Introduction to Solid State Semi-Conductors

Course No: E04-009

Credit: 4 PDH

A. Bhatia

Continuing Education and Development, Inc.
9 Greyridge Farm Court
Stony Point, NY 10980

P: (877) 322-5800
F: (877) 322-4774

info@cedengineering.com

CHAPTER 1

SEMICONDUCTOR DIODES

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the NRTC, you indicate that you have met the objectives and have learned the information. The learning objective are listed below.

Upon completion of this chapter, you should be able to do the following:

1. State, in terms of energy bands, the differences between a conductor, an insulator, and a semiconductor.
2. Explain the electron and the hole flow theory in semiconductors and how the semiconductor is affected by doping.
3. Define the term "diode" and give a brief description of its construction and operation.
4. Explain how the diode can be used as a half-wave rectifier and as a switch.
5. Identify the diode by its symbology, alphanumerical designation, and color code.
6. List the precautions that must be taken when working with diodes and describe the different ways to test them.

INTRODUCTION TO SOLID-STATE DEVICES

As you recall from previous studies in this series, semiconductors have electrical properties somewhere between those of insulators and conductors. The use of semiconductor materials in electronic components is not new; some devices are as old as the electron tube. Two of the most widely known semiconductors in use today are the JUNCTION DIODE and TRANSISTOR. These semiconductors fall under a more general heading called solid-state devices. A SOLID-STATE DEVICE is nothing more than an electronic device, which operates by virtue of the movement of electrons within a solid piece of semiconductor material.

Since the invention of the transistor, solid-state devices have been developed and improved at an unbelievable rate. Great strides have been made in the manufacturing techniques, and there is no foreseeable limit to the future of these devices. Solid-state devices made from semiconductor materials offer compactness, efficiency, ruggedness, and versatility. Consequently, these devices have invaded virtually every field of science and industry. In addition to the junction diode and transistor, a whole new family of related devices has been developed: the ZENER DIODE, LIGHT-EMITTING DIODE, FIELD EFFECT TRANSISTOR, etc. One development that has dominated solid-state technology, and probably has had a greater impact on the electronics industry than either the electron tube or transistor, is the INTEGRATED CIRCUIT. The integrated circuit is a minute piece of semiconductor material that can produce complete electronic circuit functions.

As the applications of solid-state devices mount, the need for knowledge of these devices becomes increasingly important. Personnel in the Navy today will have to understand solid-state devices if they are to become proficient in the repair and maintenance of electronic equipment. Therefore, our objective in this module is to provide a broad coverage of solid-state devices and, as a broad application, power supplies. We will begin our discussion with some background information on the development of the semiconductor. We will then proceed to the semiconductor diode, the transistor, special devices and, finally, solid-state power supplies.

SEMICONDUCTOR DEVELOPMENT

Although the semiconductor was late in reaching its present development, its story began long before the electron tube. Historically, we can go as far back as 1883 when Michael Faraday discovered that silver sulfide, a semiconductor, has a negative temperature coefficient. The term *negative temperature coefficient* is just another way of saying its resistance to electrical current flow decreases as temperature increases. The opposite is true of the conductor. It has a positive temperature coefficient. Because of this particular characteristic, semiconductors are used extensively in power-measuring equipment.

Only 2 years later, another valuable characteristic was reported by Munk A. Rosenshold. He found that certain materials have rectifying properties. Strange as it may seem, his finding was given such little notice that it had to be rediscovered 39 years later by F. Braun.

Toward the close of the 19th century, experimenters began to notice the peculiar characteristics of the chemical element SELENIUM. They discovered that in addition to its rectifying properties (the ability to convert ac into dc), selenium was also light sensitive-its resistance decreased with an increase in light intensity. This discovery eventually led to the invention of the photophone by Alexander Graham Bell. The photophone, which converted variations of light into sound, was a predecessor of the radio receiver; however, it wasn't until the actual birth of radio that selenium was used to any extent. Today, selenium is an important and widely used semiconductor.

Many other materials were tried and tested for use in communications. SILICON was found to be the most stable of the materials tested while GALENA, a crystalline form of lead sulfide, was found the most sensitive for use in early radio receivers. By 1915, Carl Beredicks discovered that GERMANIUM, another metallic element, also had rectifying capabilities. Later, it became widely used in electronics for low-power, low-frequency applications.

Although the semiconductor was known long before the electron tube was invented, the semiconductor devices of that time could not match the performance of the tube. Radio needed a device that could not only handle power and amplify but rectify and detect a signal as well. Since tubes could do all these things, whereas semiconductor devices of that day could not, the semiconductor soon lost out.

It wasn't until the beginning of World War II that interest was renewed in the semiconductor. There was a dire need for a device that could work within the ultra-high frequencies of radar. Electron tubes had interelectrode capacitances that were too high to do the job. The point-contact semiconductor diode, on the other hand, had a very low internal capacitance. Consequently, it filled the bill; it could be designed to work within the ultra-high frequencies used in radar, whereas the electron tube could not.

As radar took on greater importance and communication-electronic equipment became more sophisticated, the demands for better solid-state devices mounted. The limitations of the electron tube made necessary a quest for something new and different. An amplifying device was needed that was smaller, lighter, more efficient, and capable of handling extremely high frequencies. This was asking a

lot, but if progress was to be made, these requirements had to be met. A serious study of semiconductor materials began in the early 1940's and has continued since.

In June 1948, a significant breakthrough took place in semiconductor development. This was the discovery of POINT-CONTACT TRANSISTOR. Here at last was a semiconductor that could amplify. This discovery brought the semiconductor back into competition with the electron tube. A year later, JUNCTION DIODES and TRANSISTORS were developed. The junction transistor was found superior to the point-contact type in many respects. By comparison, the junction transistor was more reliable, generated less noise, and had higher power-handling ability than its point-contact brother. The junction transistor became a rival of the electron tube in many uses previously uncontested.

Semiconductor diodes were not to be slighted. The initial work of Dr. Carl Zener led to the development of ZENER DIODE, which is frequently used today to regulate power supply voltages at precise levels. Considerably more interest in the solid-state diode was generated when Dr. Leo Esaki, a Japanese scientist, fabricated a diode that could amplify. The device, named the TUNNEL DIODE, has amazing gain and fast switching capabilities. Although it is used in the conventional amplifying and oscillating circuits, its primary use is in computer logic circuits.

Another breakthrough came in the late 1950's when it was discovered that semiconductor materials could be combined and treated so that they functioned as an entire circuit or subassembly rather than as a circuit component. Many names have been given to this solid-circuit concept, such as INTEGRATED CIRCUITS, MICROELECTRONICS, and MICROCIRCUITRY.

So as we see, in looking back, that the semiconductor is not something new, but it has come a long way in a short time.

Q1. What is a solid-state device?

Q2. Define the term negative temperature coefficient.

SEMICONDUCTOR APPLICATIONS

In the previous paragraphs, we mentioned just a few of the many different applications of semiconductor devices. The use of these devices has become so widespread that it would be impossible to list all their different applications. Instead, a broad coverage of their specific application is presented.

Semiconductor devices are all around us. They can be found in just about every commercial product we touch, from the family car to the pocket calculator. Semiconductor devices are contained in television sets, portable radios, stereo equipment, and much more.

Science and industry also rely heavily on semiconductor devices. Research laboratories use these devices in all sorts of electronic instruments to perform tests, measurements, and numerous other experimental tasks. Industrial control systems (such as those used to manufacture automobiles) and automatic telephone exchanges also use semiconductors. Even today heavy-duty versions of the solid-state rectifier diode are being use to convert large amounts of power for electric railroads. Of the many different applications for solid-state devices, space systems, computers, and data processing equipment are some of the largest consumers.

The various types of modem military equipment are literally loaded with semiconductor devices. Many radars, communication, and airborne equipment are transistorized. Data display systems, data processing units, computers, and aircraft guidance-control assemblies are also good examples of

electronic equipments that use semiconductor devices. All of the specific applications of semiconductor devices would make a long impressive list. The fact is, semiconductors are being used extensively in commercial products, industry, and the military.

SEMICONDUCTOR COMPETITION

It should not be difficult to conclude, from what you already know, that semiconductor devices can and do perform all the conventional functions of rectification, amplification, oscillation, timing, switching, and sensing. Simply stated, these devices perform the same basic functions as the electron tube; but they perform more efficiently, economically, and for a longer period of time. Therefore, it should be no surprise to you to see these devices used in place of electron tubes. Keeping this in mind, we see that it is only natural and logical to compare semiconductor devices with electron tubes.

Physically, semiconductor devices are much smaller than tubes. You can see in figure 1-1 that the difference is quite evident. This illustration shows some commonly used tube sizes alongside semiconductor devices of similar capabilities. The reduction in size can be as great as 100:1 by weight and 1000:1 by volume. It is easy to see that size reduction favors the semiconductor device. Therefore, whenever miniaturization is required or is convenient, transistors are favored over tubes. Bear in mind, however, that the extent of practical size reduction is a big factor; many things must be considered. Miniature electron tubes, for example, may be preferred in certain applications to transistors, thus keeping size reduction to a competitive area.

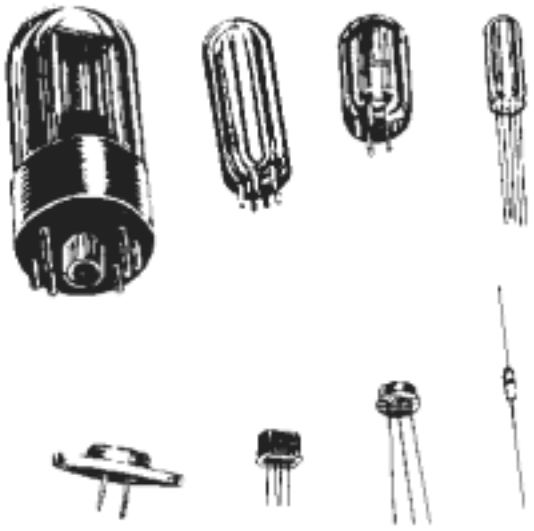


Figure 1-1.—Size comparisons of electron tubes and semiconductors.

Power is also a two-sided story. For low-power applications, where efficiency is a significant factor, semiconductors have a decided advantage. This is true mainly because semiconductor devices perform very well with an extremely small amount of power; in addition, they require no filaments or heaters as in the case of the electron tube. For example, a computer operating with over 4000 solid-state devices may require no more than 20 watts of power. However, the same number of tubes would require several kilowatts of power.

For high-power applications, it is a different story — tubes have the upper hand. The high-power tube has no equivalent in any semiconductor device. This is because a tube can be designed to operate

with over a thousand volts applied to its plate whereas the maximum allowable voltage for a transistor is limited to about 200 volts (usually 50 volts or less). A tube can also handle thousands of watts of power. The maximum power output for transistor generally ranges from 30 milliwatts to slightly over 100 watts.

When it comes to ruggedness and life expectancy, the tube is still in competition. Design and functional requirements usually dictate the choice of device. However, semiconductor devices are rugged and long-lived. They can be constructed to withstand extreme vibration and mechanical shock. They have been known to withstand impacts that would completely shatter an ordinary electron tube. Although some specially designed tubes render extensive service, the life expectancy of transistors is better than three to four times that of ordinary electronic tubes. There is no known failure mechanism (such as an open filament in a tube) to limit the semiconductor's life. However, semiconductor devices do have some limitations. They are usually affected more by temperature, humidity, and radiation than tubes are.

Q3. Name three of the largest users of semiconductor devices.

Q4. State one requirement of an electron tube, which does not exist for semiconductors, that makes the tube less efficient than the semiconductor.

SEMICONDUCTOR THEORY

To understand why solid-state devices function as they do, we will have to examine closely the composition and nature of semiconductors. This entails theory that is fundamental to the study of solid-state devices.

Rather than beginning with theory, let's first become reacquainted with some of the basic information you studied earlier concerning matter and energy (NEETS, Module 1).

ATOMIC STRUCTURE

The universe, as we know it today, is divided into two parts: matter and energy. Matter, which is our main concern at this time, is anything that occupies space and has weight. Rocks, water, air, automobiles, clothing, and even our own bodies are good examples of matter. From this, we can conclude that matter may be found in any one of three states: SOLIDS, LIQUIDS, and GASES. All matter is composed of either an element or combination of elements. As you know, an element is a substance that cannot be reduced to a simpler form by chemical means. Examples of elements with which you are in contact everyday are iron, gold, silver, copper, and oxygen. At present, there are over 100 known elements of which all matter is composed.

As we work our way down the size scale, we come to the atom, the smallest particle into which an element can be broken down and still retain all its original properties. The atoms of one element, however, differ from the atoms of all other elements. Since there are over 100 known elements, there must be over 100 different atoms, or a different atom for each element.

Now let us consider more than one element at a time. This brings us to the term "compound." A compound is a chemical combination of two or more elements. Water, table salt, ethyl alcohol, and ammonia are all examples of compounds. The smallest part of a compound, which has all the characteristics of the compound, is the molecule. Each molecule contains some of the atoms of each of the elements forming the compound.

Consider sugar, for example. Sugar in general terms is matter, since it occupies space and has weight. It is also a compound because it consists of two or more elements. Take a lump of sugar and crush

it into small particles; each of the particles still retains its original identifying properties of sugar. The only thing that changed was the physical size of the sugar. If we continue this subdividing process by grinding the sugar into a fine power, the results are the same. Even dissolving sugar in water does not change its identifying properties, in spite of the fact that the particles of sugar are now too small to be seen even with a microscope. Eventually, we end up with a quantity of sugar that cannot be further divided without its ceasing to be sugar. This quantity is known as a molecule of sugar. If the molecule is further divided, it is found to consist of three simpler kinds of matter: carbon, hydrogen, and oxygen. These simpler forms are called elements. Therefore, since elements consist of atoms, then a molecule of sugar is made up of atoms of carbon, hydrogen, and oxygen.

As we investigate the atom, we find that it is basically composed of electrons, protons, and neutrons. Furthermore, the electrons, protons, and neutrons of one element are identical to those of any other element. There are different kinds of elements because the number and the arrangement of electrons and protons are different for each element.

The electron carries a small negative charge of electricity. The proton carries a positive charge of electricity equal and opposite to the charge of the electron. Scientists have measured the mass and size of the electron and proton, and they know how much charge each possesses. Both the electron and proton have the same quantity of charge, although the mass of the proton is approximately 1,827 times that of the electron. In some atoms there exists a neutral particle called a neutron. The neutron has a mass approximately equal to that of a proton, but it has no electrical charge.

According to theory, the electrons, protons, and neutrons of the atoms are thought to be arranged in a manner similar to a miniature solar system. Notice the helium atom in figure 1-2. Two protons and two neutrons form the heavy nucleus with a positive charge around which two very light electrons revolve. The path each electron takes around the nucleus is called an orbit. The electrons are continuously being acted upon in their orbits by the force of attraction of the nucleus. To maintain an orbit around the nucleus, the electrons travel at a speed that produces a counterforce equal to the attraction force of the nucleus. Just as energy is required to move a space vehicle away from the earth, energy is also required to move an electron away from the nucleus. Like a space vehicle, the electron is said to be at a higher energy level when it travels a larger orbit. Scientific experiments have shown that the electron requires a certain amount of energy to stay in orbit. This quantity is called the electron's energy level. By virtue of just its motion alone, the electron contains kinetic energy. Because of its position, it also contains potential energy. The total energy contained by an electron (kinetic energy plus potential energy) is the main factor that determines the radius of the electron's orbit. For an electron to remain in this orbit, it must neither gain nor lose energy.

