Chapter Fourteen

SEMICONDUCTOR ELECTRONICS: MATERIALS, DEVICES AND SIMPLE CIRCUITS

14.1 Introduction

Devices in which a controlled flow of electrons can be obtained are the basic *building blocks* of all the electronic circuits. Before the discovery of transistor in 1948, such devices were mostly vacuum tubes (also called valves) like the vacuum diode which has two electrodes, viz., anode (often called plate) and cathode; triode which has three electrodes – cathode, plate and grid; tetrode and pentode (respectively with 4 and 5 electrodes). In a vacuum tube, the electrons are supplied by a heated cathode and the controlled flow of these electrons *in vacuum* is obtained by varying the voltage between its different electrodes. Vacuum is required in the inter-electrode space; otherwise the moving electrons may lose their energy on collision with the air molecules in their path. In these devices the electrons can flow only from the cathode to the anode (i.e., only in one direction). Therefore, such devices are generally referred to as *valves*. These vacuum tube devices are bulky, consume high power, operate generally at high voltages (~100 V) and have limited life and low reliability. The seed of the development of modern *solid-state semiconductor electronics* goes back to 1930's when it was realised that some solid-state semiconductors and their junctions offer the possibility of controlling the number and the direction of flow of charge carriers through them. Simple excitations like light, heat or small applied voltage can change the number of mobile charges in a semiconductor. Note that the supply

and flow of charge carriers in the semiconductor devices are *within the solid itself*, while in the earlier vacuum tubes/valves, the mobile electrons were obtained from a heated cathode and they were made to flow in an *evacuated* space or vacuum. No external heating or large evacuated space is required by the semiconductor devices. They are small in size, consume low power, operate at low voltages and have long life and high reliability. Even the Cathode Ray Tubes (CRT) used in television and computer monitors which work on the principle of vacuum tubes are being replaced by Liquid Crystal Display (LCD) monitors with supporting solid state electronics. Much before the full implications of the semiconductor devices was formally understood, a naturally occurring crystal of *galena* (Lead sulphide, PbS) with a metal point contact attached to it was used as *detector* of radio waves.

In the following sections, we will introduce the basic concepts of semiconductor physics and discuss some semiconductor devices like junction diodes (a 2-electrode device) and bipolar junction transistor (a 3-electrode device). A few circuits illustrating their applications will also be described.

14.2 Classification of Metals, Conductors and Semiconductors

On the basis of conductivity

On the basis of the relative values of electrical conductivity (σ) or resistivity ($\rho = 1/\sigma$), the solids are broadly classified as:

(i) ***Metals:*** They possess very low resistivity (or high conductivity).
$\rho \sim 10^{-2} - 10^{-8}\ \Omega$ m
$\sigma \sim 10^{2} - 10^{8}$ S m^{-1}

(ii) ***Semiconductors:*** They have resistivity or conductivity intermediate to metals and insulators.
$\rho \sim 10^{-5} - 10^{6}\ \Omega$ m
$\sigma \sim 10^{5} - 10^{-6}$ S m^{-1}

(iii) ***Insulators:*** They have high resistivity (or low conductivity).
$\rho \sim 10^{11} - 10^{19}\ \Omega$ m
$\sigma \sim 10^{-11} - 10^{-19}$ S m^{-1}

The values of ρ and σ given above are indicative of magnitude and could well go outside the ranges as well. Relative values of the resistivity are not the only criteria for distinguishing metals, insulators and semiconductors from each other. There are some other differences, which will become clear as we go along in this chapter.

Our interest in this chapter is in the study of semiconductors which could be:

(i) *Elemental semiconductors*: Si and Ge

(ii) *Compound semiconductors*: Examples are:

- Inorganic: CdS, GaAs, CdSe, InP, etc.
- Organic: anthracene, doped pthalocyanines, etc.
- Organic polymers: polypyrrole, polyaniline, polythiophene, etc.

Most of the currently available semiconductor devices are based on elemental semiconductors Si or Ge and compound *inorganic*

semiconductors. However, after 1990, a few semiconductor devices using organic semiconductors and semiconducting polymers have been developed signalling the birth of a futuristic technology of polymer-electronics and molecular-electronics. In this chapter, we will restrict ourselves to the study of inorganic semiconductors, particularly elemental semiconductors Si and Ge. The general concepts introduced here for discussing the elemental semiconductors, by-and-large, apply to most of the compound semiconductors as well.

On the basis of energy bands

According to the Bohr atomic model, in an *isolated atom* the energy of any of its electrons is decided by the orbit in which it revolves. But when the atoms come together to form a solid they are close to each other. So the outer orbits of electrons from neighbouring atoms would come very close or could even overlap. This would make the nature of electron motion in a solid very different from that in an isolated atom.

Inside the crystal each electron has a unique position and no two electrons see exactly the same pattern of surrounding charges. Because of this, each electron will have a different *energy level*. These different energy levels with continuous energy variation form what are called *energy bands*. The energy band which includes the energy levels of the valence electrons is called the *valence band*. The energy band above the valence band is called the *conduction band*. With no external energy, all the valence electrons will reside in the valence band. If the lowest level in the conduction band happens to be lower than the highest level of the valence band, the electrons from the valence band can easily move into the conduction band. Normally the conduction band is empty. But when it overlaps on the valence band electrons can move freely into it. This is the case with metallic conductors.

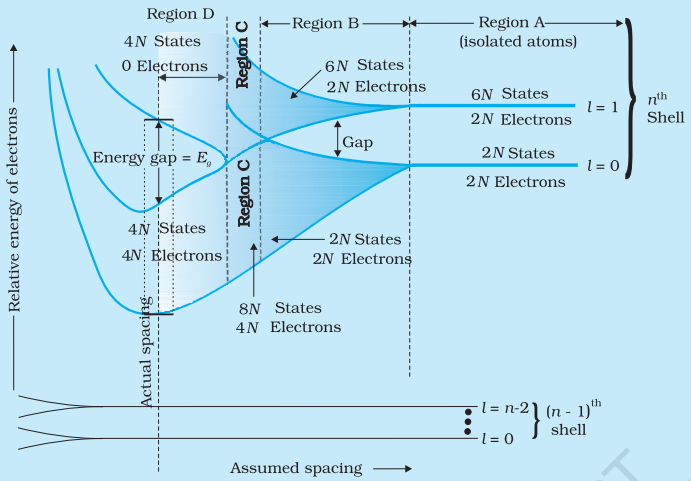
If there is some gap between the conduction band and the valence band, electrons in the valence band all remain bound and no free electrons are available in the conduction band. This makes the material an insulator. But some of the electrons from the valence band may gain external energy to cross the gap between the conduction band and the valence band. Then these electrons will move into the conduction band. At the same time they will create vacant energy levels in the valence band where other valence electrons can move. Thus the process creates the possibility of conduction due to electrons in conduction band as well as due to vacancies in the valence band.

Let us consider what happens in the case of Si or Ge crystal containing N atoms. For Si, the outermost orbit is the third orbit ($n = 3$), while for Ge it is the fourth orbit ($n = 4$). The number of electrons in the outermost orbit is 4 ($2s$ and $2p$ electrons). Hence, the total number of outer electrons in the crystal is $4N$. The maximum possible number of electrons in the outer orbit is 8 ($2s + 6p$ electrons). So, for the $4N$ valence electrons there are $8N$ available energy states. These $8N$ discrete energy levels can either form a continuous band or they may be grouped in different bands depending upon the distance between the atoms in the crystal (see box on Band Theory of Solids).

At the distance between the atoms in the crystal lattices of Si and Ge, the energy band of these $8N$ states is split apart into two which are separated by an *energy gap* E_g (Fig. 14.1). The lower band which is

completely occupied by the $4N$ valence electrons at temperature of absolute zero is the *valence band.* The other band consisting of $4N$ energy states, called the *conduction band*, is completely empty at absolute zero.

BAND THEORY OF SOLIDS



Consider that the Si or Ge crystal contains N atoms. Electrons of each atom will have discrete energies in different orbits. The electron energy will be same if all the atoms are *isolated,* i.e., separated from each other by a large distance. However, in a crystal, the atoms are close to each other (2 to 3 Å) and therefore the electrons interact with each other and also with the neighbouring atomic cores. The overlap (or interaction) will be more felt by the electrons in the outermost orbit while the inner orbit or core electron energies may remain unaffected. Therefore, for understanding electron energies in Si or Ge crystal, we need to consider the changes in the energies of the electrons in the outermost orbit only. For Si, the outermost orbit is the third orbit ($n = 3$), while for Ge it is the fourth orbit ($n = 4$). The number of electrons in the outermost orbit is 4 ($2s$ and $2p$ electrons). Hence, the total number of outer electrons in the crystal is $4N$. The maximum possible number of outer electrons in the orbit is 8 ($2s + 6p$ electrons). So, out of the $4N$ electrons, $2N$ electrons are in the $2N$ *s-states* (orbital quantum number $l = 0$) and $2N$ electrons are in the available $6N$ *p-states*. Obviously, some *p*-electron states are empty as shown in the extreme right of Figure. This is the case of well separated or isolated atoms [region A of Figure].

Suppose these atoms start coming nearer to each other to form a solid. The energies of these electrons in the outermost orbit may change (both increase and decrease) due to the interaction between the electrons of different atoms. The $6N$ states for $l = 1$, which originally had identical energies in the isolated atoms, spread out and form an *energy band* [region B in Figure]. Similarly, the $2N$ states for $l = 0$, having identical energies in the isolated atoms, split into a second band (carefully see the region B of Figure) separated from the first one by an *energy gap*.

At still smaller spacing, however, there comes a region in which the bands merge with each other. The lowest energy state that is a split from the upper atomic level appears to drop below the upper state that has come from the lower atomic level. In this region (region C in Figure), *no energy gap exists where the upper and lower energy states get mixed.*

Finally, if the distance between the atoms further decreases, the energy bands again split apart and are separated by an *energy gap* E_g (region D in Figure). The total number of available energy states $8N$ has been *re-apportioned* between the two bands ($4N$ states each in the lower and upper energy bands). Here the significant point is that there are exactly as many states in the lower band ($4N$) as there are available valence electrons from the atoms ($4N$).

Therefore, this band (called the *valence band*) is completely filled while the upper band is completely empty. The upper band is called the *conduction band.*

The lowest energy level in the conduction band is shown as E_C and highest energy level in the valence band is shown as E_V. Above E_C and below E_V there are a large number of closely spaced energy levels, as shown in Fig. 14.1.

The gap between the top of the valence band and bottom of the conduction band is called the *energy band gap* (Energy gap E_g). It may be large, small, or zero, depending upon the material. These different situations, are depicted in Fig. 14.2 and discussed below:

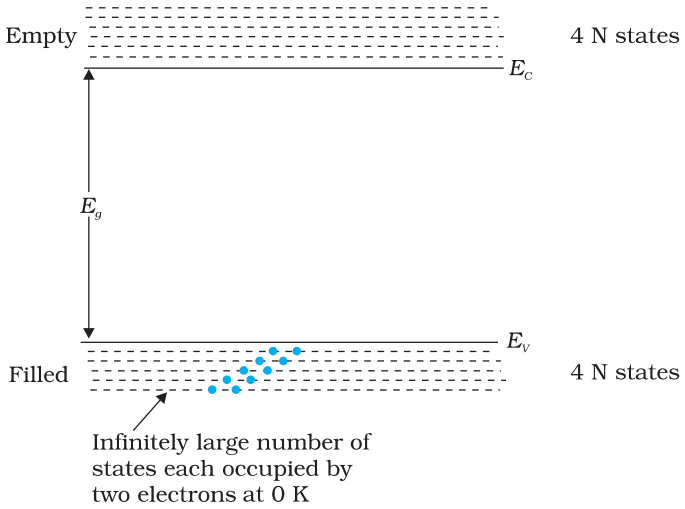


FIGURE 14.1 The energy band positions in a semiconductor at 0 K. The upper band, called the conduction band, consists of infinitely large number of closely spaced energy states. The lower band, called the valence band, consists of closely spaced completely filled energy states.

Case I: This refers to a situation, as shown in Fig. 14.2(a). One can have a metal either when the conduction band is partially filled and the balanced band is partially empty or when the conduction and valance bands overlap. When there is overlap electrons from valence band can easily move into the conduction band. This situation makes a large number of electrons available for electrical conduction. When the valence band is partially empty, electrons from its lower level can move to higher level making conduction possible. Therefore, the resistance of such materials is low or the conductivity is high.

