and the circuits which feed them. An amplifier stage working into the proper or matched load resistance may have a gain of 500, whereas the same amplifier working into an unmatched load resistance may have a gain of only 50. The emitter follower is commonly used as a resistance, or impedance-matching, device.

Tuned Amplifier

Often, in electronics, amplifiers are used to amplify one frequency. Such a circuit is called a "tuned" amplifier. This means the amplifier is tuned to a specific frequency and it amplifies only this frequency, excluding all others. In the tuned amplifier the input circuit, output circuit, or both, employ tuned circuits that are generally composed of inductive and capacitive elements. Fig. 3-16 shows a simple tuned amplifier. If a tuned circuit is placed in the collector circuit of a transistor amplifier, as shown in Fig. 3-16, the circuit will, for all practical purposes, amplify at the resonant frequency of the tuned circuit. Other input signal frequencies are not amplified. This type of circuit makes it possible to select and amplify one particular frequency, or a group of frequencies. The signal is inductively coupled from the output of one stage to the input of the next by a method called transformer coupling.

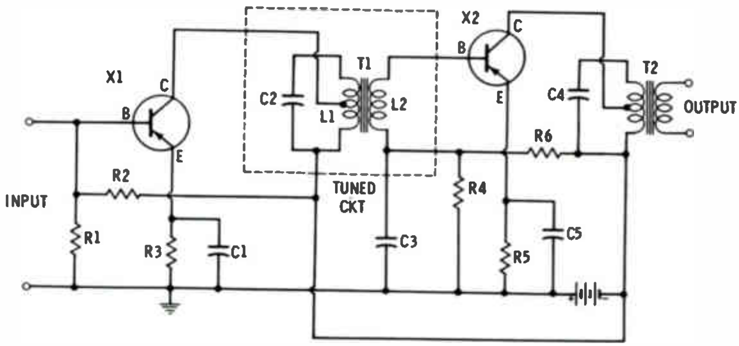


Fig. 3-16. Tuned amplifier.

OSCILLATOR CIRCUITS

An amplifier circuit oscillates when some of the output is fed back to the input in such a way that the amplifier becomes unstable. At first glance, one would think that this would be an undesirable condition. Sometimes it is; however, oscillators are used to accomplish very im-

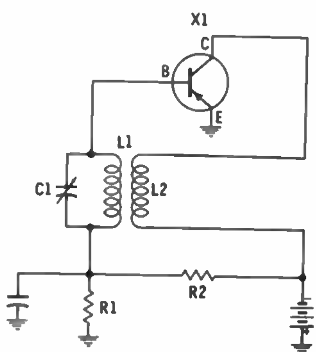


Fig. 3-17. Tuned-base oscillator.

portant functions, such as the generation of signals that act as carriers for radiocommunications.

Fig. 3-17 shows a very simple feedback oscillator. The output signal is coupled back into the input base circuit through the transformer comprised of coils L1 and L2. L1-C1 is a tuned circuit that determines the frequency of the oscillator. Energy from coil L1 is fed directly to the base of the transistor. As the collector current rises, the feedback causes the transistor to conduct more heavily, and the output current rises faster. When the collector current reaches the point where it can rise no more, the base signal begins to decay. This flywheel effect produces an output having a frequency that is determined by the values of L1 and C1.

This type of oscillator is called a tuned-base or common-emitter oscillator. The resonant elements can also be placed in the collector or emitter circuits. The frequency of an oscillator is always determined

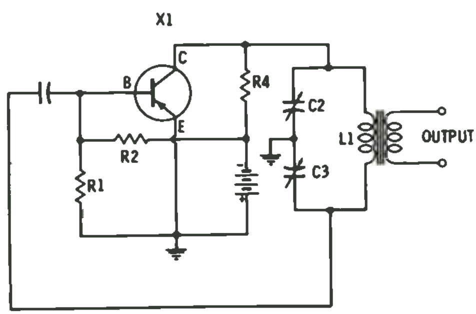


Fig. 3-18. Colpitts oscillator.

by the resonant frequency of the tuned circuit. Two other types of oscillators, the Colpitts and the Hartley, are shown in Figs. 3-18 and 3-19. The Colpitts oscillator feeds a portion of the output back to the input through a capacitive voltage divider (C2 and C3). The combination of C2, C3, and L1 makes up a tank circuit that determines the frequency of the oscillator. C2 and C3 are variable, and they determine the frequency of the oscillator. In the Hartley oscillator (Fig. 3-19) some of the output signal is fed back to the input from an inductive voltage divider (L1 and L2). C2 is variable, and its setting determines the frequency of the oscillator. These are the types of oscillators that are generally found in the high-frequency circuits of radio and TV equipment.

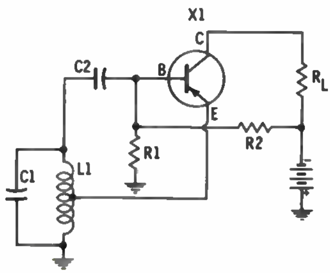


Fig. 3-19. Hartley oscillator.

FREQUENCY CONVERTERS

Very often it is desired in electronics to convert one frequency to another. For example, in a standard radio receiver the incoming radio frequency (RF) carrier is converted to a lower intermediate frequency (IF). The process of doing this is called frequency conversion, mixing, or heterodyning. The last of these terms accounts for the name "superheterodyne" applied to radio receivers.

All mixers and converters are based on the fact that when two signals of different frequencies are mixed, many other frequencies are generated. One is the sum of the original frequencies; another is the difference. For example, if the two input frequencies to a mixer are 100 kc and 75 kc, there will be four output frequencies, the two

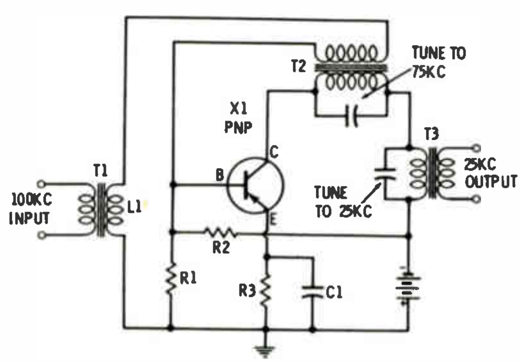


Fig. 3-20. Transistor converter stage.

original, plus the sum and the difference frequencies. These are 100 kc, 75 kc, 175 kc, and 25 kc.

To change 100 kc into 25 kc, a separate oscillator (local oscillator) would be used to generate a 75-kc signal. This would be mixed with the 100-kc input. A resonant circuit tuned to 25 kc and placed across the output would pick out the desired signal and reject the three other components, resulting in an output of 25 kc. It is possible to perform frequency conversion with one stage. Fig. 3-20 shows a practical converter circuit of the type that is often found in transistor radios. If the RF input is 100 kc and the circuit oscillates at 75 kc, the output will be 25 kc. Transistor X1 oscillates by virtue of the feedback from T2 to the base input circuit. C2 and L2 are tuned so that the circuit oscillates at 75 kc. The 100 kc is fed in through transformer T1 and mixed with the 75 kc in the base circuit. All four frequencies (100 kc, 175 kc, 75 kc, and 25 kc) are available in the collector circuit. C3 and L3 are tuned to 25 kc, and this frequency is present at the output of transformer T3.

AUDIO OUTPUT STAGES

After an RF signal has been amplified and detected, it is applied to an audio output stage. The most common example is shown in Fig. 3-21, a typical single-transistor, power-output stage. Notice that the power transistor is biased by resistors R1 and R2 in the same manner

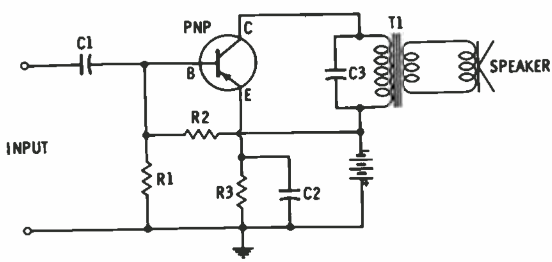
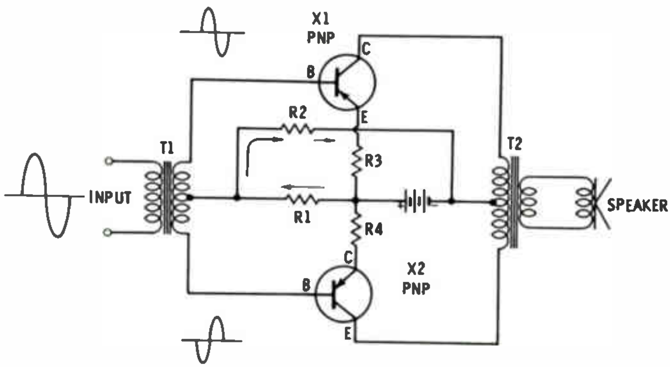


Fig. 3-21. Single-ended output stage.

as other transistor amplifiers. Fig. 3-22 illustrates a push-pull output stage. Each transistor amplifies half of the input signal; since each transistor has to work only each half cycle, it can be driven harder, thus providing an efficient output stage. The R1-R2 voltage divider provides a small amount of forward bias for X1 and X2. The secondary of T1 is center-tapped to provide opposite polarity signals for the base of transistors X1 and X2. Thus, when X1 is going positive, X2 is going negative, providing push-pull operation. The major disadvantage of



3-22. Push-pull output stage.

this circuit is the need for transformers to split the polarity of the input signal and to couple power to the speaker.

The output transformer may be eliminated from a push-pull output stage because of the low impedance of power transistors. Fig. 3-23 shows a diagram of a push-pull output stage that employs complementary transistor types. R1, R2, and R3 form a voltage divider to forward-bias transistors X1 and X2. When the input signal goes negative, transistor X1 drives the speaker in one direction; on positive signals, transistor X2 drives the speaker in the opposite direction. In this output

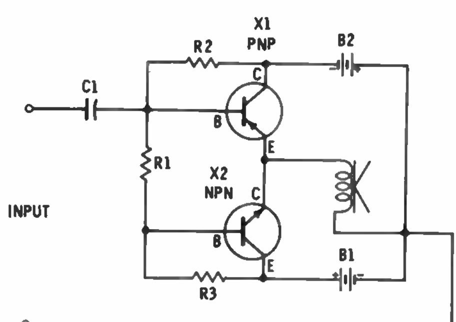


Fig. 3-23. Transformerless output stage.

stage the complementary transistors make it possible to eliminate both the driver transformer and the output transformer.

PRACTICAL RADIO RECEIVER

When a relatively complicated circuit is being discussed, it is useful to employ a block diagram, such as the one shown in Fig. 3-24. Each stage is represented by a labeled block, and the signal paths and the direction of the signals are indicated by arrows. The signals present at the output of the various stages are also included for easy reference. Fig. 3-25 shows the complete schematic of this eight-transistor radio.

L1 is an antenna transformer wound on a metallic rod; it receives the incoming signal. The tank circuit, composed of coil L1 and one

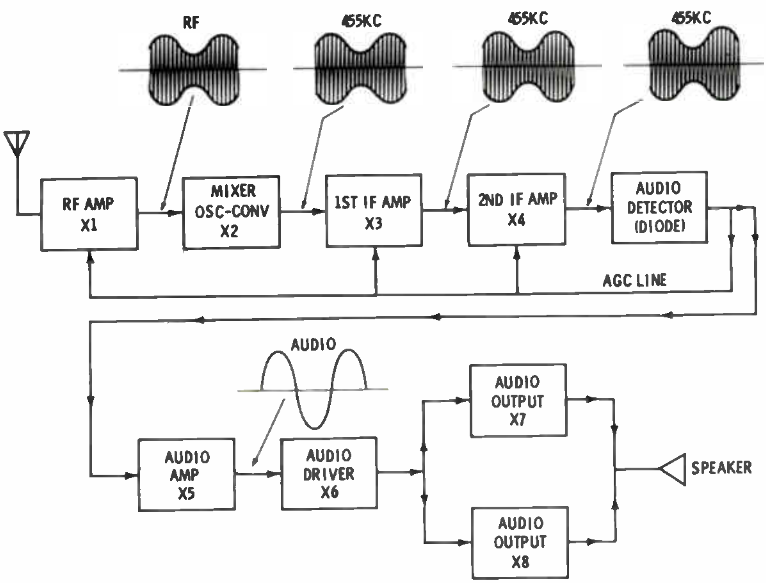


Fig. 3-24. Block diagram of a broadcast receiver.

section of capacitor M2, is tuned to the carrier frequency. Transistor X1 is an RF amplifier operated in an emitter-follower configuration. It provides an impedance match between the antenna and the converter stage. The collector of this stage is effectively signal-grounded by the bypass capacitor (C8). Resistor R3 is used as a decoupling resistor.

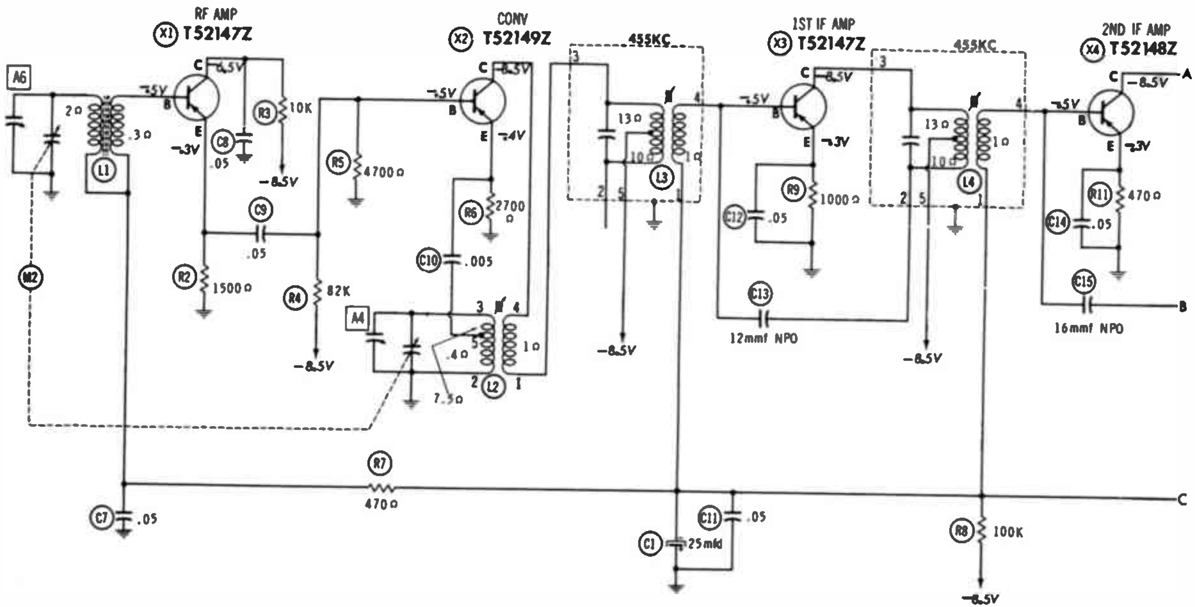
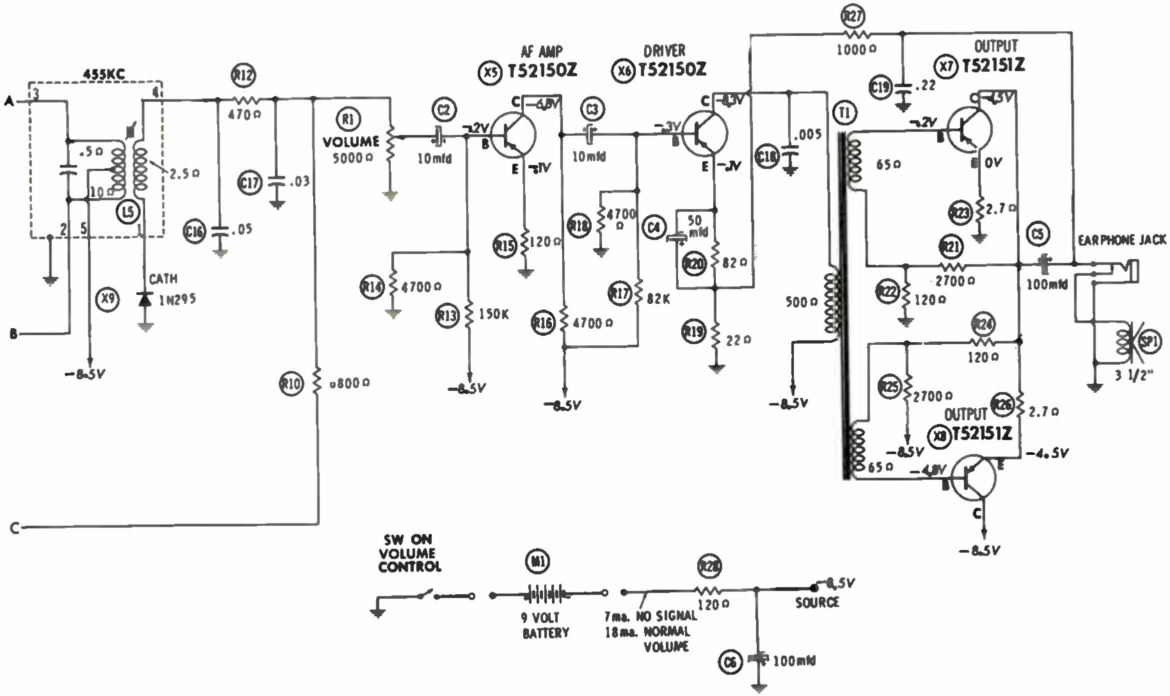


Fig. 3-25. Schematic of a

typical transistor radio.



It is also used to establish the proper voltage at the collector of transistor X1.

The RF signal is coupled to the base of the converter through coupling capacitor C9. Transistor X2 is used as an oscillator, mixer, and converter. The oscillator coil (L2) is comprised of two windings, the primary connected in series with the collector and the secondary tuned by one section of capacitor M2. The feedback signal from a tap on the tuned secondary is applied at the emitter of transistor X2. The frequency of the oscillator is maintained at 455 kc above the desired signal, and this provides an intermediate frequency of 455 kc in the collector circuit. The IF transformer (L3) in the collector of the connecter stage is tuned to accept a 455-kc signal, and it rejects all other frequencies. The secondary of this transformer (L3) applies the signal to the base of transistor X3.