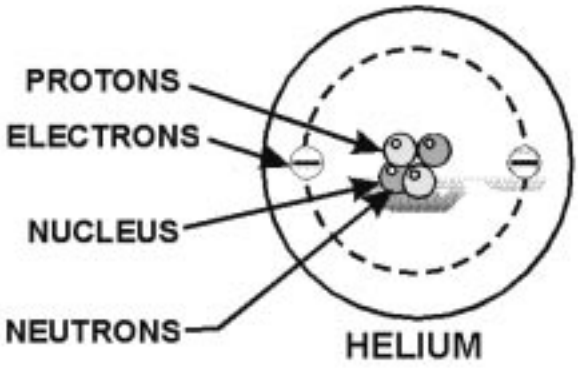


Figure 1-2.—The composition of a simple helium atom.

The orbiting electrons do not follow random paths, instead they are confined to definite energy levels. Visualize these levels as shells with each successive shell being spaced a greater distance from the nucleus. The shells, and the number of electrons required to fill them, may be predicted by using Pauli's exclusion principle. Simply stated, this principle specifies that each shell will contain a maximum of $2n^2$ electrons, where n corresponds to the shell number starting with the one closest to the nucleus. By this principle, the second shell, for example, would contain $2(2)^2$ or 8 electrons when full.

In addition to being numbered, the shells are also given letter designations starting with the shell closest to the nucleus and progressing outward as shown in figure 1-3. The shells are considered to be full, or complete, when they contain the following quantities of electrons: 2 in the K(1st) shell, 8 in the L(2nd) shell, 18 in the M(3rd) shell, and so on, in accordance with the exclusion principle. Each of these shells is a major shell and can be divided into subshells, of which there are four, labeled s, p, d, and f. Like the major shells, the subshells are also limited as to the number of electrons they contain. Thus, the "s" subshell is complete when it contains 2 electrons, the "p" subshell when it contains 6, the "d" subshell when it contains 10, and the "f" subshell when it contains 14 electrons.

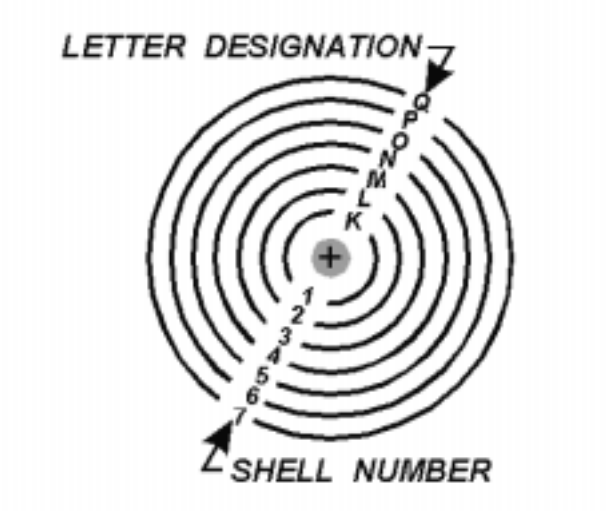


Figure 1-3.—Shell designation.

Inasmuch as the K shell can contain no more than 2 electrons, it must have only one subshell, the s subshell. The M shell is composed of three subshells: s, p, and d. If the electrons in the s, p, and d subshells are added together, their total is found to be 18, the exact number required to fill the M shell. Notice the electron configuration of copper illustrated in figure 1-4. The copper atom contains 29 electrons, which completely fill the first three shells and subshells, leaving one electron in the "s" subshell of the N shell. A list of all the other known elements, with the number of electrons in each atom, is contained in the PERIODIC TABLE OF ELEMENTS. The periodic table of elements is included in appendix 2.

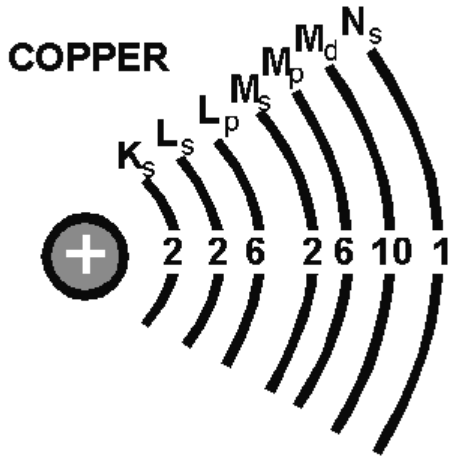


Figure 1-4.—Copper atom.

Valence is an atom's ability to combine with other atoms. The number of electrons in the outermost shell of an atom determines its valence. For this reason, the outer shell of an atom is called VALENCE SHELL, and the electrons contained in this shell are called VALENCE ELECTRONS. The valence of an atom determines its ability to gain or lose an electron, which in turn determines the chemical and electrical properties of the atom. An atom that is lacking only one or two electrons from its outer shell will easily gain electrons to complete its shell, but a large amount of energy is required to free any of its electrons. An atom having a relatively small number of electrons in its outer shell in comparison to the number of electrons required to fill the shell will easily lose these valence electrons. The valence shell always refers to the outermost shell.

Q5. Define matter and list its three different states.

Q6. What is the smallest particle into which an element can be broken down and still retain all its original properties?

Q7. What are the three particles that comprise an atom and state the type of charge they hold?

Q8. What is the outer shell of an atom called?

ENERGY BANDS

Now that you have become reacquainted with matter and energy, we will continue our discussion with electron behavior.

As stated earlier, orbiting electrons contain energy and are confined to definite energy levels. The various shells in an atom represent these levels. Therefore, to move an electron from a lower shell to a higher shell a certain amount of energy is required. This energy can be in the form of electric fields, heat, light, and even bombardment by other particles. Failure to provide enough energy to the electron, even if the energy supplied is just short of the required amount, will cause it to remain at its present energy level. Supplying more energy than is needed will only cause the electron to move to the next higher shell and the remaining energy will be wasted. In simple terms, energy is required in definite units to move electrons from one shell to the next higher shell. These units are called QUANTA (for example 1, 2, or 3 quanta).

Electrons can also lose energy as well as receive it. When an electron loses energy, it moves to a lower shell. The lost energy, in some cases, appears as heat.

If a sufficient amount of energy is absorbed by an electron, it is possible for that electron to be completely removed from the influence of the atom. This is called IONIZATION. When an atom loses electrons or gains electrons in this process of electron exchange, it is said to be ionized. For ionization to take place, there must be a transfer of energy that results in a change in the internal energy of the atom. An atom having more than its normal amount of electrons acquires a negative charge, and is called a NEGATIVE ION. The atom that gives up some of its normal electrons is left with fewer negative charges than positive charges and is called a POSITIVE ION. Thus, we can define ionization as the process by which an atom loses or gains electrons.

Up to this point in our discussion, we have spoken only of isolated atoms. When atoms are spaced far enough apart, as in a gas, they have very little influence upon each other, and are very much like lone atoms. But atoms within a solid have a marked effect upon each other. The forces that bind these atoms together greatly modify the behavior of the other electrons. One consequence of this close proximity of atoms is to cause the individual energy levels of an atom to break up and form bands of energy. Discrete (separate and complete) energy levels still exist within these energy bands, but there are many more energy levels than there were with the isolated atom. In some cases, energy levels will have disappeared. Figure 1-5 shows the difference in the energy arrangement between an isolated atom and the atom in a solid. Notice that the isolated atom (such as in gas) has energy levels, whereas the atom in a solid has energy levels grouped into ENERGY BANDS.

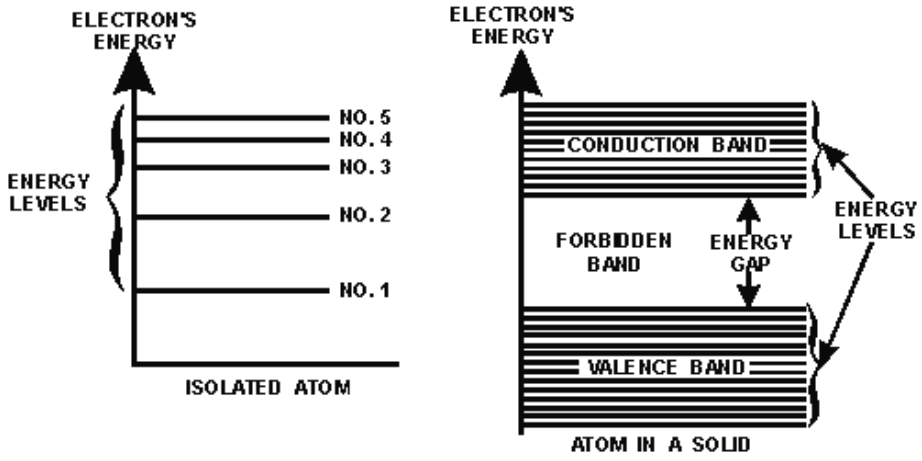


Figure 1-5.—The energy arrangement in atoms.

The upper band in the solid lines in figure 1-5 is called the CONDUCTION BAND because electrons in this band are easily removed by the application of external electric fields. Materials that have a large number of electrons in the conduction band act as good conductors of electricity.

Below the conduction band is the FORBIDDEN BAND or energy gap. Electrons are never found in this band, but may travel back and forth through it, provided they do not come to rest in the band.

The last band or VALENCE BAND is composed of a series of energy levels containing valence electrons. Electrons in this band are more tightly bound to the individual atom than the electrons in the conduction band. However, the electrons in the valence band can still be moved to the conduction band with the application of energy, usually thermal energy. There are more bands below the valence band, but they are not important to the understanding of semiconductor theory and will not be discussed.

The concept of energy bands is particularly important in classifying materials as conductors, semiconductors, and insulators. An electron can exist in either of two energy bands, the conduction band or the valence band. All that is necessary to move an electron from the valence band to the conduction band so it can be used for electric current, is enough energy to carry the electron through the forbidden band. The width of the forbidden band or the separation between the conduction and valence bands determines whether a substance is an insulator, semiconductor, or conductor. Figure 1-6 uses energy level diagrams to show the difference between insulators, semiconductors, and conductors.

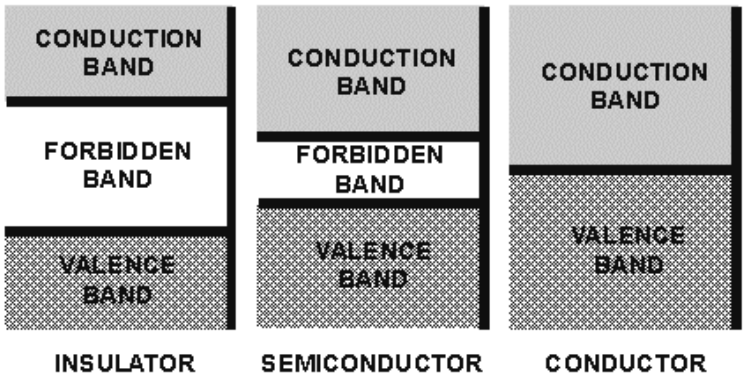


Figure 1-6.—Energy level diagram.

The energy diagram for the insulator shows the insulator with a very wide energy gap. The wider this gap, the greater the amount of energy required to move the electron from the valence band to the conduction band. Therefore, an insulator requires a large amount of energy to obtain a small amount of current. The insulator "insulates" because of the wide forbidden band or energy gap.

The semiconductor, on the other hand, has a smaller forbidden band and requires less energy to move an electron from the valence band to the conduction band. Therefore, for a certain amount of applied voltage, more current will flow in the semiconductor than in the insulator.

The last energy level diagram in figure 1-6 is that of a conductor. Notice, there is no forbidden band or energy gap and the valence and conduction bands overlap. With no energy gap, it takes a small amount of energy to move electrons into the conduction band; consequently, conductors pass electrons very easily.

Q9. What term is used to describe the definite discrete amounts of energy required to move an electron from a low shell to a higher shell?

Q10. What is a negative ion?

Q11. What is the main difference in the energy arrangement between an isolated atom and the atom in a solid?

Q12. What determines, in terms of energy bands, whether a substance is a good insulator, semiconductor, or conductor?

COVALENT BONDING

The chemical activity of an atom is determined by the number of electrons in its valence shell. When the valence shell is complete, the atom is stable and shows little tendency to combine with other atoms to form solids. Only atoms that possess eight valence electrons have a complete outer shell. These atoms are

referred to as inert or inactive atoms. However, if the valence shell of an atom lacks the required number of electrons to complete the shell, then the activity of the atom increases.

Silicon and germanium, for example, are the most frequently used semiconductors. Both are quite similar in their structure and chemical behavior. Each has four electrons in the valence shell. Consider just silicon. Since it has fewer than the required number of eight electrons needed in the outer shell, its atoms will unite with other atoms until eight electrons are shared. This gives each atom a total of eight electrons in its valence shell; four of its own and four that it borrowed from the surrounding atoms. The sharing of valence electrons between two or more atoms produces a COVALENT BOND between the atoms. It is this bond that holds the atoms together in an orderly structure called a CRYSTAL. A crystal is just another name for a solid whose atoms or molecules are arranged in a three-dimensional geometrical pattern commonly referred to as a lattice. Figure 1-7 shows a typical crystal structure. Each sphere in the figure represents the nucleus of an atom, and the arms that join the atoms and support the structure are the covalent bonds.

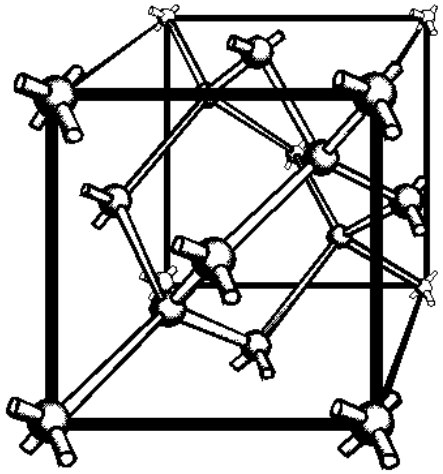


Figure 1-7.—A typical crystal structure.

As a result of this sharing process, the valence electrons are held tightly together. This can best be illustrated by the two-dimensional view of the silicon lattice in figure 1-8. The circles in the figure represent the nuclei of the atoms. The +4 in the circles is the net charge of the nucleus plus the inner shells (minus the valence shell). The short lines indicate valence electrons. Because every atom in this pattern is bonded to four other atoms, the electrons are not free to move within the crystal. As a result of this bonding, pure silicon and germanium are poor conductors of electricity. The reason they are not insulators but semiconductors is that with the proper application of heat or electrical pressure, electrons can be caused to break free of their bonds and move into the conduction band. Once in this band, they wander aimlessly through the crystal.