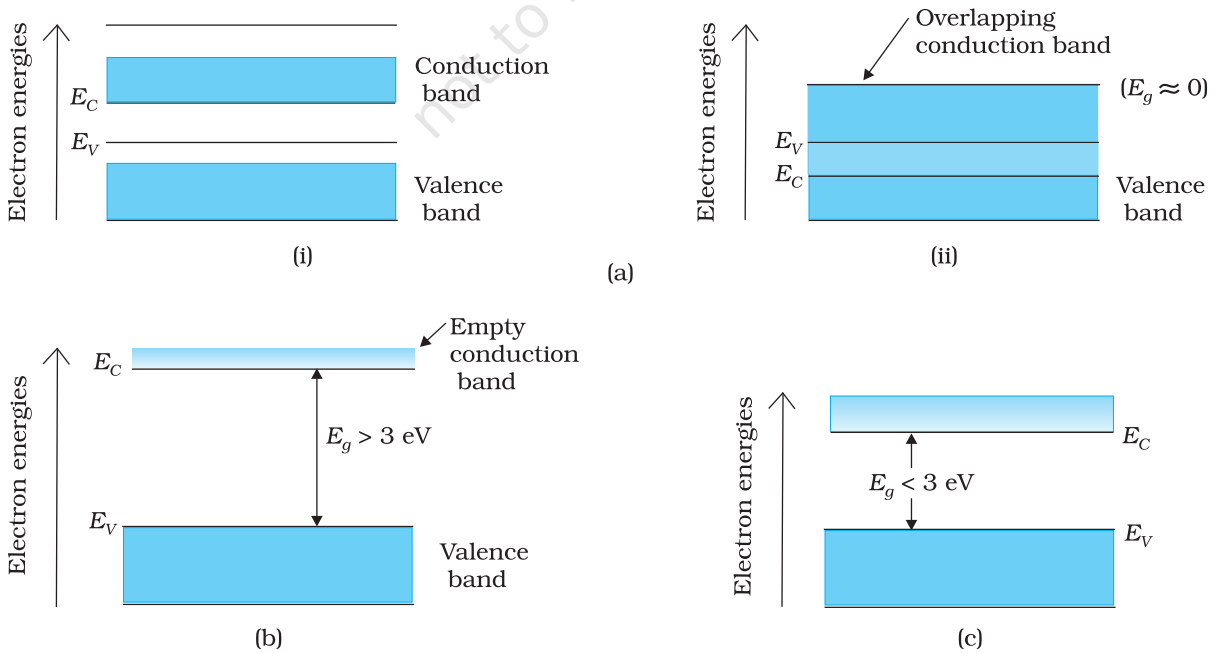


FIGURE 14.2 Difference between energy bands of (a) metals, (b) insulators and (c) semiconductors.

Case II: In this case, as shown in Fig. 14.2(b), a large band gap E_g exists ($E_g > 3$ eV). There are no electrons in the conduction band, and therefore no electrical conduction is possible. Note that the energy gap is so large that electrons cannot be excited from the valence band to the conduction band by thermal excitation. This is the case of *insulators*.

Case III: This situation is shown in Fig. 14.2(c). Here a finite but small band gap ($E_g < 3$ eV) exists. Because of the small band gap, at room temperature some electrons from valence band can acquire enough energy to cross the energy gap and enter the *conduction band*. These electrons (though small in numbers) can move in the conduction band. Hence, the resistance of *semiconductors* is not as high as that of the insulators.

In this section we have made a broad classification of metals, conductors and semiconductors. In the section which follows you will learn the conduction process in semiconductors.

14.3 Intrinsic Semiconductor

We shall take the most common case of Ge and Si whose lattice structure is shown in Fig. 14.3. These structures are called the diamond-like structures. Each atom is surrounded by four nearest neighbours. We know that Si and Ge have four valence electrons. In its crystalline structure, every Si or Ge atom tends *to share* one of its four valence electrons with each of its four nearest neighbour atoms, and also *to take share* of one electron from each such neighbour. These shared electron pairs are referred to as forming a *covalent bond* or simply a *valence bond*. The two shared electrons can be assumed to shuttle back-and-forth between the associated atoms holding them together strongly. Figure 14.4 schematically shows the 2-dimensional representation of Si or Ge structure shown in Fig. 14.3 which overemphasises the covalent bond. It shows an idealised picture in which no bonds are broken (all bonds are intact). Such a situation arises at low temperatures. As the temperature increases, more thermal energy becomes available to these electrons and some of these electrons may break–away (becoming *free* electrons contributing to conduction). The thermal energy effectively ionises only a few atoms in the crystalline lattice and creates a *vacancy* in the bond as shown in Fig. 14.5(a). The neighbourhood, from which the free electron (with charge $-q$) has come out leaves a vacancy with an effective charge ($+q$). This *vacancy* with the effective positive electronic charge is called a *hole*. The hole behaves as an *apparent free particle* with effective positive charge.

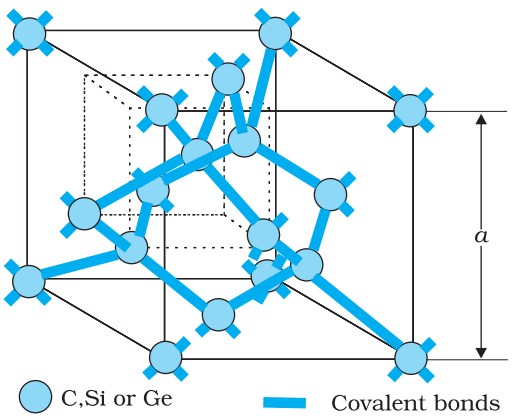


FIGURE 14.3 Three-dimensional diamond-like crystal structure for Carbon, Silicon or Germanium with respective lattice spacing a equal to 3.56, 5.43 and 5.66 Å.

In intrinsic semiconductors, the number of free electrons, n_e is equal to the number of holes, n_h. That is

$$n_e = n_h = n_i \quad (14.1)$$

where n_i is called intrinsic carrier concentration.

Semiconductors posses the unique property in which, apart from electrons, the holes also move.

Suppose there is a hole at site 1 as shown in Fig. 14.5(a). The movement of holes can be visualised as shown in Fig. 14.5(b). An electron from the covalent bond at site 2 may jump to the vacant site 1 (hole). Thus, after such a jump, the hole is at site 2 and the site 1 has now an electron. Therefore, apparently, the hole has moved from site 1 to site 2. Note that the electron originally set free [Fig. 14.5(a)] is not involved in this process of hole motion. The free electron moves completely independently as conduction electron and gives rise to an electron current, I_e under an applied electric field. Remember that the motion of hole is only a convenient way of describing the actual motion of *bound* electrons, whenever there is an empty bond anywhere in the crystal. Under the action of an electric field, these holes move towards negative potential giving the hole current, I_h. The total current, I is thus the sum of the electron current I_e and the hole current I_h:

$$I = I_e + I_h \qquad (14.2)$$

It may be noted that apart from the *process of generation* of conduction electrons and holes, a simultaneous *process of recombination* occurs in which the electrons *recombine* with the holes. At equilibrium, the rate of generation is equal to the rate of recombination of charge carriers. The recombination occurs due to an electron colliding with a hole.

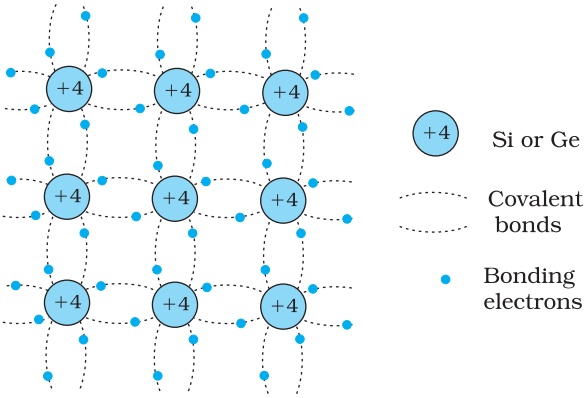


FIGURE 14.4 Schematic two-dimensional representation of Si or Ge structure showing covalent bonds at low temperature (all bonds intact). +4 symbol indicates inner cores of Si or Ge.

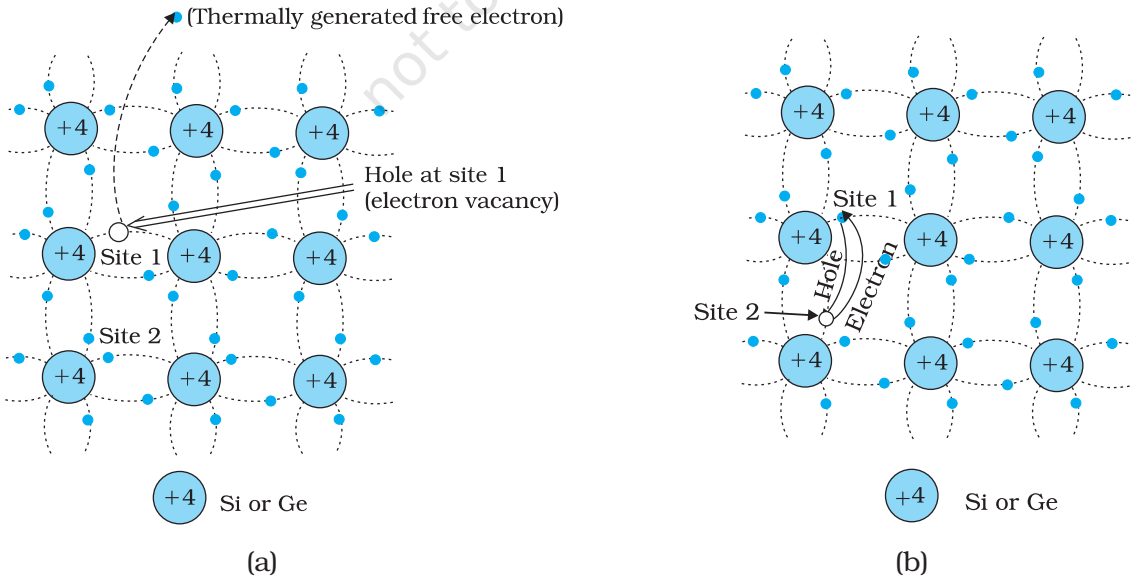


FIGURE 14.5 (a) Schematic model of generation of hole at site 1 and conduction electron due to thermal energy at moderate temperatures. (b) Simplified representation of possible thermal motion of a hole. The electron from the lower left hand covalent bond (site 2) goes to the earlier hole site1, leaving a hole at its site indicating an apparent movement of the hole from site 1 to site 2.

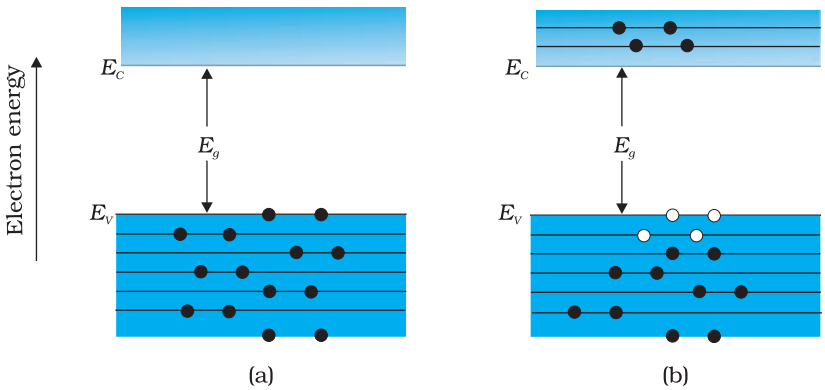


FIGURE 14.6 (a) An intrinsic semiconductor at $T = 0$ K behaves like insulator. (b) At $T > 0$ K, four thermally generated electron-hole pairs. The filled circles (•) represent electrons and empty circles (○) represent holes.

An intrinsic semiconductor will behave like an insulator at $T = 0$ K as shown in Fig. 14.6(a). It is the thermal energy at higher temperatures ($T > 0$K), which excites some electrons from the valence band to the conduction band. These thermally excited electrons at $T > 0$ K, partially occupy the conduction band. Therefore, the energy-band diagram of an intrinsic semiconductor will be as shown in Fig. 14.6(b). Here, some electrons are shown in the conduction band. These have come from the valence band leaving equal number of holes there.

EXAMPLE 14.1

Example 14.1 C, Si and Ge have same lattice structure. Why is C insulator while Si and Ge intrinsic semiconductors?

Solution The 4 bonding electrons of C, Si or Ge lie, respectively, in the second, third and fourth orbit. Hence, energy required to take out an electron from these atoms (i.e., ionisation energy E_g) will be least for Ge, followed by Si and highest for C. Hence, number of free electrons for conduction in Ge and Si are significant but negligibly small for C.

14.4 EXTRINSIC SEMICONDUCTOR

The conductivity of an intrinsic semiconductor depends on its temperature, but at room temperature its conductivity is very low. As such, no important electronic devices can be developed using these semiconductors. Hence there is a necessity of improving their conductivity. This can be done by making use of impurities.

When a small amount, say, a few parts per million (ppm), of a suitable impurity is added to the pure semiconductor, the conductivity of the semiconductor is increased manifold. Such materials are known as *extrinsic semiconductors* or *impurity semiconductors*. The deliberate addition of a desirable impurity is called *doping* and the impurity atoms are called *dopants*. Such a material is also called a *doped semiconductor*. The dopant has to be such that it does not distort the original pure semiconductor lattice. It occupies only a very few of the original semiconductor atom sites in the crystal. A necessary condition to attain this is that the sizes of the dopant and the semiconductor atoms should be nearly the same.

There are two types of dopants used in doping the tetravalent Si or Ge:

(i) Pentavalent (valency 5); like Arsenic (As), Antimony (Sb), Phosphorous (P), etc.

(ii) Trivalent (valency 3); like Indium (In), Boron (B), Aluminium (Al), etc.