The second and third IF stages in this receiver are neutralized. This is accomplished by a signal that is fed back to the base of the transistor by neutralizing capacitors C13 and C15. To prevent oscillation of the stage, the feedback signal must be of the proper phase. For this reason the values of C13 and C15 are critical.

Negative bias for the RF, first IF, and second IF stages is obtained through the 100K resistor (R8). A biasing divider network is composed of resistors R8 and R10 and volume control R1. This network is also part of the automatic gain control (AGC) system.

The signal applied to the detector diode produces a positive voltage across RF filter capacitor C16. This positive voltage is added to the negative voltage at the volume control, thus reducing the bias on the AGC line. Resistor R10 is placed in series with the volume control and the AGC filters (C1 and C11) to isolate the audio from the AGC line.

The audio signal is coupled from the center tap of the volume control to the base of the audio amplifier by electrolytic capacitor C2. Because of the low impedance of transistor circuits, it is necessary to use high-value capacitors to obtain low-frequency response. Since electrolytics are used for coupling, it is important to observe the polarity of the coupling capacitor when it is being replaced.

The output of the audio amplifier (X5) is applied to the driver transistor (X6). This stage is a second audio amplifier and is called a driver. Unlike the vacuum tube, the transistor requires both voltage and current input in order to function properly. The purpose of a driver is to provide the correct amount of voltage and current drive to the output transistors. Driver transformer T1 has two secondary windings to provide a push-pull signal to drive the output transistors (X7 and X8).

The speaker in this receiver is coupled to the output transistors by the 100-mfd capacitor (C5). This capacitor is used to block the DC from the voice coil. Tone compensation for the receiver is provided by a feedback network from the speaker to the emitter of the driver transistor. Also, to prevent an excessive rise in the impedance of the output transformer, capacitor C18 is placed across the transformer primary.

AUDIO AMPLIFIER

Another typical use of transistors is in audio amplifiers. Fig. 3-26 shows a block diagram for a Bogen public-address type amplifier that operates as a mobile unit from either a 6 or 12V power supply. It has

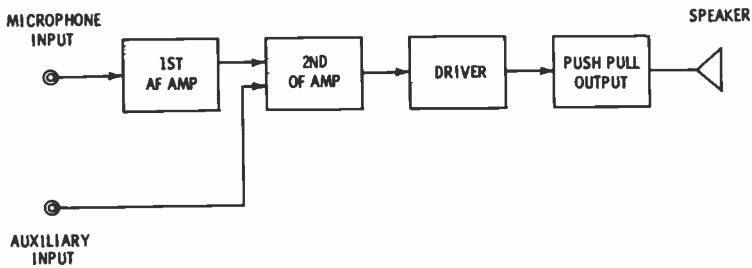


Fig. 3-26. Block diagram of an audio amplifier.

facilities for two inputs—one for a microphone and one for an auxiliary input, such as a record player, radio, etc. Since the output level of a microphone is low, an extra stage of amplification is required. The auxiliary input is connected to the second amplifier. The output of the second audio amplifier feeds an audio stage that drives the push-pull output stage. An audio frequency amplifier must be able to amplify the entire audio spectrum from approximately 100 cps to 15 kc.

Fig. 3-27 shows the schematic for this AF amplifier. The microphone input is capacity-coupled by C8 to the base of transistor X1. The first stage transistor (X1) is used as a preamplifier for the microphone.

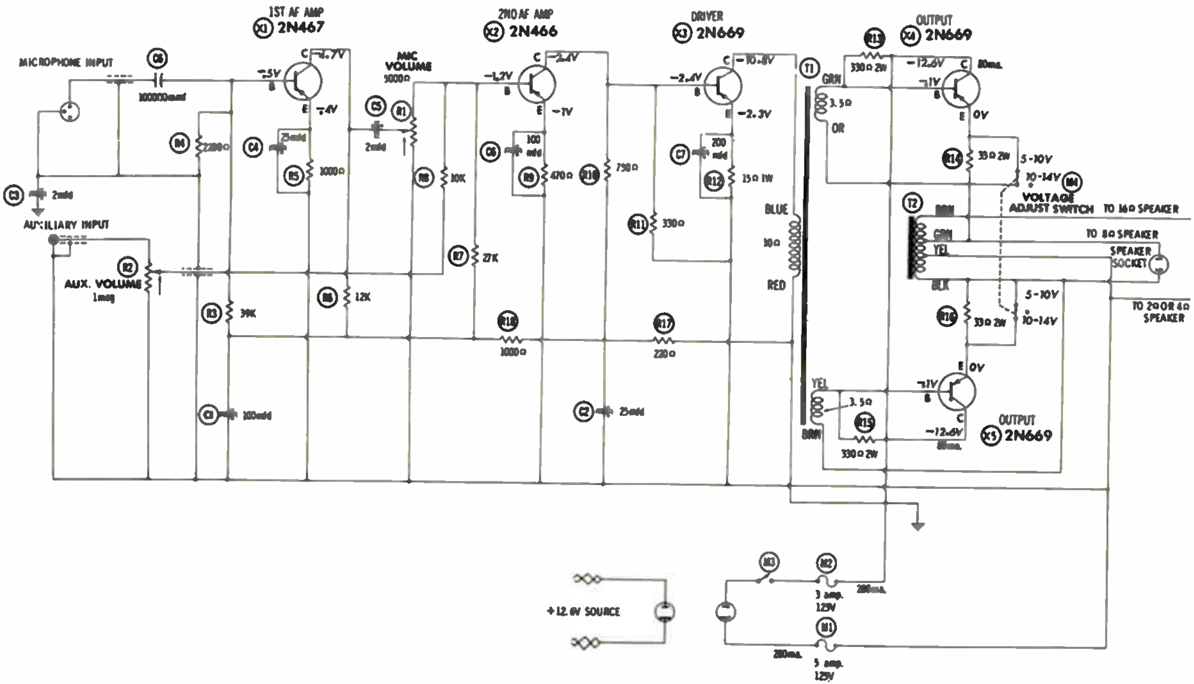


Fig. 3-27. Public address amplifier.

Bias for X1 is obtained from the power supply through the voltage divider composed of resistors R3 and R4. R1 is the volume control for the microphone (MIC) input. It is also used as part of the biasing network for transistor X2. Bias for the driver transistor (X3) is obtained directly from the collector of the first audio stage. The signal is direct-coupled from the second audio stage to the driver. The output of transistor X2 is direct-coupled to driver transistor X3. This stage feeds the driver transformer (T1) which, in turn, feeds the push-pull output transistors (X4 and X5). T2 is an impedance-matching transformer that enables the amplifier to drive a 16, 8, 4, or 2-ohm speaker by merely changing the connections.

This audio amplifier may be used with either of two supply voltages—a 6V battery or a 12V battery. When used with a 6V battery, resistors R14 and R16 are shorted. When used with a 12V battery, resistors R15 and R14 are used to stabilize the output transistors against thermal runaway.

REGULATED POWER SUPPLY

An unregulated power supply has a significant disadvantage in that the output voltage does not remain constant under varying load conditions. For example, if the open circuit voltage of a DC power supply is 20V and the application of a 2,000-ohm load resistor causes the output voltage to drop to 14V, the power supply is just not large enough to supply all the necessary current. There are many types of circuits which will not operate properly unless the supply voltage remains absolutely constant as the load changes. Transistors can be used to "regulate" power supplies so that the output voltage remains constant.

It was shown in Chapter 2 how a zener, or breakdown, diode may be used to regulate the output voltage of a power supply. Although the use of a zener diode is certainly an improvement, the regulation can be still further improved by the use of a transistor voltage regulator. Fig. 3-28 shows a schematic diagram of a basic transistor voltage regulator. The circuit function of X1 is to control the voltage drop across the output. A sensing circuit is used to measure the output voltage. If the output voltage should begin to drop, a signal is fed to the series regulator, reducing its resistance and allowing more voltage to be applied to the load, thus bringing the output voltage back to normal. Should the output voltage tend to rise, a signal is fed to the series regulator, causing its resistance to increase and thus reducing the voltage applied to the load. The net effect is to maintain the load voltage at a constant level under varying load conditions.

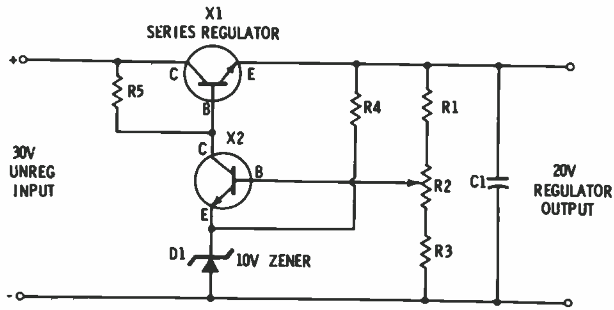


Fig. 3-28. Voltage regulator.

D1 is a zener connected in series with R4 across the output of the supply. The voltage across the zener diode is always at a constant value—in this case 10V. This constant value provides a reference voltage with which to compare the measured output voltage. The voltage divider made up of R1, R2, and R3 provides a voltage of approximately 10V that is fed into the base of transistor X2. The difference between this voltage and the zener reference voltage will determine the bias on transistor X2. R2 is a potentiometer which is used to adjust the bias on transistor X2. Resistor R5 is the load resistor for X2. The junction of this resistor and the collector of X2 is directly connected to the base of series regulator X1.

If the output voltage tends to drop, the bias on X2 decreases, reducing its conduction and increasing the voltage at the base of the current regulator. An increase in current through the series regulator returns the output voltage to the original value. If the output voltage increases, the bias on X2 increases, causing a decrease in the bias current applied to X1. A decreased current flow in the series regulator reduces the voltage to the original value.

This is a simple form of transistor regulated power supply. More complicated systems can be designed, but each is based on the principles described. The output voltage is sampled and the result is used to influence the amount of voltage applied to the load. A series regulator may be considered as a valve that is opened or closed, depending on how much voltage or current is needed for the load.

HEAT SINKS

It has been said a number of times in this book that the main disadvantage of transistors is their sensitivity to heat and temperature. If the heat generated by conducting transistors is not dissipated in some manner, it can result in the destruction of the semiconductors or other components. As a result, many devices are used to remove the heat from the vicinity of the transistors. These include fans, blowers, heat sinks, etc.

A heat sink is a piece of metal that is used to conduct the heat away from the circuit. Some heat sinks are finned for more efficient cooling. In other applications the chassis can be used as a heat sink, particularly when the transistor is mounted in direct contact with it. Fig. 3-29 shows a power transistor mounted on a typical finned heat sink. Heat sinks are generally used with power transistors, rectifiers, series regu-

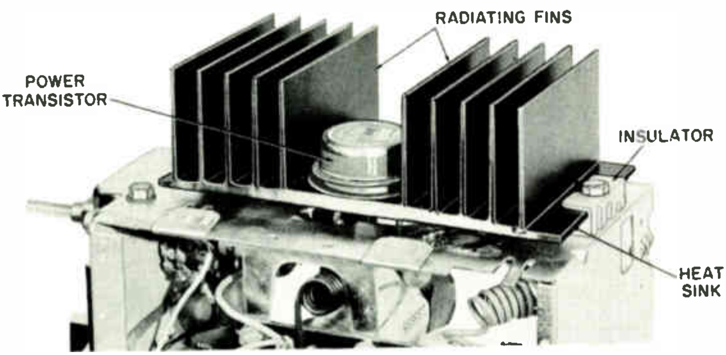


Fig. 3-29. Finned heat sink.

lators, or other semiconductors that are conducting large amounts of current.

CHAPTER 4

Other Active Linear Semiconductors

Until now, as far as linear active semiconductors are concerned we have discussed only transistors. There are other solid-state devices, however, such as tunnel diodes, photocells, and field-effect transistors; their operation is based on the material discussed in Chapter 1. This chapter is concerned with a few of these devices and some of the more common circuits in which they are used.

FIELD-EFFECT TRANSISTORS

Sometimes it seems as though field-effect transistors (FET) should have been discovered before ordinary transistors. The reason for this is that their operation is very similar to that of vacuum tubes. The FET comprises a bar of semiconductor material through which current flows, while the resistance of the bar is controlled by the application of a field. It is quite similar to a vacuum tube, where the resistance between plate and cathode is controlled by the voltage on the grid. A field is used to control the resistance in the semiconductor, thereby curtailing the flow of majority carriers—thus the name *field-effect* transistor.

You will recall that a transistor is a minority-carrier device in which minority carriers are injected into the base and are conducted to the collector. A field-effect device is a majority-carrier semiconductor. The semiconductor bar, whether P or N, is doped to provide majority carriers. These carriers flow through the material, never crossing a PN junction. The resistance of the material is controlled by applying a field; Fig. 4-1 shows the functional diagram of a field-effect transistor.

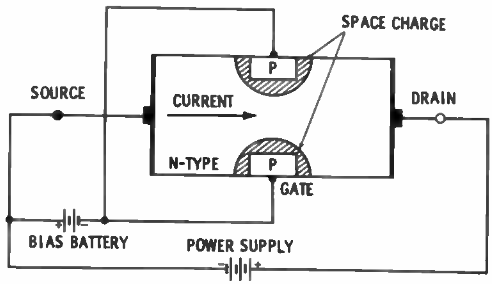


Fig. 4-1. N-type field-effect transistor.

The left side of the main block of the silicon is considered the source of current, and the right side (called the drain) is the side from which the current leaves the device. The gate is the area which controls the resistance of the bar. If the bar is N-type material, the gate is made from P-type material. It is diffused into the bar by one of the methods described in Chapter 3. The PN junction at the gate is reverse-biased, and current carriers are removed from the vicinity of the gate. This sets up a space charge that makes it difficult for carriers to get through. Another way to look at it is that the resistance of the bar is increased by removal of the majority carriers. The more reverse voltage applied to the gate, the greater the space charge becomes and the more difficult it is for electrons to get from the source to the drain. The point at which

Fig. 4-2. Schematic of N- and P-type field-effect transistors.

the field becomes great enough to prevent electrons from passing the gate is referred to as *pinch off*. The reverse voltage that causes this is called the pinch-off voltage. Fig. 4-2 shows the approximate mechanical configuration and the schematic for N- and P-type FET's. Note that there is no such thing as a PNP or NPN FET. They are either N or P types, depending on the bar material. Fig. 4-3 shows the analogy between the FET and the vacuum tube. The source can be likened to the cathode, the drain performs the same function as the plate, and the gate acts the same as the grid. Circuits using an N-type FET are very similar to vacuum-tube circuits.

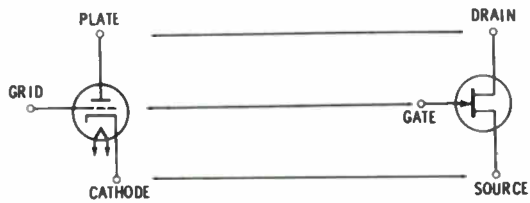


Fig. 4-3. Comparison between a vacuum tube and an FET.

As a matter of fact, the field-effect transistor and the vacuum tube are so much alike that one can use the terminology almost interchangeably and the design methods are almost identical. Bias for the FET is supplied in the same manner as it is for the vacuum tube. The input resistance of an FET is very high. The gate-circuit junction is reverse-biased and no current flows; thus the input resistance is very high. A vacuum tube also has a very high input resistance. A transistor has a very low input resistance, because its input circuit is a forward-biased emitter junction. Field-effect transistors will certainly appeal to anyone who is familiar with vacuum-tube circuits because the circuitry is almost identical. If a P-type FET is used, the polarities of the applied voltages are reversed.

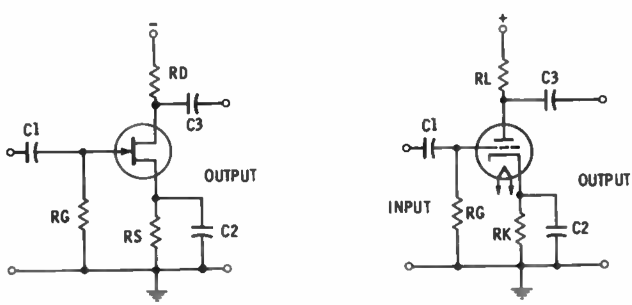


Fig. 4-4. Comparing an FET amplifier and a vacuum-tube amplifier.

Fig. 4-4 shows the similarity between a simple FET amplifier and a vacuum-tube amplifier. The FET is self-biased by the voltage drop across R_S. This makes the source positive with respect to the gate. The vacuum tube is self-biased in a similar fashion by R_K. R_G in the FET circuit stabilizes the bias by providing a return to ground for the small gate currents that exist. R_G of the tube circuit performs the same function for the grid currents of a vacuum tube. C2 is an AC bypass capacitor that keeps the load current from influencing the bias.

It is significant to remember that in transistors the current carriers flow through the three different types of semiconductor materials that make up the base, emitter, and collector. In an FET, the current travels through only one kind of semiconductor from the source to the drain. The resistance of the path can be varied by the voltage applied to the gate. Transistors create noise which is caused as holes and electrons combine.

The frequency response of transistors is limited by minority-carrier transient time. The current in an FET is a stream of electrons or carriers, the density of which is controlled by the gate. When the gate causes the resistance of the FET to increase, the current in the entire circuit changes and not just the current in the semiconductor. An FET does, however, have other problems that a transistor does not have. One problem is that a large value of input capacity tends to limit the frequency response.

These FET devices are rather new and still mostly in the pilot production stages. In the months and years to come, construction techniques will undoubtedly be developed to overcome this input capacity problem. The main applications at the present time for field-effect transistors are for semiconductor circuits having a high input impedance. In addition, the FET is capable of power gains far in excess of the ordinary transistor.

TUNNEL DIODES

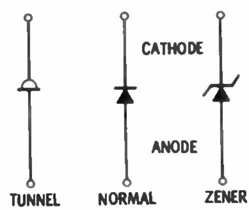
Another interesting semiconductor device is the tunnel diode. It is sometimes referred to as the "Esaki" diode because it was discovered in 1958 by the Japanese scientist Leo Esaki. The Esaki diode may be used as either a linear amplifier or a nonlinear switching device.

