CHAPTER 1

Introduction to Semiconductor Physics

Learning Objectives

After completing this chapter, you will learn the following:

- Classification of materials as insulators, conductors and semiconductors.
- Difference between intrinsic and extrinsic semiconductors.
- Difference between P-type and N-type semiconductors.
- Current density and Fermi level in a semiconductor.
- Mass action law.
- Hall effect.

This chapter covers the fundamental topics of semiconductor physics of relevance to understanding operational basics of semiconductor devices such as diodes, bipolar junction transistors (BJTs) transistors, field effect transistors (FETs) and so on covered in the latter chapters of the book. This chapter begins with a brief discussion on insulators, conductors and semiconductors and then focuses mainly on the semiconductors. The topics covered include intrinsic and extrinsic semiconductors, carrier charge densities in a semiconductor, mass action law and Hall effect. The chapter is amply illustrated with a large number of solved examples.

1.1 Insulators, Conductors and Semiconductors

Materials, in general, can be classified as insulators, conductors and semiconductors depending upon their conductivity levels. Insulators are materials that offer a large resistance to the flow of current through them. Conductors are materials that support flow of current through them when an external voltage is applied across them. In other words, insulators are materials that offer very poor conductivity while conductors have high conductivity levels. Semiconductors, on the other hand, have conductivity levels somewhere between the extremes of an insulator and a conductor. In the following paragraphs, these three types of materials are discussed in detail.

Insulators

As mentioned above, insulators are materials that offer a large resistance to the flow of current through them. The typical resistivity level of an insulator is of the order of 10^{10} to 10^{12} Ω cm. Therefore, the application of voltage across the insulator results in negligible flow of current. If one looks at the atomic structure of insulators, one finds that they have seven to eight valence electrons. (The electrons in the valence shell, that is, the outermost shell, are referred to as the valence electrons.) Valence electrons are tightly bound to

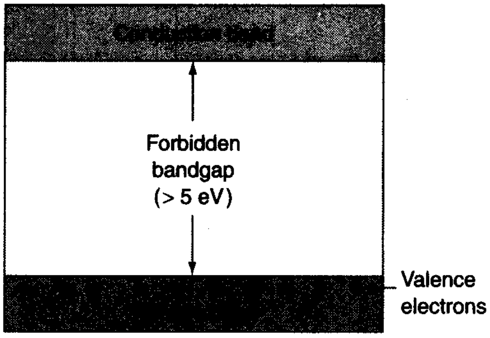


Figure 1.1 | Energy band diagram of an insulator.

the atom, so there are no free electrons that can move through the material. Some of the popular insulator materials are mica, glass, quartz, etc.

The energy-band structure of an insulator is shown in Figure 1.1. Band structure of a material defines the band of energy levels that an electron can occupy. Valence band is the range of electron energy where the electrons remain bonded to the atom and do not contribute to the electric current. It is the highest range of electron energies where the electrons are present at absolute zero temperature (0 K). Conduction band is the range of electron energies, higher than the valence band, sufficient to make the electrons free to accelerate under the influence of an external applied electric field resulting in the flow of electric current.

The energy band between the valence band and the conduction band is the forbidden bandgap. As is clear from Figure 1.1, there is a large forbidden bandgap of greater than 5 eV between the valence and the conduction energy bands of an insulator. As an example, the bandgap of diamond is approximately equal to 5.5 eV. Owing to this large forbidden bandgap, there are very few electrons in the conduction band and hence the conductivity of an insulator is poor. Even an increase in temperature or in energy of the applied electric field is insufficient to transfer the electrons from the valence band to the conduction band.

Conductors

Conductors are materials that offer very little resistance to the flow of current through them, that is, they support a generous flow of current when an external electric field is applied across their terminals. Resistivity level of conductors is of the order of 10^{-4} to 10^{-6} Ω cm. Generally, conductors have three or less than three valence electrons. These electrons are loosely bound and are free to move through the material. Metals such as copper, aluminum, gold and silver are good conductors. Figure 1.2(a) shows the atomic structure of copper. Copper has one valence electron and hence is a good conductor. Figure 1.2(b) shows the energy band structure of a conductor. The valence and conduction bands overlap and there is no energy gap for the electrons to overcome in order to move from the valence band to the conduction band. This implies that there are free electrons in the conduction band even at absolute zero temperature (0 K). Therefore, when an external electric field is applied there is a large flow of current through the conductor.

Semiconductors

Semiconductors are materials that have conductivity levels somewhere between the extremes of a conductor and an insulator. The resistivity level of semiconductors is in the range of 10 to 10^4 Ω cm. Two of the most commonly used semiconductor materials are silicon (Si) and germanium (Ge). Silicon has 14 orbiting electrons and germanium has 32 orbiting electrons as shown in Figures 1.3(a) and (b), respectively. As is evident from the figure, both silicon and germanium have four valence electrons. Materials having three and five

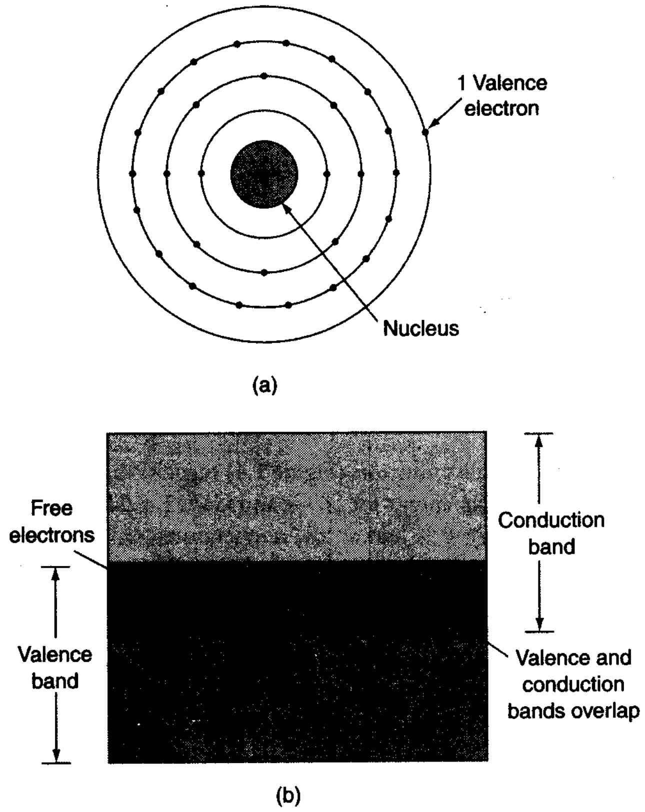
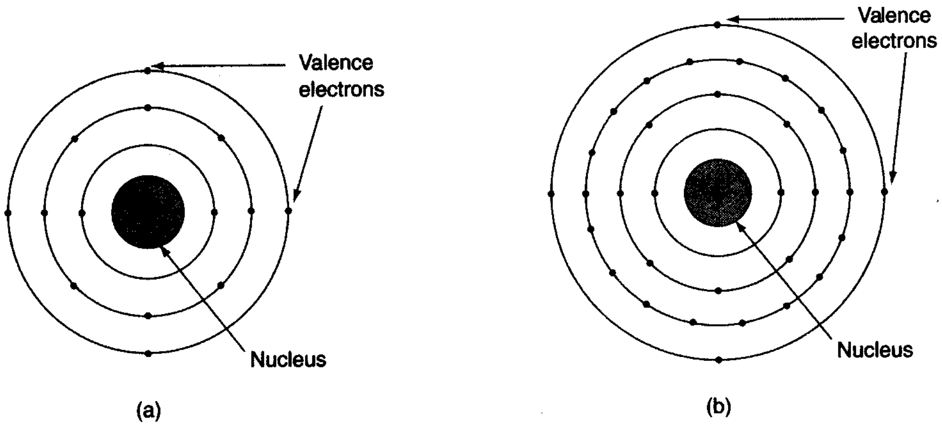


Figure 1.2 (a) Atomic structure of a conductor (copper); (b) energy band diagram of a conductor.



(a) (b)

Figure 1.3 (a) Atomic structure of silicon; (b) atomic structure of germanium.

valence electrons combine with each other to form semiconductors. Examples of such semiconductors are gallium arsenide (GaAs) and indium phosphide (InP). The valence electrons in a semiconductor are not free to move as they are in a metal and are trapped in bonds between adjacent atoms.

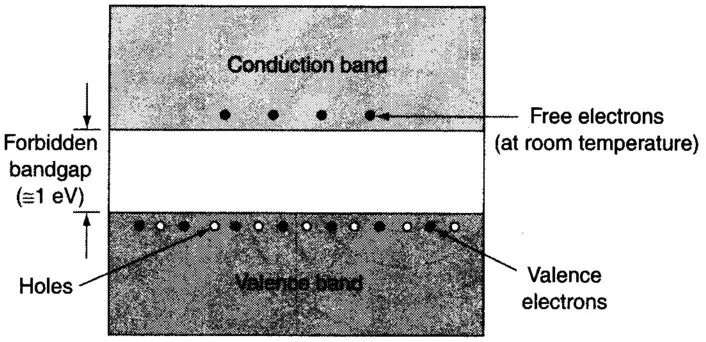


Figure 1.4 | Energy band diagram of a semiconductor.

A look at the band structure of semiconductors (Figure 1.4) suggests that the forbidden bandgap is of the order of 1 eV. For example, the bandgap energy for Si, Ge and GaAs is 1.21, 0.785 and 1.42 eV, respectively, at absolute zero temperature (0 K). At 0 K and at low temperatures, valence band electrons do not have sufficient energy to cross the forbidden bandgap and reach the conduction band. Thus, semiconductors act as insulators at 0 K and at low temperatures. As the temperature increases, a large number of valence electrons acquire sufficient energy to leave the valence band, cross the forbidden bandgap and reach the conduction band. These are now the free electrons as they can move freely under the influence of an external applied electric field. At room temperature (300 K), there are sufficient electrons in the conduction band and hence the semiconductor is capable of conducting some current at room temperature. The absence of an electron in the valence band is referred to as a hole and is represented by a small circle as shown in Figure 1.4. In the case of semiconductors both electrons and holes constitute the flow of current, whereas in the case of conductors, the current is due to the flow of electrons only.

In the discussion above, it is assumed that there are no external atoms added to the parent semiconductor material. Such semiconductors are referred to as *intrinsic semiconductors*. Certain impurity atoms when added to the intrinsic semiconductor materials increase their conductivity. Such semiconductors, with added impurity atoms, are called *extrinsic semiconductors*. Intrinsic and extrinsic semiconductors are discussed in detail in Section 1.2.

1.2 Semiconductor Types

An introduction to semiconductor materials was given in Section 1.1. In the present section, we will describe the intrinsic and extrinsic semiconductors in detail.