We shall now discuss how the doping changes the number of charge carriers (and hence the conductivity) of semiconductors. Si or Ge belongs to the fourth group in the Periodic table and, therefore, we choose the dopant element from nearby fifth or third group, expecting and taking care that the size of the dopant atom is nearly the same as that of Si or Ge. Interestingly, the pentavalent and trivalent dopants in Si or Ge give two entirely different types of semiconductors as discussed below.

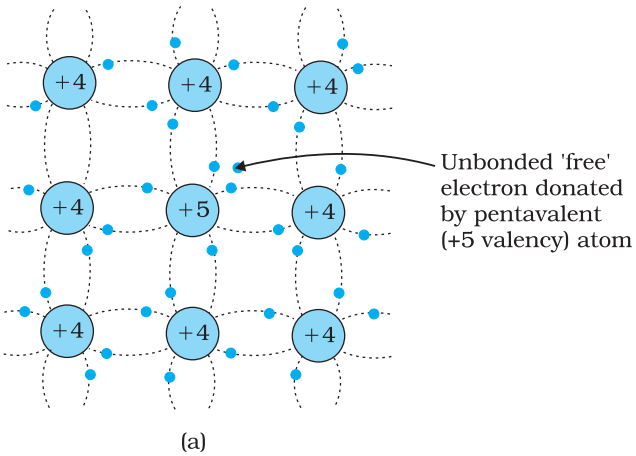


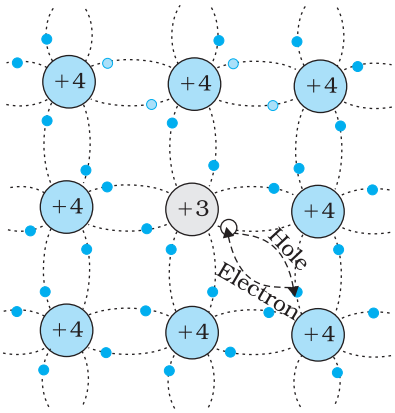
FIGURE 14.7 (a) Pentavalent donor atom (As, Sb, P, etc.) doped for tetravalent Si or Ge giving n-type semiconductor, and (b) Commonly used schematic representation of n-type material which shows only the fixed cores of the substituent donors with one additional effective positive charge and its associated extra electron.

(i) n-type semiconductor

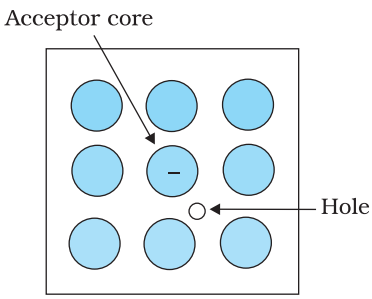
Suppose we dope Si or Ge with a pentavalent element as shown in Fig. 14.7. When an atom of +5 valency element occupies the position of an atom in the crystal lattice of Si, four of its electrons bond with the four silicon neighbours while the fifth remains very weakly bound to its parent atom. This is because the four electrons participating in bonding are seen as part of the effective core of the atom by the fifth electron. As a result the ionisation energy required to set this electron free is very small and even at room temperature it will be free to move in the lattice of the semiconductor. For example, the energy required is ~ 0.01 eV for germanium, and 0.05 eV for silicon, to separate this electron from its atom. This is in contrast to the energy required to jump the forbidden band (about 0.72 eV for germanium and about 1.1 eV for silicon) at room temperature in the intrinsic semiconductor. Thus, the pentavalent dopant is donating one extra electron for conduction and hence is known as *donor* impurity. The number of electrons made available for conduction by dopant atoms depends strongly upon the doping level and is independent of any increase in ambient temperature. On the other hand, the number of free electrons (with an equal number of holes) generated by Si atoms, increases weakly with temperature.

In a doped semiconductor the total number of conduction electrons n_e is due to the electrons contributed by donors and those generated intrinsically, while the total number of holes n_h is only due to the holes from the intrinsic source. But the rate of recombination of holes would increase due to the increase in the number of electrons. As a result, the number of holes would get reduced further.

Thus, with proper level of doping the number of conduction electrons can be made much larger than the number of holes. Hence in an extrinsic



(a)



(b)

FIGURE 14.8 (a) Trivalent acceptor atom (In, Al, B etc.) doped in tetravalent Si or Ge lattice giving p-type semiconductor. (b) Commonly used schematic representation of p-type material which shows only the fixed core of the substituent acceptor with one effective additional negative charge and its associated hole.

semiconductor doped with pentavalent impurity, electrons become the *majority carriers* and holes the *minority carriers*. These semiconductors are, therefore, known as n-*type semiconductors*. For n-type semiconductors, we have,

$$n_e >> n_h \qquad (14.3)$$

(ii) p-type semiconductor

This is obtained when Si or Ge is doped with a trivalent impurity like Al, B, In, etc. The dopant has one valence electron less than Si or Ge and, therefore, this atom can form covalent bonds with neighbouring three Si atoms but does not have any electron to offer to the fourth Si atom. So the bond between the fourth neighbour and the trivalent atom has a vacancy or hole as shown in Fig. 14.8. Since the neighbouring Si atom in the lattice wants an electron in place of a hole, an electron in the outer orbit of an atom in the neighbourhood may jump to fill this vacancy, leaving a vacancy or hole at its own site. Thus the *hole* is available for conduction. Note that the trivalent foreign atom becomes effectively negatively charged when it shares fourth electron with neighbouring Si atom. Therefore, the dopant atom of p-type material can be treated as *core of one negative charge* along with its associated hole as shown in Fig. 14.8(b). It is obvious that one *acceptor* atom gives one *hole*. These holes are in addition to the intrinsically generated holes while the source of conduction electrons is only intrinsic generation. Thus, for such a material, the holes are the majority carriers and electrons are minority carriers. Therefore, extrinsic semiconductors doped with trivalent impurity are called p-*type semiconductors*. For p-type semiconductors, the recombination process will reduce the number (n_i)of intrinsically generated electrons to n_e. We have, for p-type semiconductors

$$n_h >> n_e \qquad (14.4)$$

Note that *the crystal maintains an overall charge neutrality as the charge of additional charge carriers is just equal and opposite to that of the ionised cores in the lattice.*

In extrinsic semiconductors, because of the abundance of majority current carriers, the minority carriers produced thermally have more chance of meeting majority carriers and thus getting destroyed. Hence, the dopant, by adding a large number of current carriers of one type, which become the majority carriers, indirectly helps to reduce the intrinsic concentration of minority carriers.

The semiconductor's energy band structure is affected by doping. In the case of extrinsic semiconductors, additional energy states due to donor impurities (E_D) and acceptor impurities (E_A) also exist. In the energy band diagram of n-type Si semiconductor, the donor energy level E_D is slightly below the bottom E_C of the conduction band and electrons from this level move into the conduction band with very small supply of energy. At room temperature, most of the donor atoms get ionised but very few (~10^{12}) atoms of Si get ionised. So the conduction band will have most electrons coming from the donor impurities, as shown in Fig. 14.9(a). Similarly,

for p-type semiconductor, the acceptor energy level E_A is slightly above the top E_V of the valence band as shown in Fig. 14.9(b). With very small supply of energy an electron from the valence band can jump to the level E_A and ionise the acceptor negatively. (Alternately, we can also say that with very small supply of energy the hole from level E_A sinks down into the valence band. Electrons rise up and holes fall down when they gain external energy.) At room temperature, most of the acceptor atoms get ionised leaving holes in the valence band. Thus at room temperature the density of holes in the valence band is predominantly due to impurity in the extrinsic semiconductor. The electron and hole concentration in a semiconductor *in thermal equilibrium* is given by

$$n_e n_h = n_i^2 \qquad (14.5)$$

Though the above description is grossly approximate and hypothetical, it helps in understanding the difference between metals, insulators and semiconductors (extrinsic and intrinsic) in a simple manner. The difference in the resistivity of C, Si and Ge depends upon the energy gap between their conduction and valence bands. For C (diamond), Si and Ge, the energy gaps are 5.4 eV, 1.1 eV and 0.7 eV, respectively. Sn also is a group IV element but it is a metal because the energy gap in its case is 0 eV.

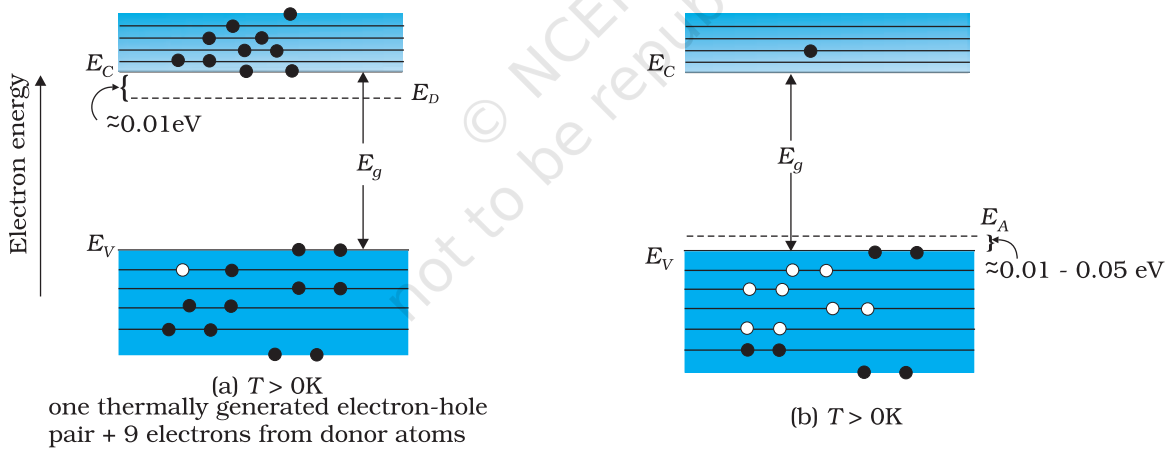


FIGURE 14.9 Energy bands of (a) n-type semiconductor at $T > 0$K, (b) p-type semiconductor at $T > 0$K.

Example 14.2 Suppose a pure Si crystal has 5×10^{28} atoms m^{-3}. It is doped by 1 ppm concentration of pentavalent As. Calculate the number of electrons and holes. Given that $n_i = 1.5 \times 10^{16}$ m^{-3}.

Solution Note that thermally generated electrons ($n_i \sim 10^{16}$ m^{-3}) are negligibly small as compared to those produced by doping.
Therefore, $n_e \approx N_D$.
Since $n_e n_h = n_i^2$, The number of holes
$n_h = (2.25 \times 10^{32})/(5 \times 10^{22})$

$\sim 4.5 \times 10^{9}$ m^{-3}

EXAMPLE 14.2

14.5 p-n Junction

A p-n junction is the basic building block of many semiconductor devices like diodes, transistor, etc. A clear understanding of the junction behaviour is important to analyse the working of other semiconductor devices. We will now try to understand how a junction is formed and how the junction behaves under the influence of external applied voltage (also called *bias*).

Formation and working of p-n junction diode http://hyperphysics.phy-astr.gsu.edu/hbase/solids/pnjun.html

PHYSICS

14.5.1 p-n junction formation

Consider a thin p-type silicon (p-Si) semiconductor wafer. By adding precisely a small quantity of pentavelent impurity, part of the p-Si wafer can be converted into n-Si. There are several processes by which a semiconductor can be formed. The wafer now contains p-region and n-region and a metallurgical junction between p-, and n- region.

Two important processes occur during the formation of a p-n junction: *diffusion* and *drift*. We know that in an n-type semiconductor, the concentration of electrons (number of electrons per unit volume) is more compared to the concentration of holes. Similarly, in a p-type semiconductor, the concentration of holes is more than the concentration of electrons. During the formation of p-n junction, and due to the concentration gradient across p-, and n- sides, holes diffuse from p-side to n-side (p → n) and electrons diffuse from n-side to p-side (n → p). This motion of charge carries gives rise to diffusion current across the junction.

When an electron diffuses from n → p, it leaves behind an ionised donor on n-side. This ionised donor (positive charge) is immobile as it is bonded to the surrounding atoms. As the electrons continue to diffuse from n → p, a layer of positive charge (or positive space-charge region) on n-side of the junction is developed.

Similarly, when a hole diffuses from p → n due to the concentration gradient, it leaves behind an ionised acceptor (negative charge) which is immobile. As the holes continue to diffuse, a layer of negative charge (or negative space-charge region) on the p-side of the junction is developed. This space-charge region on either side of the junction together is known as *depletion region* as the electrons and holes taking part in the initial movement across the junction *depleted* the region of its free charges (Fig. 14.10). The thickness of depletion region is of the order of one-tenth of a micrometre. Due to the positive space-charge region on n-side of the junction and negative space charge region on p-side of the junction, an electric field directed from positive charge towards negative charge develops. Due to this field, an electron on p-side of the junction moves to n-side and a hole on n-side of the junction moves to p-side. The motion of charge carriers due to the electric field is called drift. Thus a drift current, which is opposite in direction to the diffusion current (Fig. 14.10) starts.