A tunnel diode does not differ in physical construction from an ordinary diode, except that it is more heavily doped than the ordinary

P-N junction. As a matter of fact, this is where the tunnel diode gets its name; it is so heavily doped that there are many millions of majority carriers crowded into the vicinity of the junction. A number of these majority carriers "tunnel" through the barrier that exists at the junction if a slight bias potential is applied. If the barrier is considered a potential hill, then the carriers can be thought of as having burrowed through the hill rather than going over it. Actually, tunneling takes place in a normally doped diode, but so few carriers are involved that the effect is not noticeable. It is only when the semiconductor is heavily doped, providing many majority carriers, that the tunneling effect becomes marked.

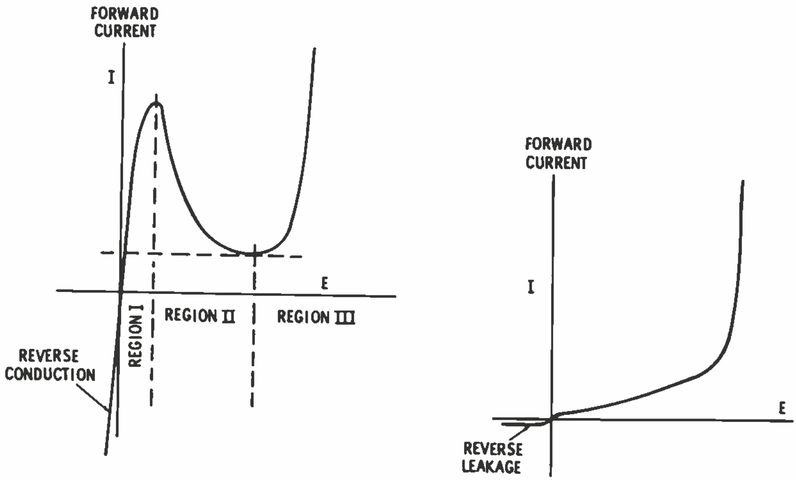
Fig. 4-5 shows the schematic symbol for a tunnel diode. Symbols for the standard PN diode and the zener diode are shown for compari-

Fig. 4-5. Comparison of schematic symbols for the normal diode, tunnel diode, and the zener diode.



son. Tunnel diodes do not constitute serious competition for transistors as amplifiers in the lower frequency ranges because they do not provide as much gain as transistors in this area. But at the higher frequencies (microwave frequencies) above those at which transistors operate, tunnel diodes will function with sufficient gain to make them of great value.

Tunnel diodes may be used as amplifiers by virtue of the negative-resistance portion of their characteristic curve. Before explaining what a negative-resistance amplifier is, let us first show how a tunnel diode exhibits a negative-resistance characteristic. Fig. 4-6 shows a tunnel-diode characteristic and a normal diode characteristic. Note that the tunnel diode (Fig. 4-6A) conducts in the reverse direction, whereas the regular diode (Fig. 4-6B) does not. This is due to the fact that there are so many current carriers in the tunnel diode that they can conduct in either direction. Region I shows the application of a small forward bias (but not enough to overcome the barrier at the junction), resulting in a high forward conduction of the diode. This is due to the many majority carriers tunneling through the potential hill rather than going



(A) Tunnel diode *(B) Regular diode*

Fig. 4-6. Characteristic curves.

over it. As the forward bias is increased (Region II), the conduction begins to decrease and the diode curve is returned to the normal operating characteristic. Further increase in the forward bias results in Region III, which is normal diode behavior and corresponds to the diode curve shown in Fig. 4-6B.

Region II is known as the negative-resistance portion of the tunnel-diode characteristic. It is so called because an increase in voltage results in a decrease in current. A normal positive resistance will behave in just the opposite fashion: an increase in voltage will cause an increase in current. This interesting characteristic of tunnel diodes can be used for many things, among which is the linear amplifier. In order to be used as a linear amplifier, the negative-resistance portion of the curve can also be used to make tunnel-diode oscillators and mixers.

The negative resistance referred to is not actually a minus resistance; it is a function of change in voltage across the diode as related to the change in current through the diode. On the tunnel-diode curve in Fig. 4-7, points A and B indicate a change in voltage of 10 millivolts and a current change of 0.7 milliamp. Note, however, that while the voltage is increasing (going in a positive direction), the current is

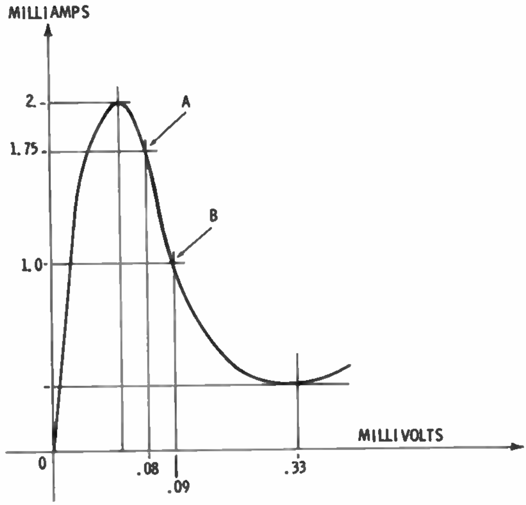


Fig. 4-7. Tunnel-diode curve.

decreasing (going in a negative direction). The formula for AC resistance can be derived from this relation. It is equal to the change in voltage divided by the change in current.

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

$$= \frac{.01}{-.00075}$$

$$R = -13.3 \text{ ohms}$$

The AC resistance of a minus 13.3 ohms represents a dynamic characteristic of the tunnel diode, whereas the actual DC resistance of the tunnel diode at point A or B is a positive resistance. This DC resistance can be calculated by dividing the positive voltage by the positive current.

The negative-resistance portion of the tunnel-diode curve can be used to amplify a signal. Fig. 4-8 shows a typical tunnel-diode amplifier complete with biasing circuitry. Bias current is supplied by A through the 2K resistor. R1 (15Ω) is the load resistor. This circuit is shown merely as an example and would not really be practical, except at very high frequencies (for example, 100 mc). The negative-resistance portion of the tunnel-diode curve may be used in the construction of oscillators, pulse circuits, and other devices. Usually tunnel diodes are not used unless frequency or size limitations necessitate their use.

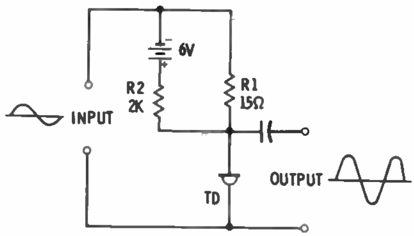


Fig. 4-8. Tunnel-diode amplifier.

HALL EFFECT

The Hall effect was discovered by E. H. Hall in 1879 at Johns Hopkins University. However, there was no practical use for the device because the output voltages available from materials at that time were very small. It was not until the advent of modern semiconductors that the Hall element became a circuit practicality. The successful operation of a Hall device depends on a semiconductor material with a large number of available current carriers. The material used is either P or N, and there are no junctions with which to be concerned.

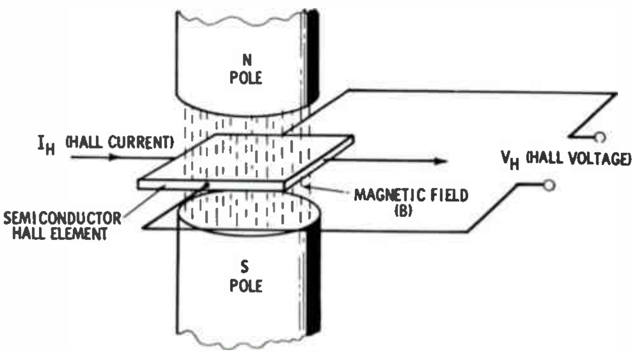


Fig. 4-9. Basic Hall effect.

Fig. 4-9 shows the basic Hall effect. If a semiconductor, such as germanium, is subjected to a magnetic field that is perpendicular to current flowing through the element, a voltage will be produced across the element. This voltage is proportional to the product of the current and the field (current × field). The Hall element can be used for many applications. For instance, it is widely used in measuring magnetic-field intensity. It can be used to measure current, as a proximity

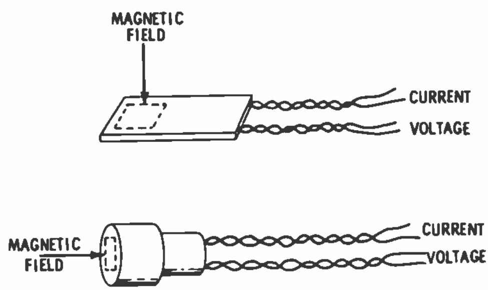
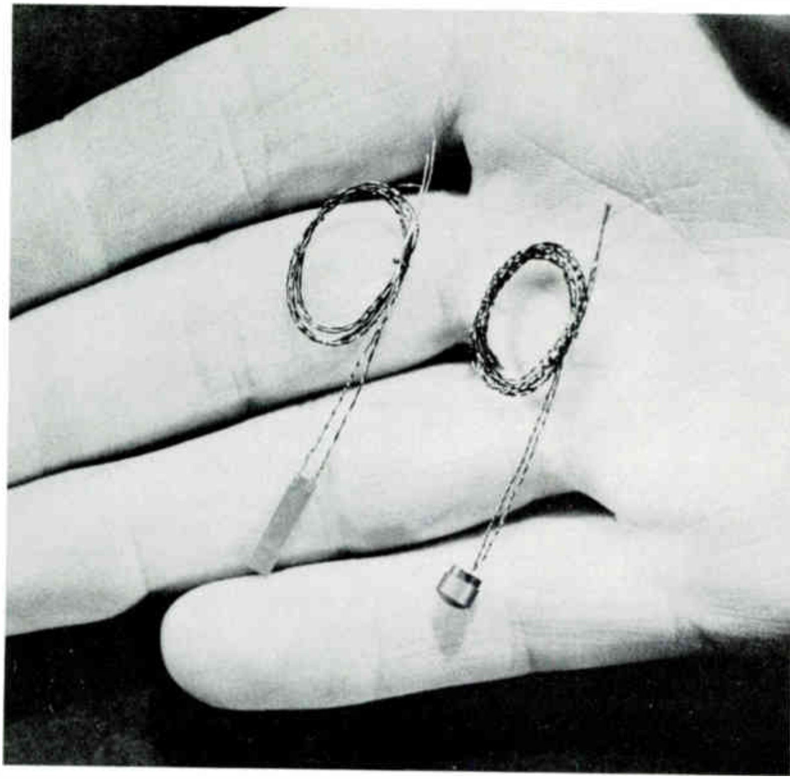


Fig. 4-10. Physical configurations for commercial types of Hall-effect devices.

detector and as a multiplier. Fig. 4-10 shows the physical configurations for commercial Hall-effect devices. There are four output leads—two for current and two for voltage. The same two devices are shown in Fig. 4-11 to demonstrate the relative physical size of the Hall-effect devices.

A simple system for measuring magnetic-flux density around a magnet is to use a Hall-effect device mounted in a probe. The current through the Hall device is maintained at a constant level, and the output voltages will then be proportional to the magnetic-field strength. The Hall voltage is amplified and indicated on a meter. By taking readings from the meter with the probe at different points, it is possible to plot the field around the magnet.

It is also possible to tell how far a Hall-effect device is from a magnet with a given field strength. This is determined by the magnitude of the voltage that is generated. Used in this fashion, it is known as a proximity, or nearness, detector. It can also be used to indicate angular position when used with two magnets. The Hall voltage will be greatest when the element is parallel to the faces of the two magnets. As the



Courtesy F. W. Bell, Inc.
Fig. 4-11. Hall-effect devices of Fig. 4-10.

element turns, the output voltage will become less. Thus, the output voltage depends on the angular position of the Hall element.

PHOTOTRANSISTORS

In Chapter 2 we discussed photosensitive diodes. The same principle may be applied to transistors. The three-element device is mounted in a standard transistor case with a lens at the top. When light is focused on the emitter junction, the transistor behaves like a photodiode, and the photoelectric action is amplified by the gain of the transistor. This often eliminates the need for a succeeding amplifier stage that is usually necessary when a photodiode is used. Fig. 4-12

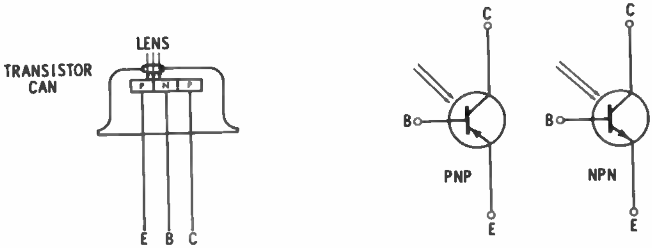


Fig. 4-12. Phototransistor.

shows the basic physical construction of a phototransistor and the schematic representation for the PNP and NPN types. The arrows in the schematic represent the incident light.

Phototransistors are used in a wide variety of applications. They are fairly linear and can be used to measure light intensity when used in a circuit like the one shown in Fig. 4-13. This circuit has disadvantages, however, which usually result in the use of some other means of measuring light intensity. The main disadvantage is that phototransistors operate over a very limited light range. In other words, it doesn't

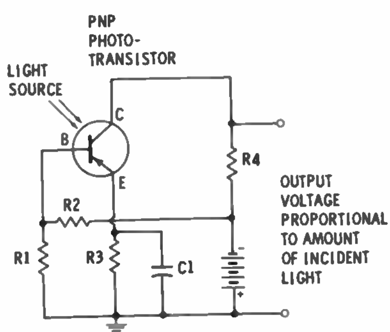


Fig. 4-13. Light level indicator using a phototransistor.

take much light to saturate the transistor. Usually density filters, or graded apertures, are used to extend the range of these devices when a wide range of light intensity must be measured.

Photodiodes, as well as phototransistors, are generally used in switching nonlinear or on-off applications.

CHAPTER 5

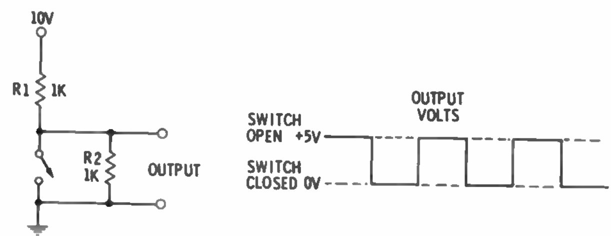
Transistors in Nonlinear Circuits

Almost every semiconductor device, such as a transistor, that is suitable for linear operation can also be used in nonlinear circuits. Nonlinear refers to the fact that the output of the device is not directly proportional to the input. For example, in a simple circuit using a transistor as a switch, the transistor will be either on or off (like a relay). That is, it is either conducting heavily, or it is cut off. When there is no base current, the transistor will be turned off. When there is base current, the transistor will be turned on and an increase in base current will not turn it on further.

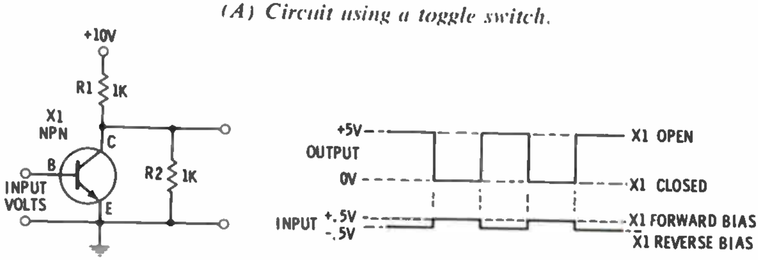
The advent of transistors provided a great step in the development of computers. Transistor switches now replace slow relays and hot, bulky, unreliable vacuum tubes. When applied to transistors and, for that matter, vacuum tubes, the word "nonlinear" can be used interchangeably with "large signal." In linear operation, a small signal input results in minimum distortion and, therefore, linear operation, whereas under large-signal conditions a large input signal overdrives the input circuit, causing distortion in the output and, therefore, nonlinear operation. It is the purpose of this chapter to show that nonlinear circuits can be put to good use. Switching transistors in computers are operated as nonlinear amplifiers, but transistors in hi-fi sets should be operated as linear amplifiers.

TRANSISTOR SWITCHES

The easiest way to understand the operation of a transistor switch is to compare it with a toggle switch. A toggle switch can be said to be either on or off. A transistor operating in a switching circuit is also designed to be either on or off. Consider Fig. 5-1A in which a resistor has been placed in series with a toggle switch. The input signal opens and closes the switch. The output depends on the position of the switch. If the switch is open, the output is connected to the positive supply through resistor R1; if the switch is closed, the output lead is shorted directly to ground; and if R2 has the same value as R1, the output voltage will be 5V when the switch is open and 0V when the switch is closed.



(A) Circuit using a toggle switch.



(B) Circuit using a transistor switch.

Fig. 5-1. Switching circuits.

As you will recall from previous chapters, the resistance of a transistor may be controlled by the amount of emitter current allowed to flow. When no emitter current flows, the transistor exhibits a very high resistance, or almost an open circuit from collector to emitter. When heavy emitter current flows, the transistor will be a very low resistance, or almost a short circuit. It thus can be used as a switch and will be open or closed, depending on the bias applied to the emitter junction.

In Fig. 5-1B a transistor is substituted for the switch in Fig. 5-1A. The transistor switch is open when the emitter junction is reverse-biased. The collector current is cut off whenever the emitter junction is reverse-biased. When the emitter junction is heavily forward-biased, the transistor becomes a very low resistance. This condition is referred to as saturation and means that the transistor is conducting just as heavily as possible. A further increase in forward bias on a saturated transistor would result in no increase in emitter-to-collector current. Further increase in the forward bias of a saturated transistor may destroy the transistor by burning out the emitter-base junction. Because of the gain of the transistor, the input voltage in Fig. 5-1B need furnish only 0.5V input to saturate the transistor. Thus, a switching transistor is one used in either of two states—saturation or cutoff. When the transistor is saturated, the switch is closed and the collector or output lead is equivalent to being connected to the emitter. When transistor X1 is reverse-biased, no output is present across resistor R2; transistor X1 is at cutoff and the collector is separated from the emitter by an open circuit. There is now a 5V output across resistor R1.