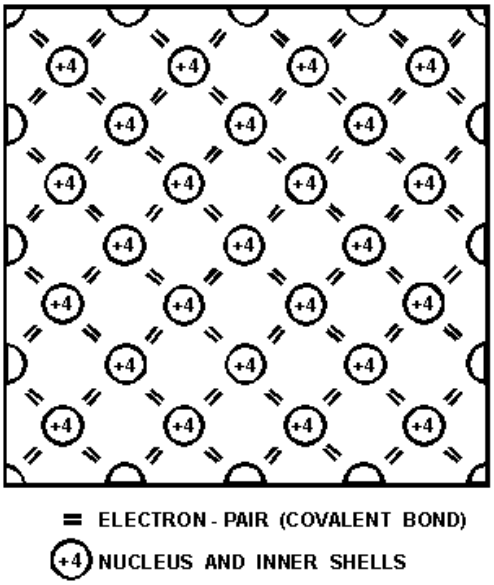


Figure 1-8.—A two-dimensional view of a silicon cubic lattice.

Q13. What determines the chemical activity of an atom?

Q14. What is the term used to describe the sharing of valence electrons between two or more atoms?

CONDUCTION PROCESS

As stated earlier, energy can be added to electrons by applying heat. When enough energy is absorbed by the valence electrons, it is possible for them to break some of their covalent bonds. Once the bonds are broken, the electrons move to the conduction band where they are capable of supporting electric current. When a voltage is applied to a crystal containing these conduction band electrons, the electrons move through the crystal toward the applied voltage. This movement of electrons in a semiconductor is referred to as electron current flow.

There is still another type of current in a pure semiconductor. This current occurs when a covalent bond is broken and a vacancy is left in the atom by the missing valence electron. This vacancy is commonly referred to as a "hole." The hole is considered to have a positive charge because its atom is deficient by one electron, which causes the protons to outnumber the electrons. As a result of this hole, a chain reaction begins when a nearby electron breaks its own covalent bond to fill the hole, leaving another hole. Then another electron breaks its bond to fill the previous hole, leaving still another hole. Each time an electron in this process fills a hole, it enters into a covalent bond. Even though an electron has moved from one covalent bond to another, the most important thing to remember is that the hole is also moving. Therefore, since this process of conduction resembles the movement of holes rather than electrons, it is termed hole flow (short for hole current flow or conduction by holes). Hole flow is very similar to electron flow except that the holes move toward a negative potential and in an opposite direction to that of the electron. Since hole flow results from the breaking of covalent bonds, which are at the valence band level, the electrons associated with this type of conduction contain only valence band energy and must remain in the valence band. However, the electrons associated with electron flow have conduction band energy and can, therefore, move throughout the crystal. A good analogy of hole flow is the movement of a hole through a tube filled with balls (figure 1-9).

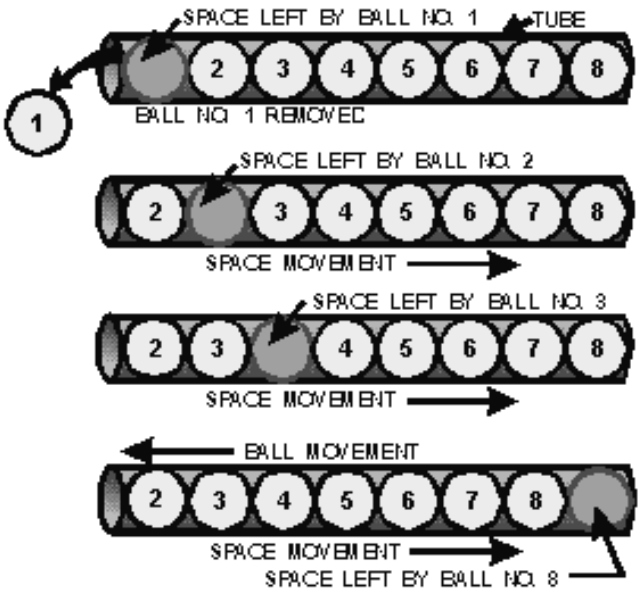


Figure 1-9.—Analogy of hole flow.

When ball number 1 is removed from the tube, a hole is left. This hole is then filled by ball number 2, which leaves still another hole. Ball number 3 then moves into the hole left by ball number 2. This causes still another hole to appear where ball 3 was. Notice the holes are moving to the right side of the tube. This action continues until all the balls have moved one space to the left in which time the hole moved eight spaces to the right and came to rest at the right-hand end of the tube.

In the theory just described, two current carriers were created by the breaking of covalent bonds: the negative electron and the positive hole. These carriers are referred to as electron-hole pairs. Since the semiconductor we have been discussing contains no impurities, the number of holes in the electron-hole pairs is always equal to the number of conduction electrons. Another way of describing this condition where no impurities exist is by saying the semiconductor is INTRINSIC. The term *intrinsic* is also used to distinguish the pure semiconductor that we have been working with from one containing impurities.

Q15. Name the two types of current flow in a semiconductor.

Q16. What is the name given to a piece of pure semiconductor material that has an equal number of electrons and holes?

DOPING PROCESS

The pure semiconductor mentioned earlier is basically neutral. It contains no free electrons in its conduction bands. Even with the application of thermal energy, only a few covalent bonds are broken, yielding a relatively small current flow. A much more efficient method of increasing current flow in semiconductors is by adding very small amounts of selected additives to them, generally no more than a few parts per million. These additives are called impurities and the process of adding them to crystals is referred to as DOPING. The purpose of semiconductor doping is to increase the number of free charges that can be moved by an external applied voltage. When an impurity increases the number of free electrons, the doped semiconductor is NEGATIVE or N TYPE, and the impurity that is added is known as an N-type impurity. However, an impurity that reduces the number of free electrons, causing more

holes, creates a POSITIVE or P-TYPE semiconductor, and the impurity that was added to it is known as a P-type impurity. Semiconductors which are doped in this manner — either with N- or P-type impurities — are referred to as EXTRINSIC semiconductors.

N-Type Semiconductor

The N-type impurity loses its extra valence electron easily when added to a semiconductor material, and in so doing, increases the conductivity of the material by contributing a free electron. This type of impurity has 5 valence electrons and is called a PENTAVALENT impurity. Arsenic, antimony, bismuth, and phosphorous are pentavalent impurities. Because these materials give or donate one electron to the doped material, they are also called DONOR impurities.

When a pentavalent (donor) impurity, like arsenic, is added to germanium, it will form covalent bonds with the germanium atoms. Figure 1-10 illustrates this by showing an arsenic atom (AS) in a germanium (GE) lattice structure. Notice the arsenic atom in the center of the lattice. It has 5 valence electrons in its outer shell but uses only 4 of them to form covalent bonds with the germanium atoms, leaving 1 electron relatively free in the crystal structure. Pure germanium may be converted into an N-type semiconductor by "doping" it with any donor impurity having 5 valence electrons in its outer shell. Since this type of semiconductor (N-type) has a surplus of electrons, the electrons are considered MAJORITY carriers, while the holes, being few in number, are the MINORITY carriers.

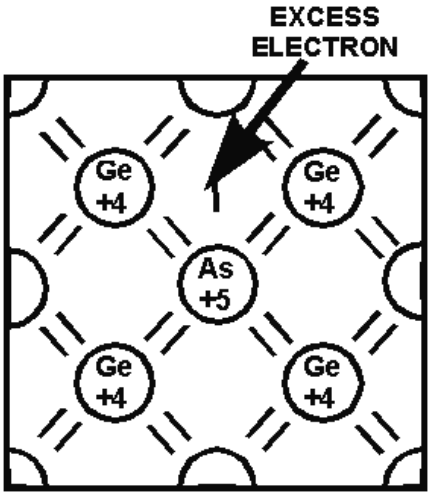


Figure 1-10.—Germanium crystal doped with arsenic.

P-Type Semiconductor

The second type of impurity, when added to a semiconductor material, tends to compensate for its deficiency of 1 valence electron by acquiring an electron from its neighbor. Impurities of this type have only 3 valence electrons and are called TRIVALENT impurities. Aluminum, indium, gallium, and boron are trivalent impurities. Because these materials accept 1 electron from the doped material, they are also called ACCEPTOR impurities.

A trivalent (acceptor) impurity element can also be used to dope germanium. In this case, the impurity is 1 electron short of the required amount of electrons needed to establish covalent bonds with 4 neighboring atoms. Thus, in a single covalent bond, there will be only 1 electron instead of 2. This arrangement leaves a hole in that covalent bond. Figure 1-11 illustrates this theory by showing what happens when germanium is doped with an indium (In) atom. Notice, the indium atom in the figure is 1

electron short of the required amount of electrons needed to form covalent bonds with 4 neighboring atoms and, therefore, creates a hole in the structure. Gallium and boron, which are also trivalent impurities, exhibit these same characteristics when added to germanium. The holes can only be present in this type semiconductor when a trivalent impurity is used. Note that a hole carrier is not created by the removal of an electron from a neutral atom, but is created when a trivalent impurity enters into covalent bonds with a tetravalent (4 valence electrons) crystal structure. The holes in this type of semiconductor (P-type) are considered the MAJORITY carriers since they are present in the material in the greatest quantity. The electrons, on the other hand, are the MINORITY carriers.

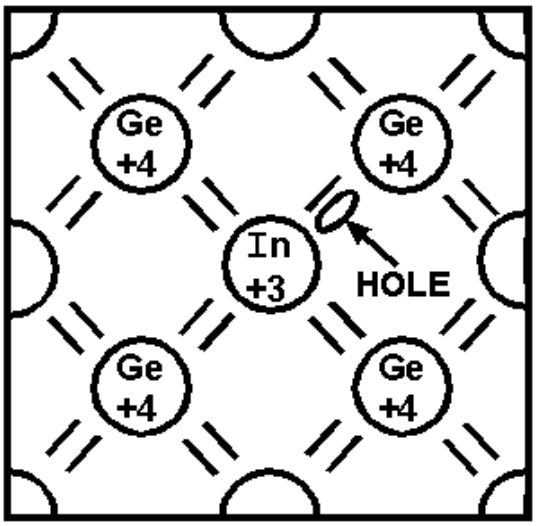


Figure 1-11.—Germanium crystal doped with indium.

Q17. What is the name given to a doped germanium crystal with an excess of free holes?

Q18. What are the majority carriers in an N-type semiconductor?

SEMICONDUCTOR DIODE

If we join a section of N-type semiconductor material with a similar section of P-type semiconductor material, we obtain a device known as a PN JUNCTION. (The area where the N and P regions meet is appropriately called the junction.) The usual characteristics of this device make it extremely useful in electronics as a diode rectifier. The diode rectifier or PN junction diode performs the same function as its counterpart in electron tubes but in a different way. The diode is nothing more than a two-element semiconductor device that makes use of the rectifying properties of a PN junction to convert alternating current into direct current by permitting current flow in only one direction. The schematic symbol of a PN junction diode is shown in figure 1-12. The vertical bar represents the cathode (N-type material) since it is the source of electrons and the arrow represents the anode. (P-type material) since it is the destination of the electrons. The label "CR1" is an alphanumerical code used to identify the diode. In this figure, we have only one diode so it is labeled CR1 (crystal rectifier number one). If there were four diodes shown in the diagram, the last diode would be labeled CR4. The heavy dark line shows electron flow. Notice it is against the arrow. For further clarification, a pictorial diagram of a PN junction and an actual semiconductor (one of many types) are also illustrated.

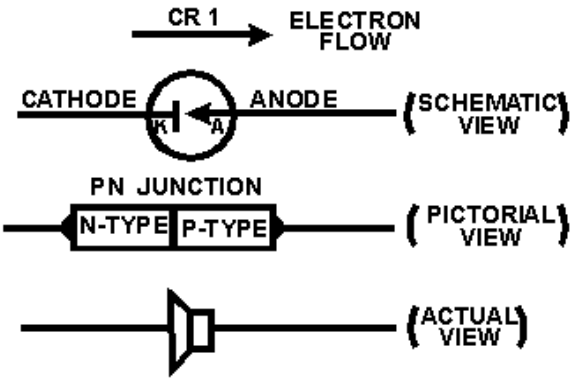


Figure 1-12.—The PN junction diode.

CONSTRUCTION

Merely pressing together a section of P material and a section of N material, however, is not sufficient to produce a rectifying junction. The semiconductor should be in one piece to form a proper PN junction, but divided into a P-type impurity region and an N-type impurity region. This can be done in various ways. One way is to mix P-type and N-type impurities into a single crystal during the manufacturing process. By so doing, a P-region is grown over part of a semiconductor's length and N-region is grown over the other part. This is called a GROWN junction and is illustrated in view A of figure 1-13. Another way to produce a PN junction is to melt one type of impurity into a semiconductor of the opposite type impurity. For example, a pellet of acceptor impurity is placed on a wafer of N-type germanium and heated. Under controlled temperature conditions, the acceptor impurity fuses into the wafer to form a P-region within it, as shown in view B of figure 1-13. This type of junction is known as an ALLOY or FUSED-ALLOY junction, and is one of the most commonly used junctions. In figure 1-14, a POINT-CONTACT type of construction is shown. It consists of a fine metal wire, called a cat whisker, that makes contact with a small area on the surface of an N-type semiconductor as shown in view A of the figure. The PN union is formed in this process by momentarily applying a high-surge current to the wire and the N-type semiconductor. The heat generated by this current converts the material nearest the point of contact to a P-type material (view B).

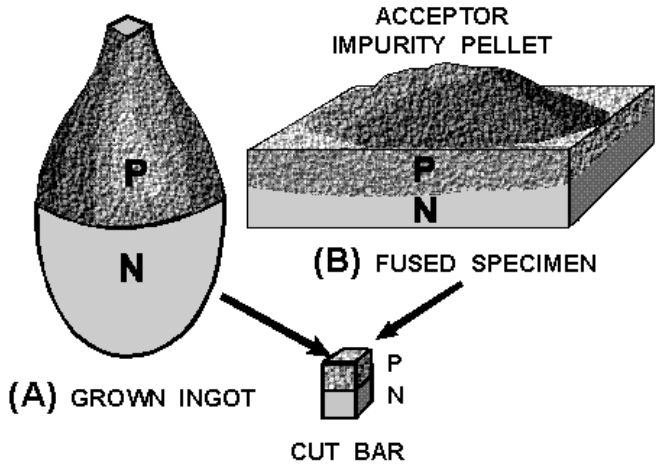


Figure 1-13.—Grown and fused PN junctions from which bars are cut.

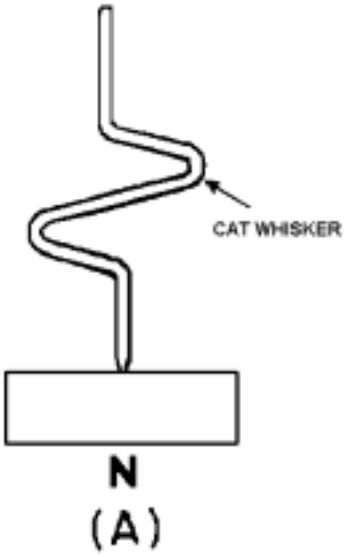


Figure 1-14A.—The point-contact type of diode construction.