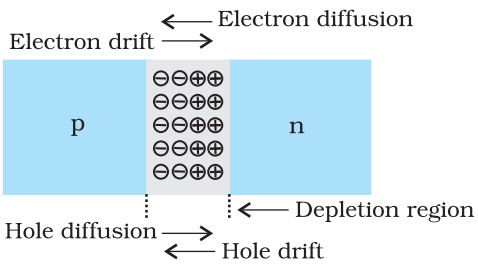


FIGURE 14.10 p-n junction formation process.

Initially, diffusion current is large and drift current is small. As the diffusion process continues, the space-charge regions on either side of the junction extend, thus increasing the electric field strength and hence drift current. This process continues until the diffusion current equals the drift current. Thus a p-n junction is formed. In a p-n junction under equilibrium there is *no net* current.

The loss of electrons from the n-region and the gain of electron by the p-region causes a difference of potential across the junction of the two regions. The polarity of this potential is such as to oppose further flow of carriers so that a condition of equilibrium exists. Figure 14.11 shows the p-n junction at equilibrium and the potential across the junction. The n-material has lost electrons, and p material has acquired electrons. The n material is thus positive relative to the p material. Since this potential tends to prevent the movement of electron from the n region into the p region, it is often called a *barrier potential*.

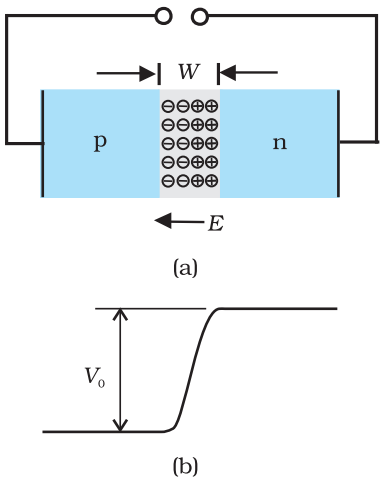


FIGURE 14.11 (a) Diode under equilibrium (V = 0), (b) Barrier potential under no bias.

EXAMPLE 14.3

Example 14.3 Can we take one slab of p-type semiconductor and physically join it to another n-type semiconductor to get p-n junction?

Solution No! Any slab, howsoever flat, will have roughness much larger than the inter-atomic crystal spacing (~2 to 3 Å) and hence *continuous contact* at the atomic level will not be possible. The junction will behave as a *discontinuity* for the flowing charge carriers.

14.6 Semiconductor Diode

A semiconductor diode [Fig. 14.12(a)] is basically a p-n junction with metallic contacts provided at the ends for the application of an external voltage. It is a two terminal device. A p-n junction diode is symbolically represented as shown in Fig. 14.12(b).

The direction of arrow indicates the conventional direction of current (when the diode is under forward bias). The equilibrium barrier potential can be altered by applying an external voltage V across the diode. The situation of p-n junction diode under equilibrium (without bias) is shown in Fig. 14.11(a) and (b).

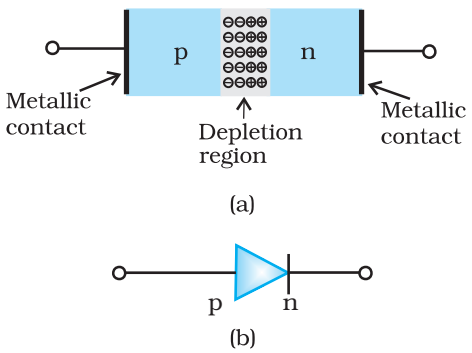


FIGURE 14.12 (a) Semiconductor diode, (b) Symbol for p-n junction diode.

14.6.1 p-n junction diode under forward bias

When an external voltage V is applied across a semiconductor diode such that p-side is connected to the positive terminal of the battery and n-side to the negative terminal [Fig. 14.13(a)], it is said to be *forward biased*.

The applied voltage mostly drops across the depletion region and the voltage drop across the p-side and n-side of the junction is negligible. (This is because the resistance of the depletion region – a region where there are no charges – is very high compared to the resistance of n-side and p-side.) The direction of the applied voltage (V) is opposite to the

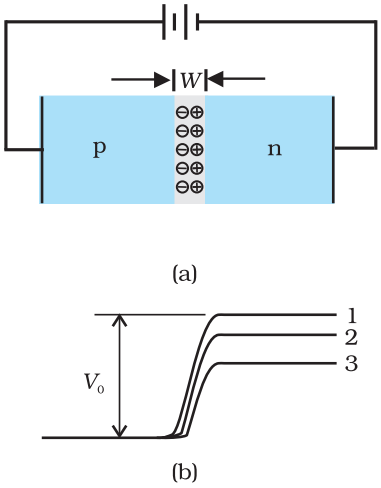


FIGURE 14.13 (a) p-n junction diode under forward bias, (b) Barrier potential (1) without battery, (2) Low battery voltage, and (3) High voltage battery.

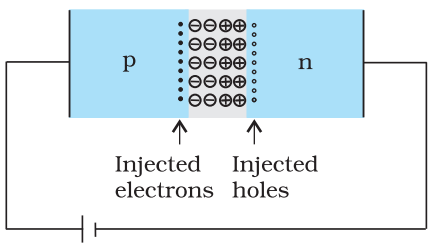


FIGURE 14.14 Forward bias minority carrier injection.

built-in potential V_0. As a result, the depletion layer width decreases and the barrier height is reduced [Fig. 14.13(b)]. The effective barrier height under forward bias is $(V_0 - V)$.

If the applied voltage is small, the barrier potential will be reduced only slightly below the equilibrium value, and only a small number of carriers in the material—those that happen to be in the uppermost energy levels—will possess enough energy to cross the junction. So the current will be small. If we increase the applied voltage significantly, the barrier height will be reduced and more number of carriers will have the required energy. Thus the current increases.

Due to the applied voltage, electrons from n-side cross the depletion region and reach p-side (where they are minority carries). Similarly, holes from p-side cross the junction and reach the n-side (where they are minority carries). This process under forward bias is known as minority carrier injection. At the junction boundary, on each side, the minority carrier concentration increases significantly compared to the locations far from the junction.

Due to this concentration gradient, the injected electrons on p-side diffuse from the junction edge of p-side to the other end of p-side. Likewise, the injected holes on n-side diffuse from the junction edge of n-side to the other end of n-side (Fig. 14.14). This motion of charged carriers on either side gives rise to current. The total diode forward current is sum of hole diffusion current and conventional current due to electron diffusion. The magnitude of this current is usually in mA.

14.6.2 p-n junction diode under reverse bias

When an external voltage (V) is applied across the diode such that n-side is positive and p-side is negative, it is said to be *reverse biased* [Fig.14.15(a)]. The applied voltage mostly drops across the depletion region. The direction of applied voltage is same as the direction of barrier potential. As a result, the barrier height increases and the depletion region widens due to the change in the electric field. The effective barrier height under reverse bias is $(V_0 + V)$, [Fig. 14.15(b)]. This suppresses the flow of electrons from n $\rightarrow$ p and holes from p $\rightarrow$ n. Thus, diffusion current, decreases enormously compared to the diode under forward bias.

The electric field direction of the junction is such that if electrons on p-side or holes on n-side in their random motion come close to the junction, they will be swept to its majority zone. This drift of carriers gives rise to current. The drift current is of the order of a few μA. This is quite low because it is due to the motion of carriers from their minority side to their majority side across the junction. The drift current is also there under forward bias but it is negligible (μA) when compared with current due to injected carriers which is usually in mA.

The diode reverse current is not very much dependent on the applied voltage. Even a small voltage is sufficient to sweep the minority carriers from one side of the junction to the other side of the junction. The current

is not limited by the magnitude of the applied voltage but is limited due to the concentration of the minority carrier on either side of the junction.

The current under reverse bias is essentially voltage independent upto a critical reverse bias voltage, known as breakdown voltage (V_{br}). When $V = V_{br}$, the diode reverse current increases sharply. Even a slight increase in the bias voltage causes large change in the current. If the reverse current is not limited by an external circuit below the rated value (specified by the manufacturer) the p-n junction will get destroyed. Once it exceeds the rated value, the diode gets destroyed due to overheating. This can happen even for the diode under forward bias, if the forward current exceeds the rated value.

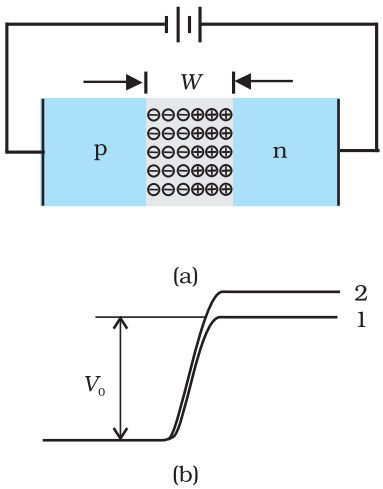


FIGURE 14.15 (a) Diode under reverse bias, (b) Barrier potential under reverse bias.

The circuit arrangement for studying the *V-I* characteristics of a diode, (i.e., the variation of current as a function of applied voltage) are shown in Fig. 14.16(a) and (b). The battery is connected to the diode through a potentiometer (or reheostat) so that the applied voltage to the diode can be changed. For different values of voltages, the value of the current is noted. A graph between *V* and *I* is obtained as in Fig. 14.16(c). Note that in forward bias measurement, we use a milliammeter since the expected current is large (as explained in the earlier section) while a micrometer is used in reverse bias to measure the current. You can see in Fig. 14.16(c) that in forward

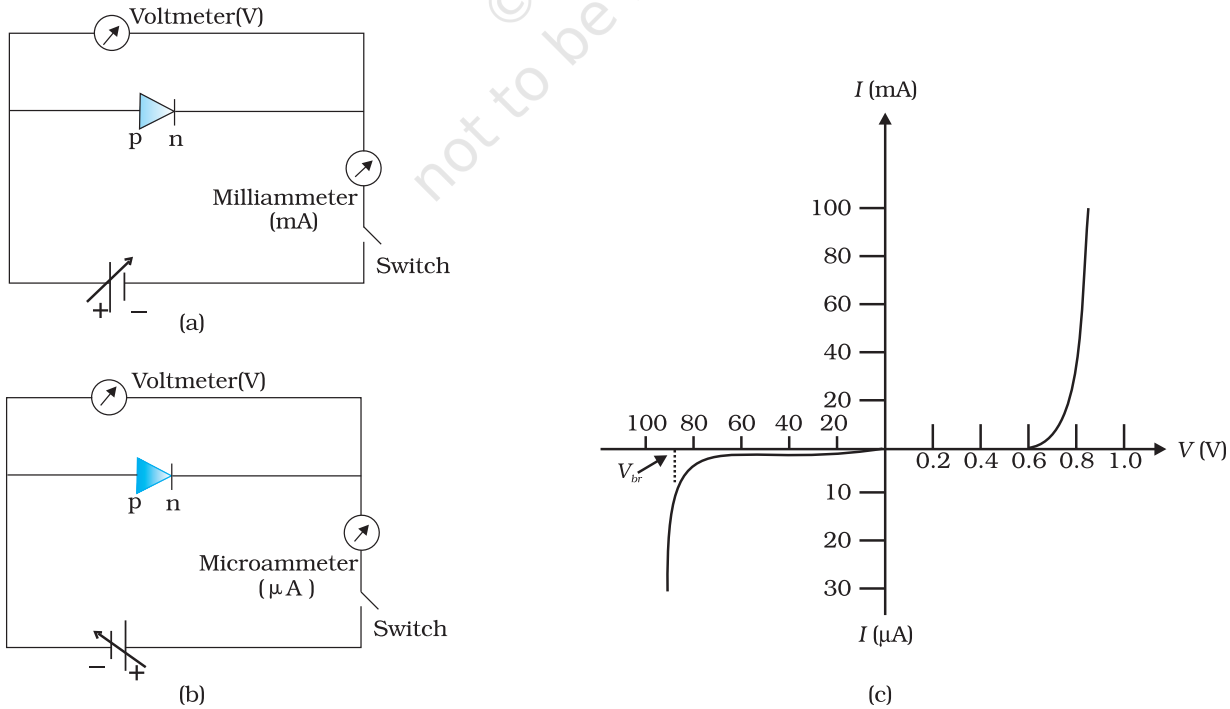


FIGURE 14.16 Experimental circuit arrangement for studying *V-I* characteristics of a p-n junction diode (a) in forward bias, (b) in reverse bias. (c) Typical *V-I* characteristics of a silicon diode.

bias, the current first increases very slowly, almost negligibly, till the voltage across the diode crosses a certain value. After the characteristic voltage, the diode current increases significantly (exponentially), even for a very small increase in the diode bias voltage. This voltage is called the *threshold voltage* or cut-in voltage (~0.2V for germanium diode and ~0.7 V for silicon diode).

For the diode in reverse bias, the current is very small (~μA) and almost remains constant with change in bias. It is called *reverse saturation current*. However, for special cases, at very high reverse bias (break down voltage), the current suddenly increases. This special action of the diode is discussed later in Section 14.8. The general purpose diode are not used beyond the reverse saturation current region.