Fig. 5-2 shows a switching circuit used for turning on an indicator lamp. When the transistor is on, the collector is effectively shorted to

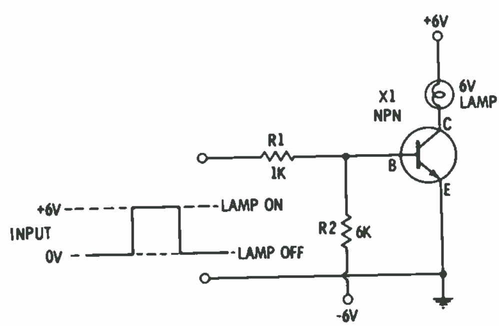


Fig. 5-2. Transistor lamp driver.

the emitter, and the lamp is connected directly across the supply and is lighted. When the transistor is open, or off, the indicator lamp will not be connected to the negative of the supply, but will be floating and the light will be off. The lamp acts as the load resistor for the transistor. When the input is at 0V, the transistor is held at cutoff, since the base of X1 will be connected through resistor R2 to the negative 6V supply. The R1-R2 voltage divider is used to ensure that the emitter junction is reverse-biased, rather than depending on zero bias to keep the transistor at cutoff. Although a transistor will not switch with zero volts on the base, there are many types that are not completely cut off if there is less than .5V of reverse bias. When the input voltage goes to +6V, the voltage at the base becomes positive and the emitter junction is heavily forward-biased. The transistor then saturates, causing the lamp to light.

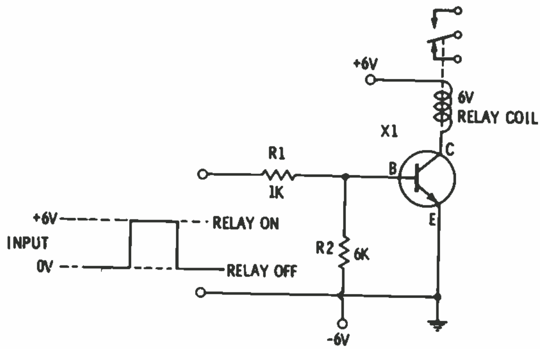


Fig. 5-3. Transistor used to operate a relay.

Fig. 5-3 shows a circuit in which a relay is used as the load for a switching transistor. A very small current at the input of the transistor can be used to turn on a high-current relay. This is possible because the transistor provides current amplification. For example, it might take only 10 milliamps to turn the transistor on and 100 milliamps to turn the relay on. Using the circuit in Fig. 5-2 and a transistor switch with a current gain of 10, the relay can be turned on with 10 milliamps.

TRANSISTOR GATES

You will recall that Chapter 2 provided an introduction to AND OR gates and diode logic. Gates may also be constructed with transistors as the active elements. Fig. 5-4 shows a two-input AND gate using transistors connected in an emitter-follower configuration. Both transistors are biased to full conduction, the current being supplied through resistors R2 and R4. The output voltage remains at −6V until both X1 and X2 are cut off by the application of +6V at signal inputs A and B. Fig. 5-5 shows a transistor OR gate. In this case, current supplied through resistors R3 and R4 maintains both transistors at cutoff.

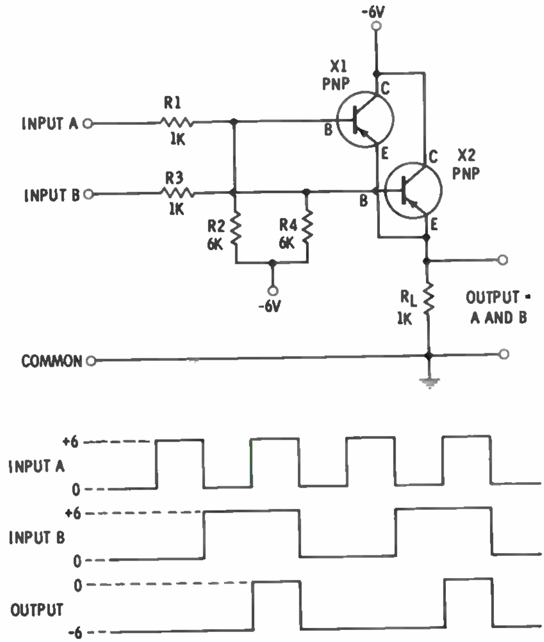


Fig. 5-4. Emitter-follower AND gate.

When either transistor is turned on, the output rises. The output is 0V when both inputs are 0V.

These circuits will function as AND and OR gates, but they are neither the easiest nor the most economical to work with. First, it is easy to see that they require more expensive components than the diode gates mentioned in Chapter 2. Second, notice that the output voltage levels of Fig. 5-4 (−6, 0) are different from the input levels (0, +6) of Fig. 5-5. This makes it impossible to cascade two such stages without using a level-changing circuit between the two gates. Figs. 5-4 and 5-5 do, however, serve to introduce the idea of transistor gates.

In practice, gates like those shown in Fig. 5-6 are used. It is desirable to use transistors rather than diode gates because of the amplification feature. When passive diode gates are used, the output pulses degenerate and lose sharpness after passing through a few gates. This does not happen when transistors are used, because the pulses are re-

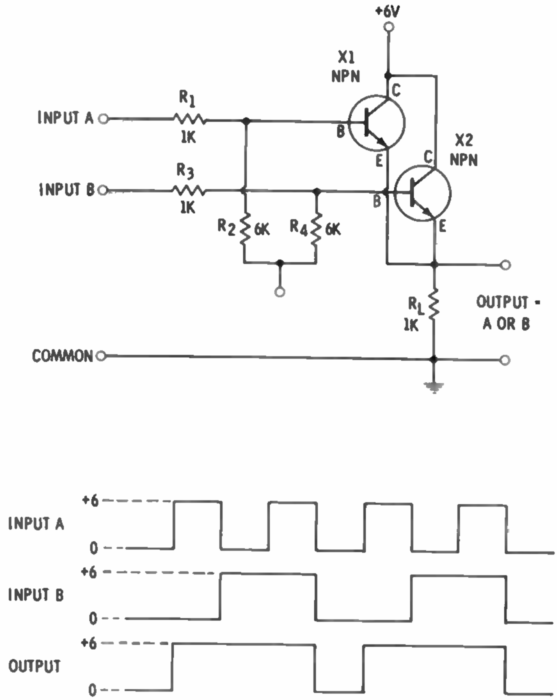
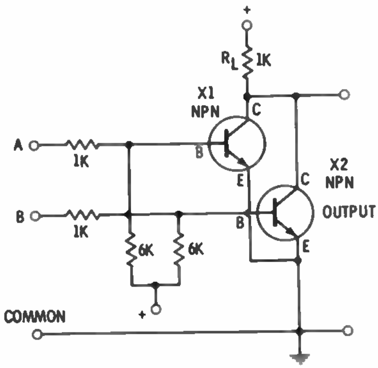


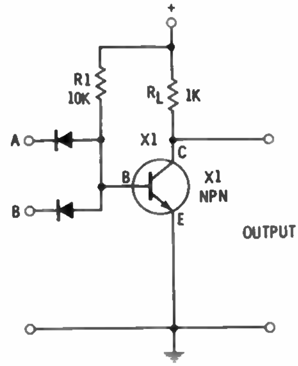
Fig. 5-5. Emitter-follower OR gate.

amplified in each gate. However, each time a common-emitter transistor is used, there is a phase inversion. This means the output is opposite, or inverted, from that at the input. The waveshapes in Fig. 5-6C show this inversion or negation. The dot (•) means AND. Thus A•B is read as "A and B." A plus (+) means OR. Thus A+B is read as "A or B." The bar over a group means NOT or that the letter or group is to be inverted. Thus, $\overline{A \bullet B}$ is read as NOT A and B. Figs. 5-6A and 5-6B are NOT AND gates. That is, they are AND gates where the AND function is inverted. One gate uses transistors and resistors; the other uses diodes, transistors, and resistors.

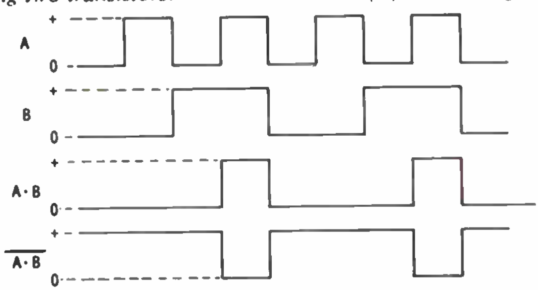
Other common gates are shown in Fig. 5-7. Fig. 5-7A is a NOT AND gate. All three transistors must be turned on in order to make the out-



(A) Gate using two transistors.

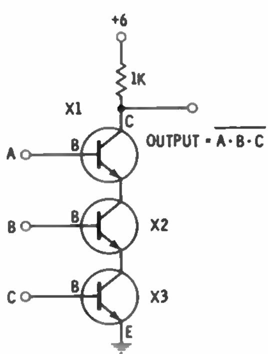


(B) Gate using one transistor.

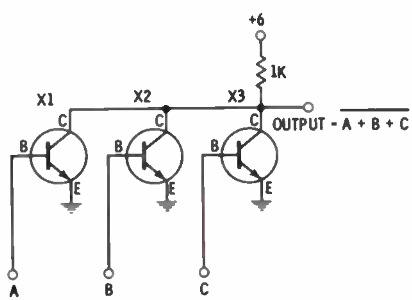


(C) Waveshapes for AND gates.

Fig. 5-6. AND gates.



(A) NOT AND gate.



(B) NOT OR gate.

Fig. 5-7. Typical transistor gate circuits.

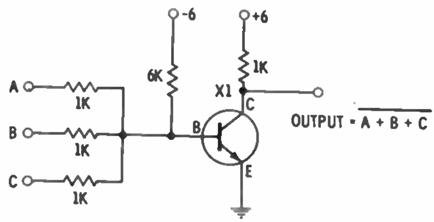
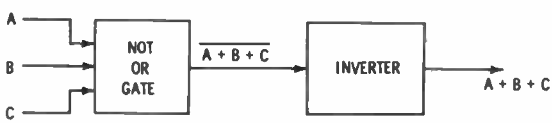
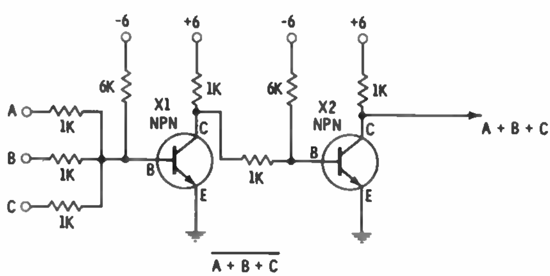


Fig. 5-7C. Another typical NOT OR gate circuit.

put fall to zero. Fig. 5-7B is a NOT OR gate. Turning on any one of the three transistors will short the output to ground. A very popular NOT OR gate is shown in Fig. 5-7C. It is commonly called a NOR circuit. Its popularity is derived from the fact that it uses only one transistor and no diodes. Since resistors are relatively inexpensive, it is an economical circuit. It is well to note that the bar, or NOT, can be



(A) Block diagram of a gate and inverter.



(B) Schematic diagram of a gate and inverter.

Fig. 5-8. Inversion of NOT OR function.

removed from any of these logic gates by merely feeding the output through another inverter (Fig. 5-8A). A schematic diagram of the NOT OR gate and the inverting stage is shown in Fig. 5-8B.

TRANSISTOR MULTIVIBRATORS

A transistor multivibrator is a square-wave oscillator made up of two transistor switches that are alternately turned on and off. When one transistor is on, the other transistor is off, and vice versa.

Astable Multivibrator

There are three basic variations of the multivibrator—the astable, bistable, and monostable. A typical detailed schematic is shown in Fig. 5-9 of a multivibrator that operates at approximately 30 kc. X1 and X2 are the two transistor switches. Notice that the output from the collector of X1 is coupled through C1 to the base of X2, and the output from the collector of X2 is connected through C2 to the base of X1. The purpose of this arrangement is to provide feedback to keep the device oscillating. When X1 turns on, the output signal is fed back to the base of transistor X2 and turns it off. When X2 turns on, its output signal turns X1 off.

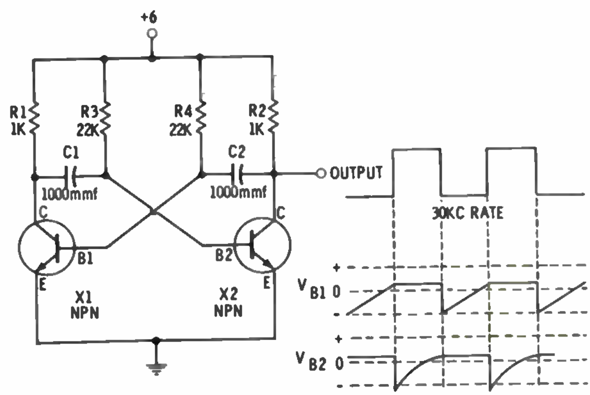


Fig. 5-9. Astable multivibrator.

Let us take a detailed look at what happens in a multivibrator. To begin the oscillating cycle, some inequality in the transistor switches causes one to conduct more than the other. Suppose X1 has just turned on; this means the collector potential went toward ground. This happens so fast that C1 does not have a chance to charge up; and the entire negative change appears across R3, which reverse-biases X2, thus cutting it off. X2 stays off and X1 stays on until C1 charges to the point where X2 is no longer reverse-biased. When this happens, X2 begins to conduct and its collector potential starts toward ground. This feeds a negative pulse through C2 to the base of X1, turning it off. In a like manner, X1 remains off until C2 charges completely. The frequency of oscillation is determined by the resistance and capacitance values because they determine the charge and discharge time. The output of this multivibrator is a series of pulses on each collector.

The waveshapes in Fig. 5-9 illustrate how the oscillations are maintained. Only one switch (either X1 or X2) is on at a time. While one transistor is on, the other one is off. The off transistor remains off until the capacitor in the base circuit charges to the point where the emitter junction is no longer reverse-biased. Then when this transistor begins to conduct, it in turn turns off the other transistor, and the entire cycle is repeated. Multivibrators are used in all kinds of electronic equipment—particularly in timing circuits, where the output pulses are used to synchronize the operation of other circuits.

Bistable Multivibrators

Once the basic multivibrator circuit is understood, a few simple alterations produce some interesting and useful circuits. The multivibrator circuit just discussed is a free-running type; that is, it is not stable in either of two states. It is not stable because the feedback

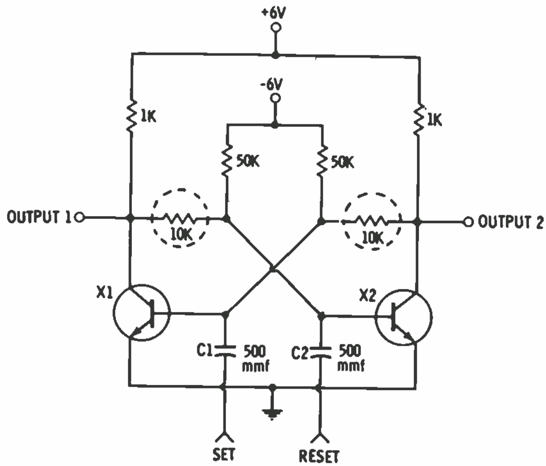


Fig. 5-10. Bistable multivibrator.

coupling from one switch to the other is capacitive and a capacitor can transfer only voltage and current changes. Thus the circuit keeps switching back and forth. By replacing the capacitor with a resistor (Fig. 5-10), the multivibrator can be made into what is known as a bistable multivibrator. This means that it is stable with either side conducting. However, by pulsing the circuit in the proper way, it can be made to change from one side to the other. This circuit is also known as a flip-flop.

If X1 is on, X2 will be off and the flip-flop will remain in this state until a pulse is entered, which starts the process discussed for the free-running multivibrator. A pulse on the set terminal will turn on transistor X1 and turn off transistor X2. A pulse on the reset line will cause the bistable to flip, turning on transistor X2 and turning off transistor X1.

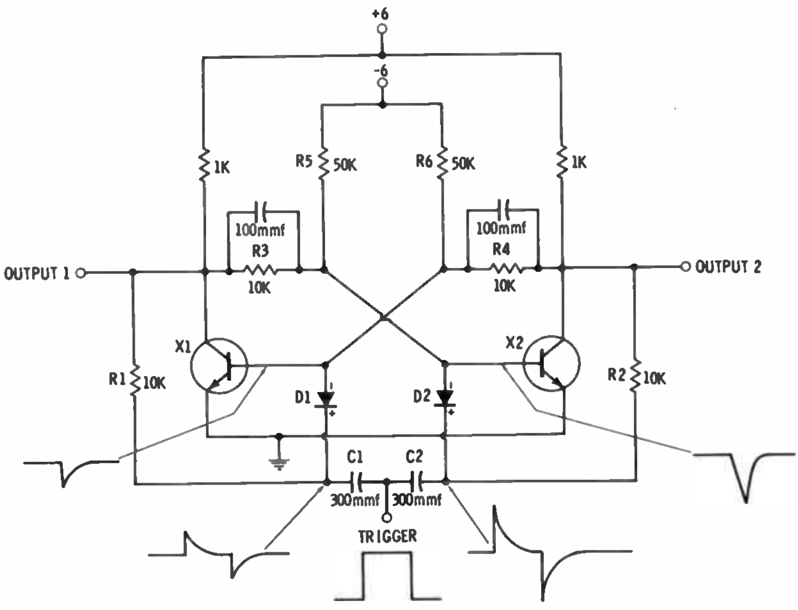


Fig. 5-11. Triggered flip-flop.

When its state changes, the multivibrator remains in that state until another pulse comes along that will again change its state. Fig. 5-11 shows a practical triggered flip-flop circuit. In this circuit a trigger pulse will cause a change of state, regardless of the present condition. R1-R2, C1-C2, and D1-D2 make up a "steering" network that directs the pulse to the proper transistor. A large negative pulse is fed to the on transistor, causing it to cut off; as a result, the device flips to the other state. Two capacitors across the feedback resistors (R3 and R4) speed up the switching time by providing an easy path for the high-frequency portions of the pulses. They are known as commutating capacitors because they help to speed the commutation of the flip-flop from one state to the other.