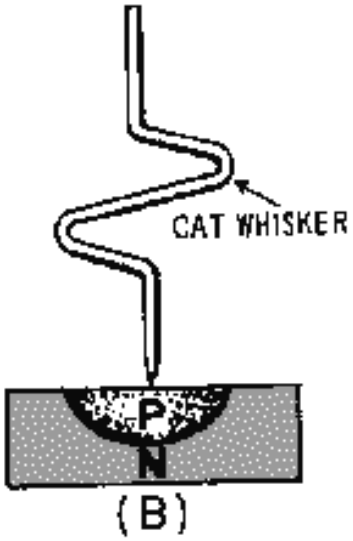


Figure 1-14B.—The point-contact type of diode construction.

Still another process is to heat a section of semiconductor material to near melting and then diffuse impurity atoms into a surface layer. Regardless of the process, the objective is to have a perfect bond everywhere along the union (interface) between P and N materials. Proper contact along the union is important because, as we will see later, the union (junction or interface) is the rectifying agent in the diode.

Q19. What is the purpose of a PN junction diode?

Q20. In reference to the schematic symbol for a diode, do electrons flow toward or away from the arrow?

Q21. What type of PN diode is formed by using a fine metal wire and a section of N-type semiconductor material?

PN JUNCTION OPERATION

Now that you are familiar with P- and N-type materials, how these materials are joined together to form a diode, and the function of the diode, let us continue our discussion with the operation of the PN junction. But before we can understand how the PN junction works, we must first consider current flow in the materials that make up the junction and what happens initially within the junction when these two materials are joined together.

Current Flow in the N-Type Material

Conduction in the N-type semiconductor, or crystal, is similar to conduction in a copper wire. That is, with voltage applied across the material, electrons will move through the crystal just as current would flow in a copper wire. This is shown in figure 1-15. The positive potential of the battery will attract the free electrons in the crystal. These electrons will leave the crystal and flow into the positive terminal of the battery. As an electron leaves the crystal, an electron from the negative terminal of the battery will enter the crystal, thus completing the current path. Therefore, the majority current carriers in the N-type material (electrons) are repelled by the negative side of the battery and move through the crystal toward the positive side of the battery.

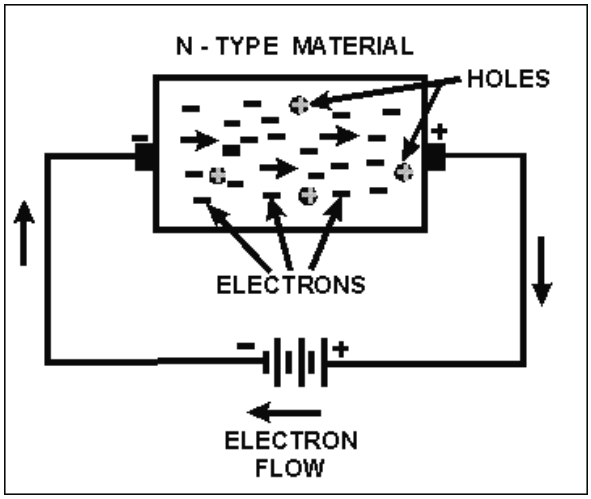


Figure 1-15.—Current flow In the N-type material.

Current Flow in the P-Type Material

Current flow through the P-type material is illustrated in figure 1-16. Conduction in the P material is by positive holes, instead of negative electrons. A hole moves from the positive terminal of the P material to the negative terminal. Electrons from the external circuit enter the negative terminal of the material and fill holes in the vicinity of this terminal. At the positive terminal, electrons are removed from the covalent bonds, thus creating new holes. This process continues as the steady stream of holes (hole current) moves toward the negative terminal.

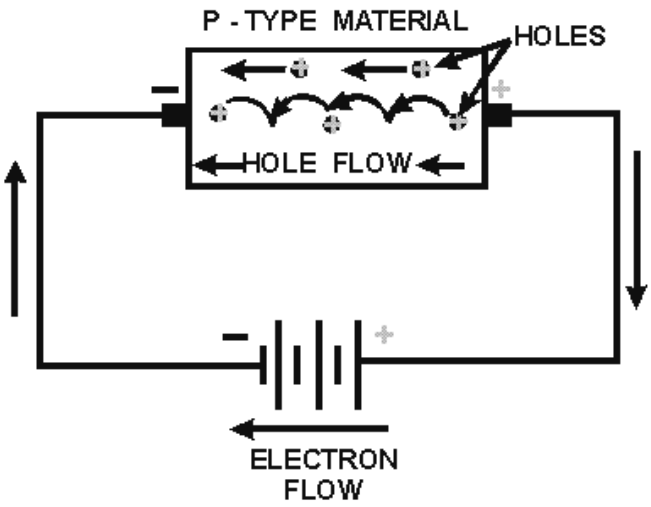


Figure 1-16.—Current flow In the P-type material.

Notice in both N-type and P-type materials, current flow in the external circuit consists of electrons moving out of the negative terminal of the battery and into the positive terminal of the battery. Hole flow, on the other hand, only exists within the material itself.

Q22. What are the majority carriers in a P-type semiconductor?

Q23. Conduction in which type of semiconductor material is similar to conduction in a copper wire?

Junction Barrier

Although the N-type material has an excess of free electrons, it is still electrically neutral. This is because the donor atoms in the N material were left with positive charges after free electrons became available by covalent bonding (the protons outnumbered the electrons). Therefore, for every free electron in the N material, there is a corresponding positively charge atom to balance it. The end result is that the N material has an overall charge of zero.

By the same reasoning, the P-type material is also electrically neutral because the excess of holes in this material is exactly balanced by the number of electrons. Keep in mind that the holes and electrons are still free to move in the material because they are only loosely bound to their parent atoms.

It would seem that if we joined the N and P materials together by one of the processes mentioned earlier, all the holes and electrons would pair up. On the contrary, this does not happen. Instead the electrons in the N material diffuse (move or spread out) across the junction into the P material and fill some of the holes. At the same time, the holes in the P material diffuse across the junction into the N material and are filled by N material electrons. This process, called JUNCTION RECOMBINATION, reduces the number of free electrons and holes in the vicinity of the junction. Because there is a depletion, or lack of free electrons and holes in this area, it is known as the DEPLETION REGION.

The loss of an electron from the N-type material created a positive ion in the N material, while the loss of a hole from the P material created a negative ion in that material. These ions are fixed in place in the crystal lattice structure and cannot move. Thus, they make up a layer of fixed charges on the two sides of the junction as shown in figure 1-17. On the N side of the junction, there is a layer of positively charged ions; on the P side of the junction, there is a layer of negatively charged ions. An electrostatic field, represented by a small battery in the figure, is established across the junction between the oppositely

charged ions. The diffusion of electrons and holes across the junction will continue until the magnitude of the electrostatic field is increased to the point where the electrons and holes no longer have enough energy to overcome it, and are repelled by the negative and positive ions respectively. At this point equilibrium is established and, for all practical purposes, the movement of carriers across the junction ceases. For this reason, the electrostatic field created by the positive and negative ions in the depletion region is called a barrier.

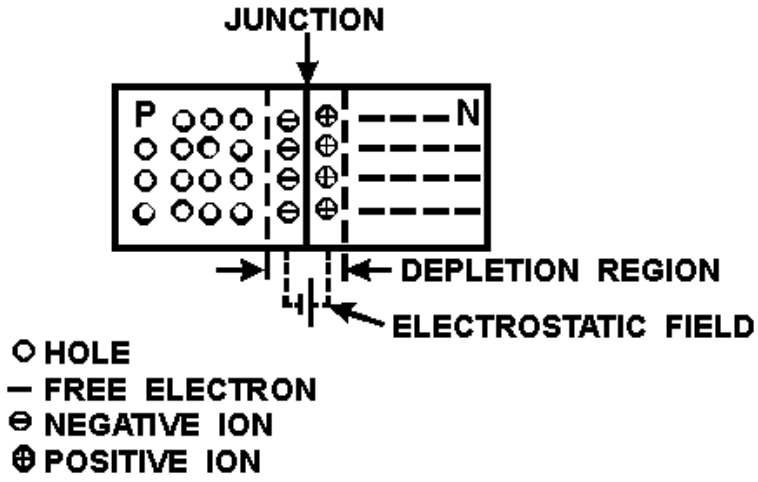


Figure 1-17.—The PN junction barrier formation.

The action just described occurs almost instantly when the junction is formed. Only the carriers in the immediate vicinity of the junction are affected. The carriers throughout the remainder of the N and P material are relatively undisturbed and remain in a balanced condition.

FORWARD BIAS.—An external voltage applied to a PN junction is call BIAS. If, for example, a battery is used to supply bias to a PN junction and is connected so that its voltage opposes the junction field, it will reduce the junction barrier and, therefore, aid current flow through the junction. This type of bias is known as forward bias, and it causes the junction to offer only minimum resistance to the flow of current.

Forward bias is illustrated in figure 1-18. Notice the positive terminal of the bias battery is connected to the P-type material and the negative terminal of the battery is connected to the N-type material. The positive potential repels holes toward the junction where they neutralize some of the negative ions. At the same time the negative potential repels electrons toward the junction where they neutralize some of the positive ions. Since ions on both sides of the barrier are being neutralized, the width of the barrier decreases. Thus, the effect of the battery voltage in the forward-bias direction is to reduce the barrier potential across the junction and to allow majority carriers to cross the junction. Current flow in the forward-biased PN junction is relatively simple. An electron leaves the negative terminal of the battery and moves to the terminal of the N-type material. It enters the N material, where it is the majority carrier and moves to the edge of the junction barrier. Because of forward bias, the barrier offers less opposition to the electron and it will pass through the depletion region into the P-type material. The electron loses energy in overcoming the opposition of the junction barrier, and upon entering the P material, combines with a hole. The hole was produced when an electron was extracted from the P material by the positive potential of the battery. The created hole moves through the P material toward the junction where it combines with an electron.

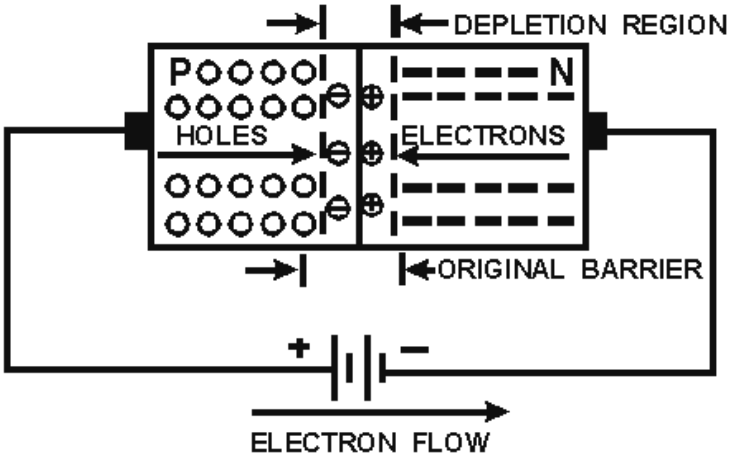


Figure 1-18.—Forward-biased PN junction.

It is important to remember that in the forward biased condition, conduction is by MAJORITY current carriers (holes in the P-type material and electrons in the N-type material). Increasing the battery voltage will increase the number of majority carriers arriving at the junction and will therefore increase the current flow. If the battery voltage is increased to the point where the barrier is greatly reduced, a heavy current will flow and the junction may be damaged from the resulting heat.

REVERSE BIAS.—If the battery mentioned earlier is connected across the junction so that its voltage aids the junction, it will increase the junction barrier and thereby offer a high resistance to the current flow through the junction. This type of bias is known as reverse bias.

To reverse bias a junction diode, the negative battery terminal is connected to the P-type material, and the positive battery terminal to the N-type material as shown in figure 1-19. The negative potential attracts the holes away from the edge of the junction barrier on the P side, while the positive potential attracts the electrons away from the edge of the barrier on the N side. This action increases the barrier width because there are more negative ions on the P side of the junction, and more positive ions on the N side of the junction. Notice in the figure the width of the barrier has increased. This increase in the number of ions prevents current flow across the junction by majority carriers. However, the current flow across the barrier is not quite zero because of the minority carriers crossing the junction. As you recall, when the crystal is subjected to an external source of energy (light, heat, etc.), electron-hole pairs are generated. The electron-hole pairs produce minority current carriers. There are minority current carriers in both regions: holes in the N material and electrons in the P material. With reverse bias, the electrons in the P-type material are repelled toward the junction by the negative terminal of the battery. As the electron moves across the junction, it will neutralize a positive ion in the N-type material. Similarly, the holes in the N-type material will be repelled by the positive terminal of the battery toward the junction. As the hole crosses the junction, it will neutralize a negative ion in the P-type material. This movement of minority carriers is called MINORITY CURRENT FLOW, because the holes and electrons involved come from the electron-hole pairs that are generated in the crystal lattice structure, and not from the addition of impurity atoms.

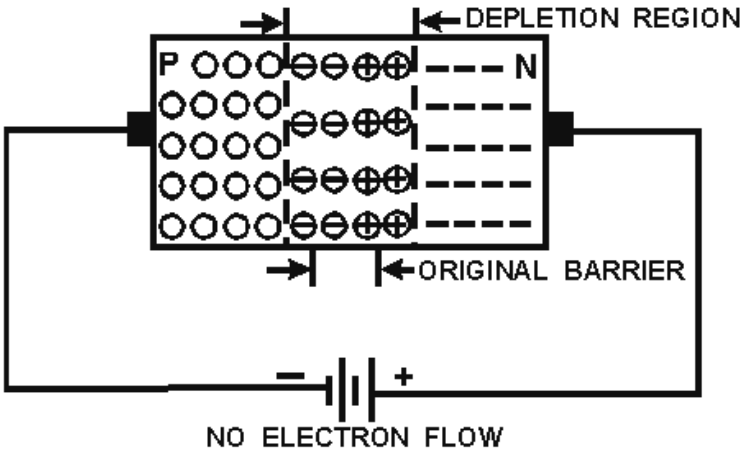


Figure 1-19.—Reverse-biased PN junction.

Therefore, when a PN junction is reverse biased, there will be no current flow because of majority carriers but a very small amount of current because of minority carriers crossing the junction. However, at normal operating temperatures, this small current may be neglected.

In summary, the most important point to remember about the PN junction diode is its ability to offer very little resistance to current flow in the forward-bias direction but maximum resistance to current flow when reverse biased. A good way of illustrating this point is by plotting a graph of the applied voltage versus the measured current. Figure 1-20 shows a plot of this voltage-current relationship (characteristic curve) for a typical PN junction diode.

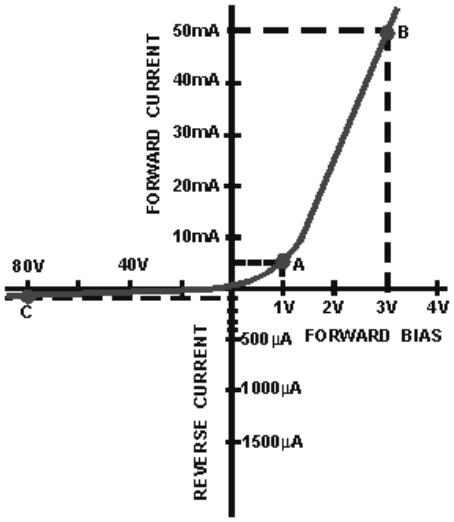


Figure 1-20.—PN junction diode characteristic curve.