The above discussion shows that the p-n junction diode primerly allows the flow of current only in one direction (forward bias). The forward bias resistance is low as compared to the reverse bias resistance. This property is used for rectification of ac voltages as discussed in the next section. For diodes, we define a quantity called *dynamic resistance* as the ratio of small change in voltage ΔV to a small change in current ΔI:

$$r_d = \frac{\Delta V}{\Delta I} \qquad (14.6)$$

EXAMPLE 14.4

Example 14.4 The *V-I* characteristic of a silicon diode is shown in the Fig. 14.17. Calculate the resistance of the diode at (a) I_D = 15 mA and (b) V_D = –10 V.

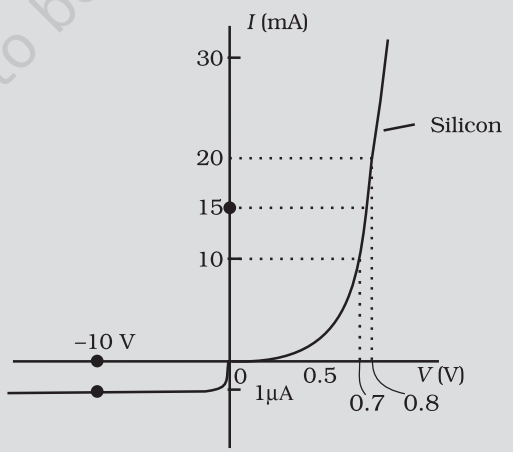


FIGURE 14.17

Solution Considering the diode characteristics as a straight line between *I* = 10 mA to *I* = 20 mA passing through the origin, we can calculate the resistance using Ohm's law.

(a) From the curve, at *I* = 20 mA, *V* = 0.8 V; *I* = 10 mA, *V* = 0.7 V
$r_{fb} = \Delta V/\Delta I$ = 0.1V/10 mA = 10 Ω

(b) From the curve at V = –10 V, *I* = –1 μA,
Therefore,
r_{rb} = 10 V/1μA= 1.0×10^7 Ω

14.7 Application of Junction Diode as a Rectifier

From the *V-I* characteristic of a junction diode we see that it allows current to pass only when it is forward biased. So if an alternating voltage is applied across a diode the current flows only in that part of the cycle when the diode is forward biased. This property is used *to rectify* alternating voltages and the circuit used for this purpose is called a *rectifier*.

If an alternating voltage is applied across a diode in series with a load, a pulsating voltage will appear across the load only during the half cycles of the ac input during which the diode is forward biased. Such rectifier circuit, as shown in Fig. 14.18, is called a *half-wave rectifier*. The secondary of a transformer supplies the desired ac voltage across terminals A and B. When the voltage at A is positive, the diode is forward biased and it conducts. When A is negative, the diode is reverse-biased and it does not conduct. The reverse saturation current of a diode is negligible and can be considered equal to zero for practical purposes. (The reverse breakdown voltage of the diode must be sufficiently higher than the peak ac voltage at the secondary of the transformer to protect the diode from reverse breakdown.)

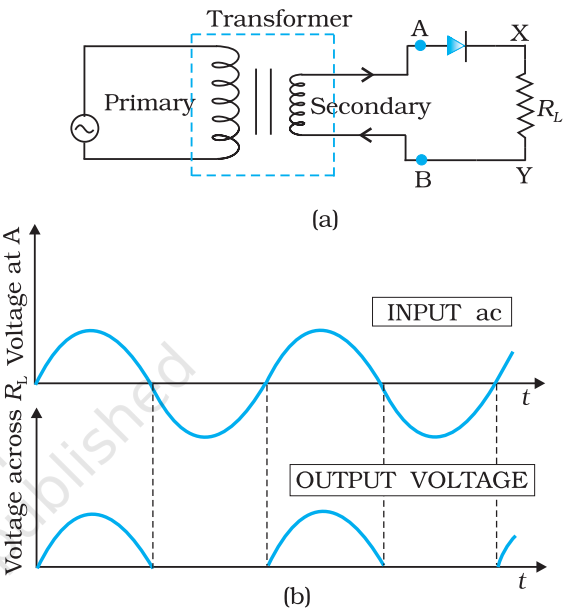


FIGURE 14.18 (a) Half-wave rectifier circuit, (b) Input ac voltage and output voltage waveforms from the rectifier circuit.

Therefore, in the positive *half-cycle* of ac there is a current through the load resistor R_L and we get an output voltage, as shown in Fig. 14.18(b), whereas there is no current in the negative half-cycle. In the next positive half-cycle, again we get the output voltage. Thus, the output voltage, though still varying, is restricted to *only one direction* and is said to be *rectified*. Since the rectified output of this circuit is only for half of the input ac wave it is called as *half-wave rectifier*.

The circuit using two diodes, shown in Fig. 14.19(a), gives output rectified voltage corresponding to both the positive as well as negative half of the ac cycle. Hence, it is known as *full-wave rectifier*. Here the p-side of the two diodes are connected to the ends of the secondary of the transformer. The n-side of the diodes are connected together and the output is taken between this common point of diodes and the midpoint of the secondary of the transformer. So for a full-wave rectifier the secondary of the transformer is provided with a centre tapping and so it is called *centre-tap transformer*. As can be seen from Fig.14.19(c) the voltage rectified by each diode is only half the total secondary voltage. Each diode rectifies only for half the cycle, but the two do so for alternate cycles. Thus, the output between their common terminals and the centre-tap of the transformer becomes a full-wave rectifier output. (Note that there is another circuit of full wave rectifier which does not need a centre-tap transformer but needs four diodes.) Suppose the input voltage to A

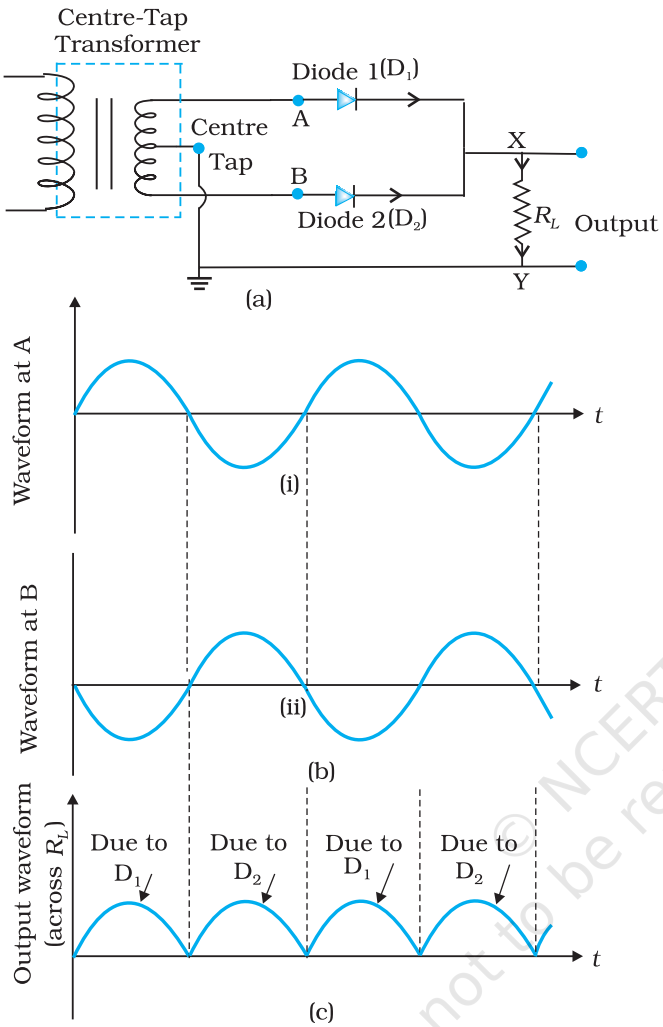


FIGURE 14.19 (a) A Full-wave rectifier circuit; (b) Input wave forms given to the diode D_1 at A and to the diode D_2 at B; (c) Output waveform across the load R_L connected in the full-wave rectifier circuit.

with respect to the centre tap at any instant is positive. It is clear that, at that instant, voltage at B being out of phase will be negative as shown in Fig.14.19(b). So, diode D_1 gets forward biased and conducts (while D_2 being reverse biased is not conducting). Hence, during this positive half cycle we get an output current (and a output voltage across the load resistor R_L) as shown in Fig.14.19(c). In the course of the ac cycle when the voltage at A becomes negative with respect to centre tap, the voltage at B would be positive. In this part of the cycle diode D_1 would not conduct but diode D_2 would, giving an output current and output voltage (across R_L) during the negative half cycle of the input ac. Thus, we get output voltage during both the positive as well as the negative half of the cycle. Obviously, this is a more efficient circuit for getting rectified voltage or current than the half-wave rectifier.

The rectified voltage is in the form of pulses of the shape of half sinusoids. Though it is unidirectional it does not have a steady value. To get steady dc output from the pulsating voltage normally a capacitor is connected across the output terminals (parallel to the load R_L). One can also use an inductor in series with R_L for the same purpose. Since these additional circuits appear to *filter* out the *ac ripple* and give a *pure dc* voltage, so they are called filters.

Now we shall discuss the role of capacitor in filtering. When the voltage across the capacitor is rising, it gets charged. If there is no external load, it remains charged to the peak voltage of the rectified output. When there is a load, it gets discharged through the load and the voltage across it begins to fall. In the next half-cycle of rectified output it again gets charged to the peak value (Fig. 14.20). The rate of fall of the voltage across the capacitor depends inversely upon the product of capacitance C and the effective resistance R_L used in the circuit and is called the *time constant*. To make the time constant large value of C should be large. So capacitor input filters use large capacitors. The *output voltage* obtained by using capacitor input filter is nearer to the *peak voltage* of the rectified voltage. This type of filter is most widely used in power supplies.

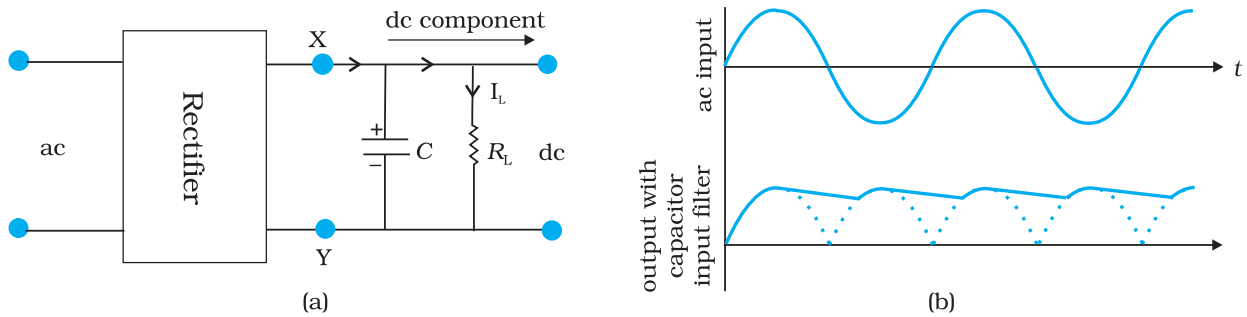


FIGURE 14.20 (a) A full-wave rectifier with capacitor filter, (b) Input and output voltage of rectifier in (a).

14.8 Special Purpose p-n Junction Diodes

In the section, we shall discuss some devices which are basically junction diodes but are developed for different applications.

14.8.1 Zener diode

It is a special purpose semiconductor diode, named after its inventor C. Zener. It is designed to operate under reverse bias in the breakdown region and used as a voltage regulator. The symbol for Zener diode is shown in Fig. 14.21(a).

Zener diode is fabricated by heavily doping both p-, and n- sides of the junction. Due to this, depletion region formed is very thin ($<10^{-6}$ m) and the electric field of the junction is extremely high ($\sim 5\times10^{6}$ V/m) even for a small reverse bias voltage of about 5V. The I-V characteristics of a Zener diode is shown in Fig. 14.21(b). It is seen that when the applied reverse bias voltage(V) reaches the breakdown voltage (V_z) of the Zener diode, there is a large change in the current. Note that after the breakdown voltage V_z, a large change in the current can be produced by almost insignificant change in the reverse bias voltage. In other words, Zener voltage remains constant, even though current through the Zener diode varies over a wide range. This property of the Zener diode is used for regulating supply voltages so that they are constant.

Let us understand how reverse current suddenly increases at the breakdown voltage. We know that reverse current is due to the flow of electrons (minority carriers) from p $\rightarrow$ n and holes from n $\rightarrow$ p. As the reverse bias voltage is increased, the electric field at the junction becomes significant. When the reverse bias voltage V = V_z, then the electric field strength is high enough to pull valence electrons from the host atoms on the p-side which are accelerated to n-side. These electrons account for high current observed at the breakdown. The emission of electrons from the host atoms due to the high electric field is known as internal field emission or field ionisation. The electric field required for field ionisation is of the order of 10^{6} V/m.