Intrinsic Semiconductors

Intrinsic semiconductors are semiconductors with very low level of impurity concentration. They are essentially as pure as can be available through modern technology. The purity levels are of the order of 1 part in 10 billion. Conduction in intrinsic semiconductors is either due to thermal excitation or due to crystal defects. Si and Ge are the two most important semiconductors used. Other examples include GaAs, indium antimonide (InSb) and so on.

Structure

Let us consider the case of silicon. Silicon has 14 orbiting electrons. The innermost shell can hold two electrons, the middle shell eight electrons and the outermost shell four electrons. Therefore, silicon has four valence electrons and is referred to as a tetravalent atom. These four electrons are shared by four neighboring atoms in the crystal. It is the sharing of these four electrons of an atom with their respective neighboring atoms that constitutes a total of eight electrons in its valence shell. This bonding of atoms due to sharing of electrons is called covalent bonding. Figure 1.5(a) shows the crystal structure of silicon at

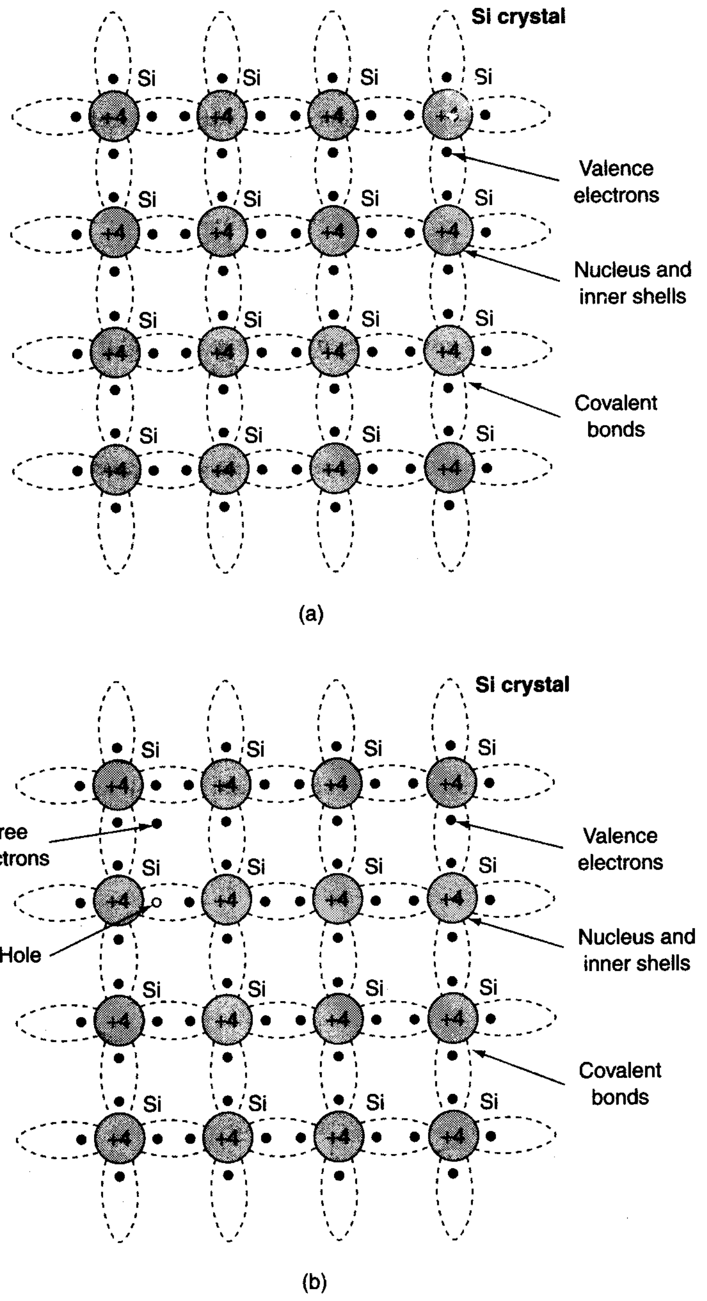


Figure 1.5 | Crystal structure of silicon at (a) absolute zero temperature and (b) room temperature.

absolute zero temperature (0 K). Owing to the covalent bonding, the valence electrons are tightly bound to the nucleus and hence the crystal has poor conductivity at low temperatures of the order of 0 K.

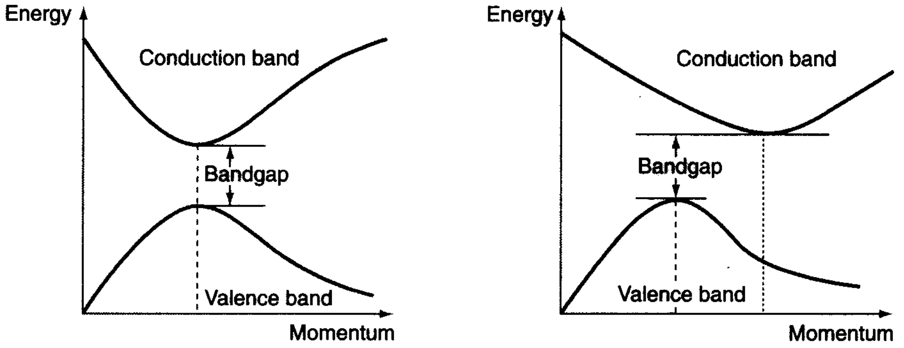
At room temperature, the thermal energy is sufficient enough to break some of the covalent bonds as shown in Figure 1.5(b). The electrons are raised to the conduction band and are referred to as free electrons that are available for conduction. The absence of electron in the covalent bond means that the atom now has a positive charge referred to as a hole [represented by a small circle in Figure 1.5(b)]. Holes serve as a carrier of electricity in a manner similar to free electrons. In fact, the motion of hole in one direction is equivalent to the motion of negative charge in the opposite direction.

Germanium also has four electrons in the valence shell. Intrinsic semiconductors are also formed by combination of atoms having three valence electrons and atoms having five valence electrons. Examples include GaAs and InSb. In GaAs, gallium atom has three valence electrons and arsenic atom has five valence electrons. Other combinations are also possible, for example mercury, cadmium and tellurium bond to form mercury cadmium telluride (HgCdTe). Detailed description of these semiconductors is outside the scope of the book. However, the discussion for silicon semiconductors holds good for these intrinsic semiconductors also.

In a nutshell, it can be said that at low temperatures of the order of 0 K, the intrinsic semiconductor behaves as an insulator as no free carriers of electricity are available.

Types

Intrinsic semiconductors can be further classified as ***direct bandgap semiconductors*** and ***indirect bandgap semiconductors***. In a direct bandgap semiconductor, the maximum energy of the valence band occurs at the same momentum value as the minimum energy of the conduction band [Figure 1.6(a)]. Thus, in a direct bandgap semiconductor, electrons present at the minimum of conduction band combine with holes present at the maximum of valence band while conserving momentum. The energy released due to recombination is emitted in the form of photon of light. Therefore, these semiconductors are used in making light-emitting diodes (LED) and laser diodes. Examples of direct bandgap semiconductors include GaAs and HgCdTe. In an indirect bandgap semiconductor, the maximum energy of the valence band occurs at a different momentum value than the minimum energy of the conduction band [Figure 1.6(b)]. Hence, a direct transition across the bandgap does not conserve momentum and does not emit photons of light. Instead, the energy in this case is released in the form of heat. Silicon and germanium are examples of indirect bandgap semiconductors.



(a) (b)

Figure 1.6 (a) Direct bandgap intrinsic semiconductor; (b) indirect bandgap intrinsic semiconductor.

Charge Concentration

In an intrinsic semiconductor, the number of holes is equal to the number of electrons. Hole and electron pairs are generated by thermal agitation and disappear due to recombination. Therefore, in an intrinsic semiconductor

$$n = p = n_\mathrm{i} \tag{1.1}$$

where n is the electron concentration (number of electrons/cm^3), p is the hole concentration (number of holes/cm^3) and n_i is the intrinsic concentration.

The value of n_i is given by

$$n_\mathrm{i}^2 = AT^3 \exp(-E_{\mathrm{G0}}/kT) \tag{1.2}$$

where T is the temperature in Kelvin, E_{G0} is the energy gap at 0 K, k is the Boltzmann constant in eV/K and A is a constant. It is clear from Eq. (1.2) that the intrinsic concentration n_i increases with increase in temperature.

Electrical Properties

A semiconductor is a bipolar device, that is, both electrons and holes contribute to the flow of current. It may be mentioned here that metals are unipolar devices, that is, only electrons act as current carriers. In a semiconductor there are two different mechanisms of current flow, namely, the "electron flow in the conduction band" and the "hole flow in the valence band". When an external electric field is applied, the free electron may either contribute to the current by drifting through the crystal or combine with a hole in the valence band. The first component constitutes the electron flow in the conduction band. When an electron combines with a hole, it leaves a hole in its initial position. This hole may now be filled by an electron from another covalent bond creating a hole in its position and the process continues. This results in the motion of holes in the valence band in the direction opposite to the direction of motion of electrons.

The mathematical expression for the current density in any material is given by

$$J = (n\mu_\mathrm{n} + p\mu_\mathrm{p})q\varepsilon \tag{1.3}$$

where J is the current density in A/cm^2, n is the electron concentration (number of electrons/cm^3), p is the hole concentration (number of holes/cm^3), μ_n is the mobility of an electron in the material in cm^2/Vs, μ_p is the mobility of a hole in the material in cm^2/Vs, q is the charge of an electron (1.6×10^{-19} C) and ε is the applied electric field in V/cm.

This current is due to the potential gradient created by the applied electric field and is referred to as drift current density. The expression for conductivity (σ) is given by

$$\sigma = (n\mu_\mathrm{n} + p\mu_\mathrm{p})q \tag{1.4}$$

Since in an intrinsic semiconductor, $n = p = n_\mathrm{i}$, therefore from Eqs. (1.3) and (1.4), respectively, we obtain

$$J = (\mu_\mathrm{n} + \mu_\mathrm{p})n_\mathrm{i}q\varepsilon \tag{1.5}$$

$$\sigma = (\mu_\mathrm{n} + \mu_\mathrm{p})n_\mathrm{i}q \tag{1.6}$$

Energy Bandgap

The forbidden bandgap (also called energy bandgap) of a semiconductor depends on its temperature and decreases with increase in temperature.

For silicon, the energy bandgap [E_G (T)] at temperature T (K) is given by

$$E_\mathrm{G}(T) = 1.21 - 3.60 \times 10^{-4} T \tag{1.7}$$

and for germanium it is given by

$$E_G(T) = 0.785 - 2.23 \times 10^{-4}T \tag{1.8}$$

where T is the temperature in Kelvin. At room temperature (taken as 300 K), the bandgap for silicon and germanium are 1.1 and 0.72 eV, respectively.