A pulse introduced at the trigger lead will be differentiated by capacitors C1 and C2 (Fig. 5-12). This results in a positive and a

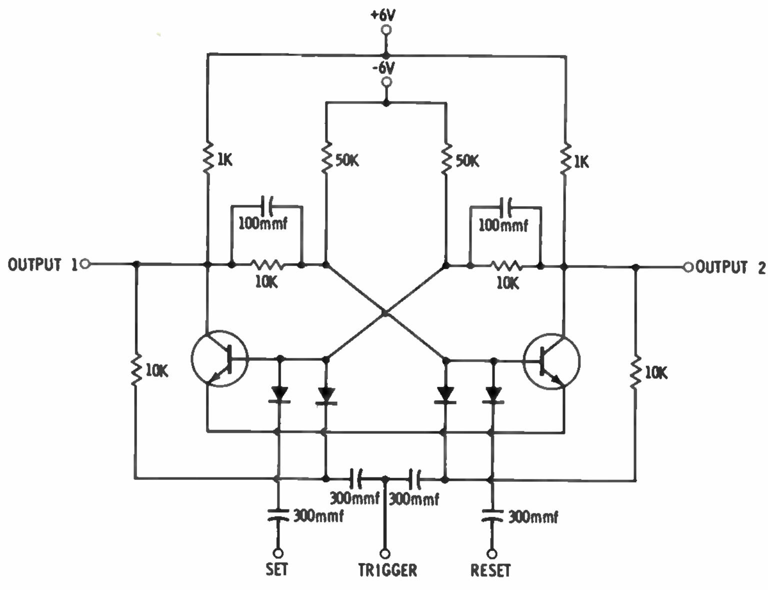


Fig. 5-12. Set-reset trigger flip-flop.

negative spike. The two diodes permit only the negative spike to reach the bases. Suppose that when this happens, transistor X2 is conducting and transistor X1 is off. When X2 is on, its collector is at zero volts. This means that R4 and R6 form a voltage divider between −6V and 0V, thus placing the base of X1 at some positive voltage that keeps it cut off. The negative pulse that reaches the base of X1 is not large enough to cause an effect. The pulse presented to the base of X2, however, is large enough to cut it off. When X2 cuts off, feedback to X1 is such that it is turned on. After the capacitors have charged and discharged, the circuit remains in the state where X1 is conducting and X2 is off. Fig. 5-12 displays the schematic of a flip-flop that contains both trigger and set-reset capabilities. This multivibrator contains both features of the two previous bistable units.

Monostable Multivibrators

Another circuit in the multivibrator family is called the monostable multivibrator. We have discussed the astable multivibrator and the bistable multivibrator. The monostable multivibrator is one which is stable in only one mode rather than both. This means that when you trigger a monostable, it will change states and then revert to the stable state. It is not free-running, but neither is it bistable; it returns to its original state. As a result, the monostable multivibrator becomes an extremely useful single-pulse generator. Again, this function is accomplished by merely adjusting the feedback circuit to the bases of the two transistors in the basic multivibrator. By having a capacitor in one feedback line and a resistor in the other feedback line, the device will be stable in one state and unstable in the other.

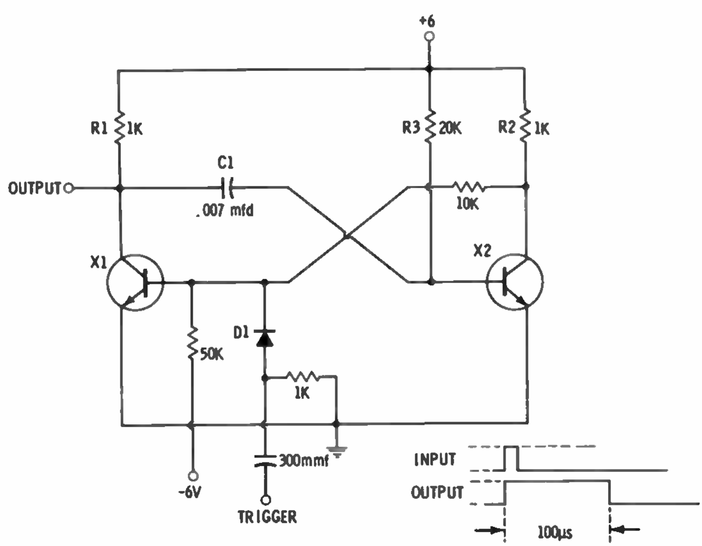


Fig. 5-13. Monostable multivibrator.

Fig. 5-13 shows a circuit for a 100 μsec multivibrator. Assume transistor X1 is on and transistor X2 is off. The input trigger circuit is such that an incoming positive pulse will turn X1 off. This will turn X2 on, which tends to hold X1 off. X2 will stay on only until C1 has discharged. Thus the size of C1 and R3 determines the width of the output pulse. As C1 discharges, transistor X2 turns off again, and this action turns X1 back on. The device will settle back to its original position, with the result that the multivibrator action produces an output pulse having a different width from that of the input pulse. The monostable multivibrator is often used as a pulse stretcher. Every time it receives an input pulse, it produces an output pulse having a width that is determined by the time constant of the circuit.

COUNTERS

The bistable multivibrator is often used as a counting device or a memory element. Each flip-flop can count up to two. We shall see shortly that if you consider one output, the number of output pulses will be half the number of input trigger pulses. By using a number of flip-flops, the circuit can be made to count as high as desired by two's. This system of counting by two's is called the "binary" system. Remember that in the decimal system we count by ten's. Electronic counters use the binary system simply because a two-state device (bistable) is a lot easier to build than a ten-state device. The flip-flop can also be used as a memory device, because if its state is changed, the flip-flop will remember this change until it is changed back. If a flip-flop is used as a temporary memory in a computer system, the computer can come back much later and look at the flip-flop and see that its state was changed.

Fig. 5-14 shows how a flip-flop divides by two. It takes two input pulses to get one complete output pulse. The "0" and "1" are merely a convenient way of indicating that the transistor switches are on or off. For every two input pulses, the flip-flop will be back to where it was in the beginning. If we put flip-flops following the flip-flop, as is shown in Fig. 5-15, the output of each flip-flop will be divided by two in the next stage. It can thus be seen that every pulse causes flip-flop 1 to make a transition. Flip-flop 2 makes a transition for every other pulse, flip-flop 3 counts every fourth pulse, and flip-flop 4 makes a

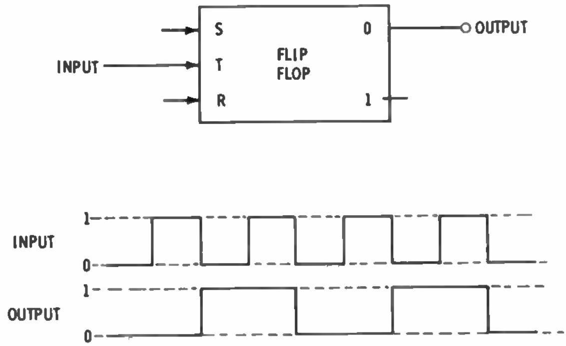


Fig. 5-14. A flip-flop divider.

transition for every eighth pulse. Using four flip-flops in this fashion, one feeding the other, it is possible to count from 1 to 15. After the fifteenth count, all the flip-flops go back to the zero state and the count will start all over again. Each additional flip-flop enables the counter to count twice as high. One flip-flop can count to 2; two can count to 4; three can count to 8; four can count to 16; five can count to 32, etc.

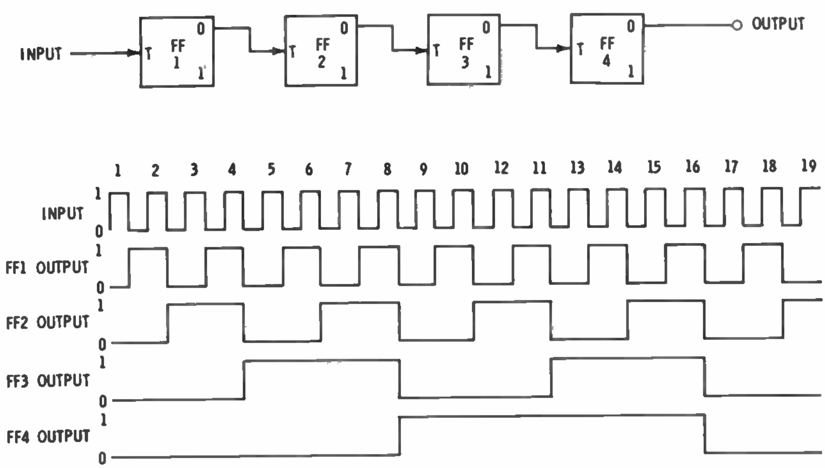


Fig. 5-15. Four-stage flip-flop divider.

Counters similar to this scheme are used in many places in modern computers. It is with devices very similar to these that all the addition, subtraction, multiplication, and division in computers are accomplished.

Since many systems exist where hundreds and perhaps even thousands of flip-flops are used, each with identical schematic diagrams, it would be senseless to draw the entire schematic more than once for the drawings and blueprints of the system. This is particularly true for computer systems. When a drawing is made up for a computer, not only are flip-flops shown as blocks, but so are other logic elements, such as AND gates, OR gates, amplifiers, etc. Each one of these blocks is

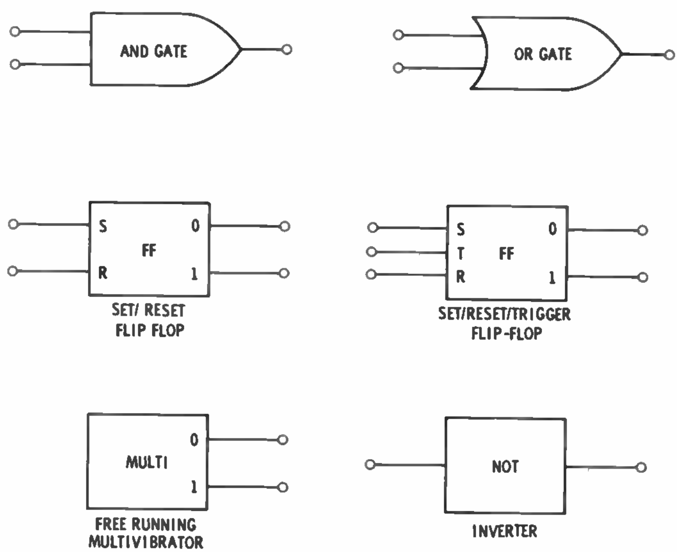


Fig. 5-16. Block diagram of logic elements.

identical to others of the same type, and therefore nothing would be gained from repeating the schematic. Fig. 5-16 shows some of the block representations for logic elements. A system diagram made up of these blocks is called a logic diagram. Each block performs a certain logic function. The logic designer is not concerned so much with the circuitry that goes into the blocks as he is with the interconnection of the blocks to perform a certain job. We will see a small example of this shortly.

If the circuit diagram were drawn out each time one of these devices were used, the result would be so complicated that it would be almost impossible to decipher. Therefore, logic diagrams are used. Consider a simple example of how these devices are used; suppose that an operator of a certain computer pushes two buttons periodically to perform separate operations. Let's call them buttons A and B. We want the computer to keep track of how many times button A is pushed. After it has been pushed three times, we want a flashing yellow lamp to light on the front panel of the computer. After either button A has

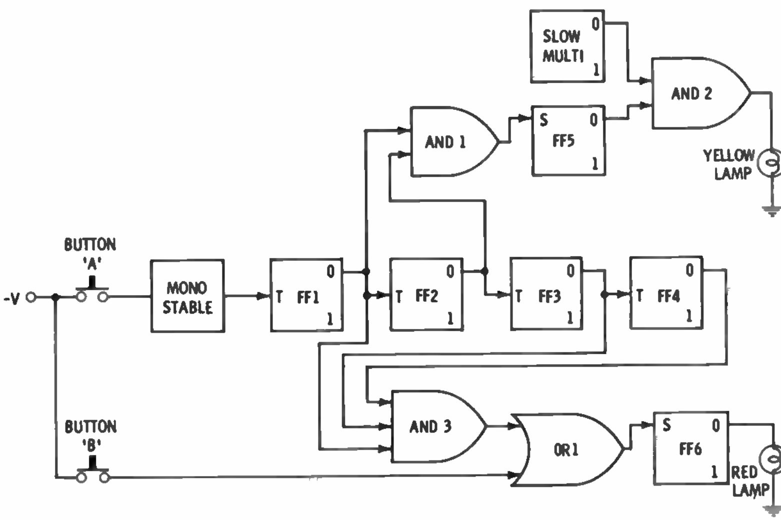


Fig. 5-17. Logic diagram.

been pushed thirteen times or button B has been pushed once, we want a red light to be turned on. Fig. 5-17 shows a logic diagram for performing these functions. FF1, FF2, FF3, and FF4 make up a counter for keeping track of the number of times button A is pushed. Each time it is pushed, the monostable generates a pulse which is counted. After the operator has pushed the button three times, FF1 and FF2 both have a "one" in the output. This energizes AND-1. Thus, this gate will have an output when the count is 3. FF5 is then set, enabling AND-2, allowing the pulses to get through to the yellow lamp. Since AND-2 will have an output only when the multivibrator is on, the lamp will flash in synchronization with it. The purpose of FF5 is to remember that count "3" has been reached, even after it has passed. When the count for button A is 13, AND-3 will have an output that passes through OR-1, setting FF6 and turning on the red lamp. The red lamp will also be turned on if button B is pushed. The purpose of FF6 is to remember that the "13" count was reached, even after it has been passed. If FF5 and FF6 were not in the diagram, the lamps would stay lighted for only one count. To start all over, a reset pulse would be entered into the reset leads of all FF's.

When it comes to understanding computer operations, the problem is not tied up in the circuitry; it is in understanding the logic. Most circuits are simple, but the way the logic blocks are interconnected is the difficult part. The computer then can be considered as a lot of simple blocks interconnected to perform a complex function. Each of these blocks contains transistor circuitry.

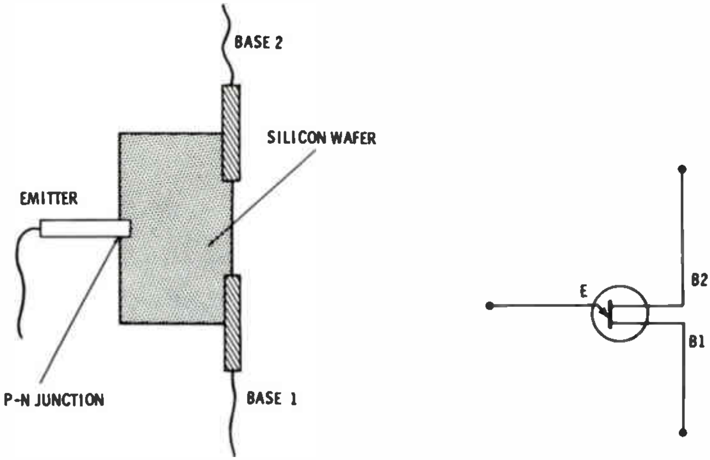
CHAPTER 6

Special Semiconductor Devices

Research into the basic fundamentals of semiconductor theory has produced a variety of devices, each of which performs a particular, specialized function. Among these is the unijunction transistor that is used primarily as a pulse generator, as well as the four-layer device used as a switch or controlled rectifier.

UNIJUNCTION TRANSISTOR

The unijunction transistor is so named because it has only one junction. Its operation differs considerably from that of the ordinary transistor. Fig. 6-1A shows a pictorial diagram of a unijunction transistor. It consists of two bases and an emitter. The two bases are connected to a semiconductor bar of N-type silicon, and the emitter forms a P-N junction with the silicon. When the voltage applied across bases B1 and B2 is 20V and the emitter junction is located halfway between the two base contacts, the silicon bar will act as a voltage divider and +10V will exist at the emitter. The resistance between the emitter and B1 will be relatively high. In order to forward-bias the emitter junction, an input voltage of −10V must be applied between the emitter and base 1. When a voltage sufficient to overcome this reverse bias is applied to the emitter, the unijunction transistor *fires* and the resistance between the emitter and B1 becomes very small.



(A) Pictorial. *(B) Schematic.*

Fig. 6-1. Unijunction transistor.

The best way to understand the operation of a unijunction transistor is to consider it in a simple application. Fig. 6-2 shows the unijunction transistor in a typical relaxation oscillator circuit. Assume that resistor

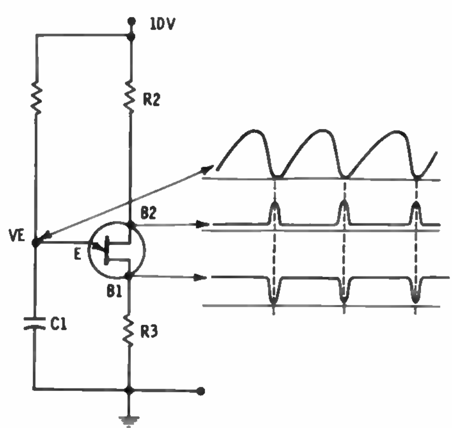


Fig. 6-2. Unijunction relaxation oscillator.

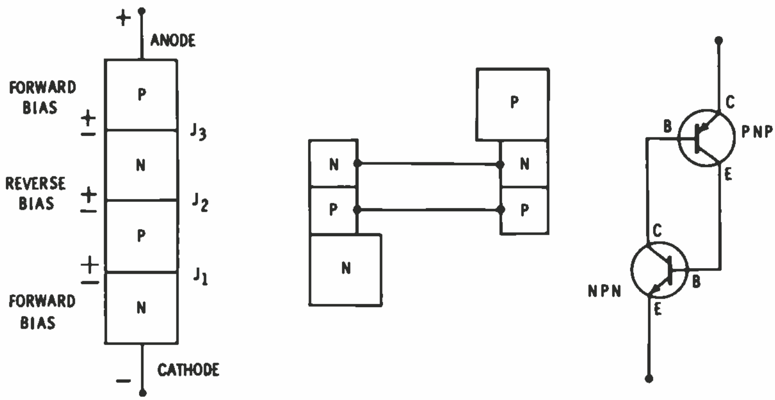
R1 and capacitor C1 are not in the circuit. A portion of the positive voltage will appear across the silicon bar, and the emitter junction will be reverse-biased. In order to overcome this reverse bias, it is necessary that the emitter become more positive than the N side of the P-N junction. R1 and C1 are used to produce this voltage. C1 will charge through R1 from the +10V supply. When the voltage across C1 reaches a point where it forward-biases the emitter junction, the resistance of the bar from emitter to base 1 becomes very low and this circuit conducts heavily. Capacitor C1 discharges through the unijunction and resistor R3, providing an output pulse across R3. When there is a positive pulse at B1, B2 will produce a negative pulse because the B1-B2 resistance is lowered, thus bringing B2 closer to ground. After the discharge of capacitor C1, the voltage at the emitter is nearly zero, the junction is reverse-biased, and the process repeats. The waveshapes shown in Fig. 6-2 indicate how C1 charges through R1 until the unijunction fires and C1 discharges through R3. The frequency of the output pulses is determined by the values of resistor R1 and capacitor C1.