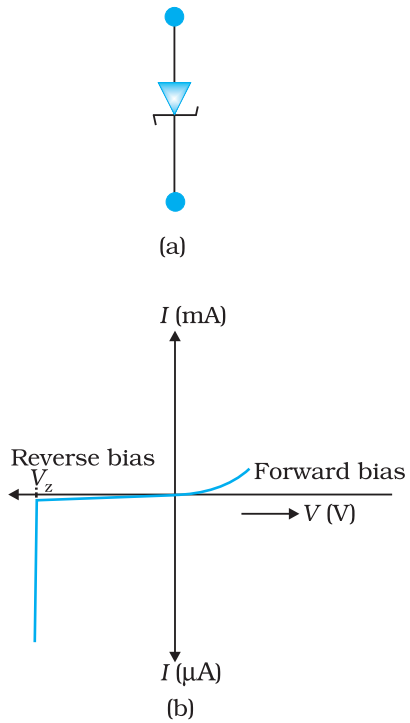


FIGURE 14.21 Zener diode, (a) symbol, (b) *I-V* characteristics.

Zener diode as a voltage regulator

We know that when the ac input voltage of a rectifier fluctuates, its rectified output also fluctuates. To get a constant dc voltage from the dc unregulated output of a rectifier, we use a Zener diode. The circuit diagram of a voltage regulator using a Zener diode is shown in Fig. 14.22.

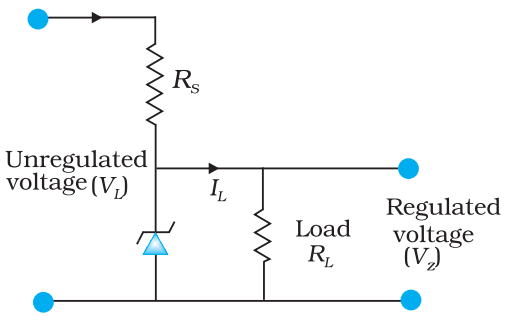


FIGURE 14.22 Zener diode as DC voltage regulator

The unregulated dc voltage (filtered output of a rectifier) is connected to the Zener diode through a series resistance R_s such that the Zener diode is reverse biased. If the input voltage increases, the current through R_s and Zener diode also increases. This increases the voltage drop across R_s without any change in the voltage across the Zener diode. This is because in the breakdown region, Zener voltage remains constant even though the current through the Zener diode changes. Similarly, if the input voltage decreases, the current through R_s and Zener diode also decreases. The voltage drop across R_s decreases without any change in the voltage across the Zener diode. Thus any increase/decrease in the input voltage results in, increase/decrease of the voltage drop across R_s without any change in voltage across the Zener diode. Thus the Zener diode acts as a voltage regulator. We have to select the Zener diode according to the required output voltage and accordingly the series resistance R_s.

EXAMPLE 14.5

Example 14.5 In a Zener regulated power supply a Zener diode with V_Z = 6.0 V is used for regulation. The load current is to be 4.0 mA and the unregulated input is 10.0 V. What should be the value of series resistor R_S?

Solution

The value of R_S should be such that the current through the Zener diode is much larger than the load current. This is to have good load regulation. Choose Zener current as five times the load current, i.e., I_Z = 20 mA. The total current through R_S is, therefore, 24 mA. The voltage drop across R_S is 10.0 – 6.0 = 4.0 V. This gives $R_S = 4.0\text{V}/(24 \times 10^{-3})\ \text{A} = 167\ \Omega$. The nearest value of carbon resistor is 150 Ω. So, a series resistor of 150 Ω is appropriate. Note that slight variation in the value of the resistor does not matter, what is important is that the current I_Z should be sufficiently larger than I_L.

14.8.2 Optoelectronic junction devices

We have seen so far, how a semiconductor diode behaves under applied electrical inputs. In this section, we learn about semiconductor diodes in which carriers are generated by photons (photo-excitation). All these devices are called *optoelectronic devices*. We shall study the functioning of the following optoelectronic devices:

(i) *Photodiodes* used for detecting optical signal (photodetectors).
(ii) *Light emitting diodes* (LED) which convert electrical energy into light.
(iii) *Photovoltaic devices* which convert optical radiation into electricity (*solar cells*).

(i) *Photodiode*

A Photodiode is again a special purpose p-n junction diode fabricated with a transparent window to allow light to fall on the diode. It is operated under reverse bias. When the photodiode is illuminated with light (photons) with energy ($h\nu$) greater than the energy gap (E_g) of the semiconductor, then electron-hole pairs are generated due to the absorption of photons. The diode is fabricated such that the generation of *e-h* pairs takes place in or near the depletion region of the diode. Due to electric field of the junction, electrons and holes are separated before they recombine. The direction of the electric field is such that electrons reach n-side and holes reach p-side. Electrons are collected on n-side and holes are collected on p-side giving rise to an emf. When an external load is connected, current flows. The magnitude of the photocurrent depends on the intensity of incident light (photocurrent is proportional to incident light intensity).

It is easier to observe the change in the current with change in the light intensity, if a reverse bias is applied. Thus photodiode can be used as a photodetector to detect optical signals. The circuit diagram used for the measurement of *I-V* characteristics of a photodiode is shown in Fig. 14.23(a) and a typical *I-V* characteristics in Fig. 14.23(b).

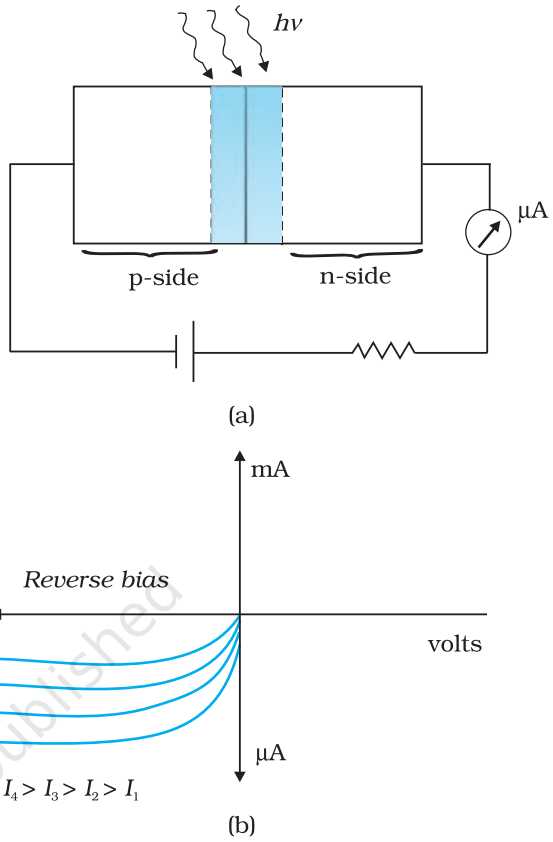


FIGURE 14.23 (a) An illuminated photodiode under reverse bias , (b) *I-V* characteristics of a photodiode for different illumination intensity $I_4 > I_3 > I_2 > I_1$.

Example 14.6 The current in the forward bias is known to be more (~mA) than the current in the reverse bias (~μA). What is the reason then to operate the photodiodes in reverse bias?

Solution Consider the case of an n-type semiconductor. Obviously, the majority carrier density (n) is considerably larger than the minority hole density p (i.e., $n >> p$). On illumination, let the excess electrons and holes generated be Δn and Δp, respectively:

$$n' = n + \Delta n$$

$$p' = p + \Delta p$$

Here n' and p' are the electron and hole concentrations* at any particular illumination and n and p are carriers concentration when there is no illumination. Remember $\Delta n = \Delta p$ and $n >> p$. Hence, the

EXAMPLE 14.6

* Note that, to create an e-h pair, we spend some energy (photoexcitation, thermal excitation, etc.). Therefore when an electron and hole recombine the energy is released in the form of light (radiative recombination) or heat (non-radiative recombination). It depends on semiconductor and the method of fabrication of the p-n junction. For the fabrication of LEDs, semiconductors like GaAs, GaAs-GaP are used in which radiative recombination dominates.

Example 14.6

fractional change in the majority carriers (i.e., $\Delta n/n$) would be much less than that in the minority carriers (i.e., $\Delta p/p$). In general, we can state that the fractional change due to the photo-effects on the *minority carrier dominated reverse bias current* is more easily measurable than the fractional change in the forward bias current. Hence, photodiodes are preferably used in the reverse bias condition for measuring light intensity.

(ii) *Light emitting diode*

It is a heavily doped p-n junction which under forward bias emits spontaneous radiation. The diode is encapsulated with a transparent cover so that emitted light can come out.

When the diode is forward biased, electrons are sent from n → p (where they are minority carriers) and holes are sent from p → n (where they are minority carriers). At the junction boundary the concentration of minority carriers increases compared to the equilibrium concentration (i.e., when there is no bias). Thus at the junction boundary on either side of the junction, excess minority carriers are there which recombine with majority carriers near the junction. On recombination, the energy is released in the form of photons. Photons with energy equal to or slightly less than the band gap are emitted. When the forward current of the diode is small, the intensity of light emitted is small. As the forward current increases, intensity of light increases and reaches a maximum. Further increase in the forward current results in decrease of light intensity. LEDs are biased such that the light emitting efficiency is maximum.

The *V-I* characteristics of a LED is similar to that of a Si junction diode. But the threshold voltages are much higher and slightly different for each colour. The reverse breakdown voltages of LEDs are very low, typically around 5V. So care should be taken that high reverse voltages do not appear across them.

LEDs that can emit red, yellow, orange, green and blue light are commercially available. The semiconductor used for fabrication of visible LEDs must at least have a band gap of 1.8 eV (spectral range of visible light is from about 0.4 μm to 0.7 μm, i.e., from about 3 eV to 1.8 eV). The compound semiconductor Gallium Arsenide – Phosphide ($GaAs_{1-x}P_x$) is used for making LEDs of different colours. $GaAs_{0.6}P_{0.4}$ (E_g ~ 1.9 eV) is used for red LED. GaAs (E_g ~ 1.4 eV) is used for making infrared LED. These LEDs find extensive use in remote controls, burglar alarm systems, optical communication, etc. Extensive research is being done for developing white LEDs which can replace incandescent lamps.

LEDs have the following advantages over conventional incandescent low power lamps:

(i) Low operational voltage and less power.
(ii) Fast action and no warm-up time required.
(iii) The bandwidth of emitted light is 100 Å to 500 Å or in other words it is nearly (but not exactly) monochromatic.
(iv) Long life and ruggedness.
(v) Fast on-off switching capability.

(iii) *Solar cell*

A solar cell is basically a p-n junction which generates emf when solar radiation falls on the p-n junction. It works on the same principle (photovoltaic effect) as the photodiode, except that no external bias is applied and the junction area is kept much larger for solar radiation to be incident because we are interested in more power.

A simple p-n junction solar cell is shown in Fig. 14.24.

A p-Si wafer of about 300 μm is taken over which a thin layer (~0.3 μm) of n-Si is grown on one-side by diffusion process. The other side of p-Si is coated with a metal (back contact). On the top of n-Si layer, metal finger electrode (or metallic grid) is deposited. This acts as a front contact. The metallic grid occupies only a very small fraction of the cell area (<15%) so that light can be incident on the cell from the top.

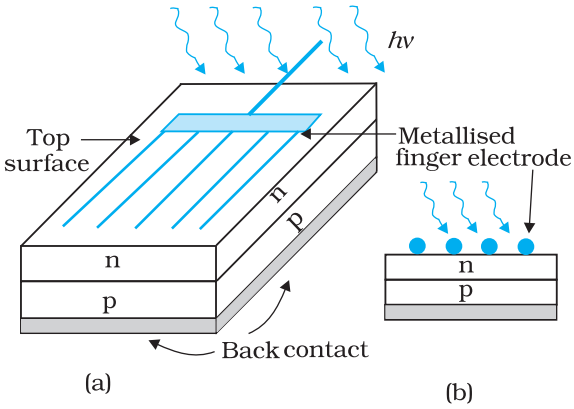


FIGURE 14.24 (a) Typical p-n junction solar cell; (b) Cross-sectional view.

The generation of emf by a solar cell, when light falls on, it is due to the following three basic processes: generation, separation and collection— (i) generation of e-h pairs due to light (with $h\nu > E_g$) close to the junction; (ii) separation of electrons and holes due to electric field of the depletion region. Electrons are swept to n-side and holes to p-side; (iii) the electrons reaching the n-side are collected by the front contact and holes reaching p-side are collected by the back contact. Thus p-side becomes positive and n-side becomes negative giving rise to *photovoltage*.

When an external load is connected as shown in the Fig. 14.25(a) a photocurrent I_L flows through the load. A typical *I-V* characteristics of a solar cell is shown in the Fig. 14.25(b).

Note that the $I - V$ characteristics of solar cell is drawn in the fourth quadrant of the coordinate axes. This is because a solar cell does not draw current but supplies the same to the load.

Semiconductors with band gap close to 1.5 eV are ideal materials for solar cell fabrication. Solar cells are made with semiconductors like Si (E_g = 1.1 eV), GaAs (E_g = 1.43 eV), CdTe (E_g = 1.45 eV), $CuInSe_2$ (E_g = 1.04 eV), etc. The important criteria for the selection of a material for solar cell fabrication are (i) band gap (~1.0 to 1.8 eV), (ii) high optical absorption (~10^4 cm^{-1}), (iii) electrical conductivity, (iv) availability of the raw material, and (v) cost. Note that sunlight is not always required for a solar cell. *Any* light with photon energies greater than the bandgap will do. Solar cells are used to power electronic devices in satellites and space vehicles and also as power supply to some calculators. Production of low-cost photovoltaic cells for large-scale solar energy is a topic for research.