Fermi Level

The probability that an energy level in a semiconductor is occupied by an electron is given by

$$f(E) = \frac{1}{1 + \exp[(E - E_F)/kT]} \tag{1.9}$$

where $f(E)$ is the Fermi–Dirac probability function (i.e., probability of finding an electron in the energy state E), k is the Boltzmann constant (8.642×10^{-5} eV/K), T is the temperature in Kelvin and E_F is the Fermi level in eV.

Fermi level represents the energy state with 50% probability of being filled by an electron if no forbidden energy bandgap exists. In an intrinsic semiconductor at absolute zero temperature, the probability of finding an electron in the valence band is 100% and the probability of finding the electron in the conduction band is 0%. The Fermi level in an intrinsic semiconductor at absolute zero temperature lies at the center of the forbidden bandgap [Figure 1.7(a)].

As the temperature increases, some of the electrons are excited to higher energy levels. They leave the valence band and jump to the conduction band. Thus, the probability of finding an electron in the valence band decreases and the probability of finding an electron in the conduction band increases [Figure 1.7(b)]. The Fermi level remains at the center of the forbidden bandgap.

EXAMPLE 1.1 *Find the electrical conductivity and resistivity of copper, given that density of copper is 8.96 g/cm³, atomic weight is 63.546 and mobility of electron in copper is 43 cm²/Vs.*

Solution

1. The concentration of atoms in any material is given by

$$\text{Atom concentration} = \frac{6.023 \times 10^{23} \times \text{Density of material}}{\text{Atomic weight}}$$

2. $$\text{Atom concentration in copper} = \frac{6.023 \times 10^{23} \times 8.96}{63.546}$$
$$= 0.849 \times 10^{23}\ \text{atoms/cm}^3$$

3. Since each copper atom contributes one free electron, therefore the concentration of free electrons in copper is 0.849×10^{23}.

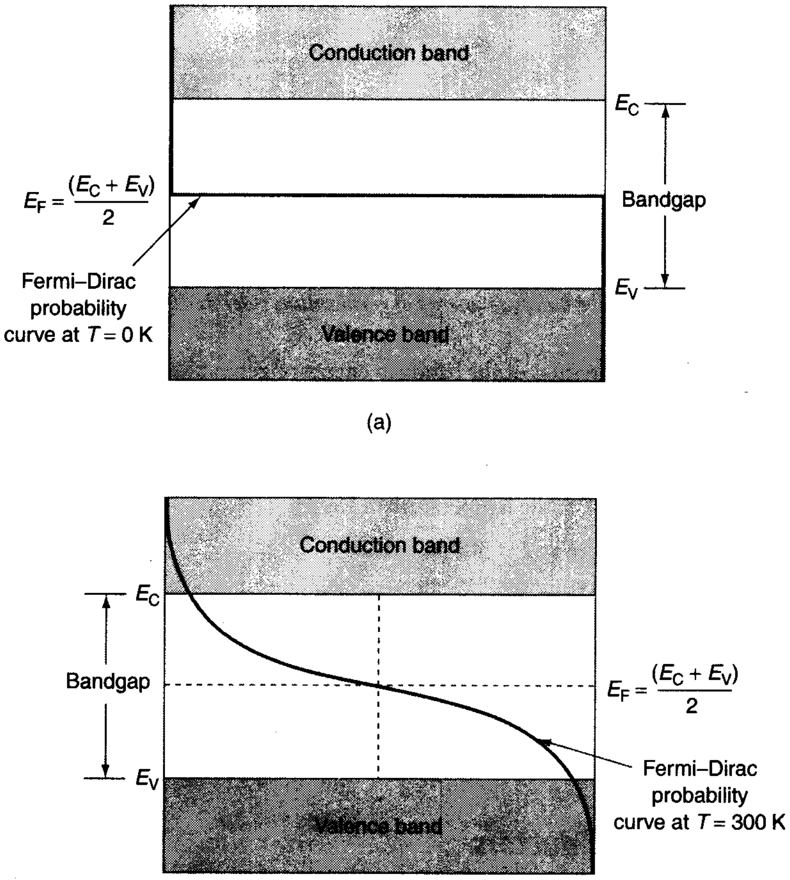
4. In metals, only electrons contribute to the flow of current. Therefore the conductivity of a metal is given by $\sigma = n\mu_n q$.

   So, the conductivity of copper $= 0.849 \times 10^{23} \times 43 \times 1.6 \times 10^{-19}$
   $= 58.4 \times 10^4\ (\Omega\ \text{cm})^{-1}$

5. $\text{Resistivity} = \dfrac{1}{\text{Conductivity}}$

$$\begin{aligned}\text{Resistivity of copper} &= 1/[58.4 \times 10^4\ (\Omega\ \text{cm})^{-1}] \\ &= 0.017 \times 10^{-4}\ \Omega\ \text{cm} \\ &= 17 \times 10^{-9}\ \Omega\ \text{m} \\ &= 17\ \text{n}\Omega\ \text{m}\end{aligned}$$

Answer: The conductivity and resistivity of copper are $58.4 \times 10^4\ (\Omega\ \text{cm})^{-1}$ and $17\ \text{n}\Omega$ m, respectively.



(b)

Figure 1.7 Fermi–Dirac probability function of an intrinsic semiconductor (a) at absolute zero temperature and (b) at 300 K.

EXAMPLE 1.2 *Find the resistivity of intrinsic silicon at 300 K, given that the intrinsic concentration of silicon is 1.5 × 10^{10} atoms/cm^3 and the mobility of electrons and holes is 1300 cm^2/Vs and 500 cm^2/Vs, respectively. Charge of an electron can be assumed to be 1.6 × 10^{-19} C.*

Solution

1. The value of conductivity of an intrinsic semiconductor is given by

$$\sigma = (\mu_\mathrm{n} + \mu_\mathrm{p}) \times n_\mathrm{i} q$$

Therefore, conductivity of intrinsic Si is given by

$$(1300 + 500) \times 1.5 \times 10^{10} \times 1.6 \times 10^{-19} = 4.32 \times 10^{-6}\ (\Omega\ \mathrm{cm})^{-1}$$

2. Resistivity of intrinsic Si is

$$\begin{aligned}\text{Resistivity} &= \frac{1}{\text{Conductivity}} \\ &= \frac{1}{4.32 \times 10^{-6}\ (\Omega\ \mathrm{cm})^{-1}} \\ &= 231.481\ \mathrm{k\Omega\ cm}\end{aligned}$$

Answer: The resistivity is 231.481 kΩ cm.

EXAMPLE 1.3 *Calculate the bandgap energy of germanium (Ge) at 300 K.*

Solution

1. The variation of the bandgap energy of germanium with temperature is given by the relationship

$$E_\mathrm{G}(T) = 0.785 - (2.23 \times 10^{-4} \times T)$$

where T is the temperature in Kelvin.

2. Therefore at $T = 300$ K,

$$\begin{aligned}E_\mathrm{G}(T) &= 0.785 - (2.23 \times 10^{-4} \times 300) \\ &= 0.785 - 0.0669 \\ &= 0.7181\ \mathrm{eV}\end{aligned}$$

Answer: The bandgap energy of germanium at 300 K is 0.7181 eV.

N-Type Extrinsic Semiconductors

Intrinsic semiconductors have very limited applications as they conduct a very small amount of current. However, the electrical characteristics of an intrinsic semiconductor are changed significantly by adding impurity atoms to the pure semiconductor material. The impurities added are of the order of 1 part in 10^5 parts to 1 part in 10^8 parts. However, this small alteration results in a large change in the semiconductor material properties. For example, the conductivity is increased about 1000 times. This process of adding impurities is called doping and the resultant semiconductor is called an extrinsic semiconductor. If the added impurity is a pentavalent atom, then the resultant semiconductor is called an N-type semiconductor; and if the impurity added is trivalent in nature, then it is called a P-type semiconductor. N-type semiconductors are discussed in this sub-section, while the P-type semiconductors are covered in the next sub-section.

An N-type semiconductor material is created by adding approximately 1 part in 10^8 parts of pentavalent impurities to the intrinsic semiconductor material. Pentavalent atoms are those atoms that have five valence electrons. Some examples of pentavalent atoms are phosphorus, antimony, arsenic, etc. Pentavalent impurity atoms are called the donor atoms. Figure 1.8 shows the crystal structure of an N-type semiconductor material where four of the five electrons of the pentavalent impurity atom (antimony) form covalent bonds with four intrinsic semiconductor atoms; the fifth electron is loosely bound to the pentavalent atom, is relatively free to move within the crystal and is referred to as the free electron. The energy required to detach this fifth electron from the atom is very small, of the order of 0.01 eV for germanium and 0.05 eV for silicon.

The effect of doping creates a discrete energy level called donor energy level in the forbidden bandgap with energy level (E_D) slightly less than the conduction band (Figure 1.9). The difference between the energy

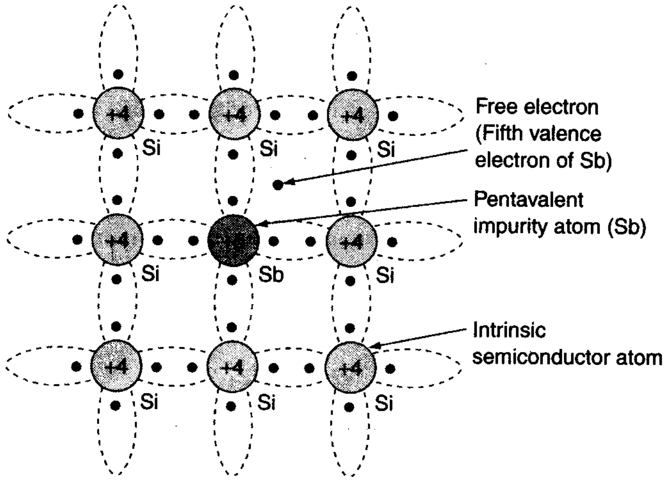


Figure 1.8 | Crystal structure of an N-type semiconductor.

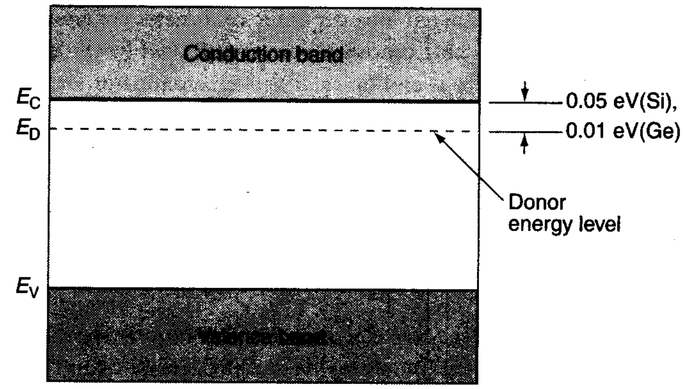


Figure 1.9 | Energy band diagram of an N-type semiconductor.