To determine the resistance from the curve in this figure we can use Ohm's law:

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

For example at point A the forward-bias voltage is 1 volt and the forward-bias current is 5 milliamperes. This represents 200 ohms of resistance (1 volt/5mA = 200 ohms). However, at point B the voltage is 3 volts and the current is 50 milliamperes. This results in 60 ohms of resistance for the diode. Notice that when the forward-bias voltage was tripled (1 volt to 3 volts), the current increased 10 times (5mA to 50 mA). At the same time the forward-bias voltage increased, the resistance decreased from 200 ohms to 60 ohms. In other words, when forward bias increases, the junction barrier gets smaller and its resistance to current flow decreases.

On the other hand, the diode conducts very little when reverse biased. Notice at point C the reverse bias voltage is 80 volts and the current is only 100 microamperes. This results in 800 k ohms of resistance, which is considerably larger than the resistance of the junction with forward bias. Because of these unusual features, the PN junction diode is often used to convert alternating current into direct current (rectification).

Q24. What is the name of the area in a PN junction that has a shortage of electrons and holes?

Q25. In order to reverse bias in a PN junction, what terminal of a battery is connected to the P material?

Q26. What type of bias opposes the PN junction barrier?

PN JUNCTION APPLICATION

Until now, we have mentioned only one application for the diode-rectification, but there are many more applications that we have not yet discussed. Variations in doping agents, semiconductor materials, and manufacturing techniques have made it possible to produce diodes that can be used in many different applications. Examples of these types of diodes are signal diodes, rectifying diodes, Zener diodes (voltage protection diodes for power supplies), varactors (amplifying and switching diodes), and many more. Only applications for two of the most commonly used diodes, the signal diode and rectifier diode, will be presented in this chapter. The other diodes will be explained later on in this module.

Half-Wave Rectifier

One of the most important uses of a diode is rectification. The normal PN junction diode is well-suited for this purpose as it conducts very heavily when forward biased (low-resistance direction) and only slightly when reverse biased (high-resistance direction). If we place this diode in series with a source of ac power, the diode will be forward and reverse biased every cycle. Since in this situation current flows more easily in one direction than the other, rectification is accomplished. The simplest rectifier circuit is a half-wave rectifier (fig. 1-21 view A and view B) which consists of a diode, an ac power source, and a load resister.

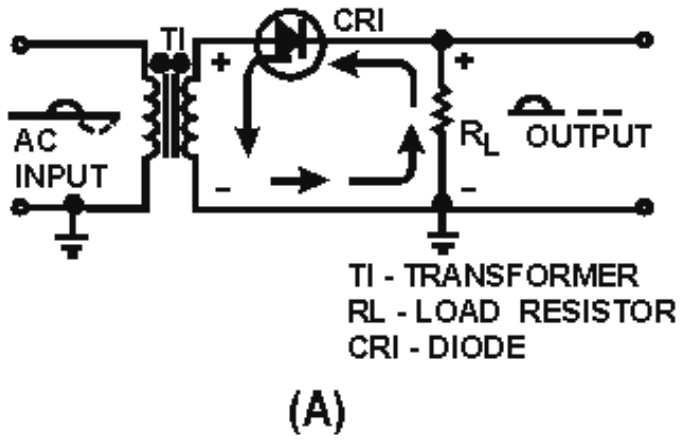


Figure 1-21A.—Simple half-wave rectifier.

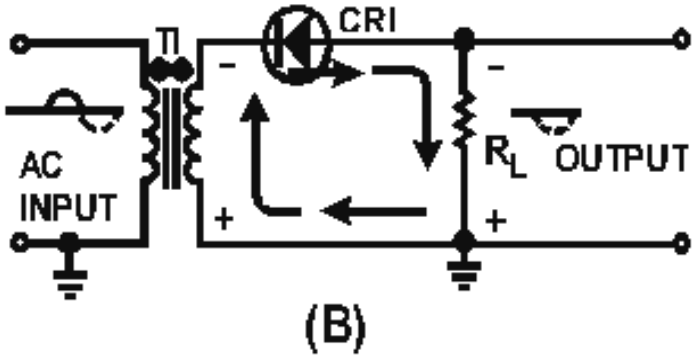


Figure 1-21B.—Simple half-wave rectifier.

The transformer (T1) in the figure provides the ac input to the circuit; the diode (CR1) provides the rectification; and the load resistor (R_L) serves two purposes: it limits the amount of current flow in the circuit to a safe level, and it also develops the output signal because of the current flow through it.

Before describing how this circuit operates, the definition of the word "load" as it applies to power supplies must be understood. Load is defined as any device that draws current. A device that draws little current is considered a light load, whereas a device that draws a large amount of current is a heavy load. Remember that when we speak of "load," we are speaking about the device that draws current from the power source. This device may be a simple resistor, or one or more complicated electronic circuits.

During the positive half-cycle of the input signal (solid line) in figure 1-21 view A, the top of the transformer is positive with respect to ground. The dots on the transformer indicate points of the same polarity. With this condition the diode is forward biased, the depletion region is narrow, the resistance of the diode is low, and current flows through the circuit in the direction of the solid lines. When this current flows through the load resistor, it develops a negative to positive voltage drop across it, which appears as a positive voltage at the output terminal.

When the ac input goes in a negative direction (fig. 1-21 view A), the top of the transformer becomes negative and the diode becomes reverse biased. With reverse bias applied to the diode, the depletion region increases, the resistance of the diode is high, and minimum current flows through the diode. For all

practical purposes, there is no output developed across the load resistor during the negative alternation of the input signal as indicated by the broken lines in the figure. Although only one cycle of input is shown, it should be realized that the action described above continually repeats itself, as long as there is an input. Therefore, since only the positive half-cycles appear at the output this circuit converted the ac input into a positive pulsating dc voltage. The frequency of the output voltage is equal to the frequency of the applied ac signal since there is one pulse out for each cycle of the ac input. For example, if the input frequency is 60 hertz (60 cycles per second), the output frequency is 60 pulses per second (pps).

However, if the diode is reversed as shown in view B of figure 1-21, a negative output voltage would be obtained. This is because the current would be flowing from the top of R_L toward the bottom, making the output at the top of R_L negative with respect to the bottom or ground. Because current flows in this circuit only during half of the input cycle, it is called a half-wave rectifier.

The semiconductor diode shown in the figure can be replaced by a metallic rectifier and still achieve the same results. The metallic rectifier, sometimes referred to as a dry-disc rectifier, is a metal-to-semiconductor, large-area contact device. Its construction is distinctive; a semiconductor is sandwiched between two metal plates, or electrodes, as shown in figure 1-22. Note in the figure that a barrier, with a resistance many times greater than that of the semiconductor material, is constructed on one of the metal electrodes. The contact having the barrier is a rectifying contact; the other contact is nonrectifying. Metallic rectifiers act just like the diodes previously discussed in that they permit current to flow more readily in one direction than the other. However, the metallic rectifier is fairly large compared to the crystal diode as can be seen in figure 1-23. The reason for this is: metallic rectifier units are stacked (to prevent inverse voltage breakdown), have large area plates (to handle high currents), and usually have cooling fins (to prevent overheating).

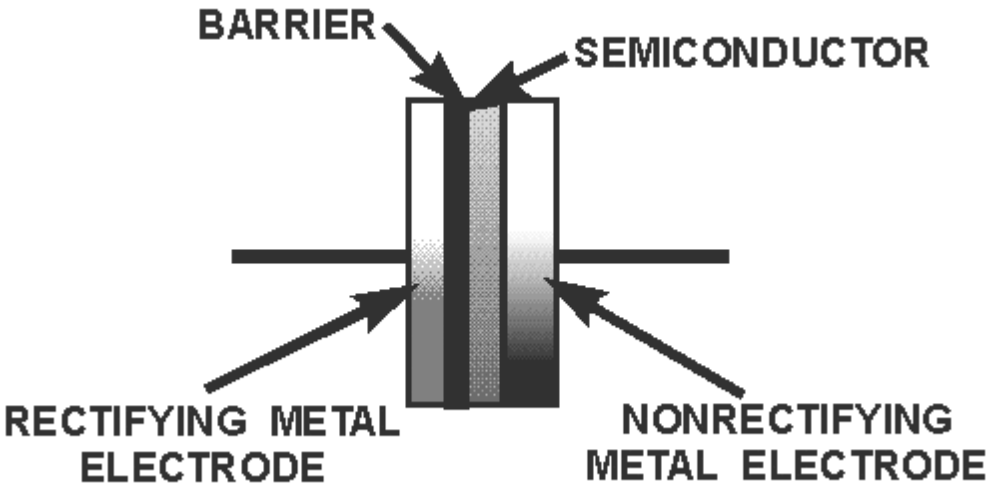


Figure 1-22.—A metallic rectifier.

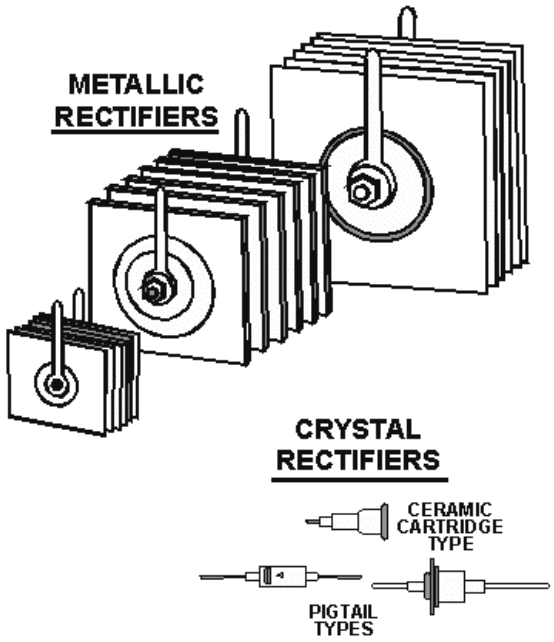


Figure 1-23.—Different types of crystal and metallic rectifiers.

There are many known metal-semiconductor combinations that can be used for contact rectification. Copper oxide and selenium devices are by far the most popular. Copper oxide and selenium are frequently used over other types of metallic rectifiers because they have a large forward current per unit contact area, low forward voltage drop, good stability, and a lower aging rate. In practical application, the selenium rectifier is used where a relatively large amount of power is required. On the other hand, copper-oxide rectifiers are generally used in small-current applications such as ac meter movements or for delivering direct current to circuits requiring not more than 10 amperes.

Since metallic rectifiers are affected by temperature, atmospheric conditions, and aging (in the case of copper oxide and selenium), they are being replaced by the improved silicon crystal rectifier. The silicon rectifier replaces the bulky selenium rectifier as to current and voltage rating, and can operate at higher ambient (surrounding) temperatures.

Diode Switch

In addition to their use as simple rectifiers, diodes are also used in circuits that mix signals together (mixers), detect the presence of a signal (detector), and act as a switch "to open or close a circuit." Diodes used in these applications are commonly referred to as "signal diodes." The simplest application of a signal diode is the basic diode switch shown in figure 1-24.

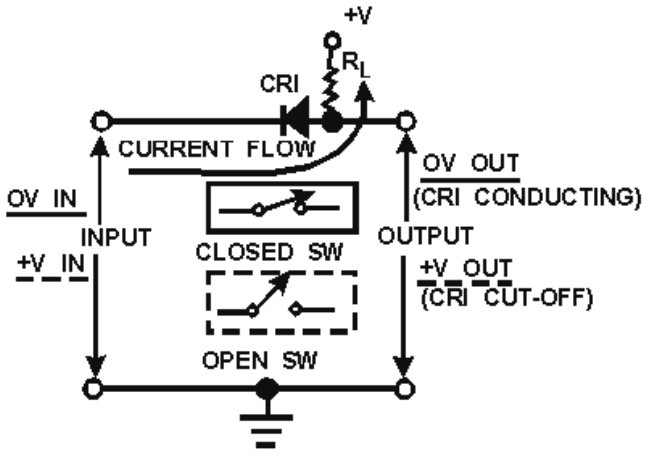


Figure 1-24.—Basic diode switch.

When the input to this circuit is at zero potential, the diode is forward biased because of the zero potential on the cathode and the positive voltage on the anode. In this condition, the diode conducts and acts as a straight piece of wire because of its very low forward resistance. In effect, the input is directly coupled to the output resulting in zero volts across the output terminals. Therefore, the diode, acts as a closed switch when its anode is positive with respect to its cathode.

If we apply a positive input voltage (equal to or greater than the positive voltage supplied to the anode) to the diode's cathode, the diode will be reverse biased. In this situation, the diode is cut off and acts as an open switch between the input and output terminals. Consequently, with no current flow in the circuit, the positive voltage on the diode's anode will be felt at the output terminal. Therefore, the diode acts as an open switch when it is reverse biased.

Q27. What is a load?

Q28. What is the output of a half-wave rectifier?

Q29. What type of rectifier is constructed by sandwiching a section of semiconductor material between two metal plates?

Q30. What type of bias makes a diode act as a closed switch?

DIODE CHARACTERISTICS

Semiconductor diodes have properties that enable them to perform many different electronic functions. To do their jobs, engineers and technicians must be supplied with data on these different types of diodes. The information presented for this purpose is called DIODE CHARACTERISTICS. These characteristics are supplied by manufacturers either in their manuals or on specification sheets (data sheets). Because of the scores of manufacturers and numerous diode types, it is not practical to put before you a specification sheet and call it typical. Aside from the difference between manufacturers, a single manufacturer may even supply specification sheets that differ both in format and content. Despite these differences, certain performance and design information is normally required. We will discuss this information in the next few paragraphs.

A standard specification sheet usually has a brief description of the diode. Included in this description is the type of diode, the major area of application, and any special features. Of particular interest is the specific application for which the diode is suited. The manufacturer also provides a drawing of the diode which gives dimension, weight, and, if appropriate, any identification marks. In addition to the above data, the following information is also provided: a static operating table (giving spot values of parameters under fixed conditions), sometimes a characteristic curve similar to the one in figure 1-20 (showing how parameters vary over the full operating range), and diode ratings (which are the limiting values of operating conditions outside which could cause diode damage).

Manufacturers specify these various diode operating parameters and characteristics with "letter symbols" in accordance with fixed definitions. The following is a list, by letter symbol, of the major electrical characteristics for the rectifier and signal diodes.

RECTIFIER DIODES

DC BLOCKING VOLTAGE [V_R]—the maximum reverse dc voltage that will not cause breakdown.

AVERAGE FORWARD VOLTAGE DROP [$V_{F(AV)}$]—the average forward voltage drop across the rectifier given at a specified forward current and temperature.

AVERAGE RECTIFIER FORWARD CURRENT [$I_{F(AV)}$]—the average rectified forward current at a specified temperature, usually at 60 Hz with a resistive load.

AVERAGE REVERSE CURRENT [$I_{R(AV)}$]—the average reverse current at a specified temperature, usually at 60 Hz.

PEAK SURGE CURRENT [I_{SURGE}]—the peak current specified for a given number of cycles or portion of a cycle.

SIGNAL DIODES

PEAK REVERSE VOLTAGE [PRV]—the maximum reverse voltage that can be applied before reaching the breakdown point. (PRV also applies to the rectifier diode.)