The unijunction is often used as a pulse generator in timing applications. Its main advantage over a transistor pulse generator is that it requires fewer components, as can be seen by comparing it with the circuit of a free-running multivibrator. In addition, it is easy to calculate the frequency of the output pulses from the RC time constant of R1 and C1. Unijunctions are often used in the triggering and firing circuits of silicon-controlled rectifiers.

FOUR-LAYER DEVICES

Fig. 6-3 shows the pictorial schematic symbol for a P-N-P-N four-layer device. Voltage drops across the junctions in Fig. 6-3A are caused by leakage currents that forward-bias junctions J1 and J3 and reverse-bias junction J2. Until J2 becomes forward-biased, the device will present a very high resistance. When the voltage applied across the device becomes large enough, junction J2 breaks down and the device becomes a short circuit.

Fig. 6-3B shows the transistor equivalent to the four-layer device. If enough forward voltage is applied from the emitter of the P-N-P to the emitter of the N-P-N, the transistor pair can be made to conduct



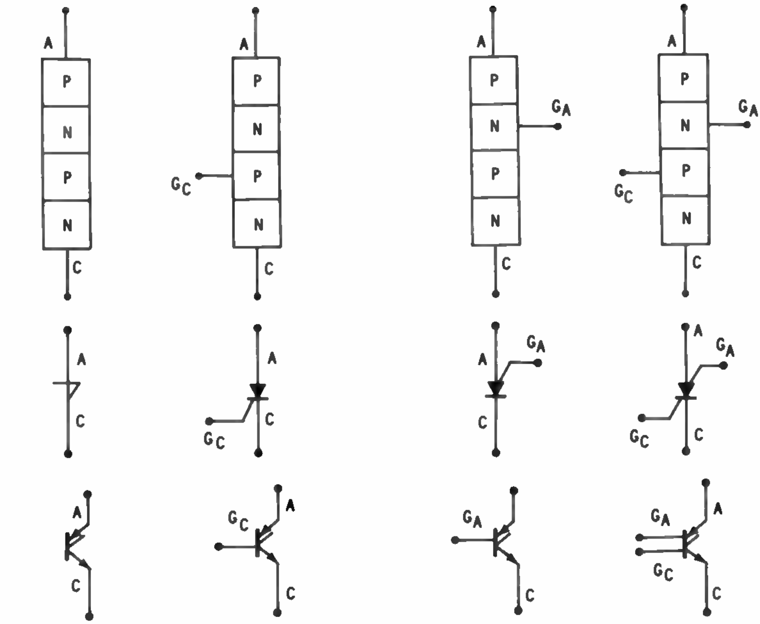
(A) Description of structure.

(B) Transistor equivalent of 4-layer device.

Fig. 6-3. 4-layer semiconductor device.

heavily. As mentioned, the natural voltage drop across the junctions causes J1 to be forward-biased along with J3. By increasing the voltage across the entire device, more electrons crowd into the P-side of the J2 region and more holes crowd into the N-side of J2, finally providing enough minority carriers to cause J2 to start conducting. At this point the device fires. Junction J2 becomes forward-biased because of heavy current carrier traffic and the device exhibits a very low resistance from anode to cathode. Once the device is conducting heavily it is necessary to reverse-bias J2 to turn it off. The most common way to turn off a four-layer semiconductor device is to remove the anode potential. It will, however, turn off if the current is allowed to drop below a certain minimum value.

Fig. 6-4 shows the possible terminal connections of four-layer P-N-P-N devices. The diagrams directly below the pictorials are the symbols generally used in schematic diagrams. The schematics at the bottom are those that are used to a lesser extent. As shown in Fig. 6-3, if a large enough voltage is applied from the anode to the cathode, the four-layer device may be made to short-circuit and conduct very heavily. It is possible to control the firing point more accurately and with more flexibility with a lead on one of the other layers. These



(A) Shockley, or 4-layer diode. *(B) Silicon-controlled rectifier.* *(C) Complementary SCR.* *(D) Silicon-controlled switch.*

Fig. 6-4. Types of 4-layer devices.

are called gate leads. A lead can be put on the N-layer, the P-layer, or both. This accounts for all possible types. The gates enable the introduction of current carriers and thus give greater control over the point at which the device will fire.

The four-layer diode shown in Fig. 6-4A is sometimes referred to as a Shockley diode. It has no gate leads and is turned on when the applied voltage exceeds the breakover voltage; it is turned off when the supply voltage is removed.

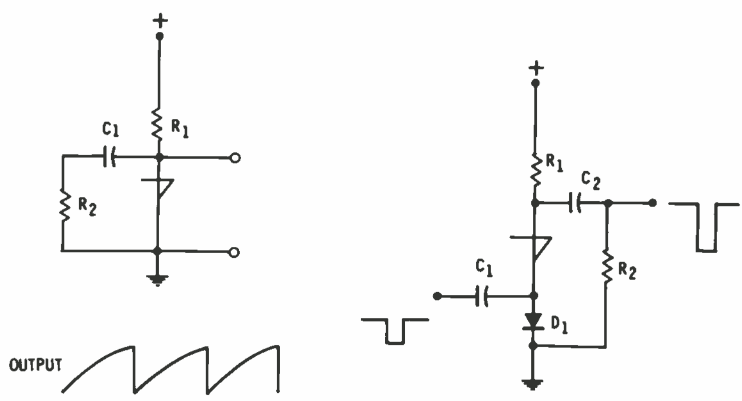
Fig. 6-4B shows a diagram for a silicon-controlled rectifier. This type of rectifier is nothing more than a four-layer diode with a gate lead connected to the second layer.

Fig. 6-4C shows what is called a complementary SCR. The gate is on an N-type material instead of a P-type material. This means that the opposite type of current carriers used for the gating functions.

The SCR is in wide use in a multitude of circuits. It is noted for its ability to handle large currents at relatively high voltages and is widely used in AC power circuits.

The device represented in Fig. 6-4D is a silicon-controlled switch (SCS). It embodies all the advantages of the other three devices. By using or not using the gate leads it is possible to employ this device in place of any of the others. In addition, it has a certain flexibility that enables it to be used in unique circuits where others cannot be used.

The silicon-controlled rectifier is the solid-state counterpart of the thyratron vacuum tube. If you are familiar with the thyratron you will remember that it is a gas-filled tube across which a voltage is applied. A firing voltage is then applied to the grid, causing the tube to conduct. During conduction, the voltage on the grid has little or no control over the current flowing through the tube. The SCR performs in the same way. A voltage is applied across the SCR and the voltage on the gate lead controls the point at which the SCR breaks down and starts conduction. When the SCR has started conduction, the voltage on the gate lead has no effect and the only way the device can be turned off is to remove the anode voltage to the SCR.



(A) Sawtooth generator. *(B) Pulse amplifier.*

Fig. 6-5. 4-layer-diode circuits.

Fig. 6-5 shows two typical circuits for a Shockley diode. The sawtooth generator of Fig. 6-5A operates very simply. C1 charges through R1 and R2 until the breakover voltage of the four-layer diode is exceeded. The device fires and C1 discharges; the diode shuts off and the process repeats. Fig. 6-5B shows a pulse amplifier. A negative input pulse causes the diode to fire and produce a high-power output pulse. This type of circuit is often used for triggering squibs, an explosive charge used in missiles and rockets to separate unwanted parts. For example, on launch, the umbilical cord is blown away by a squib, and during flight, sections are separated by squibs.

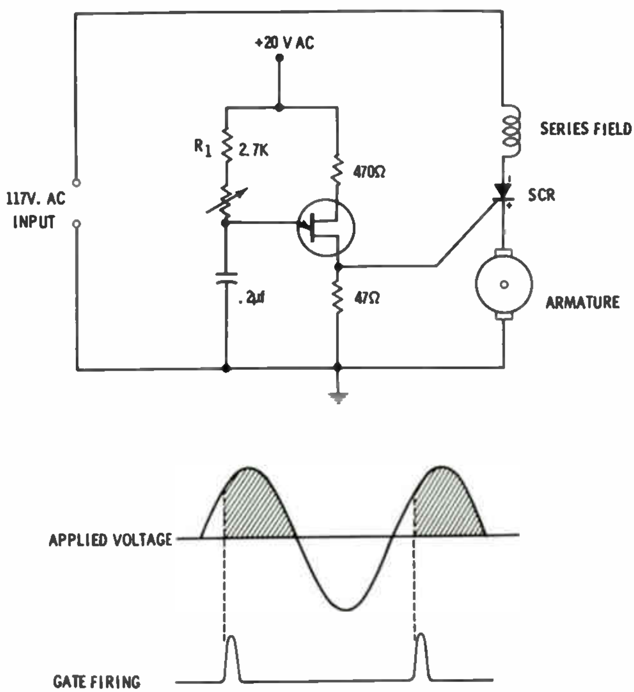


Fig. 6-6. Speed control for motor.

Fig. 6-6 shows a simple SCR circuit used for a motor speed control. This control is accomplished by varying the firing angle of the applied voltage with R1. Adjusting R1 changes the trigger point of the unijunction oscillator, and determines the point at which the SCR begins conduction and supplies voltage to the motor. The SCR can be made to fire at any point on the positive half cycle of input voltage. Thus, the motor speed can be controlled over a given speed range by adjustment of R1. The negative half cycles serves to turn off the SCR until the next positive half cycle.

PHOTOTRANSISTOR USED AS A SWITCH

Fig 6-7 shows a circuit for a simple light flasher employing a phototransistor. A phototransistor is equivalent to a normal transistor plus a photocell. In Fig. 6-7, when the lamp illuminates the phototran-

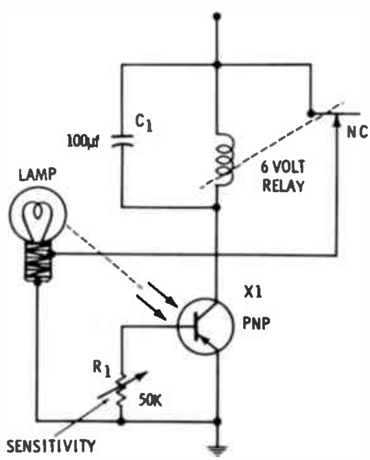


Fig. 6-7. Flasher using phototransistor.

sistor, it causes transistor X1 to become forward-biased and switch on. This pulls in the relay, opening the normally closed contacts. The lamp extinguishes, the photocell resistance increases, transistor X1 turns off, the relay drops out, and the lamp turns on to start the cycle again. R1 is the sensitivity control and capacitor C1 slows down the relay action so that the flasher frequency is not too high.

TUNNEL DIODE USED AS A SWITCH

In Chapter 4 it was mentioned that the best linear applications for tunnel diodes are at higher frequencies, where other components will not operate properly. This same idea holds true for tunnel-diode switching circuits. At present their main use is in high-frequency computer circuits.

Referring to Fig. 4-6, a tunnel diode may be made to switch from Region I to Region III and back to Region I again. Used this way it can perform the function of bistable and monostable circuits, as well as a square-wave multivibrator. Fig. 6-8 shows a simple tunnel-diode

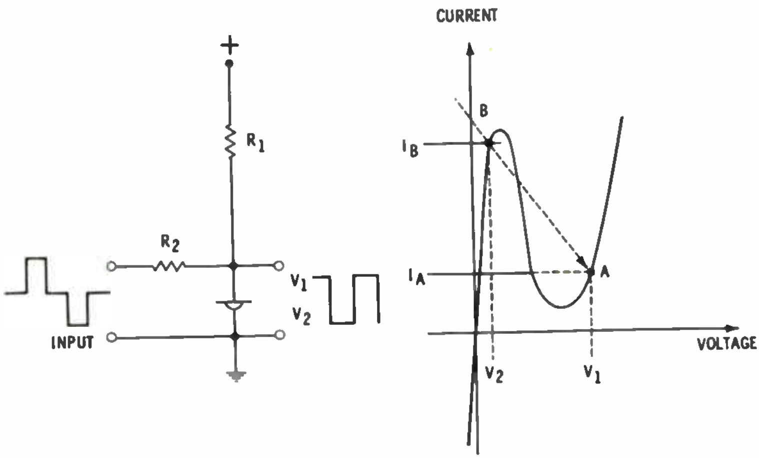


Fig. 6-8. Tunnel-diode circuit and characteristic curve.

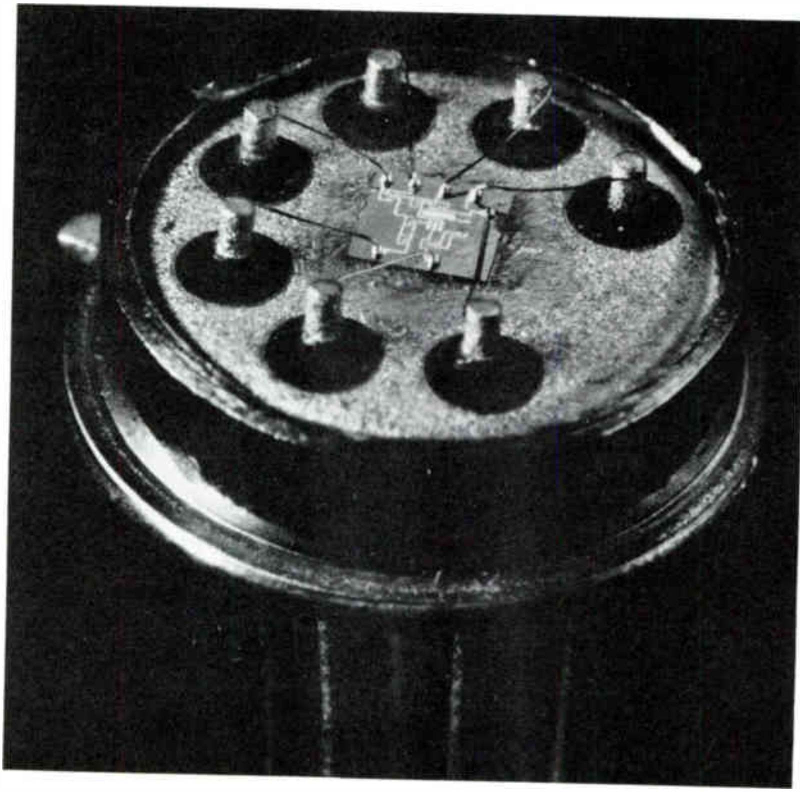
bistable circuit along with the characteristic curve to facilitate the explanation of switching. If the voltage and current are such that the diode operates at point A on the characteristic curve and a positive pulse is entered at the input, momentarily increasing the current and decreasing the voltage, the operating point on the curve will change to point B. Thus the trigger pulse switches the diode from one stable state to another. It can be switched back by a negative trigger pulse, causing the values to switch back to those represented by point A.

CHAPTER 7

Integrated Circuits

We are all familiar with the standard transistor package and have to admit that, compared to vacuum tubes, transistors are quite small. The discovery of transistors has brought startling reductions in the size of electronic equipment: radios are truly vest-pocket size, hearing aids are now so small that "even your friends won't know," and television receivers are actually portable. And yet, if we take a close look at the cutaway sketches of transistor envelopes in Chapter 3, we will conclude that in the tiny transistor envelope the greatest percentage of the space is wasted. The actual transistor occupies only a small fraction of the available room. The only factor determining the size of the transistor is the method by which it will be handled. People and machines have to be able to pick it up, attach leads to it and place it into a circuit. With modern techniques steadily improving, the smallest possible transistor gets smaller every day.

Integrated circuit techniques enable dozens and even scores of components to be mounted in the space ordinarily occupied by a single transistor. This means that entire circuits can be placed on a semiconductor chip hardly big enough to pick up. The word "integrated" is derived from the fact that all the components, transistors, resistors, and capacitors are fashioned from semiconductor material. There are no discrete components; that is, there are no resistors or capacitors that can be put in or taken out of the circuit. They are all integrated into one circuit function. There are no solder joints, and the only connections involved are those leads which connect the integrated circuit to the associated circuits. Fig. 7-1 is a photograph

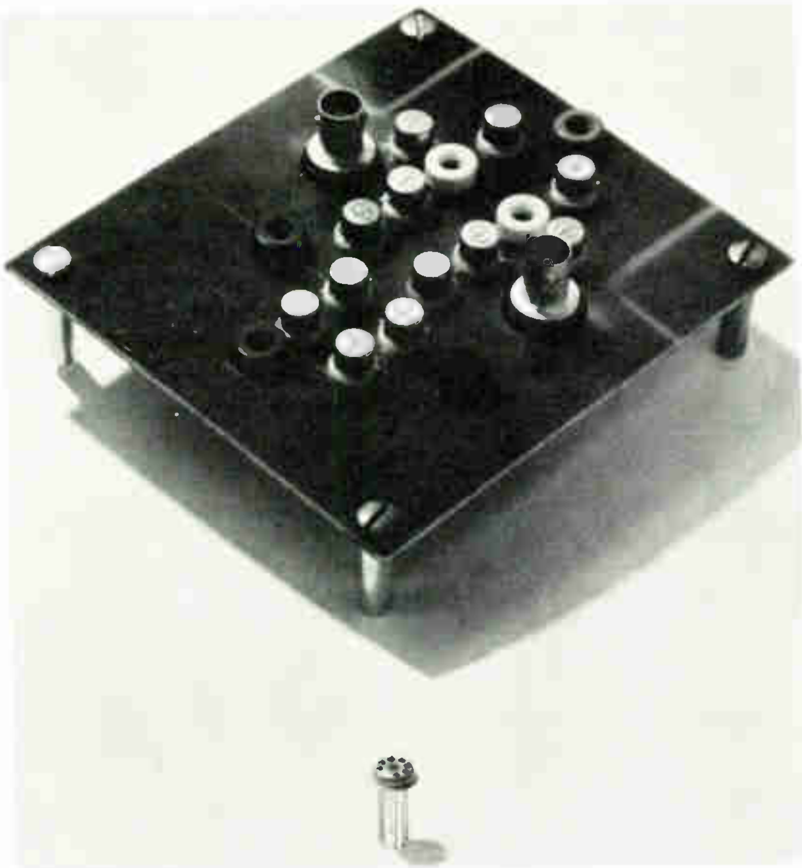


Courtesy Fairchild Semiconductor

Fig. 7-1. Microcircuit attached to a 7-lead heater.

of a typical microcircuit mounted on a seven-lead transistor-type header. The top of the can has been cut away.