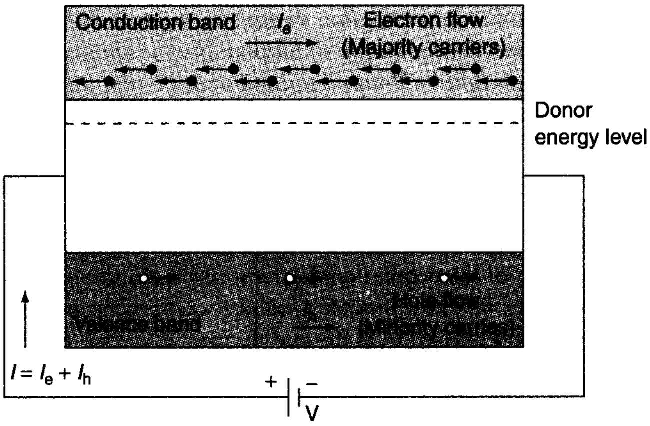


Figure 1.10 Current flow in an N-type semiconductor. Here I is total conventional current flow, I_e the current flow due to electrons and I_h the current flow due to holes.

levels of the conduction band and the donor energy level is the energy required to free the electron (0.01 eV for germanium and 0.05 eV for silicon). At room temperature, almost all the "fifth" electrons from the donor material are raised to the conduction band and hence the number of electrons in the conduction band increases significantly.

In an N-type semiconductor, the number of electrons increases and the number of holes decreases compared to those available in an intrinsic semiconductor. The decrease in the number of holes is attributed to the increase in the rate of recombination of electrons with holes. The current in the N-type semiconductor is dominated by electrons, which are referred to as majority carriers. Holes are the minority carriers in the N-type semiconductor (Figure 1.10).

Electrical Properties

Semiconductor materials are electrically neutral. According to the law of electrical neutrality, in an electrically neutral material the magnitude of positive charge concentration is equal to that of negative charge concentration. Let us consider a semiconductor that has N_D donor atoms per cubic centimeter and N_A acceptor atoms per cubic centimeter, that is, the concentrations of donor and acceptor atoms are N_D and N_A, respectively. Therefore, N_D positively charged ions per cubic centimeter are contributed by the donor atoms and N_A negatively charged ions per cubic centimeter are contributed by the acceptor atoms. Let us assume that the concentration of free electrons and holes in the semiconductor are n and p, respectively. Therefore according to the law of electrical neutrality,

$$N_D + p = N_A + n \tag{1.10}$$

For an N-type semiconductor $N_A = 0$. Also the concentration of free electrons (n) is much greater than the concentration of holes (p). Therefore, for an N-type semiconductor Eq. (1.10) reduces to

$$n \cong N_D \tag{1.11}$$

Hence, for an N-type semiconductor, the free electron concentration is approximately equal to the concentration of donor atoms and vice versa. Therefore, the current density in an N-type semiconductor is given by

$$J \cong N_D \mu_n q \varepsilon \tag{1.12}$$

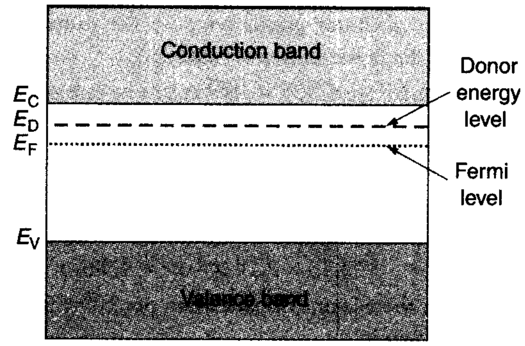


Figure 1.11 | Fermi level in an N-type semiconductor.

where μ_n is the mobility of free electrons in the semiconductor (cm²/Vs) and ε the applied electric field (V/cm). The expression for conductivity in an N-type semiconductor is

$$\sigma \cong N_D \mu_n q \tag{1.13}$$

Fermi Level

The expression for Fermi–Dirac probability function for an extrinsic semiconductor is same as that for the intrinsic semiconductor. The only change that occurs is in the Fermi level. The Fermi level (Figure 1.11) in an N-type semiconductor is raised and is closer to the conduction band as there is a significant increase in the number of electrons in the conduction band and there are fewer holes in the valence band. As the temperature increases, more electron–hole pairs are generated and the Fermi level shifts toward the center of the forbidden energy bandgap.

The Fermi level is given by

$$E_F = E_C - kT \ln\left(\frac{N_C}{N_D}\right) \tag{1.14}$$

where E_C is the energy at the bottom of the conduction band, k the Boltzmann constant in eV/K (8.642 × 10^{-5} eV/K), T the temperature in Kelvin, N_C the density of states in the conduction band, which is constant for a material at a given temperature, and N_D the donor atom concentration (number of atoms/cm³).

The value of N_C is given by

$$N_C = 2\left(\frac{2\pi m_n kTq}{h^2}\right)^{3/2}$$

where m_n is the effective mass of an electron, T the temperature in Kelvin, h the Plank's constant and q the electronic charge (1.6 × 10^{-19} C).

EXAMPLE 1.4 | *A pentavalent impurity is added to an intrinsic silicon semiconductor with 1 part in 10^8 parts. Find the concentration of donor atoms and the resistivity of the semiconductor. Also find the Fermi level of the semiconductor. It is given that atomic weight of silicon is 28.1, density is 2.33 g/cm³, Avogadro's number is 6.023 × 10^{23}, effective mass of an electron is 1.08 × mass of an electron, mass of the electron is 9.11 × 10^{-31} kg, mobility of electron is 1300 cm²/Vs, Boltzmann constant is 8.642 × 10^{-5} eV/K, Plank's constant is 6.626 × 10^{-34} Js and temperature is 300 K.*

Solution

1. Concentration of donor atoms refers to the number of donor atoms per cubic centimeter of the semiconductor material.
2. As we know, there are 6.023×10^{23} atoms in one mole of an element.
3. Atomic weight of silicon is 28.1. Therefore, in 1 g of silicon the number of atoms is
$$6.023 \times 10^{23}/28.1 = 2.14 \times 10^{22}$$
4. Density of silicon is 2.33 g/cm^3. Therefore, the number of atoms of silicon in 1 cm^3 is
$$2.33 \times 2.14 \times 10^{22} = 4.986 \times 10^{22}$$
5. It is given that there is one dopant atom per 10^8 silicon atoms. Therefore, the dopant concentration is
$$4.986 \times 10^{22}/1 \times 10^8 = 4.986 \times 10^{14} \text{ atoms/cm}^3$$
6. The conductivity of an N-type semiconductor is given by $\sigma \cong N_D \mu_n q$. That is conductivity is
$$4.986 \times 10^{14} \times 1300 \times 1.6 \times 10^{-19} = 0.103\ (\Omega\ \text{cm})^{-1}$$
7. Therefore, resistivity is 1/conductivity, that is
$$1/0.103 = 9.642\ \Omega\ \text{cm}$$
8. The expression for Fermi level in an N-type semiconductor is given by
$$E_F = E_C - kT \ln\left(\frac{N_C}{N_D}\right)$$
9. $$N_C = 2\left(\frac{2\pi m_n kTq}{h^2}\right)^{3/2}$$
$$= 2\left(\frac{2 \times 3.14 \times 1.08 \times 9.11 \times 10^{-31} \times 8.642 \times 10^{-5} \times 300 \times 1.6 \times 10^{-19}}{(6.626 \times 10^{-34})^2}\right)^{3/2}$$
$$= 2.8 \times 10^{25} \text{ atoms/m}^3$$
$$= 2.8 \times 10^{19} \text{ atoms/cm}^3$$

   Therefore,
$$E_F = E_C - 8.642 \times 10^{-5} \times 300 \ln\left(\frac{2.8 \times 10^{19}}{4.986 \times 10^{14}}\right)$$
$$= E_C - 0.284 \text{ eV}$$
10. Thus, the Fermi level is 0.284 eV below the conduction band.

Answer: The concentration of dopant atoms is 4.986×10^{14} atoms/cm^3, resistivity is 9.642 Ω cm and the Fermi level is 0.284 eV below the conduction band.

P-Type Extrinsic Semiconductors

A P-type semiconductor is created by adding approximately 1 part in 10^5 parts of trivalent impurity to the intrinsic semiconductor. Trivalent atoms have three electrons in their valence shell and are called acceptor atoms in the context of semiconductor devices. Examples of trivalent impurities include boron (B), indium (In) and gallium (Ga). As there are three electrons in the valence shell of these trivalent impurity atoms, only three covalent bonds can be formed with the neighboring intrinsic semiconductor atoms and a vacancy exists in the fourth bond as shown in Figure 1.12. This vacancy is referred to as the hole and is represented by a small circle. The hole is ready to accept an electron from the neighboring atom, thereby creating a hole in the neighboring atom. This hole in turn is ready to accept an electron, thereby creating another hole. In this way, the hole moves through the crystal.

The effect of doping creates a discrete energy level called acceptor level in the forbidden bandgap with energy level (E_A) just above the valence band (Figure 1.13). The difference between the energy levels of the acceptor band (E_A) and the valence band (E_V) is the energy required by an electron to leave the valence band and occupy the acceptor band, thereby leaving a hole in the valence band. This difference ($E_A - E_V$) is of the order of 0.08 eV for silicon and 0.01 eV for germanium. Since very small energy is required for the electron to leave the valence band and occupy the acceptor energy level, large number of electrons jump to the acceptor energy level resulting in a large number of holes in the valence band.

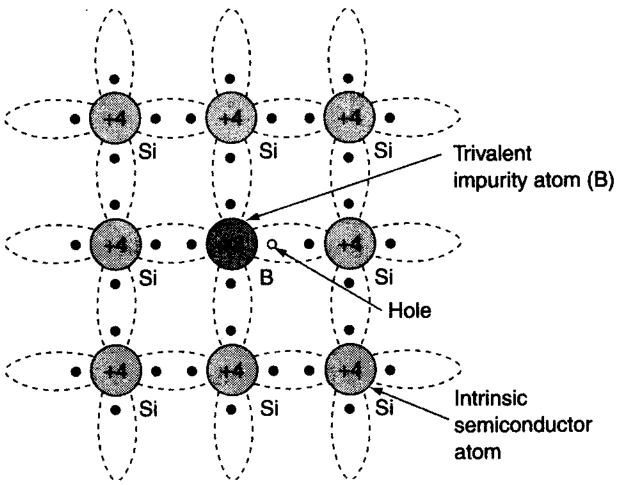


Figure 1.12 | Crystal structure of a P-type semiconductor.