REVERSE CURRENT [I_R]—the small value of direct current that flows when a semiconductor diode has reverse bias.

MAXIMUM FORWARD VOLTAGE DROP AT INDICATED FORWARD CURRENT [$V_F@I_F$]—the maximum forward voltage drop across the diode at the indicated forward current.

REVERSE RECOVERY TIME [t_{rr}]—the maximum time taken for the forward-bias diode to recover its reverse bias.

The ratings of a diode (as stated earlier) are the limiting values of operating conditions, which if exceeded could cause damage to a diode by either voltage breakdown or overheating. The PN junction diodes are generally rated for: MAXIMUM AVERAGE FORWARD CURRENT, PEAK RECURRENT FORWARD CURRENT, MAXIMUM SURGE CURRENT, and PEAK REVERSE VOLTAGE.

Maximum average forward current is usually given at a special temperature, usually 25º C, (77º F) and refers to the maximum amount of average current that can be permitted to flow in the forward direction. If this rating is exceeded, structure breakdown can occur.

Peak recurrent forward current is the maximum peak current that can be permitted to flow in the forward direction in the form of recurring pulses.

Maximum surge current is the maximum current permitted to flow in the forward direction in the form of nonrecurring pulses. Current should not equal this value for more than a few milliseconds.

Peak reverse voltage (PRV) is one of the most important ratings. PRV indicates the maximum reverse-bias voltage that may be applied to a diode without causing junction breakdown.

All of the above ratings are subject to change with temperature variations. If, for example, the operating temperature is above that stated for the ratings, the ratings must be decreased.

Q31. What is used to show how diode parameters vary over a full operating range?

Q32. What is meant by diode ratings?

DIODE IDENTIFICATION

There are many types of diodes varying in size from the size of a pinhead (used in subminiature circuitry) to large 250-ampere diodes (used in high-power circuits). Because there are so many different types of diodes, some system of identification is needed to distinguish one diode from another. This is accomplished with the semiconductor identification system shown in figure 1-25. This system is not only used for diodes but transistors and many other special semiconductor devices as well. As illustrated in this figure, the system uses numbers and letters to identify different types of semiconductor devices. The first number in the system indicates the number of junctions in the semiconductor device and is a number, one less than the number of active elements. Thus 1 designates a diode; 2 designates a transistor (which may be considered as made up of two diodes); and 3 designates a tetrode (a four-element transistor). The letter "N" following the first number indicates a semiconductor. The 2- or 3-digit number following the letter "N" is a serialized identification number. If needed, this number may contain a suffix letter after the last digit. For example, the suffix letter "M" may be used to describe matching pairs of separate semiconductor devices or the letter "R" may be used to indicate reverse polarity. Other letters are used to indicate modified versions of the device which can be substituted for the basic numbered unit. For example, a semiconductor diode designated as type 1N345A signifies a two-element diode (1) of semiconductor material (N) that is an improved version (A) of type 345.

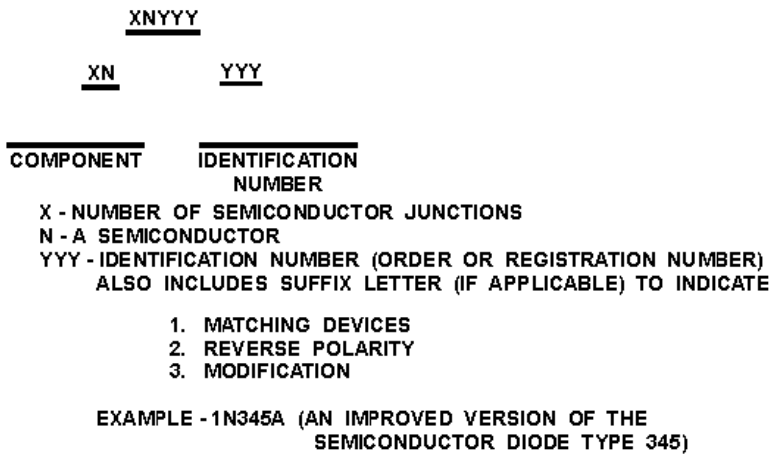


Figure 1-25.—Semiconductor identification codes.

When working with these different types of diodes, it is also necessary to distinguish one end of the diode from the other (anode from cathode). For this reason, manufacturers generally code the cathode end of the diode with a "k," "+," "cath," a color dot or band, or by an unusual shape (raised edge or taper) as shown in figure 1-26. In some cases, standard color code bands are placed on the cathode end of the diode. This serves two purposes: (1) it identifies the cathode end of the diode, and (2) it also serves to identify the diode by number.

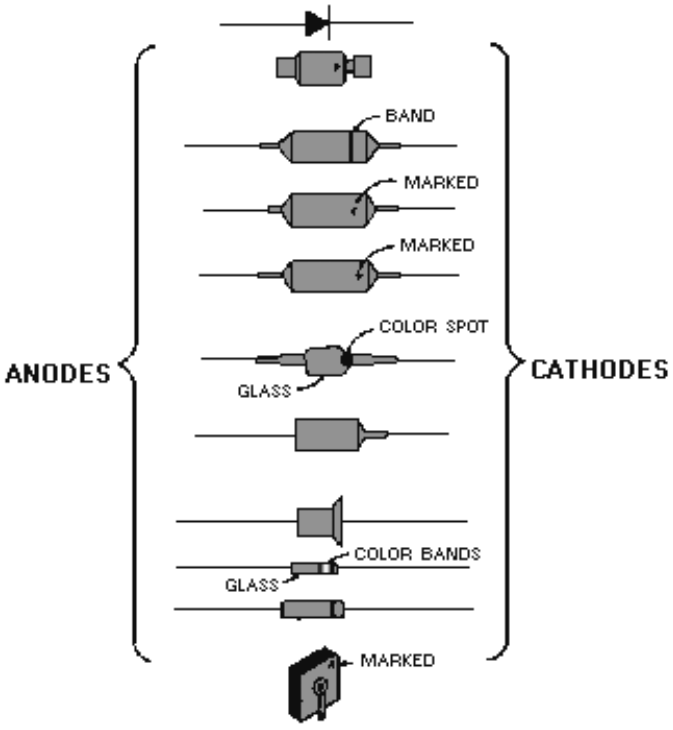
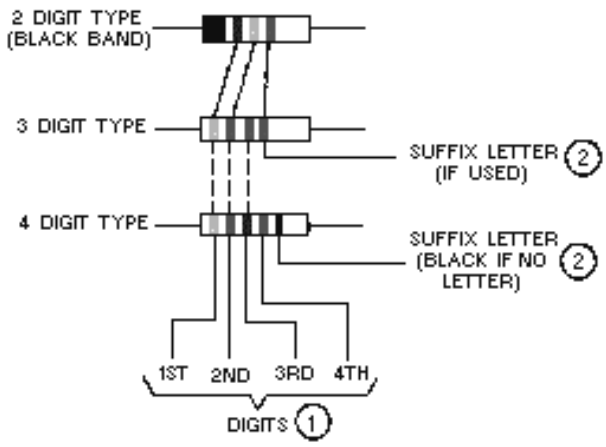


Figure 1-26.—Semiconductor diode markings.

The standard diode color code system is shown in figure 1-27. Take, for example, a diode with brown, orange, and white bands at one terminal and figure out its identification number. With brown

being a "1," orange a "3," and white "9," the device would be identified as a type 139 semiconductor diode, or specifically 1N139.



COLOR	① DIGIT	② DIODE SUFFIX LETTER
BLACK	0	-
BROWN	1	A
RED	2	B
ORANGE	3	C
YELLOW	4	D
GREEN	5	E
BLUE	6	F
VIOLET	7	G
GRAY	8	H
WHITE	9	J
SILVER	-	-
GOLD	-	-
NONE	-	-

Figure 1-27.—Semiconductor diode color code system.

Keep in mind, whether the diode is a small crystal type or a large power rectifier type, both are still represented schematically, as explained earlier, by the schematic symbol shown in figure 1-12.

Q33. What does the letter "N" indicate in the semiconductor identification system?

Q34. What type of diode has orange, blue, and gray bands?

DIODE MAINTENANCE

Diodes are rugged and efficient. They are also expected to be relatively trouble free. Protective encapsulation processes and special coating techniques have even further increased their life expectancies. In theory, a diode should last indefinitely. However, if diodes are subjected to current overloads, their junctions will be damaged or destroyed. In addition, the application of excessively high operating voltages can damage or destroy junctions through arc-over, or excessive reverse currents. One of the greatest dangers to the diode is heat. Heat causes more electron-hole pairs to be generated, which in turn increases current flow. This increase in current generates more heat and the cycle repeats itself until

the diode draws excessive current. This action is referred to as THERMAL RUNAWAY and eventually causes diode destruction. Extreme caution should be used when working with equipment containing diodes to ensure that these problems do not occur and cause irreparable diode damage.

The following is a list of some of the special safety precautions that should be observed when working with diodes:

- Never remove or insert a diode into a circuit with voltage applied.
- Never pry diodes to loosen them from their circuits.
- Always be careful when soldering to ensure that excessive heat is not applied to the diode.
- When testing a diode, ensure that the test voltage does not exceed the diode's maximum allowable voltage.
- Never put your fingers across a signal diode because the static charge from your body could short it out.
- Always replace a diode with a direct replacement, or with one of the same type.
- Ensure a replacement diode is put into a circuit in the correct direction.

If a diode has been subjected to excessive voltage or temperature and is suspected of being defective, it can be checked in various ways. The most convenient and quickest way of testing a diode is with an ohmmeter (fig. 1-28). To make the check, simply disconnect one of the diode leads from the circuit wiring, and make resistance measurements across the leads of the diode. The resistance measurements obtained depend upon the test-lead polarity of the ohmmeter; therefore, two measurements must be taken. The first measurement is taken with the test leads connected to either end of the diode and the second measurement is taken with the test leads reversed on the diode. The larger resistance value is assumed to be the reverse (back) resistance of the diode, and the smaller resistance (front) value is assumed to be the forward resistance. Measurement can be made for comparison purposes using another identical-type diode (known to be good) as a standard. Two high-value resistance measurements indicate that the diode is open or has a high forward resistance. Two low-value resistance measurements indicate that the diode is shorted or has a low reverse resistance. A normal set of measurements will show a high resistance in the reverse direction and a low resistance in the forward direction. The diode's efficiency is determined by how low the forward resistance is compared with the reverse resistance. That is, it is desirable to have as great a ratio (often known as the front-to-back ratio or the back-to-front ratio) as possible between the reverse and forward resistance measurements. However, as a rule of thumb, a small signal diode will have a ratio of several hundred to one, while a power rectifier can operate satisfactorily with a ratio of 10 to 1.

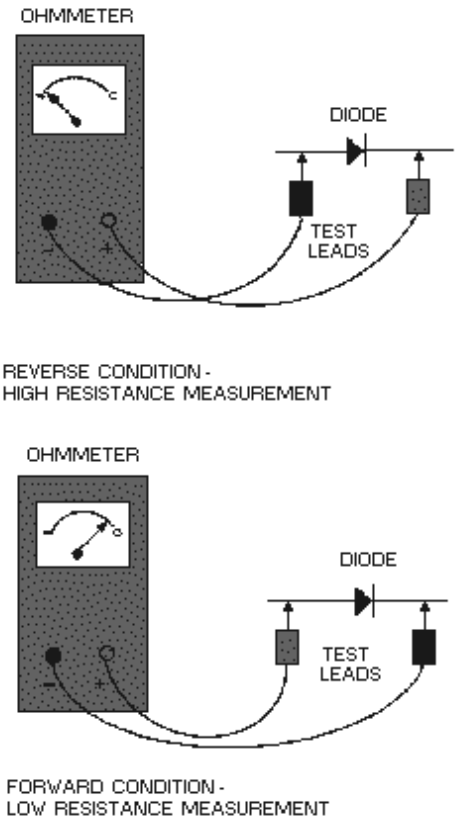


Figure 1-28.—Checking a diode with an ohmmeter.

One thing you should keep in mind about the ohmmeter check-it is not conclusive. It is still possible for a diode to check good under this test, but break down when placed back in the circuit. The problem is that the meter used to check the diode uses a lower voltage than the diode usually operates at in the circuit.

Another important point to remember is that a diode should not be condemned because two ohmmeters give different readings on the diode. This occurs because of the different internal resistances of the ohmmeters and the different states of charge on the ohmmeter batteries. Because each ohmmeter sends a different current through the diode, the two resistance values read on the meters will not be the same.

Another way of checking a diode is with the substitution method. In this method, a good diode is substituted for a questionable diode. This technique should be used only after you have made voltage and resistance measurements to make certain that there is no circuit defect that might damage the substitution diode. If more than one defective diode is present in the equipment section where trouble has been localized, this method becomes cumbersome, since several diodes may have to be replaced before the trouble is corrected. To determine which stages failed and which diodes are not defective, all of the removed diodes must be tested. This can be accomplished by observing whether the equipment operates correctly as each of the removed diodes is reinserted into the equipment.

In conclusion, the only valid check of a diode is a dynamic electrical test that determines the diode's forward current (resistance) and reverse current (resistance) parameters. This test can be accomplished using various crystal diode test sets that are readily available from many manufacturers.

Q35. What is the greatest threat to a diode?

Q36. When checking a diode with an ohmmeter, what is indicated by two high resistance measurements?

SUMMARY

Now that we have completed this chapter, a short review of the more important points covered in the chapter will follow. You should be thoroughly familiar with these points before continuing on to chapter 2.

The **UNIVERSE** consists of two main parts-matter and energy.

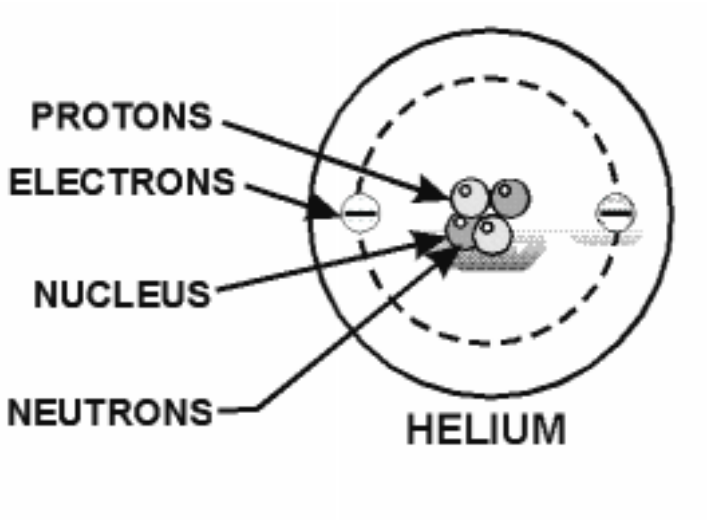
MATTER is anything that occupies space and has weight. Rocks, water, and air are examples of matter. Matter may be found in any one of three states: solid, liquid and gaseous. It can also be composed of either an element or a combination of elements.

An **ELEMENT** is a substance that cannot be reduced to a simpler form by chemical means. Iron, gold, silver, copper, and oxygen are all good examples of elements.

A **COMPOUND** is a chemical combination of two or more elements. Water, table salt, ethyl alcohol, and ammonia are all examples of compounds.