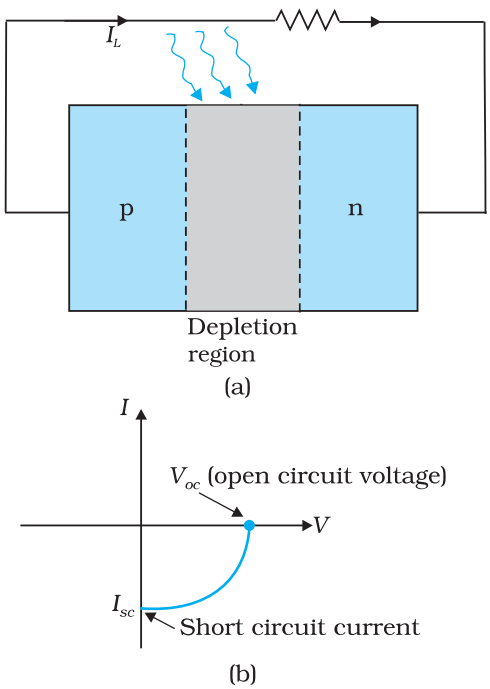


FIGURE 14.25 (a) A typical illuminated p-n junction solar cell; (b) *I-V* characteristics of a solar cell.

EXAMPLE 14.7

Example 14.7 Why are Si and GaAs are preferred materials for solar cells?

Solution The solar radiation spectrum received by us is shown in Fig. 14.26.

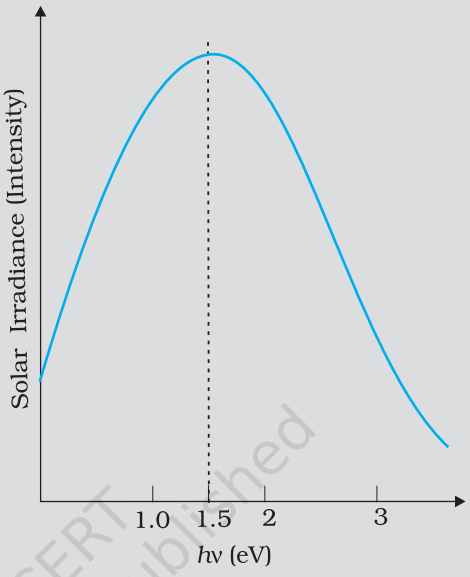


FIGURE 14.26

The maxima is near 1.5 eV. For photo-excitation, $h\nu > E_g$. Hence, semiconductor with band gap ~1.5 eV or lower is likely to give better solar conversion efficiency. Silicon has E_g ~ 1.1 eV while for GaAs it is ~1.53 eV. In fact, GaAs is better (in spite of its higher band gap) than Si because of its relatively higher absorption coefficient. If we choose materials like CdS or CdSe (E_g ~ 2.4 eV), we can use only the high energy component of the solar energy for photo-conversion and a significant part of energy will be of no use.

The question arises: why we do not use material like PbS (E_g ~ 0.4 eV) which satisfy the condition $h\nu > E_g$ for ν maxima corresponding to the solar radiation spectra? If we do so, most of the solar radiation will be absorbed on the *top-layer* of solar cell and will not reach in or near the depletion region. For effective electron-hole separation, due to the junction field, we want the photo-generation to occur in the junction region only.

14.9 Digital Electronics and Logic Gates

In electronics circuits like amplifiers, oscillators, introduced to you in earlier sections, the signal (current or voltage) has been in the form of continuous, time-varying voltage or current. Such signals are called continuous or *analog signals*. A typical analog signal is shown in Figure. 14.27(a). Fig. 14.27(b) shows a *pulse waveform* in which only discrete values of voltages are possible. It is convenient to use binary numbers to represent such signals. A binary number has only two digits '0' (say, 0V) and '1' (say, 5V). In digital electronics we use only these two levels of voltage as shown in Fig. 14.27(b). Such signals are called *Digital Signals*.

In digital circuits only two values (represented by 0 or 1) of the input and output voltage are permissible.

This section is intended to provide the first step in our understanding of digital electronics. We shall restrict our study to some basic building blocks of digital electronics (called *Logic Gates*) which process the digital signals in a specific manner. Logic gates are used in calculators, digital watches, computers, robots, industrial control systems, and in telecommunications.

A light switch in your house can be used as an example of a digital circuit. The light is either ON or OFF depending on the switch position. When the light is ON, the output value is '1'. When the light is OFF the output value is '0'. The inputs are the position of the light switch. The switch is placed either in the ON or OFF position to activate the light.

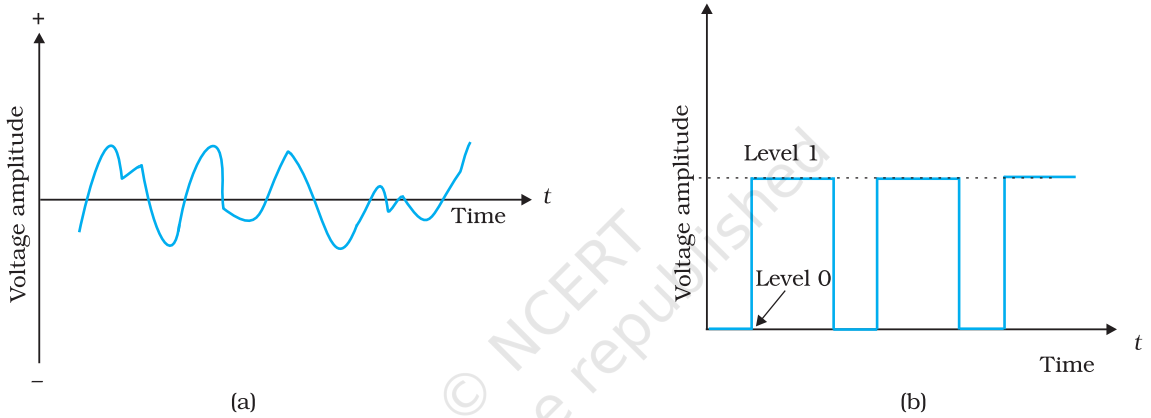


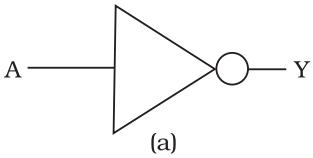
FIGURE 14.27 (a) Analog signal, (b) Digital signal.

14.9.1 Logic gates

A gate is a digital circuit that follows curtain *logical* relationship between the input and output voltages. Therefore, they are generally known as *logic gates* — gates because they control the flow of information. The five common logic gates used are NOT, AND, OR, NAND, NOR. Each logic gate is indicated by a symbol and its function is defined by a *truth table* that shows all the possible input logic level combinations with their respective output logic levels. Truth tables help understand the behaviour of logic gates. These logic gates can be realised using semiconductor devices.

(i) NOT gate

This is the most basic gate, with one input and one output. It produces a '1' output if the input is '0' and vice-versa. That is, it produces an inverted version of the input at its output. This is why it is also known as an *inverter*. The commonly used symbol together with the truth table for this gate is given in Fig. 14.28.



(a)

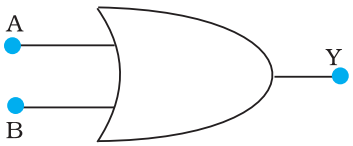
Input	Output
A	Y
0	1
1	0

(b)

FIGURE 14.28 (a) Logic symbol, (b) Truth table of NOT gate.

(ii) OR Gate

An *OR* gate has two or more inputs with one output. The logic symbol and truth table are shown in Fig. 14.29. The output Y is 1 when either input A *or* input B *or* both are 1s, that is, if any of the input is high, the output is high.



(a)

Input		Output
A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

(b)

FIGURE 14.29 (a) Logic symbol (b) Truth table of OR gate.

Apart from carrying out the above mathematical logic operation, this *gate* can be used for modifying the pulse waveform as explained in the following example.

EXAMPLE 14.8

Example 14.8 Justify the output waveform (Y) of the OR gate for the following inputs A and B given in Fig. 14.30.

Solution Note the following:

- At $t < t_1$; A = 0, B = 0; Hence Y = 0
- For t_1 to t_2; A = 1, B = 0; Hence Y = 1
- For t_2 to t_3; A = 1, B = 1; Hence Y = 1
- For t_3 to t_4; A = 0, B = 1; Hence Y = 1
- For t_4 to t_5; A = 0, B = 0; Hence Y = 0
- For t_5 to t_6; A = 1, B = 0; Hence Y = 1
- For $t > t_6$; A = 0, B = 1; Hence Y = 1

Therefore the waveform Y will be as shown in the Fig. 14.30.

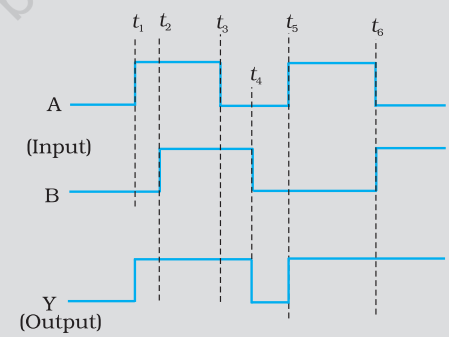


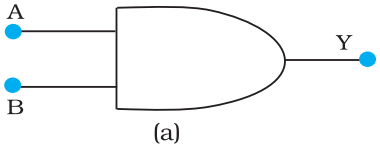
FIGURE 14.30

Input		Output
A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

(b)

(iii) AND Gate

An *AND* gate has two or more inputs and one output. The output Y of AND gate is 1 only when input A *and* input B are both 1. The logic symbol and truth table for this gate are given in Fig. 14.31



(a)

FIGURE 14.31 (a) Logic symbol, (b) Truth table of AND gate.

Example 14.9 Take A and B input waveforms similar to that in Example 14.8. Sketch the output waveform obtained from AND gate.

Solution

- For $t \leq t_1$; A = 0, B = 0; Hence Y = 0
- For t_1 to t_2; A = 1, B = 0; Hence Y = 0
- For t_2 to t_3; A = 1, B = 1; Hence Y = 1
- For t_3 to t_4; A = 0, B = 1; Hence Y = 0
- For t_4 to t_5; A = 0, B = 0; Hence Y = 0
- For t_5 to t_6; A = 1, B = 0; Hence Y = 0
- For $t > t_6$; A = 0, B = 1; Hence Y = 0

Based on the above, the output waveform for AND gate can be drawn as given below.

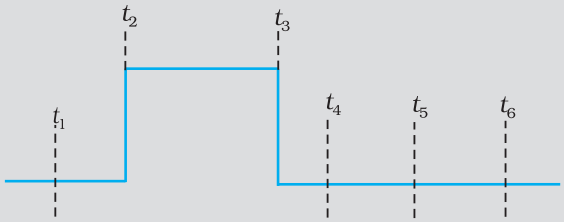


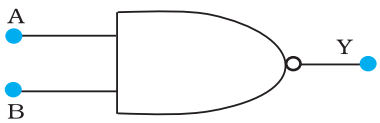
FIGURE 14.32

EXAMPLE 14.9

(iv) NAND Gate

This is an AND gate followed by a NOT gate. If inputs A *and* B are both '1', the output Y is *not* '1'. The gate gets its name from this NOT AND behaviour. Figure 14.33 shows the symbol and truth table of NAND gate.

NAND gates are also called *Universal Gates* since by using these gates you can realise other basic gates like OR, AND and NOT (Exercises 14.12 and 14.13).



(a)

Input		Output
A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

(b)

FIGURE 14.33 (a) Logic symbol, (b) Truth table of NAND gate.

Example 14.10 Sketch the output Y from a NAND gate having inputs A and B given below:

Solution

- For $t < t_1$; A = 1, B = 1; Hence Y = 0
- For t_1 to t_2; A = 0, B = 0; Hence Y = 1
- For t_2 to t_3; A = 0, B = 1; Hence Y = 1
- For t_3 to t_4; A = 1, B = 0; Hence Y = 1

EXAMPLE 14.10

Example 14.10

- For t_4 to t_5; A = 1, B = 1; Hence Y = 0
- For t_5 to t_6; A = 0, B = 0; Hence Y = 1
- For $t > t_6$; A = 0, B = 1; Hence Y = 1

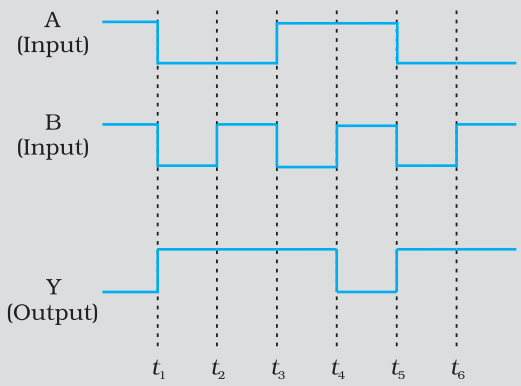
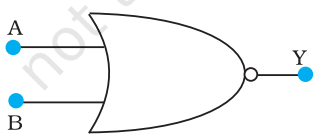


FIGURE 14.34

(v) NOR Gate

It has two or more inputs and one output. A NOT- operation applied *after* OR gate gives a NOT-OR gate (or simply NOR gate). Its output Y is '1' only when both inputs A and B are '0', i.e., neither one input *nor* the other is '1'. The symbol and truth table for NOR gate is given in Fig. 14.35.