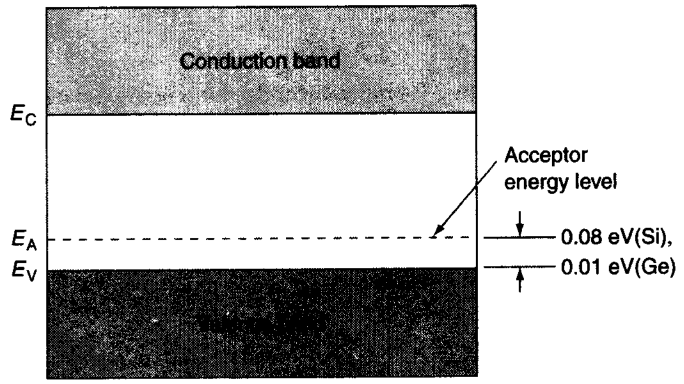


Figure 1.13 | Energy band diagram of a P-type semiconductor.

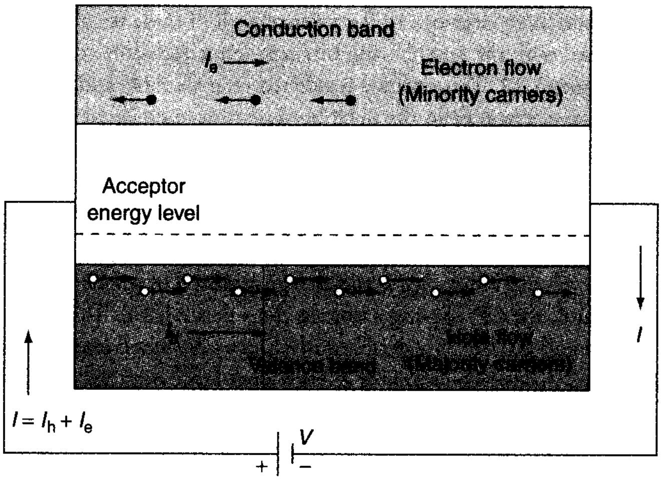


Figure 1.14 Current flow in a P-type semiconductor. Here I is the total conventional current flow, I_e the current flow due to electrons and I_h the current flow due to holes.

In a P-type semiconductor, the number of electrons decreases and the number of holes increases compared to those available in an intrinsic semiconductor. The decrease in the number of electrons is attributed to the increase in the rate of recombination of electrons with holes. The current in the P-type semiconductor is dominated by holes, which are referred to as majority carriers. Electrons are the minority carriers in a P-type semiconductor material (Figure 1.14). It may be mentioned here that the conductivity of an N-type semiconductor is higher than that of a P-type semiconductor as the mobility of electrons is greater than that of holes. For the same level of doping in the N-type and the P-type semiconductors, the conductivity of an N-type semiconductor is around twice that of a P-type semiconductor. Also, it may be noted in practical semiconductors that the concentration of dopants is greater in P-type semiconductors (approximately 1 part in 10^5 parts) than in N-type semiconductors (approximately 1 part in 10^8 parts).

Electrical Properties

For a P-type semiconductor, $N_D = 0$. Also the concentration of free electrons (n) is much less than the concentration of holes (p). Therefore, for a P-type semiconductor, the hole concentration is approximately equal to the acceptor atom concentration, that is,

$$p \cong N_A \tag{1.15}$$

Alternatively, the number of holes in a P-type semiconductor is approximately equal to the number of acceptor atoms. Therefore, the current density in a P-type semiconductor is given by

$$J \cong N_A \mu_p q \varepsilon \tag{1.16}$$

where μ_p is the mobility of holes in the semiconductor (cm²/Vs) and ε the applied electric field (V/cm).

The expression for conductivity is

$$\sigma \cong N_A \mu_p q \tag{1.17}$$

Fermi Level

As mentioned before, the Fermi–Dirac probability function for an extrinsic semiconductor is same as that for an intrinsic semiconductor. The only change that occurs is the change in the Fermi level. The Fermi level

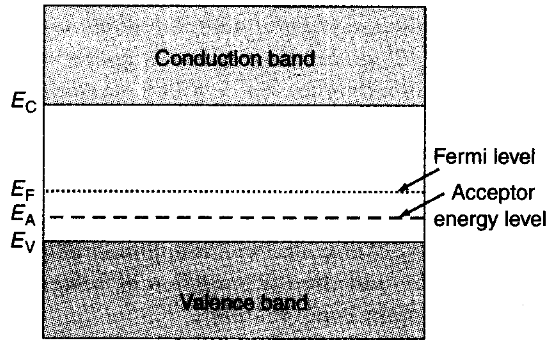


Figure 1.15 Fermi level in a P-type semiconductor.

in a P-type semiconductor (Figure 1.15) is lower than that of an intrinsic semiconductor and is closer to the valence band; this is because there is a significant increase in the number of holes in the valence band and decrease in the number of electrons in the conduction band. As the temperature increases, more electron-hole pairs are generated and the Fermi level shifts toward the center of the energy gap.

The Fermi level can be derived using the expression

$$E_F = E_V + kT \ln\left(\frac{N_V}{N_A}\right) \tag{1.18}$$

where E_V is the energy at the top of the valence band, N_V the density of states in the valence band, which is constant for a material at a given temperature, N_A the acceptor atom concentration (number of atoms/cm^3) and T the temperature in Kelvin.

The value of N_V can be calculated using

$$N_V = 2\left(\frac{2\pi m_p kTq}{h^2}\right)^{3/2}$$

where m_p is the effective mass of a hole, h the Plank's constant, T the temperature in Kelvin, q the electronic charge (1.6×10^{-19} C) and k the Boltzmann constant in eV/K (= 8.642×10^{-5} eV/K).

EXAMPLE 1.5

Find the concentration of holes in a P-type silicon at 300 K if its conductivity is 1 (Ω cm)$^{-1}$, given that the mobility of holes in silicon is 500 cm^2/Vs.

Solution

1. For a P-type semiconductor, the expression for conductivity is $\sigma = N_A \mu_p q$.
2. Therefore, $N_A = 1/(500 \times 1.6 \times 10^{-19})$
   $= 1/(8 \times 10^{-17})$
   $= 1.25 \times 10^{16}$ atoms/cm^3
3. In a P-type semiconductor, hole concentration $\cong N_A$. Therefore, hole concentration = 1.25×10^{16} holes/cm^3.

Answer: The concentration of holes is 1.25×10^{16} holes/cm^3.

EXAMPLE 1.6

A sample of germanium is doped with both donor and acceptor impurities with donor concentration of 10^{14} donor atoms/cm^3 and acceptor concentration of 10^{15} acceptor atoms/cm^3. Find the resistivity of the semiconductor material. Also find the conduction current density for an applied electric field of 1.5 V/cm. It is given that the mobility of holes and electrons in germanium is 1800 cm^2/Vs and 3800 cm^2/Vs, respectively.

Solution

1. The given semiconductor is doped with both donor and acceptor impurities. The concentration of free electrons is approximately equal to the donor impurity concentration and the concentration of holes is approximately equal to the acceptor impurity concentration.
2. Therefore, the concentration of free electrons (n) = 10^{14} and the concentration of holes (p) = 10^{15}. The conductivity is

$$\begin{aligned}\sigma &= (n\mu_\text{n} + p\mu_\text{p}) \times q \\ &= (10^{14} \times 3800 + 10^{15} \times 1800) \times 1.6 \times 10^{-19} \\ &= 0.3488\ (\Omega\ \text{cm})^{-1}\end{aligned}$$

3. Now resistivity is given by

$$\begin{aligned}\text{Resistivity} &= \frac{1}{\text{Conductivity}} \\ &= \frac{1}{0.3488} \\ &= 2.867\ \Omega\ \text{cm}\end{aligned}$$

4. Current density in a semiconductor is given by

$$\begin{aligned}J &= \sigma\varepsilon \\ J &= 0.3488 \times 1.5\ \text{A/cm}^2 \\ &= 0.5232\ \text{A/cm}^2\end{aligned}$$

Answer: The value of resistivity is 2.867 Ω cm and the current density is 0.5232 A/cm^2.

1.3 Law of Mass Action

As discussed in the previous section, in an N-type semiconductor electrons are the majority carriers and holes are the minority carriers. For the P-type semiconductor, holes are the majority carriers and electrons are the minority carriers. The concentration of holes and electrons in a semiconductor is governed by the law of mass action. According to the law of mass action, the product of free electron concentration and hole concentration in any semiconductor is constant and is given by

$$np = n_\text{i}^2 \tag{1.19}$$

where n is the free electron concentration (negatively charged carriers), p the hole concentration (positively charged carriers) and n_i the intrinsic concentration. Therefore, the product of concentration of negative and positive charge carriers in a semiconductor is independent of the type and amount of doping and is equal to the square of the intrinsic concentration. Hence, in an N-type semiconductor as the number of electrons increases the number of holes decreases and in a P-type semiconductor as the number of holes increases the number of electrons decreases.

For an N-type semiconductor,

$$n \cong N_D$$

Therefore,

$$p \cong \frac{n_i^2}{N_D} \tag{1.20}$$

For a P-type semiconductor,

$$p \cong N_A$$

Therefore,

$$n \cong \frac{n_i^2}{N_A} \tag{1.21}$$

EXAMPLE 1.7

For the semiconductor in Example 1.5, find the concentration of electrons. Also, determine the ratio of holes to the free electrons. It is given that the intrinsic concentration of silicon is 1.5 × 10^10^.

Solution

1. As found out in Example 1.5, the concentration of holes is 1.25×10^{16} holes/cm^3.
2. Using the mass action law, the concentration of free electrons is given by

$$\begin{aligned} n &= \frac{n_i^2}{p} \\ &= \frac{(1.5 \times 10^{10})^2}{1.25 \times 10^{16}} \\ &= 18000 \end{aligned}$$

3. The concentration of free electrons is 18000 electrons/cm^3, which is much less than the concentration of holes.
4. $$\begin{aligned} \frac{\text{Number of holes}}{\text{Number of free electrons}} &= \frac{1.25 \times 10^{16}}{18000} \\ &= 6.95 \times 10^{11} \end{aligned}$$

Answer: The number of electrons is 18000 electrons/cm^3 and the ratio of holes to electrons is 6.95×10^{11}.

1.4 Drift and Diffusion Carriers

The current in a semiconductor is the sum of the drift current and the diffusion current as opposed to a conductor where the current flow is only due to the drift phenomenon. As discussed in Section 1.2, drift current is due to the potential gradient in the semiconductor and is given by the expression

$$J = (n\mu_n + p\mu_p)q\varepsilon \tag{1.22}$$

where J is the current density in A/cm^2, n the free electron concentration in the material (number of free electrons/cm^3), p the hole concentration in the material (number of holes/cm^3), μ_n the mobility of an electron in the material (cm^2/Vs), μ_p the mobility of a hole in the material (cm^2/Vs), q the charge of an electron (1.6×10^{-19} C) and ε the applied electric field (V/cm).