A **MOLECULE** is the smallest part of a compound that has all the characteristics of the compound. Each molecule contains some of the atoms of each of the elements forming the compound.

The **ATOM** is the smallest particle into which an element can be broken down and still retain all its original properties. An atom is made up of electrons, protons, and neutrons. The number and arrangement of these particles determine the kind of element.



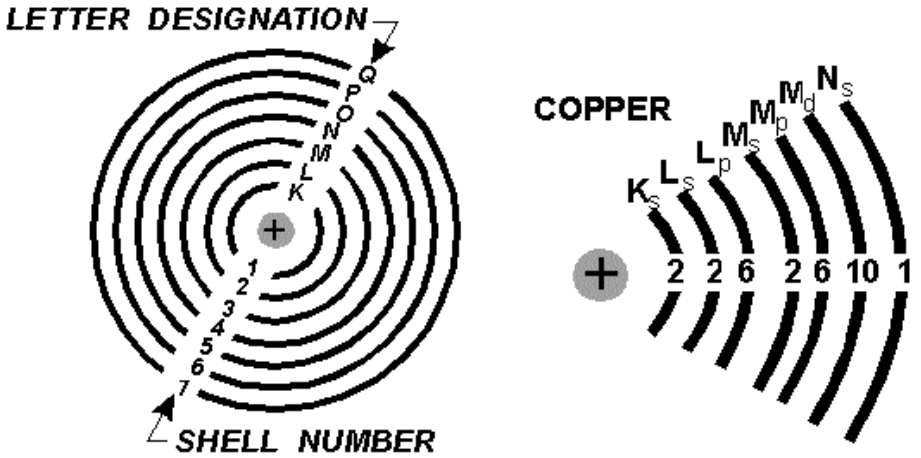
An **ELECTRON** carries a small negative charge of electricity.

The **PROTON** carries a positive charge of electricity that is equal and opposite to the charge of the electron. However, the mass of the proton is approximately 1,837 times that of the electron.

The **NEUTRON** is a neutral particle in that it has no electrical charge. The mass of the neutron is approximately equal to that of the proton.

An **ELECTRON'S ENERGY LEVEL** is the amount of energy required by an electron to stay in orbit. Just by the electron's motion alone, it has kinetic energy. The electron's position in reference to the nucleus gives it potential energy. An energy balance keeps the electron in orbit and as it gains or loses energy, it assumes an orbit further from or closer to the center of the atom.

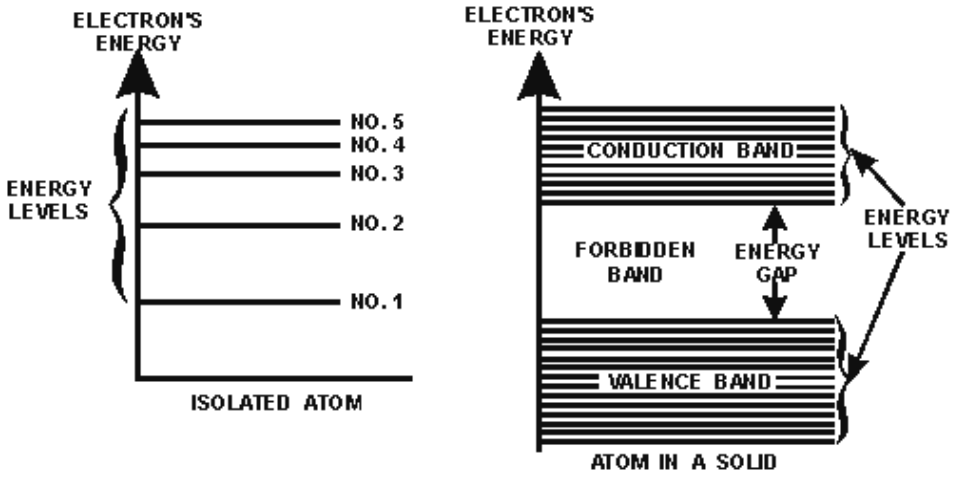
SHELLS and **SUBSHELLS** are the orbits of the electrons in an atom. Each shell can contain a maximum number of electrons, which can be determined by the formula $2n^2$. Shells are lettered K through Q, starting with K, which is the closest to the nucleus. The shell can also be split into four subshells labeled s, p, d, and f, which can contain 2, 6, 10, and 14 electrons, respectively.



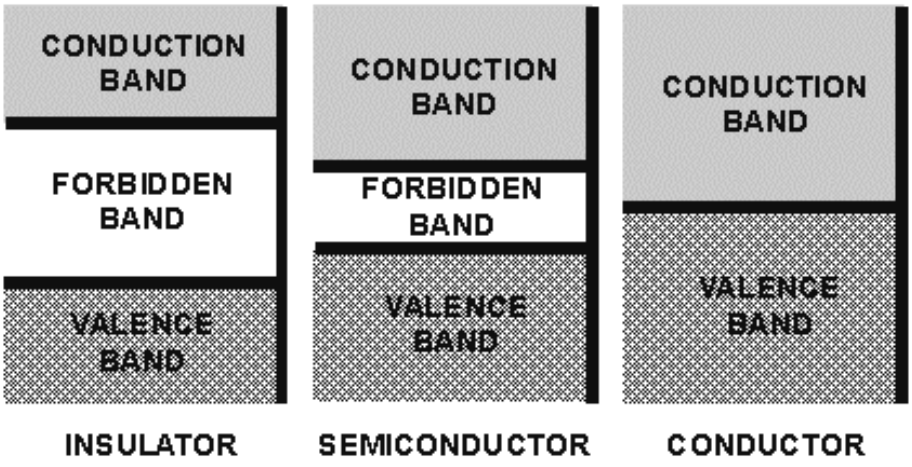
VALENCE is the ability of an atom to combine with other atoms. The valence of an atom is determined by the number of electrons in the atom's outermost shell. This shell is referred to as the VALENCE SHELL. The electrons in the outermost shell are called VALENCE ELECTRONS.

IONIZATION is the process by which an atom loses or gains electrons. An atom that loses some of its electrons in the process becomes positively charged and is called a POSITIVE ION. An atom that has an excess number of electrons is negatively charged and is called a **NEGATIVE ION**.

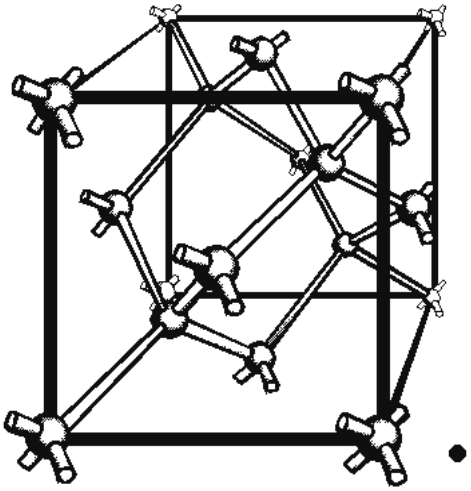
ENERGY BANDS are groups of energy levels that result from the close proximity of atoms in a solid. The three most important energy bands are the CONDUCTION BAND, FORBIDDEN BAND, and VALENCE BAND.



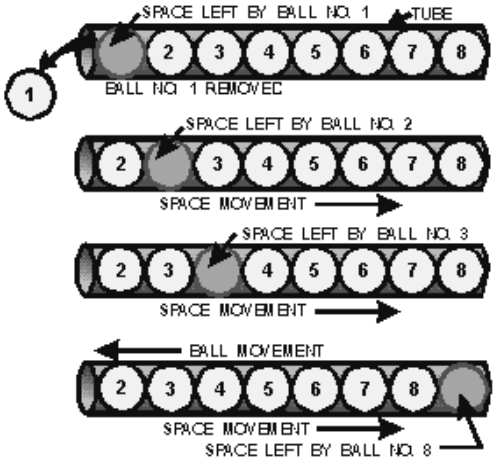
CONDUCTORS, **SEMICONDUCTORS**, and **INSULATORS** are categorized as such by using the energy band concept. It is the width of the forbidden band that determines whether a material is an insulator, a semiconductor, or a conductor. A CONDUCTOR has a very narrow forbidden band or none at all. A SEMICONDUCTOR has a medium width forbidden band. An INSULATOR has a wide forbidden band.



COVALENT BONDING is the sharing of valence electrons between two or more atoms. It is this bonding that holds the atoms together in an orderly structure called a CRYSTAL.



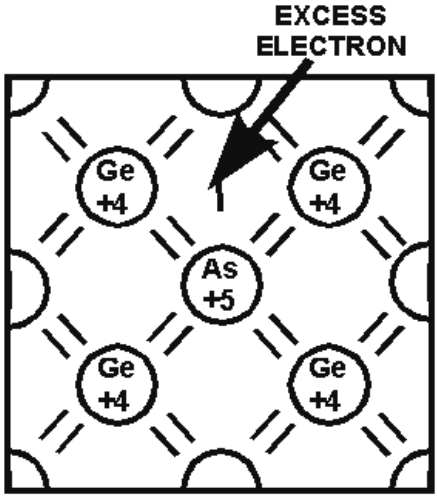
The **CONDUCTION PROCESS** in a **SEMICONDUCTOR** is accomplished by two different types of current flow: **HOLE FLOW** and **ELECTRON FLOW**. Hole flow is very similar to electron flow except that holes (positive charges) move toward a negative potential and in an opposite direction to that of the electrons. In an **INTRINSIC** semiconductor (one which does not contain any impurities), the number of holes always equals the number of conducting electrons.



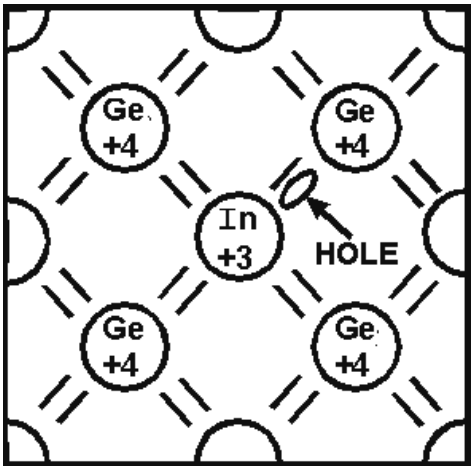
DOPING is the process by which small amounts of selected additives, called impurities, are added to semiconductors to increase their current flow. Semiconductors that undergo this treatment are referred to as **EXTRINSIC SEMICONDUCTORS**.

An **N-TYPE SEMICONDUCTOR** is one that is doped with an **N-TYPE** or donor impurity (an impurity that easily loses its extra electron to the semiconductor causing it to have an excess number of

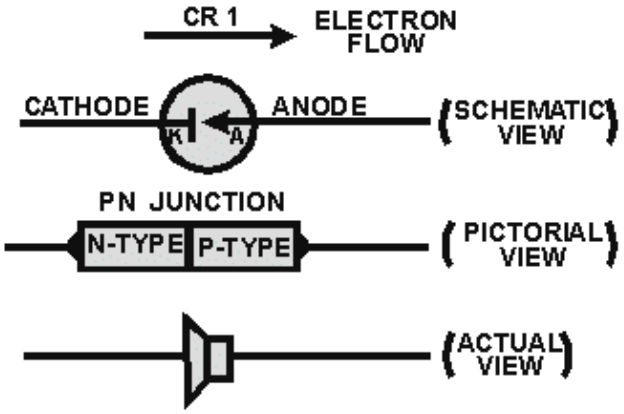
free electrons). Since this type of semiconductor has a surplus of electrons, the electrons are considered the majority current carriers, while the holes are the minority current carriers.



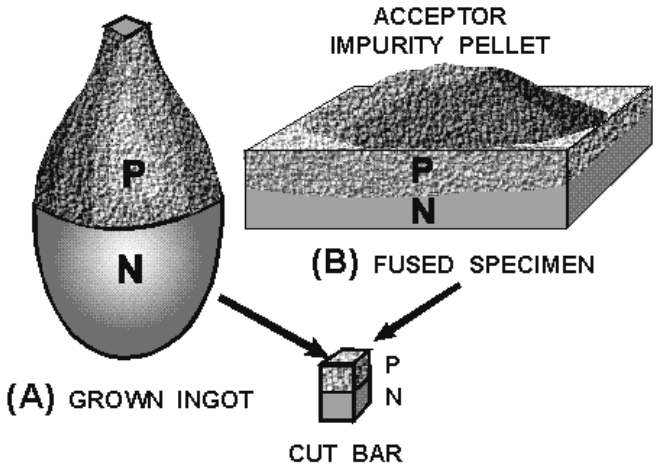
A **P-TYPE SEMICONDUCTOR** is one which is doped with a P-TYPE or acceptor impurity (an impurity that reduces the number of free electrons causing more holes). The holes in this type semiconductor are the majority current carriers since they are present in the greatest quantity while the electrons are the minority current carriers.



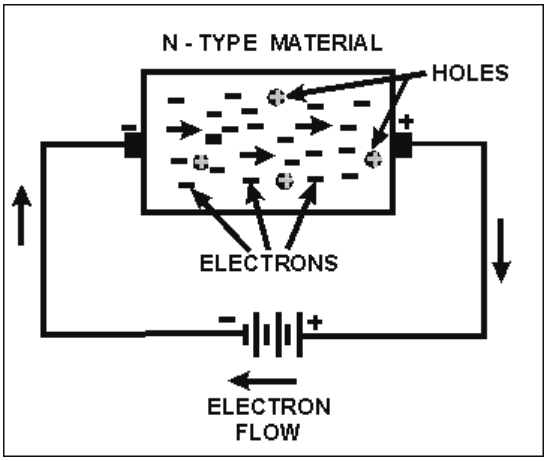
The **SEMICONDUCTOR DIODE**, also known as **a PN JUNCTION DIODE**, is a two-element semiconductor device that makes use of the rectifying properties of a PN junction to convert alternating current into direct current by permitting current flow in only one direction.



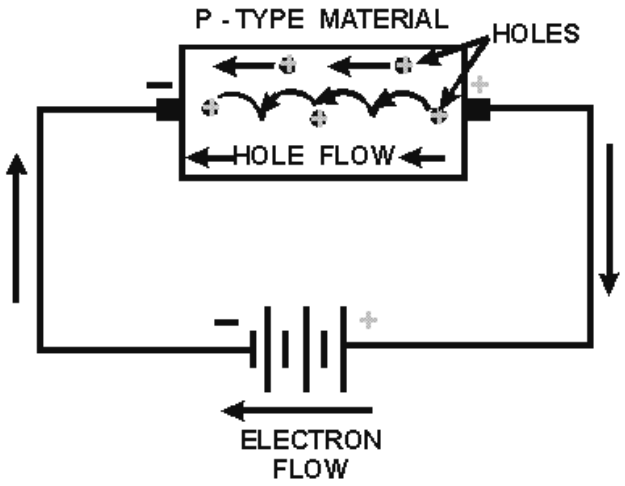
A **PN JUNCTION CONSTRUCTION** varies from one manufacturer to the next. Some of the more commonly used manufacturing techniques are: **GROWN**, **ALLOY** or **FUSED-ALLOY**, **DIFFUSED**, and **POINT-CONTACT**.