As in the case of putting many vacuum tubes into one envelope, it would seem that a respectable goal would be to put more than one transistor in a can. The transistor manufacturers went one better; they put complete circuits and even groups of circuits in the space previously reserved for one transistor. As a matter of fact, as many as 15 or 20 transistors and all the associated resistors and capacitors can be put onto a semiconductor chip measuring a few square millimeters. And to add to what already seems to be a remarkable achievement, these integrated circuits are far more reliable than the corresponding



Courtesy Fairchild Semiconductor

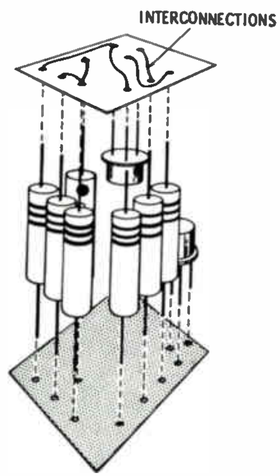
Fig. 7-2. Microcircuit shown beside the conventional circuit it replaces.

circuit built from standard components and, in most cases, they are less costly! Fig. 7-2 demonstrates one example of the circuitry that can be packaged in a single unit. The solid-state circuit in the upper portion of Fig. 7-2 is replaced by the microcircuit shown below it.

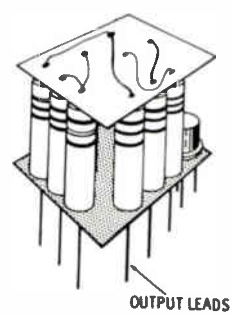
The saving of size alone would not justify these small packaging concepts if cost and reliability had to be sacrificed. Consider the fact that microcircuits are basically more reliable than standard circuitry, the number of solder connections and welds are reduced, and the resulting circuit is much more reliable. The techniques which have been developed and mastered by the manufacturers have resulted in mass-production methods which turn out circuits economically competitive with standard discrete component circuits.

The fully integrated circuit is one where the components, transistors, resistors, and capacitors are all made from semiconductor or thin film materials inseparably associated with each other on a silicon wafer or chip. Unlike standard circuitry, it is impossible to remove a component from an integrated circuit. This, of course, means that defective units must be thrown away and replaced with a new one.

A number of microelectronic approaches preceded the fully integrated circuit. The great-grandfather of the integrated circuit involves



(A) Exploded view.



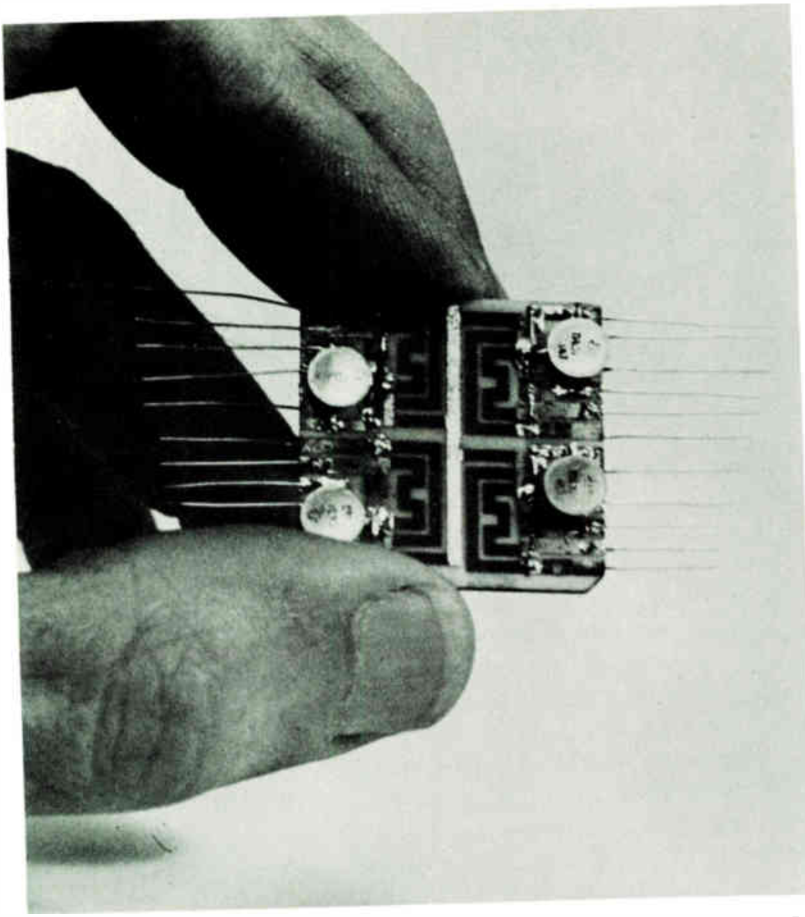
(B) Normal.

Fig. 7-3. Cordwood module.

high-density packaging using discrete components. This generally takes the form of what is known as the "cordwood" approach to packaging. Fig. 7-3 shows a sketch of a cordwood module. It uses standard components placed as close together as possible. It gets its name from the fact that the components are stacked the same as cordwood. The cordwood method is still in wide use and for many applications it is the best package.

The next advance was to use uncased transistors, or transistor "chips" as they are called. This substantially reduced the package size by eliminating the "wasted" space inside the transistor can.

The next step closer to fully integrated circuits is the hybrid circuit,



Courtesy Lockheed Missiles and Space Co.

Fig. 7-4. Titanium thin-film circuits on a single substrate.

such as the one show in Fig. 7-4. These use discrete active components (transistors) and thin-film passive components. Thin metallic films are deposited on a substrate, usually some kind of ceramic, to make up resistors and capacitors. The active components are then mounted on the substrate to complete the module. Resistors are made by controlling the shape and resistance of the metallic film. Capacitors are formed by placing a nonconductive dielectric between two layers of

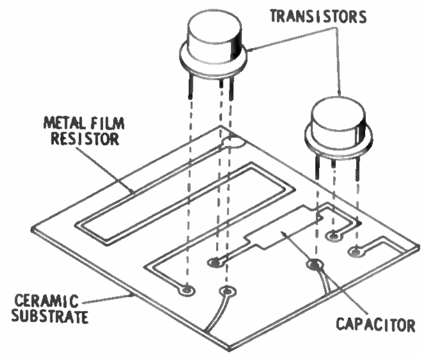
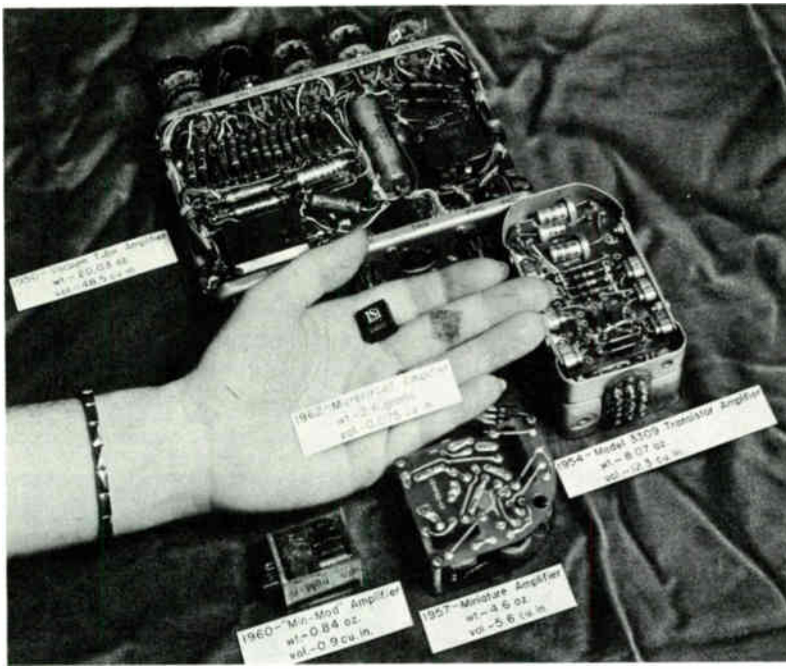


Fig. 7-5. Hybrid thin-film circuits.

deposited thin metallic films. Fig. 7-5 shows how the basic thin-film hybrid circuits are fabricated. Fig. 7-6 shows how the size of an airborne servo amplifier has been reduced over the years by advances in microcircuitry techniques. The original amplifier was a vacuum tube version and the present one is a thin-film hybrid.

Many of the approaches to microelectronics are still very much in use. As a matter of fact, advancements and improvements are being made within the confines of each concept, but the ultimate in small size, reliability, and cost is the fully integrated circuit.

Let us take a closer look at how a fully integrated circuit is processed. It is constructed using the planar-passivated method described in Chapter 3. Fig. 7-7 shows a more detailed picture of how this process works. It begins with an N-type silicon wafer that is oxidized. The impurity material will not diffuse into the passivated layer and therefore the oxide is first etched away over the area where the diffusion is to take place. A PN junction is then diffused into the wafer and another layer of oxide is grown. A certain portion of this is in turn etched away to prepare for the next diffusion. After the complete NPN transistor has been constructed, the base, emitter, and collector contacts are metallized and leads are attached.



Courtesy Lear Siegler, Inc.

Fig. 7-6. Size reduction of an amplifier.

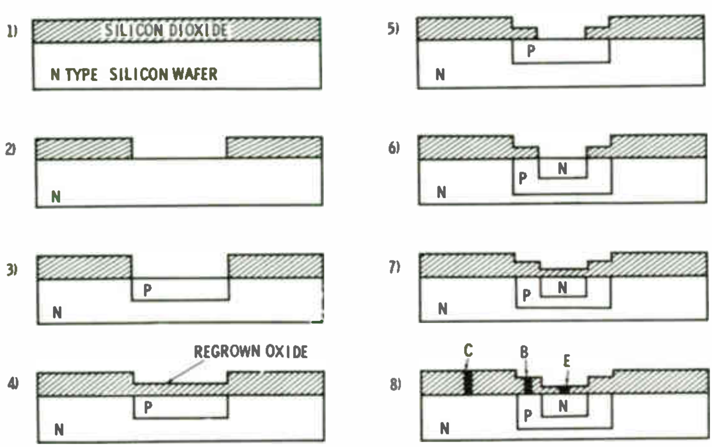
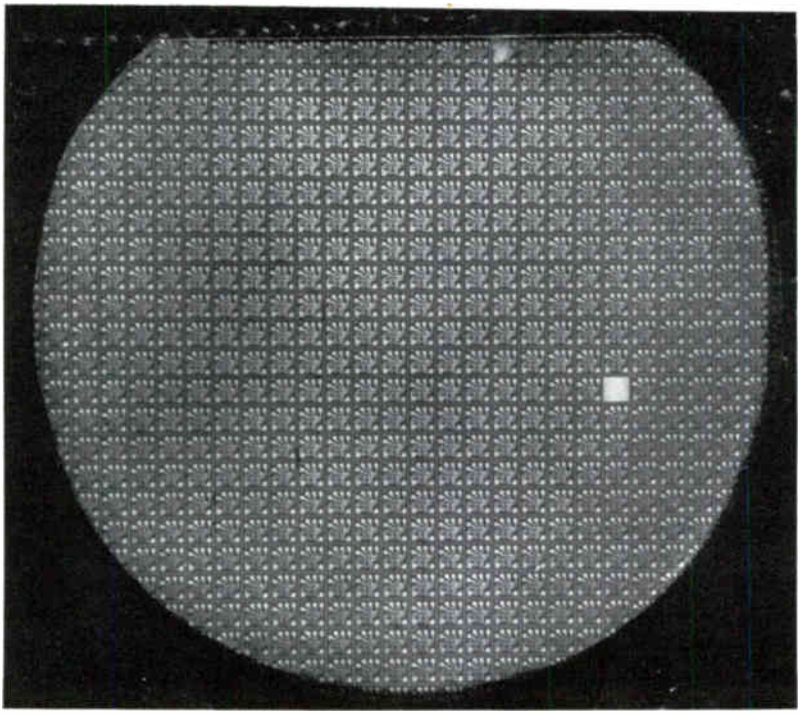


Fig. 7-7. Planar-passivated process.

The method of etching away the oxide at those places where you wish to put the next diffused junction is done by a photographic process. The oxide-coated surface is covered with a chemical that hardens when exposed to light. A mask is placed over the surface to block out the portions where the next diffusion is to take place. After exposure to light, the chemical under the masked part of the surface does not become hard and it washes off. The wafer is then placed in another chemical which etches away only the exposed oxide. It is then placed in the diffusion ovens and the new junction formed. This process can be repeated as many times as desired. The junctions can be made as small as it is possible to make the photographic masks. In actual practice, hundreds of junctions can be formed simultaneously. Fig. 7-8 shows a mass-produced wafer containing more than 400 individual circuits. After completion, the larger wafers are then cut up or "diced"



Courtesy Fairchild Semiconductor

Fig. 7-8. Mass-produced microcircuits.

into individual circuits, leads are attached, and the microcircuits are then mounted for convenient use.

The basic wafer on which the integrated circuit is built is called a substrate. Fig. 7-9 shows five diodes built into a common substrate. This is a very simple circuit and would require only one diffusion.

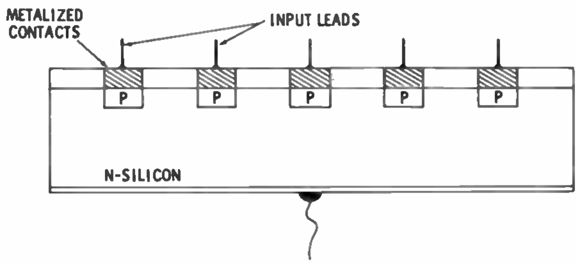


Fig. 7-9. Integrated circuit containing five diodes.

Thus, all of the diodes necessary for five inputs to an AND or OR gate are contained in a very small space with no interconnecting wires to bother with, and at a competitive price. Fig. 7-10 illustrates how these five diodes can be built into the output of a transistor circuit all on the

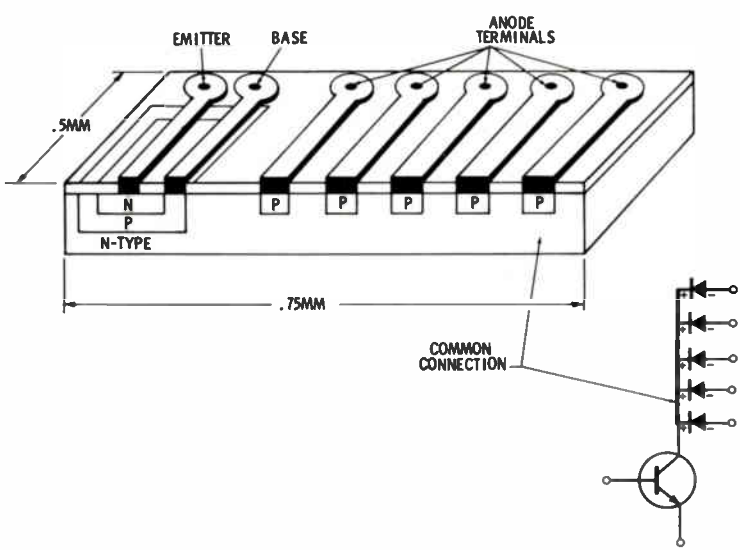
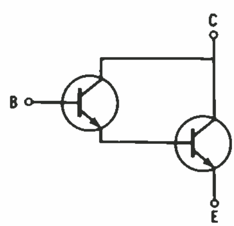


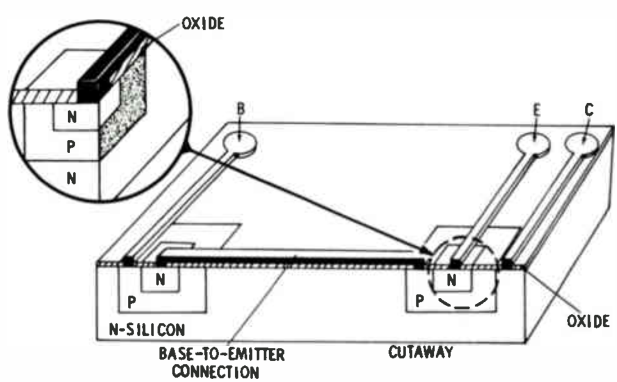
Fig. 7-10. Transistor-diode gate.

same substrate. This is a slightly more complicated integrated circuit and would require two diffusings. The light lines seen on the top of the crystal, around the transistor junctions, are oxide growth marks where the thickness of the oxide changes due to the etching and regrowing. The collector of the transistor and the cathodes of the diodes are common to each other, being formed out of the same piece of semiconductor. The base of the transistor and all the diode anodes would be formed by the first diffusion. The emitter of the transistor is a second diffusion. The aluminum contacts are made by evaporating aluminum onto the surface of the wafer through a thin metallic shadow-mask. Holes are punched in the mask at the places where the aluminum is desired. Leads are connected to these areas.

The simplest example of two transistors in one integrated circuit is that of the Darlington amplifier (Fig. 7-11A). In this circuit two



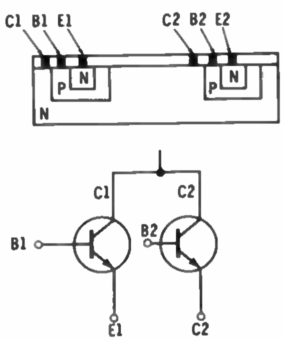
(A) Integrated circuit.



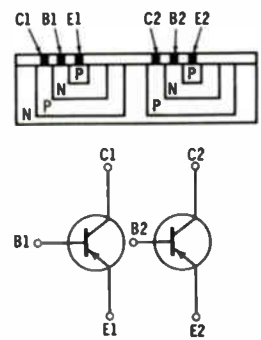
(B) Schematic.