The drift current density due to holes is given by

$$J_p = p\mu_p q\varepsilon \tag{1.23}$$

The drift current density due to electrons is given by

$$J_n = n\mu_n q\varepsilon \tag{1.24}$$

Diffusion current is caused by the concentration gradient in the semiconductor, that is, when there is non-uniform concentration of charge particles in a semiconductor.

The hole diffusion current density is given by the expression

$$J_p = -qD_p \frac{dp}{dx} \tag{1.25}$$

where D_p is the diffusion constant of holes in cm^2/s and dp/dx the variation in hole concentration with distance x (it is positive when the hole concentration increases with distance and is negative when the hole concentration decreases with distance).

Similarly, the electron diffusion current density is

$$J_n = qD_n \frac{dn}{dx} \tag{1.26}$$

where D_n is the diffusion constant of electrons in cm^2/s and dn/dx is the variation of electron concentration with distance x (it is positive when the concentration of electrons increases with distance and is negative if the concentration of electrons decreases with distance).

The diffusion constant of a carrier is related to its mobility and is given by

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = V_T \tag{1.27}$$

where V_T is the volt equivalent of temperature, and is equal to kT (k is the Boltzmann constant in eV/K and T is the temperature in Kelvin). The total current is the sum of the diffusion and drift currents. The total hole current density is given by the expression

$$J_p = p\mu_p q\varepsilon - qD_p \frac{dp}{dx} \tag{1.28}$$

The total electron current density is given by the expression

$$J_n = n\mu_n q\varepsilon + qD_n \frac{dn}{dx} \tag{1.29}$$

1.5 Hall Effect

Hall effect is the phenomenon by which a potential difference is created on the opposite sides of a conductor placed in a magnetic field, with the current flowing in perpendicular direction to the magnetic field. The potential created is perpendicular to the direction of both the magnetic field and the current. In other words, if a conductor or a semiconductor carrying current (I) is placed in a transverse magnetic field (B) as shown in Figure 1.16, an electric field (ε) is induced in a direction perpendicular to both B and I. Edwin Hall discovered this effect in the year 1879.

If the current (I) is in the positive X direction and the transverse magnetic field B is in the positive Z direction, a force will be exerted on the current carriers in the negative Y direction. Thus, the carriers will accumulate on the side B as shown in Figure 1.16. For a P-type semiconductor, the holes will accumulate on side B and thus side B will be more positive than side A. Similarly for N-type semiconductors and conductors, electrons will accumulate on side B; and thus side A will be more positive than side B. The magnitude of the voltage will depend on the carrier concentration. Thus, Hall effect can be used to determine the carrier concentration and also whether the semiconductor is a P-type or an N-type semiconductor.

The Hall voltage (V_{H}) is given by

$$V_{\mathrm{H}} = \frac{BIR_{\mathrm{H}}}{d} \tag{1.30}$$

where B is the magnetic field in Tesla, I the current in amperes, R_{H} the Hall coefficient and d the width of the conductor or semiconductor in the direction of the magnetic field in meters.

For conductors the value of Hall coefficient R_{H} is given by

$$R_{\mathrm{H}} = \frac{1}{nq} \tag{1.31}$$

where n is the electron concentration and q the electron charge.

For semiconductors with both positive and negative carriers the value of Hall coefficient R_{H} is given by

$$R_{\mathrm{H}} = \frac{n\mu_{\mathrm{n}}^2 - p\mu_{\mathrm{p}}^2}{q(n\mu_{\mathrm{n}} + p\mu_{\mathrm{p}})^2} \tag{1.32}$$

where μ_{n} is the electron mobility, μ_{p} the hole mobility, n the electron concentration, p the hole concentration and q the electron charge.

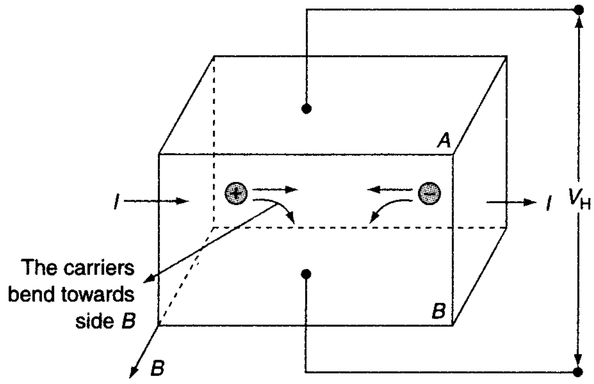


Figure 1.16 | Hall effect.

Hall effect is used in design of instruments such as magnetic field meter, Hall-effect multiplier, etc. Magnetic field meters are used to measure the magnetic field. Hall-effect multipliers give an output proportional to the product of two signals. Here, I and B are made proportional to the two signals.

KEY TERMS

Band structure
Conduction band
Conductor
Diffusion current
Doping
Drift current
Extrinsic semiconductor
Fermi level
Forbidden bandgap
Hall effect
Insulator
Intrinsic semiconductor
Law of mass action
N-type semiconductor
P-type semiconductor
Semiconductor
Valence band

OBJECTIVE-TYPE EXERCISES

Multiple-Choice Questions

1. Doping of semiconductor is
   a. the process of purifying semiconductor materials.
   b. the process of adding certain impurities to the semiconductor material in controlled amounts.
   c. the process of converting semiconductor material into some form of active device such as FET, UJT, etc.
   d. one of the steps used in fabrication of ICs.
2. Referring to energy level diagram of semiconductor materials, the width of forbidden bandgap is about
   a. 10 eV.
   b. 100 eV.
   c. 1 eV.
   d. 0.1 eV.
3. The forbidden bandgap of the semiconductor material
   a. increases with increase in temperature.
   b. decreases with increase in temperature.
   c. does not vary with temperature.
   d. can increase or decrease with increase in temperature depending upon the semiconductor material.
4. One of the following is not a semiconductor.
   a. Gallium arsenide
   b. Indium
   c. Germanium
   d. Silicon
5. One of the following statements justifies the extensive use of semiconductor materials.
   a. It is because of their low forbidden bandgap.
   b. It is because of their resistance value, which lies between that of a good conductor and an insulator.
   c. It is because of ease of fabrication of semiconductor material into practical active and passive devices.
   d. It is because of the fact they exhibit some wide ranging characteristics when certain specified impurities are added to them in controlled amounts.
6. Which of the following statements is false?
   a. The resistivity of the semiconductor is of the order of $10^{-3}\ \Omega$ cm.
   b. Silicon and germanium are semiconductors.
   c. Indium is an acceptor impurity.
   d. Arsenic is a donor impurity.
7. The Fermi level of an intrinsic semiconductor is
   a. in the center of the forbidden bandgap.
   b. in the valence band.
   c. in the conduction band.

   d. anywhere in the valence, conduction, forbidden bandgap.

8. According to the law of mass action:
   a. The product of free electron concentration and hole concentration in an extrinsic semiconductor is equal to the intrinsic concentration in an intrinsic semiconductor.
   b. The product of free electron concentration and hole concentration in an extrinsic semiconductor is equal to the square of the intrinsic concentration in an intrinsic semiconductor.
   c. The product of free electron concentration and hole concentration in an extrinsic semiconductor is equal to the square root of the intrinsic concentration in an intrinsic semiconductor.
   d. None of these.
9. Which of the following statements is true?
   a. An N-type semiconductor has excess of electrons and hence has a net negative charge.
   b. A P-type semiconductor has excess of holes and hence has a net positive charge.
   c. An N-type semiconductor has excess of electrons and a P-type semiconductor has excess of holes but both of them are neutral.
   d. None of these.
10. According to Hall effect, the Hall voltage is proportional to
    a. the product of *B* and *I*.
    b. inverse of the product of *B* and *I*.
    c. *I* only.
    d. *B* only.

    (where *B* is the magnetic field and *I* the current.)

State whether True or False

1. Conductivity of silicon is less than that of germanium at room temperature (300 K).
2. As the temperature increases, the Fermi level of both N-type and P-type semiconductor materials moves toward the center of the forbidden bandgap.
3. For the same level of doping, the conductivity of an N-type semiconductor is same as that of a P-type semiconductor.
4. Under thermal equilibrium conditions, the product of concentration of free electrons and concentration of holes is constant, and is independent of the amount of doping by donor and acceptor impurities.
5. Ratio of majority to minority carriers in an intrinsic semiconductor is very large.

REVIEW QUESTIONS

1. What is a semiconductor material? How does it differ from a conductor and an insulator?
2. Differentiate between
   a. N-type and P-type semiconductors
   b. Extrinsic and intrinsic semiconductors
   c. Drift and diffusion currents
   d. Insulators, conductors and semiconductors
3. Give reasons for the following:
   a. Why the conductivity of intrinsic semiconductors increases with increase in temperature?
   b. Extrinsic semiconductors are used invariably in all applications in contrast to intrinsic semiconductors.
   c. Why are direct bandgap semiconductors used in making light-emitting diodes (LED) and laser diodes?
   d. The Fermi level of an N-type semiconductor is near the conduction band whereas that of a P-type semiconductor is near the valence band?
4. What is Fermi level? Explain how does the Fermi level of a semiconductor change with doping?
5. "*An N-type semiconductor has excess of electrons whereas a P-type semiconductor has excess of holes.*" Comment.

6. *"An N-type semiconductor has excess of electrons and a P-type semiconductor has excess of holes."* Does that mean N-type materials have a net negative charge and P-type materials a net positive charge? Comment.
7. What is the law of mass action and what does it imply?
8. Explain the following terms:
    a. Majority carriers
    b. Minority carriers
    c. Drift current
    d. Diffusion current
9. What is Hall effect? State any two applications.
10. *"An intrinsic semiconductor behaves like an insulator at absolute zero temperature."* Comment.

PROBLEMS

1. Find the conductivity and resistivity of the silver metal. (Given that density of silver is 19.3 g/cm^3, atomic weight is 196.96, mobility of electron in silver is 47 cm^2/Vs.)
2. Find the energy bandgap of silicon at 1000 K.
3. An N-type germanium semiconductor has donor concentration of 10^{14} atoms/cm^3. Find the conductivity of the material. What is the concentration of acceptor impurities required to achieve the same conductivity level in a P-type semiconductor? (Given that the mobility of holes and electrons in germanium is 1800 cm^2/Vs and 3800 cm^2/Vs, respectively.)
4. Find the Fermi energy level of an N-type silicon semiconductor having donor concentration of 10^{14} atoms/cm^3. How would the Fermi level change if the donor concentration were changed to 10^{16} atoms/cm^3? (Given that the effective mass of an electron is 1.08 × mass of an electron, mass of the electron is 9.11×10^{-31} kg, Boltzmann constant is 8.642×10^{-5} eV/K, Plank's constant is 6.626×10^{-34} Js and temperature is 300 K.)
5. By what amount would the Fermi level for the N-type silicon semiconductor having donor concentration of 10^{14} atoms/cm^3 shift for a temperature change from 300 K to 500 K. (Given that the effective mass of an electron is 9.84×10^{-31} kg, Boltzmann constant is 8.642×10^{-5} eV/K, Plank's constant is 6.626×10^{-34} Js.)