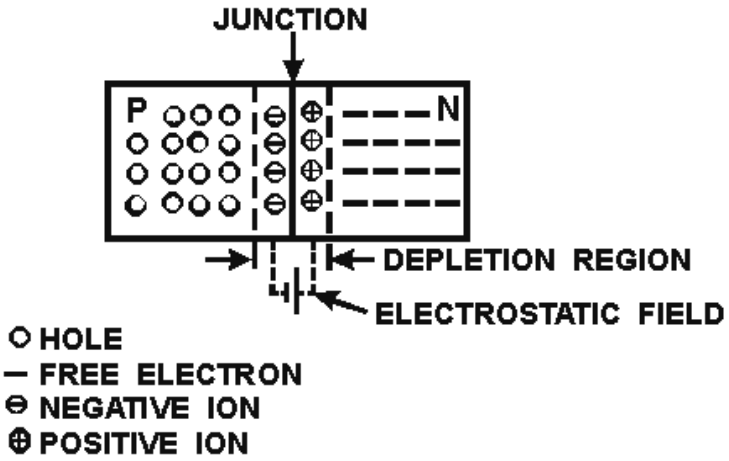
CURRENT FLOW in an **N-TYPE MATERIAL** is similar to conduction in a copper wire. That is, with voltage applied across the material, electrons will move through the crystal toward the positive terminal just like current flows in a copper wire.



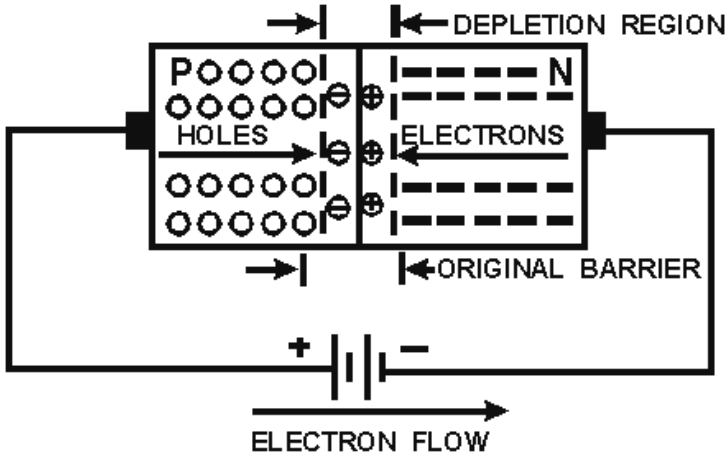
CURRENT FLOW in a **P-TYPE MATERIAL** is by positive holes, instead of negative electrons. Unlike the electron, the hole moves from the positive terminal of the P material to the negative terminal.



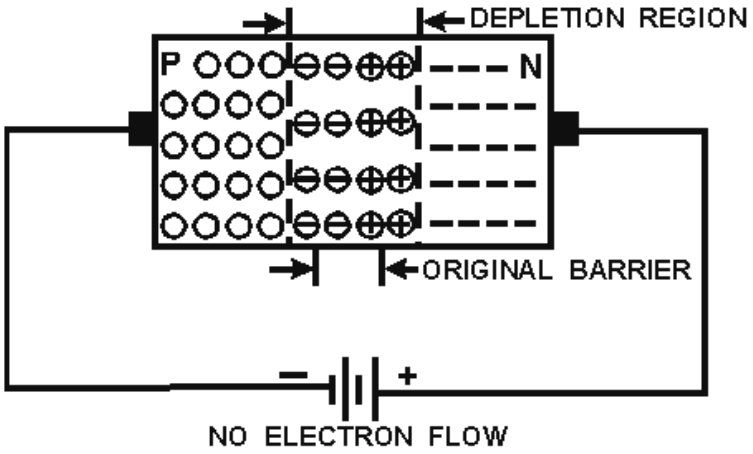
JUNCTION BARRIER is an electrostatic field that has been created by the joining of a section of N material with a section of P material. Since holes and electrons must overcome this field to cross the junction, the electrostatic field is commonly called a BARRIER. Because there is a lack or depletion of free electrons and holes in the area around the barrier, this area has become known as the DEPLETION REGION.



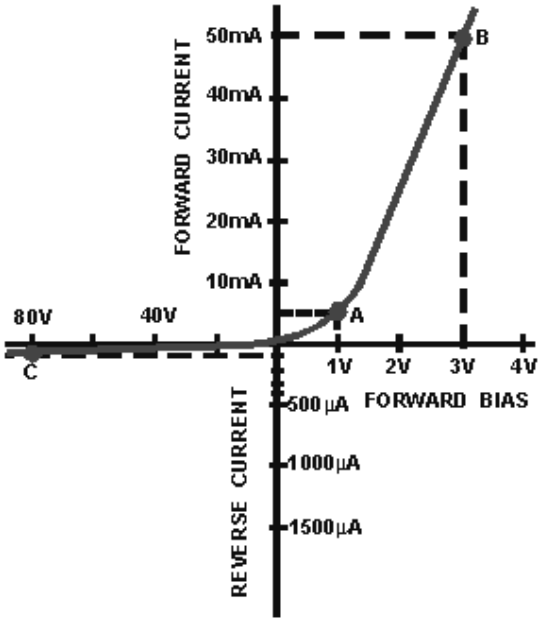
FORWARD BIAS is an external voltage that is applied to a PN junction to reduce its barrier and, therefore, aid current flow through the junction. To accomplish this function, the external voltage is connected so that it opposes the electrostatic field of the junction.



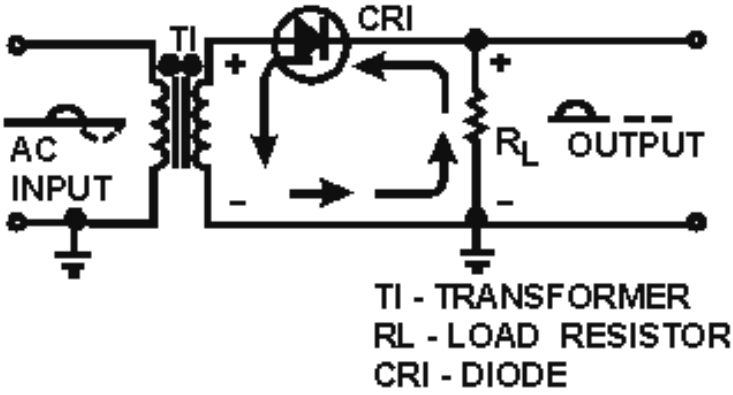
REVERSE BIAS is an external voltage that is connected across a PN junction so that its voltage aids the junction and, thereby, offers a high resistance to the current flow through the junction.



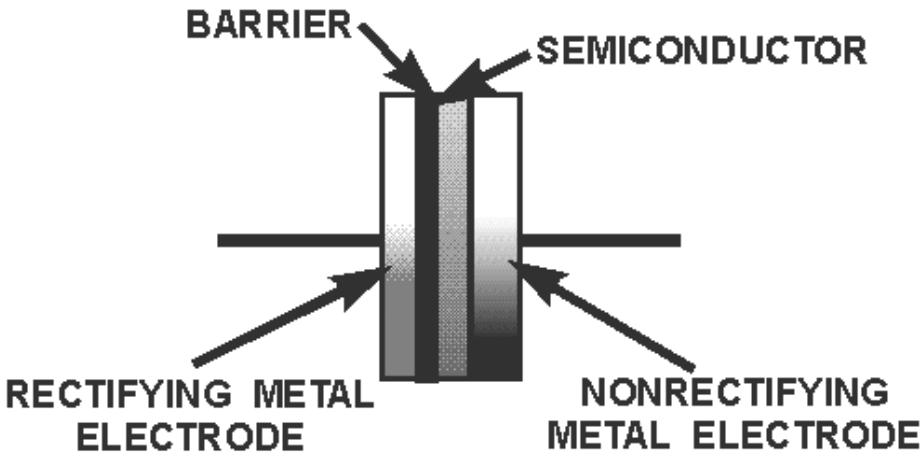
The **PN JUNCTION** has a unique ability to offer very little resistance to current flow in the forward-bias direction, but maximum resistance to current flow when reverse biased. For this reason, the PN junction is commonly used as a diode to convert ac to dc.



The **PN JUNCTION'S APPLICATION** expands many different areas-from a simple voltage protection device to an amplifying diode. Two of the most commonly used applications for the PN junction are the **SIGNAL DIODE** (mixing, detecting, and switching signals) and the **RECTIFYING DIODE** (converting ac to dc).



The **METALLIC RECTIFIER** or dry-disc rectifier is a metal-to-semiconductor device that acts just like a diode in that it permits current to flow more readily in one direction than the other. Metallic rectifiers are used in many applications where a relatively large amount of power is required.



DIODE CHARACTERISTICS is the information supplied by manufacturers on different types of diodes, either in their manuals or on specification sheets.

DIODE RATINGS are the limiting value of operating conditions of a diode. Operation of the diode outside of its operating limits could damage the diode. Diodes are generally rated for: **MAXIMUM AVERAGE FORWARD CURRENT**, **PEAK RECURRENT FORWARD CURRENT**, **MAXIMUM SURGE CURRENT**, and **PEAK REVERSE VOLTAGE**.

The **SEMICONDUCTOR IDENTIFICATION SYSTEM** is an alphanumerical code used to distinguish one semiconductor from another. It is used for diodes, transistors, and many other special semiconductor devices.

XNYYY

XN YYY

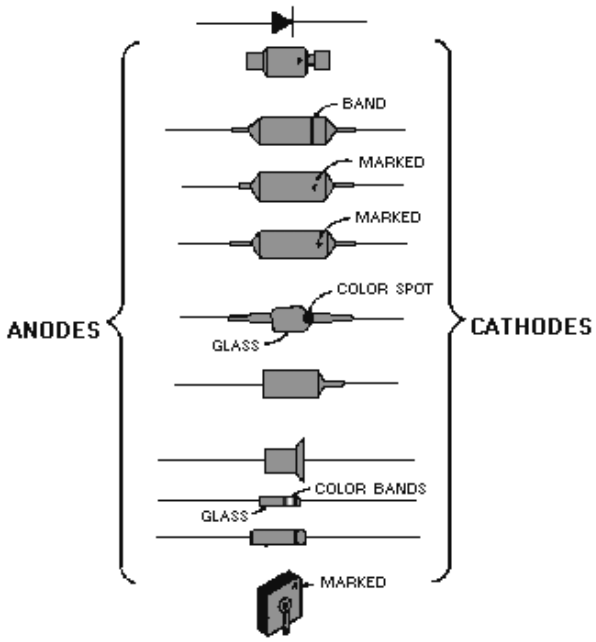
COMPONENT IDENTIFICATION NUMBER

X - NUMBER OF SEMICONDUCTOR JUNCTIONS
N - A SEMICONDUCTOR
YYY - IDENTIFICATION NUMBER (ORDER OR REGISTRATION NUMBER)
ALSO INCLUDES SUFFIX LETTER (IF APPLICABLE) TO INDICATE

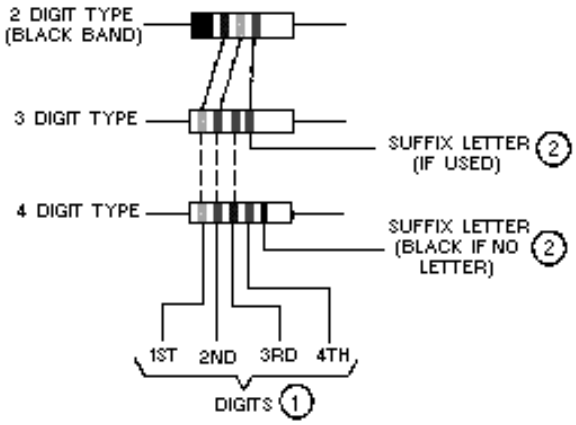
1. MATCHING DEVICES
2. REVERSE POLARITY
3. MODIFICATION

EXAMPLE - 1N345A (AN IMPROVED VERSION OF THE SEMICONDUCTOR DIODE TYPE 345)

DIODE MARKINGS are letters and symbols placed on the diode by manufacturers to distinguish one end of the diode from the other. In some cases, an unusual shape or the addition of color code bands is used to distinguish the cathode from the anode.



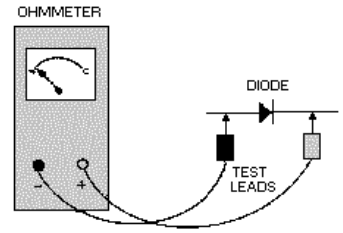
The **STANDARD DIODE COLOR CODE SYSTEM** serves two purposes when it is used: (1) it identifies the cathode end of the diode, and (2) it also serves to identify the diode by number.

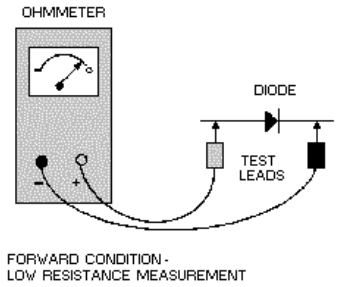


COLOR	(1) DIGIT	(2) DIODE SUFFIX LETTER
BLACK	0	-
BROWN	1	A
RED	2	B
ORANGE	3	C
YELLOW	4	D
GREEN	5	E
BLUE	6	F
VIOLET	7	G
GRAY	8	H
WHITE	9	J
SILVER	-	-
GOLD	-	-
NONE	-	-

DIODE MAINTENANCE is the procedures or methods used to keep a diode in good operating condition. To prevent diode damage, you should observe standard diode safety precautions and ensure that diodes are not subjected to heat, current overloads, and excessively high operating voltages.

TESTING A DIODE can be accomplished by using an ohmmeter, the substitution method, or a dynamic diode tester. The most convenient and quickest way of testing a diode is with an ohmmeter.





ANSWERS TO QUESTIONS Q1. THROUGH Q36.

A1. An electronic device that operates by virtue of the movement of electrons within a solid piece of semiconductor material.

A2. It is the decrease in a semiconductor's resistance as temperature rises.

A3. Space systems, computers, and data processing equipment.

A4. The electron tube requires filament or heater voltage, whereas the semiconductor device does not; consequently, no power input is spent by the semiconductor for conduction.

A5. Anything that occupies space and has weight. Solid, liquid, and gas.

A6. The atom.

A7. Electrons-negative, protons-positive, and neutrons-neutral.

A8. The valence shell.

A9. Quanta.

A10. A negatively charged atom having more than its normal amount of electrons.

A11. The energy levels of an atom in a solid group together to form energy bands, whereas the isolated atom does not.

A12. The width of the forbidden band.

A13. The number of electrons in the valence shell.

A14. Covalent bonding.

A15. Electron flow and hole flow.

A16. Intrinsic.

A17. P-type crystal.

A18. Electrons.

A19. To convert alternating current into direct current.

A20. Toward the arrow.

A21. Point-contact.

A22. Holes.

A23. N-type material.

A24. Depletion region.

A25. Negative.

A26. Forward.

A27. Any device that draws current.

A28. A pulsating dc voltage.

A29. Metallic rectifier.

A30. Forward bias.

A31. A characteristic curve.

A32. They are the limiting values of operating conditions outside which operations could cause diode damage.

A33. A semiconductor.

A34. 1N368.

A35. Heat.

A36. The diode is open or has a high-forward resistance.