(a)

Input		Output
A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

(b)

FIGURE 14.35 (a) Logic symbol, (b) Truth table of NOR gate.

NOR gates are considered as *universal* gates because you can obtain all the gates like AND, OR, NOT by using only NOR gates (Exercises 14.14 and 14.15).

FASTER AND SMALLER: THE FUTURE OF COMPUTER TECHNOLOGY

The *Integrated Chip* (IC) is at the heart of all computer systems. In fact ICs are found in almost all electrical devices like cars, televisions, CD players, cell phones etc. The miniaturisation that made the modern personal computer possible could never have happened without the IC. ICs are electronic devices that contain many transistors, resistors, capacitors, connecting wires – all in one package. You must have heard of the

microprocessor. The microprocessor is an IC that processes all information in a computer, like keeping track of what keys are pressed, running programmes, games etc. The IC was first invented by Jack Kilky at Texas Instruments in 1958 and he was awarded Nobel Prize for this in 2000. ICs are produced on a piece of semiconductor crystal (or chip) by a process called *photolithography*. Thus, the entire Information Technology (IT) industry hinges on semiconductors. Over the years, the complexity of ICs has increased while the size of its features continued to shrink. In the past five decades, a dramatic miniaturisation in computer technology has made modern day computers *faster and smaller*. In the 1970s, Gordon Moore, co-founder of INTEL, pointed out that the memory capacity of a chip (IC) approximately doubled every one and a half years. This is popularly known as *Moore's law*. The number of transistors per chip has risen exponentially and each year computers are becoming more powerful, yet cheaper than the year before. It is intimated from current trends that the computers available in 2020 will operate at 40 GHz (40,000 MHz) and would be much smaller, more efficient and less expensive than present day computers. The explosive growth in the semiconductor industry and computer technology is best expressed by a famous quote from Gordon Moore: "If the auto industry advanced as rapidly as the semiconductor industry, a Rolls Royce would get half a million miles per gallon, and it would be cheaper to throw it away than to park it".

SUMMARY

1. Semiconductors are the basic materials used in the present solid state electronic devices like diode, transistor, ICs, etc.
2. Lattice structure and the atomic structure of constituent elements decide whether a particular material will be insulator, metal or semiconductor.
3. Metals have low resistivity (10^{-2} to 10^{-8} Ωm), insulators have very high resistivity (>10^{8} Ω m^{-1}), while semiconductors have intermediate values of resistivity.
4. Semiconductors are elemental (Si, Ge) as well as compound (GaAs, CdS, etc.).
5. Pure semiconductors are called 'intrinsic semiconductors'. The presence of charge carriers (electrons and holes) is an 'intrinsic' property of the material and these are obtained as a result of thermal excitation. The number of electrons (n_e) is equal to the number of holes (n_h) in intrinsic conductors. Holes are essentially electron vacancies with an effective positive charge.
6. The number of charge carriers can be changed by 'doping' of a suitable impurity in pure semiconductors. Such semiconductors are known as extrinsic semiconductors. These are of two types (n-type and p-type).
7. In n-type semiconductors, $n_e >> n_h$ while in p-type semiconductors $n_h >> n_e$.
8. n-type semiconducting Si or Ge is obtained by doping with pentavalent atoms (donors) like As, Sb, P, etc., while p-type Si or Ge can be obtained by doping with trivalent atom (acceptors) like B, Al, In etc.
9. $n_e n_h = n_i^2$ in all cases. Further, the material possesses an *overall charge neutrality*.

10. There are two distinct band of energies (called valence band and conduction band) in which the electrons in a material lie. Valence band energies are low as compared to conduction band energies. All energy levels in the valence band are filled while energy levels in the conduction band may be fully empty or partially filled. The electrons in the conduction band are free to move in a solid and are responsible for the conductivity. The extent of conductivity depends upon the energy gap (E_g) between the top of valence band (E_V) and the bottom of the conduction band E_C. The electrons from valence band can be excited by heat, light or electrical energy to the conduction band and thus, produce a change in the current flowing in a semiconductor.
11. For insulators $E_g > 3$ eV, for semiconductors E_g is 0.2 eV to 3 eV, while for metals $E_g \approx 0$.
12. p-n junction is the 'key' to all semiconductor devices. When such a junction is made, a 'depletion layer' is formed consisting of immobile ion-cores devoid of their electrons or holes. This is responsible for a junction potential barrier.
13. By changing the external applied voltage, junction barriers can be changed. In forward bias (n-side is connected to negative terminal of the battery and p-side is connected to the positive), the barrier is decreased while the barrier increases in reverse bias. Hence, forward bias current is more (mA) while it is very small (μA) in a p-n junction diode.
14. Diodes can be used for rectifying an ac voltage (restricting the ac voltage to one direction). With the help of a capacitor or a suitable filter, a dc voltage can be obtained.
15. There are some special purpose diodes.
16. Zener diode is one such special purpose diode. In reverse bias, after a certain voltage, the current suddenly increases (breakdown voltage) in a Zener diode. This property has been used to obtain *voltage regulation*.
17. p-n junctions have also been used to obtain many photonic or optoelectronic devices where one of the participating entity is 'photon': (a) Photodiodes in which photon excitation results in a change of reverse saturation current which helps us to measure light intensity; (b) Solar cells which convert photon energy into electricity; (c) Light Emitting Diode and Diode Laser in which electron excitation by a bias voltage results in the generation of light.
18. There are some special circuits which handle the digital data consisting of 0 and 1 levels. This forms the subject of Digital Electronics.
19. The important digital circuits performing special logic operations are called logic gates. These are: OR, AND, NOT, NAND, and NOR gates.

POINTS TO PONDER

1. The energy bands (E_C or E_V) in the semiconductors are space delocalised which means that these are not located in any specific place inside the solid. The energies are the overall averages. When you see a picture in which E_C or E_V are drawn as straight lines, then they should be respectively taken simply as the *bottom* of conduction band energy levels and *top* of valence band energy levels.
2. In elemental semiconductors (Si or Ge), the n-type or p-type semiconductors are obtained by introducing 'dopants' as defects. In compound semiconductors, the change in relative stoichiometric ratio can also change the type of semiconductor. For example, in ideal GaAs

the ratio of Ga:As is 1:1 but in Ga-rich or As-rich GaAs it could respectively be $Ga_{1.1}$ $As_{0.9}$ or $Ga_{0.9}$ $As_{1.1}$. In general, the presence of defects control the properties of semiconductors in many ways.

3. In modern day circuit, many logical gates or circuits are integrated in one single 'Chip'. These are known as Intgrated circuits (IC).

EXERCISES

14.1 In an n-type silicon, which of the following statement is true:
(a) Electrons are majority carriers and trivalent atoms are the dopants.
(b) Electrons are minority carriers and pentavalent atoms are the dopants.
(c) Holes are minority carriers and pentavalent atoms are the dopants.
(d) Holes are majority carriers and trivalent atoms are the dopants.

14.2 Which of the statements given in Exercise 14.1 is true for p-type semiconductos.

14.3 Carbon, silicon and germanium have four valence electrons each. These are characterised by valence and conduction bands separated by energy band gap respectively equal to $(E_g)_C$, $(E_g)_{Si}$ and $(E_g)_{Ge}$. Which of the following statements is true?
(a) $(E_g)_{Si} < (E_g)_{Ge} < (E_g)_C$
(b) $(E_g)_C < (E_g)_{Ge} > (E_g)_{Si}$
(c) $(E_g)_C > (E_g)_{Si} > (E_g)_{Ge}$
(d) $(E_g)_C = (E_g)_{Si} = (E_g)_{Ge}$

14.4 In an unbiased p-n junction, holes diffuse from the p-region to n-region because
(a) free electrons in the n-region attract them.
(b) they move across the junction by the potential difference.
(c) hole concentration in p-region is more as compared to n-region.
(d) All the above.

14.5 When a forward bias is applied to a p-n junction, it
(a) raises the potential barrier.
(b) reduces the majority carrier current to zero.
(c) lowers the potential barrier.
(d) None of the above.

14.6 In half-wave rectification, what is the output frequency if the input frequency is 50 Hz. What is the output frequency of a full-wave rectifier for the same input frequency.

14.7 A p-n photodiode is fabricated from a semiconductor with band gap of 2.8 eV. Can it detect a wavelength of 6000 nm?

ADDITIONAL EXERCISES

14.8 The number of silicon atoms per m^3 is 5×10^{28}. This is doped simultaneously with 5×10^{22} atoms per m^3 of Arsenic and 5×10^{20} per m^3 atoms of Indium. Calculate the number of electrons and holes. Given that $n_i = 1.5 \times 10^{16}$ m^{-3}. Is the material n-type or p-type?

14.9 In an intrinsic semiconductor the energy gap E_g is 1.2eV. Its hole mobility is much smaller than electron mobility and independent of temperature. What is the ratio between conductivity at 600K and that at 300K? Assume that the temperature dependence of intrinsic carrier concentration n_i is given by

$$n_i = n_0 \exp\left(-\frac{E_g}{2k_B T}\right)$$

where n_0 is a constant.

14.10 In a p-n junction diode, the current I can be expressed as

$$I = I_0 \exp\left(\frac{eV}{2k_B T} - 1\right)$$

where I_0 is called the reverse saturation current, V is the voltage across the diode and is positive for forward bias and negative for reverse bias, and I is the current through the diode, k_B is the Boltzmann constant (8.6×10^{-5} eV/K) and T is the absolute temperature. If for a given diode $I_0 = 5 \times 10^{-12}$ A and T = 300 K, then

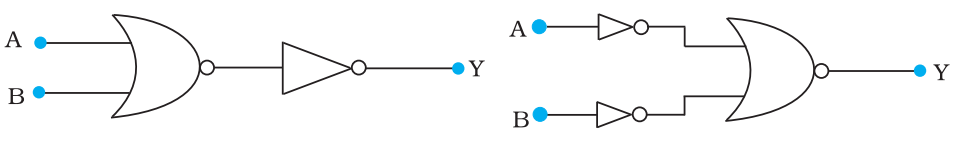
(a) What will be the forward current at a forward voltage of 0.6 V?

(b) What will be the increase in the current if the voltage across the diode is increased to 0.7 V?

(c) What is the dynamic resistance?

(d) What will be the current if reverse bias voltage changes from 1 V to 2 V?

14.11 You are given the two circuits as shown in Fig. 14.36. Show that circuit (a) acts as OR gate while the circuit (b) acts as AND gate.



(a) (b)

FIGURE 14.36

14.12 Write the truth table for a NAND gate connected as given in Fig. 14.37.

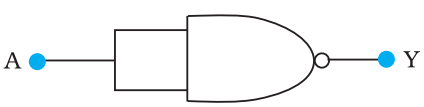
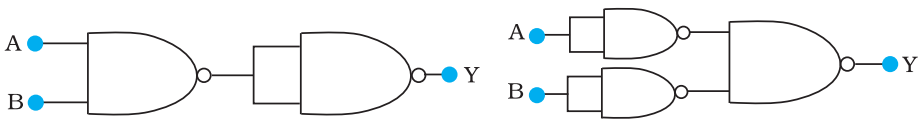


FIGURE 14.37

Hence identify the exact logic operation carried out by this circuit.

14.13 You are given two circuits as shown in Fig. 14.38, which consist of NAND gates. Identify the logic operation carried out by the two circuits.



(a) (b)

FIGURE 14.38

14.14 Write the truth table for circuit given in Fig. 14.39 below consisting of NOR gates and identify the logic operation (OR, AND, NOT) which this circuit is performing.

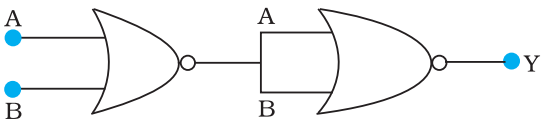
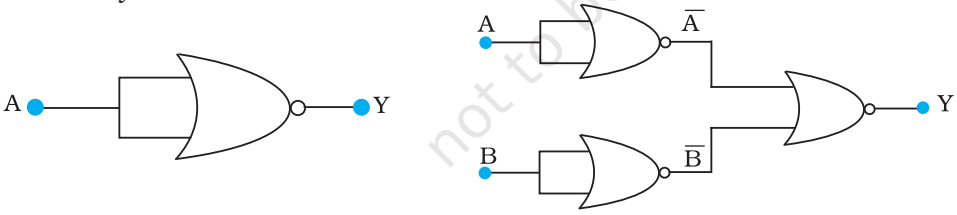


FIGURE 14.39

(Hint: A = 0, B = 1 then A and B inputs of second NOR gate will be 0 and hence Y=1. Similarly work out the values of Y for other combinations of A and B. Compare with the truth table of OR, AND, NOT gates and find the correct one.)

14.15 Write the truth table for the circuits given in Fig. 14.40 consisting of NOR gates only. Identify the logic operations (OR, AND, NOT) performed by the two circuits.



(a) (b)

FIGURE 14.40