ANSWERS

Multiple-Choice Questions

1. (b)
2. (c)
3. (b)
4. (b)
5. (d)
6. (a)
7. (a)
8. (b)
9. (c)
10. (a)

State whether True or False

1. True
2. True
3. False
4. True
5. False

Problems

1. 44.38×10^4 (Ω cm)$^{-1}$, 22.53 nΩ m
2. 0.85 eV
3. 0.0608 (Ω cm)$^{-1}$, 2.11×10^{14} atoms/cm^3
4. 0.325 eV below the conduction band, 0.205 eV below the conduction band
5. 0.25 eV toward the center of the forbidden energy bandgap

CHAPTER 2

Semiconductor Diodes

Learning Objectives

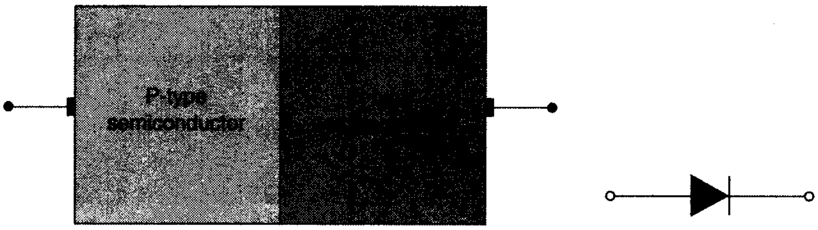
After completing this chapter, you will learn the following:

- Basics of diode construction and operation.
- Diode V–I characteristics.
- Diode parameters and their significance.
- Different diode packages and lead identification.
- How to test a diode.
- Different types of diodes.
- Connecting diodes in series and parallel.

Diodes are the simplest of all the semiconductor devices. They are used in a variety of applications including communication systems, radio, TV, computers, power supplies and so on. The focus in this chapter is on semiconductor diodes. The topics covered include fundamental topics such as diode construction and operation, characteristic curves, diode parameters and diode equivalent circuits. A brief description of different types of diodes – including varactor diodes, tunnel diodes, Schottky diodes, power diodes, light-emitting diodes (LEDs) and photodiodes – and their working principle is given. Also, topics of practical interest such as diode packages and lead identification, connection of diodes in series and parallel and diode testing are covered. The concepts are explained with the help of a large number of solved examples.

2.1 P–N Junction

A semiconductor diode is a polarity-sensitive two-terminal device comprising a P–N junction formed between a P-type semiconductor material and an N-type semiconductor material [Figure 2.1(a)]. As discussed in Chapter 1, the N-type semiconductor is formed by introducing pentavalent dopant impurity atoms while the P-type semiconductor is formed by introducing trivalent dopant impurity atoms into the intrinsic semiconductor material. Also, in an N-type semiconductor, electrons are the majority carriers and holes are the minority carriers, whereas in a P-type semiconductor, holes are the majority carriers and electrons are the minority carriers. The P–N junction is formed by introducing the donor impurities on one side and acceptor impurities on the other side of a single crystal of a semiconductor. Figure 2.1(b) shows the circuit symbol of a P–N junction diode. The arrow is associated with the P-region and the vertical line with the N-region. The P- and N-regions are referred to as the anode and the cathode, respectively. Silicon and germanium are the most commonly used materials for fabricating semiconductor diodes.



(a) (b)

Figure 2.1 (a) P–N junction; (b) symbol of a P–N junction diode.

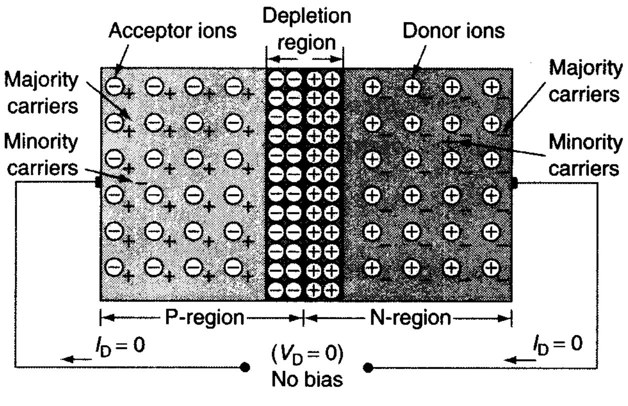
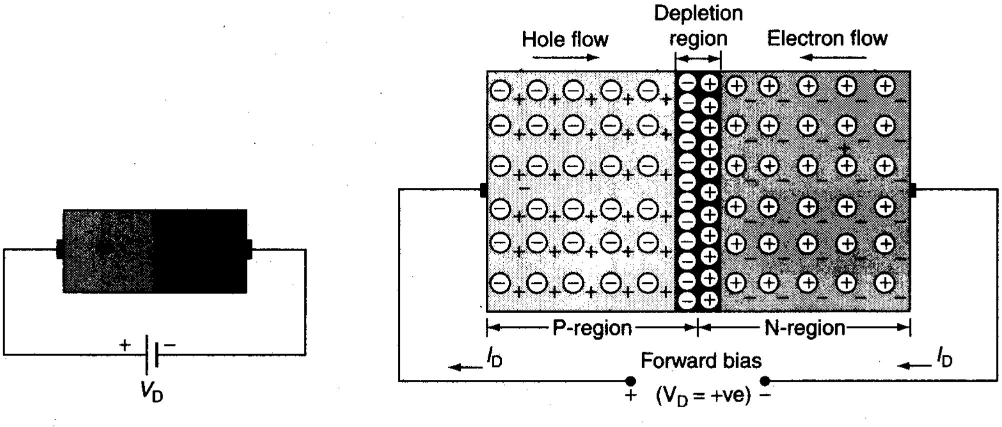


Figure 2.2 P–N junction with no applied bias.

The electrons in the N-region and the holes in the P-region combine near the junction, resulting in a region near the junction that is devoid of free electrons and holes. This region of uncovered positive and negative ions is called the *depletion region* because of the depletion of free carriers in this region. The thickness of this region is of the order of 0.5 μm.

Electrons in the N-region (majority carriers) and negatively charged ions in the P-region repel each other near the junction. Similarly, holes in the P-region (majority carriers) and positively charged ions in the N-region also repel each other near the junction. An effective potential of the order of few tenths of a volt, referred to as the *contact potential* or the *barrier potential*, is developed across the depletion region. However, some of the holes and electrons have sufficient kinetic energy to overcome the contact potential and be able to pass through the depletion region. This results in a flow of electrons from the N-region to the P-region and flow of holes from the P-region to the N-region. This constitutes the majority carrier flow vector.

Also, holes that are present in the depletion region of the N-region (minority carriers) will pass to the P-region. Similarly, electrons that are present in the depletion region of the P-region (minority carriers) will pass to the N-region. This constitutes the minority carrier flow vector. The relative magnitudes of the minority and the majority flow vectors are such that the net flow in either direction is zero. This is referred to as the open-circuit condition of the semiconductor diode where no bias voltage is applied to the diode. In other words, in the absence of an applied bias voltage, the net flow of current in a semiconductor diode is zero. Figure 2.2 shows the P–N junction with no applied bias.



(a)

(b)

Figure 2.3 Forward-biased P–N junction.

In the subsequent paragraphs, we will discuss the response of the semiconductor diode under forward-bias and reverse-bias conditions.

Forward-Bias Condition

A semiconductor diode is forward-biased by applying a positive potential to the P-region and a negative potential to the N-region as shown in Figure 2.3(a). This applied potential causes the electrons in the N-region and the holes in the P-region to combine with positive and negative ions, respectively, in the depletion region. This results in a reduction of the width of the depletion region [Figure 2.3(b)] and a decrease in the potential barrier at the junction. As the applied bias is increased in magnitude, the width of the depletion region decreases until a point is reached where there is a sharp rise in the number of majority carriers crossing the junction. In other words, a large number of holes cross the junction from the P-region to the N-region and a large number of electrons cross the junction in the reverse direction, that is, from the N-region to the P-region. It may be mentioned here that holes traveling from left to right constitute a current in the same direction as the electrons traveling from right to left. This results in exponential rise in the current due to the majority carriers. The current due to the majority carriers is referred to as the *forward current* and is in the range of few tens of milliamperes (except for power diodes where the current is of the order of few amperes). Typically, the voltage across the forward-biased diode is less than 1 V and depends upon the diode material. As an example, the forward voltage for silicon and germanium diodes is typically 0.7 V and 0.3 V, respectively.

The flow of the minority carriers remains the same as in the case of diode with no applied bias. The current contributed by the minority carriers is referred to as the *reverse saturation current* or *reverse leakage current* and is of the order of a few nanoamperes to a few microamperes. The reverse saturation current is in the opposite direction to the forward current. However, its magnitude is negligible as compared to the forward current. V–I characteristics of the diode are discussed in detail in Section 2.4.

Reverse-Bias Condition

A diode is said to be reverse-biased when an external potential applied across it is such that the positive terminal is connected to the N-region and the negative terminal is connected to the P-region [Figure 2.4(a)].

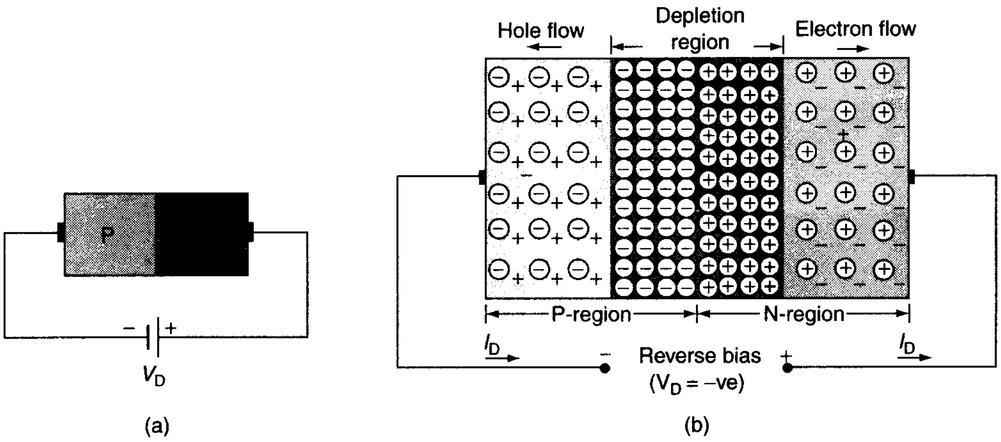


Figure 2.4 | Reverse-biased P–N junction.