Fig. 7-11. Darlington amplifier.

transistors are direct-coupled to provide increased gain. If the transistors in a Darlington pair had a gain of 100 each, the total gain of the circuit would be 10,000. Thus, you can see the possible applications of such a circuit. Fig. 7-11B shows the integrated circuit layout for a Darlington amplifier. The substrate is N-type silicon, which forms the collectors of both transistors. Then, by appropriate maskings, etchings, and diffusions the bases and emitters are formed. It is necessary to connect the emitter of one transistor to the base of the other. This is done with an aluminum conductor evaporated onto the surface. This aluminum conductor is deposited at the same time as the metallizing contacts. A double-cut view through the emitter connector is shown to illustrate how contact is made to the emitter and not to the base. The oxide serves as the insulator.



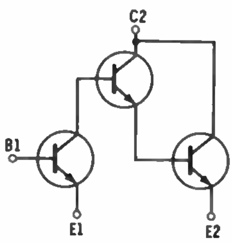
(A) Common collectors.



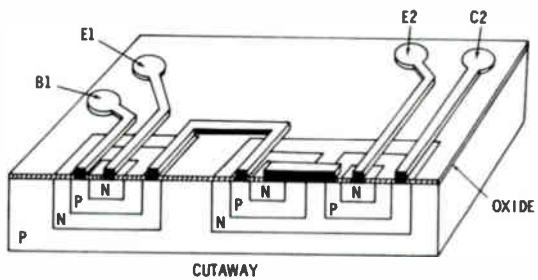
(B) Isolated collectors.

Fig. 7-12. Two transistors on one substrate.

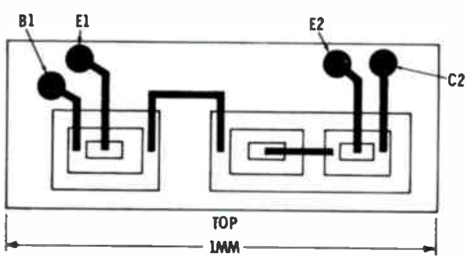
The collector of a transistor must carry the maximum current in the device and, therefore, must be capable of dissipating the most heat. For this reason, the collector of an integrated transistor is always the substrate. The obvious question then becomes, how can two or more transistors be put on one substrate without having their collectors connected. A circuit containing two transistors with common collectors is shown in Fig. 7-12A. Fig. 7-12B shows a configuration with two isolated transistors. Notice that another junction has been added. This junction is then reverse-biased to prevent current flow between



(A) Two transistors with common collectors.



(B) Two isolated transistors.

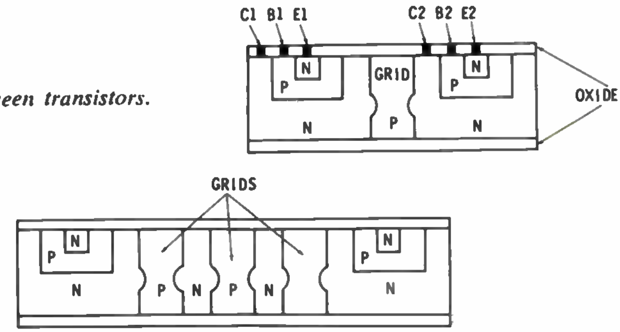


(C) Darlington pair.

Fig. 7-13. Three transistors on one substrate.

collector 1 and collector 2. Fig. 7-13A shows a circuit containing a Darlington amplifier and a driving transistor. In Fig. 7-13B note that the collector of the first transistor is isolated from the collectors of the Darlington pair by the reverse-biased isolation junctions. Notice that the top view (Fig. 7-13C) shows the transistor as three concentric squares. This is a good way to pick out the transistors on a microphotograph of an integrated circuit. They are not always squares, but three shapes (one inside the other) usually indicate a transistor.

(A) A grid between transistors.



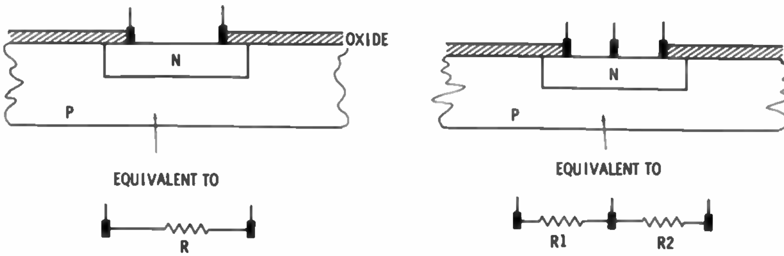
(B) Multiple grids.

Fig. 7-14. Grid isolation.

Another method of isolating adjacent transistors from each other involves diffusing a junction completely through the basic wafer (Fig. 7-14A). This provides two reverse-biased junctions, called a grid, between the transistors. The effective amount of isolation can be increased by increasing the number of grids, as in Fig. 7-14B. The main disadvantages of this method are that long diffusion times are required and maskings must be performed on both sides of the wafer. Both of these isolation methods have advantages for certain applications, and they are used in circuits.

Now that we can put as many transistors on a substrate as we wish, let us discuss methods for integrating other components, such as resistors and capacitors. Resistors are fairly easy to make since a semiconductor is a resistor whose resistance depends on the amount of doping. In addition, the length, height, and width of the resistive path will determine the resistance. If a certain length of semiconductor has a resistance of, say, 50,000 ohms, it will have twice as much resistance if the path for current carriers is cut in half and four times the resistance if the path is divided by four. Thus, different thicknesses and lengths of doped semiconductor may be used to provide different resistances.

There are two basic methods of using this principle to form resistors in integrated circuits. One method uses the semiconductor material itself, while the other involves depositing resistive elements on the surface of the wafer very similar to the thin-film resistors described earlier. In the case of the former (Fig. 7-15A) a channel is cut through

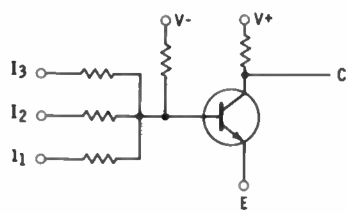


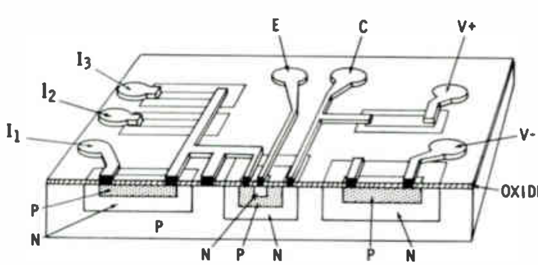
(A) *Single resistor.* (B) *Tapped resistor.*

Fig. 7-15. Integrated resistors.

the oxide and contact is made at the ends of a previously formed resistor strip. The resistance depends on the length, width, depth, and doping level of the strip. Fig. 7-15B shows how the addition of a conductor forms a resistor voltage divider. The ratio of the two resistances will be determined by the placement of the conductors.

(A) Schematic.



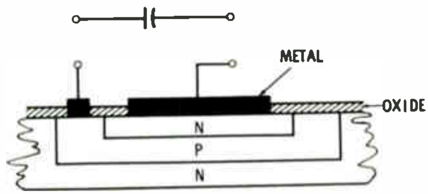


(B) Integrated circuit.

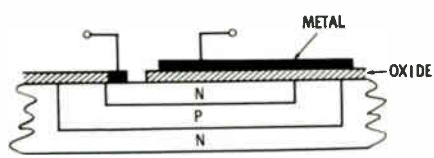
Fig. 7-16. NOR gate.

The application of these principles is shown in Fig. 7-16. The circuit for a three-input NOR gate is shown in Fig. 7-16A, and Fig. 7-16B shows a cutaway of the integrated circuit. This particular circuit could easily be put onto a 1-millimeter square. A number of such circuits can be mounted in a single transistor envelope.

Thin-film resistors can be deposited in the same manner as the metallized contacts (through a shadow-mask), but they are usually applied by photo methods. That is, the entire wafer is coated with the film and then the desired pattern is left after etching. These thin-film resistors are not semiconductors, but are made of metals such as nichrome ,a mixture of nickel and chromium) and cermet (a mixture of ceramic and metal). In the case of thin-film resistors, there are no tell-tale oxide growth marks, but often they appear as thin lines running back and forth across the chip, as in the case of the hybrid circuits.



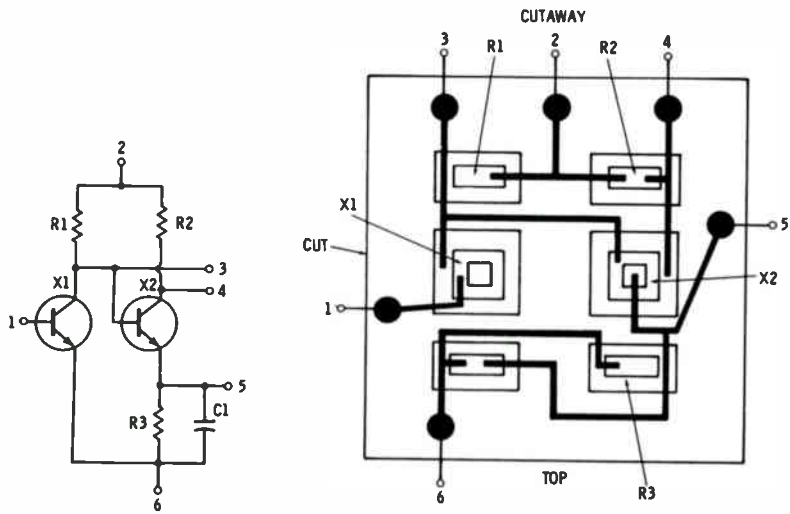
(A) Junction capacitors.



(B) Oxide capacitors.

Fig. 7-17. Integrated capacitors.

There are two basic ways of making capacitors for integrated circuits. The first utilizes the inherent capacity of a PN junction (Fig. 7-17A). The second uses the silicon dioxide as a dielectric (Fig. 7-17B). The junction type has two main disadvantages: first, the capacitance value is sensitive to the voltage across the junction; and second, a PN junction will conduct in one direction and, therefore, the junction capacitors must be reverse-biased in order to operate. The oxide dielectric capacitor is more stable than the junction type, but only a few hundred picofarads of capacity can be formed by this method. A conductor is placed on top of the oxide layer and connection is made to the semiconductor under it. Thus the oxide acts as the dielectric for the capacitor.

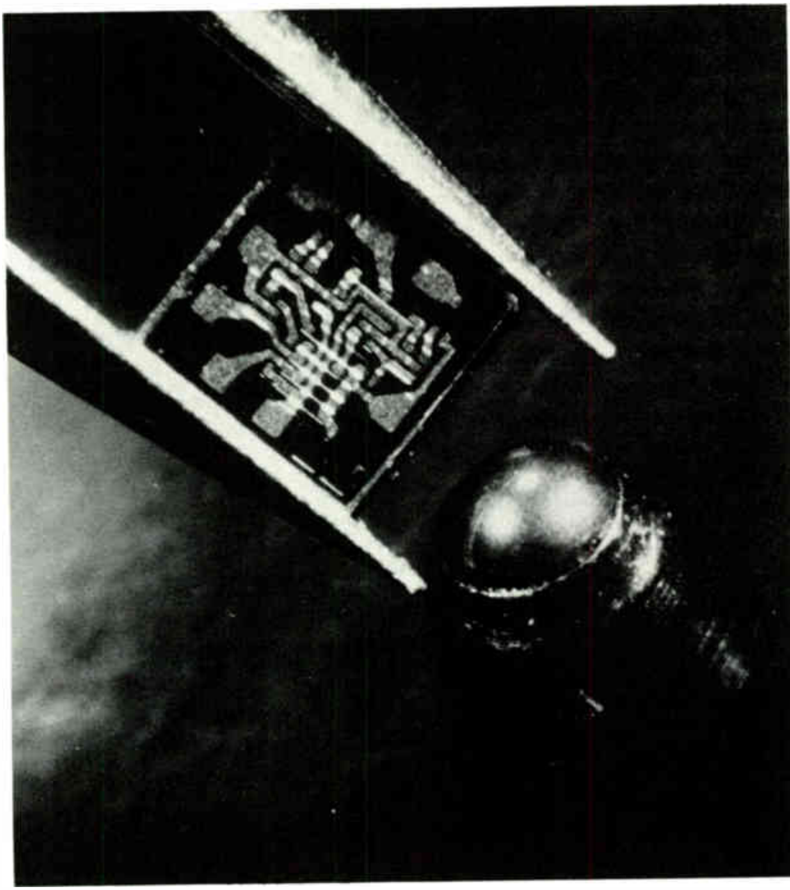


(A) Schematic. *(B) Circuit on substrate.*

Fig. 7-18. Typical integrated circuit.

Fig. 7-18B shows the layout of a typical integrated circuit containing two transistors, three resistors, and one capacitor. The schematic diagram of this circuit is shown in Fig. 7-18A. All the components are made by the photographic masking techniques mentioned earlier. The components are electrically separated on a single substrate by isolation diffusion. The transistors are NPN types, the resistors are N-type semiconductor strips, and the capacitor is an oxide type. Fig. 7-19 is a microphotograph of an integrated circuit, showing the relative size compared to that of a ball-point pen.

Another important semiconductor device which lends itself to easy integration is the field-effect transistor. Fig. 7-20 shows how this is accomplished. Recall that the source and the drain are connected to different ends of the same type of material. The gate is used to restrict



Courtesy Fairchild Semiconductor

Fig. 7-19. Microphotograph of a microcircuit.

the flow of current carriers from the source to the drain as a function of the voltage applied to the gate. The same isolation techniques are used for FET's as for regular transistors. The methods of constructing FET integrated circuits are completely compatible with those for normal transistors; thus, there is no problem in building both into the same circuit.

Another structure that lends itself to the use of integrated circuits, although it does not use devices or techniques other than those we

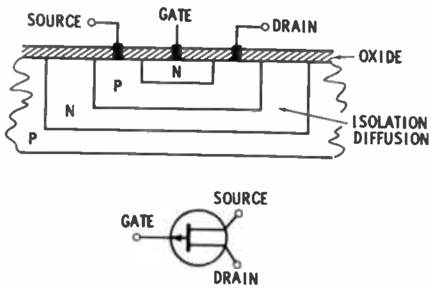


Fig. 7-20. Integrated field-effect transistor.

have already mentioned, is the complementary circuit. This requires two complementary transistors (an NPN and a PNP) in the same circuit. The collectors are direct-coupled to the bases of the opposite transistor type (Fig. 7-21). Three junctions are used to construct an isolated NPN transistor, and the basic P substrate is used as the col-

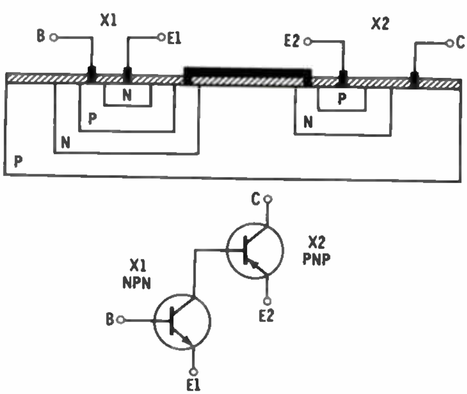


Fig. 7-21. Integrated circuit and schematic of complementary transistors.

lector of the PNP transistor. Fig. 7-22 shows the schematic and top view of an integrated circuit of a typical differential amplifier.

You are probably wondering why we have neglected to discuss integrated inductances. The reason is very simple—there is no good way to integrate inductance. For this reason, integrated circuits have found their best acceptance in the area of digital circuitry. Linear communications-type circuits require tuned circuits, and as yet there

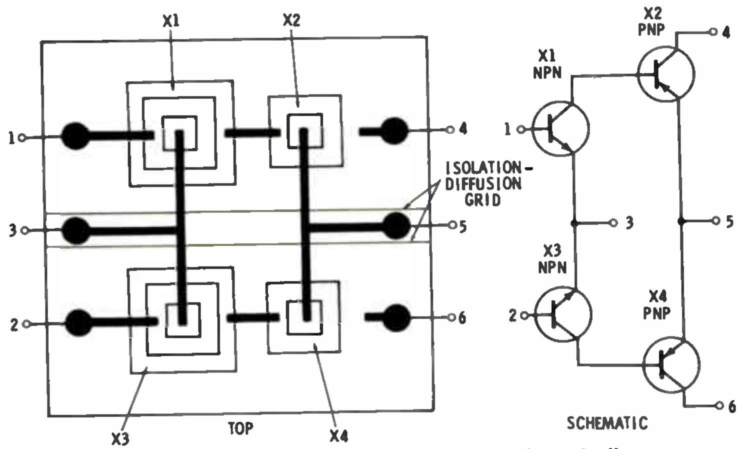


Fig. 7-22. Integrated differential amplifier and schematic diagram.

is no way to build an inductance out of semiconductor materials or junctions. A high-quality RF or IF stage cannot yet be made in integrated form. In addition to this, the tolerances of resistors and capacitors that must be lived with at the present are in many cases outside the realm of desirability for good linear amplifiers.

Index

Practical Transistor Theory

by
E. PATRICK WIESI

A practical understanding of basic semiconductor theory is needed before the real importance of semiconductor devices can be appreciated. *Practical Transistor Theory* provides an insight into the basic operation of the simplest forms of semiconductor devices and guides the reader progressively through the construction and operation of more intricate semiconductor devices.

The technical aspects of semiconductor theory are presented in terms that can be understood by nontechnical persons; however, much of the content covers areas of interest to more technically advanced technicians and engineers.

In this book, semiconductor devices are classified into active, passive, linear, and nonlinear categories—four, easily remembered classes into which all semiconductor devices can be grouped. This classification system is useful in understanding the later chapters, which provide detailed descriptions of radio, hi-fi, switching, and logic circuits.

A glimpse of the future is given in the last chapter, where the miniature world of *microelectronics* is described. The methods used in this new technology make it possible to produce complete circuits no larger than the head of a pin.

ABOUT THE AUTHOR

Mr. Wiesner, a graduate of Canisius College, is Midwest Editor for Electrical Design News. He has taught at DeVry Technical Institute in Chicago and served as a design engineer for ITT-Kellogg. His technical training and practical experience combine to make this book valuable to everyone concerned with transistors.

HOWARD W. SAMS & CO., INC.
THE BOBBS-MERRILL COMPANY, INC.

$2.50
PTW-1