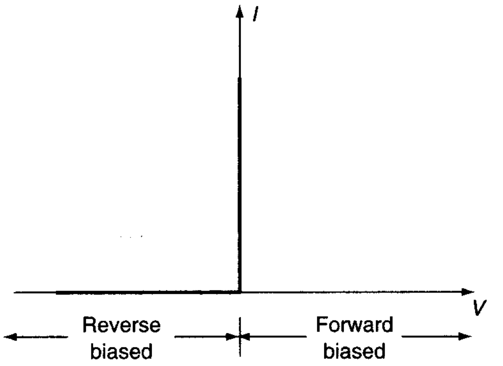


Figure 2.5 | V–I characteristics of an ideal diode.

This results in widening of the depletion region as electrons and holes are drawn away from the junction due to the polarity of the applied voltage [Figure 2.4(b)]. Widening of the depletion region reduces the flow of majority carriers to approximately zero.

The minority carrier flow remains the same as in case of diode with no applied bias. As mentioned before, this current is referred to as the reverse saturation current and is of the order of a few nanoamperes to a few microamperes. The reverse saturation current does not significantly change with change in the reverse-bias potential. However, it is a strong function of the diode temperature and increases with increase in diode temperature. When the applied reverse-bias is increased beyond the breakdown voltage of the diode, there is a sharp increase in the reverse current. This is discussed in detail in Section 2.4.

2.2 Ideal Diode

An ideal diode behaves like a switch that conducts current only in one direction, from anode to cathode. An ideal diode acts as a short circuit when forward-biased and as an open circuit when reverse-biased. Thus, the resistance of the forward-biased diode is zero and the resistance of the reverse-biased diode is infinite. Figure 2.5 shows the V–I characteristics of an ideal diode.

2.3 Practical Diode

The actual diode differs from the ideal diode described in Section 2.2. In the forward-bias condition, the ideal diode acts as a closed switch, with zero ON-resistance that allows the current to flow in one direction, that is, from anode to cathode. However, practical diodes do not conduct until a certain value of forward voltage is applied to them. This voltage, referred to as the ***cut-in voltage*** or the ***knee voltage*** or the ***threshold voltage***, is of the order of less than 1 V for semiconductor diodes. Also, the ON-resistance of the practical diode is not zero and varies from few ohms to few hundreds of ohms. In the reverse-bias state, the practical diode differs from the ideal open switch as in this condition a small amount of current, referred to as the reverse saturation current, flows through the diode. Also, there is sharp increase in the reverse current when the applied reverse-bias voltage exceeds the reverse breakdown voltage of the diode.

2.4 V–I Characteristics of a Diode

The V–I characteristics of a semiconductor diode both in the forward-bias and reverse-bias conditions are expressed by the universal diode equation also referred to as the ***Shockley's diode equation*** [Eq. (2.1)]:

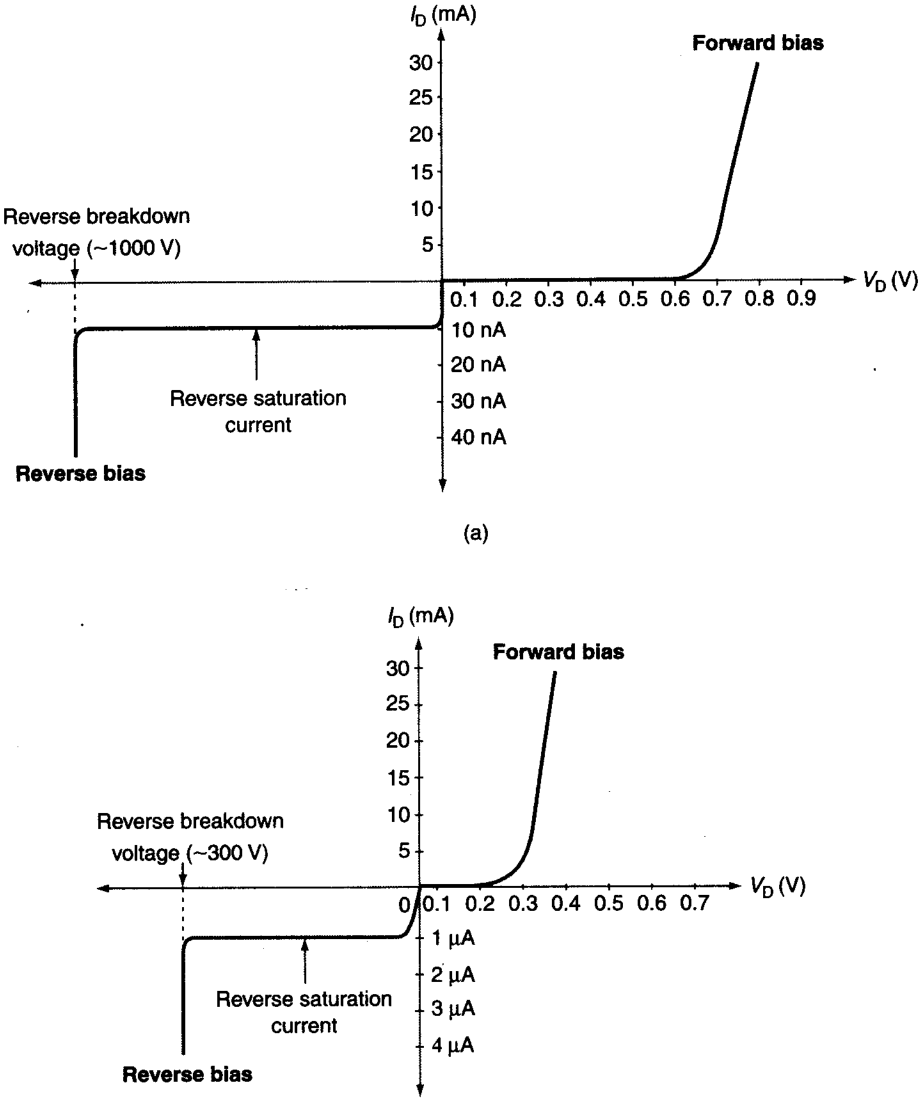
$$I_D = I_0(e^{V_D/\eta V_T} - 1) \tag{2.1}$$

where V_D is the voltage across the diode (in V); I_D the diode current (in mA); I_0 the reverse saturation current (in mA); $\eta = 1$ for germanium and silicon (for relatively higher values of diode current) and $\eta = 2$ for silicon at relatively low levels of diode current, that is below the cut-in-voltage or the knee-point of the diode characteristics; V_T the volt equivalent of temperature (in V).

It may be mentioned here that, the value of $V_T = kT/q$, where k is the Boltzmann constant (8.642×10^{-5} eV/K); q the electron charge (1.6×10^{-19} C); T the temperature (in K); V_T the volt equivalent of temperature (in V). Also, diode voltage (V_D) and diode current (I_D) are positive when the diode is forward-biased and negative when the diode is reverse-biased. The V–I characteristics of a silicon P–N junction diode are shown in Figure 2.6(a) and that of a germanium P–N junction diode in Figure 2.6(b). As is evident from the figures, when the diode is forward-biased there is a minimum voltage that must be exceeded before there is sufficient conduction of current through the diode. In other words, current flows through the diode when it is forward-biased, with the applied voltage greater than the cut-in voltage (V_γ) of the diode. The cut-in voltage is 0.7 V in the case of silicon diodes and 0.3 V in the case of germanium diodes.

When the applied forward voltage exceeds the cut-in voltage, there is a sharp rise in the current through the diode. In other words, a very small increment in the forward voltage (V_D) results in a very large increase in the forward current (I_D). For positive values of V_D, we can see from Eq. (2.1) that the first term of the equation will grow exponentially and overpower the effect of the second term. The first term corresponds to the forward current through the diode and the second term corresponds to the reverse saturation current. Thus, the current through the diode varies exponentially with the applied voltage, provided that the applied voltage is greater than the cut-in voltage. The forward current is measured in milliamperes and is generally in the range of few tens of milliamperes.

In the reverse-bias mode, the small current that flows is the reverse saturation current. It is of the order of few nanoamperes for silicon diodes and typically 1 μA for germanium diodes. This current is independent of the applied reverse voltage until the semiconductor junction breaks down at a voltage known as the ***reverse breakdown voltage*** or the ***peak inverse voltage***. The breakdown of the junction results in a sudden rise of current that ends up in damaging the diode. Hence, when the diodes are operated in the reverse-bias mode, their operating voltage should be less than the breakdown voltage. Some diodes known as breakdown diodes are designed to operate in the breakdown region. Breakdown diodes are discussed in detail in Section 2.11.



(a)

(b)

Figure 2.6 | V–I characteristics of (a) silicon diode and (b) germanium diode.

2.5 Temperature Dependence of the V–I Characteristics

Temperature has a significant effect on the V–I characteristics of the diode. Figure 2.7 shows the variation in the diode characteristic curve with change in temperature. As is evident from the figure, the reverse saturation current, reverse breakdown voltage, cut-in voltage and the diode's forward voltage are strong functions of the diode temperature.

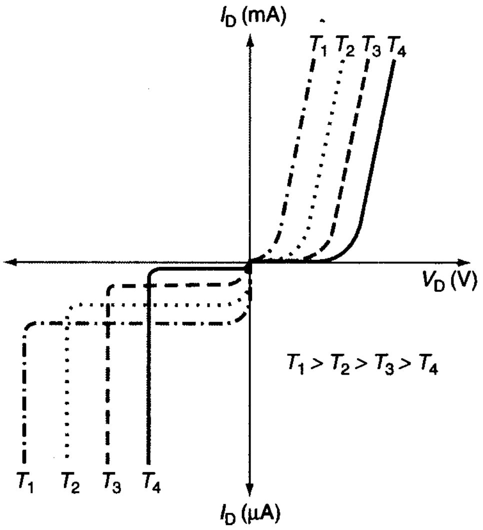


Figure 2.7 Temperature dependence of the diode V–I characteristics.

As an approximation it can be said that reverse saturation current doubles itself for every 10°C rise in diode temperature. As an example, the reverse saturation current of the germanium diode is of the order of 1 μA at 25°C and increases to around 100 μA at 100°C. The variation of the reverse saturation current with temperature is given by

$$I_0(T) = I_0(T_1) \times 2^{(T-T_1)/10} \tag{2.2}$$

where $I_0(T)$ is the reverse saturation current at temperature T and $I_0(T_1)$ the reverse saturation current at temperature T_1.

The reverse breakdown voltage of the diode increases with increase in temperature. Also, the cut-in-voltage (V_γ) and the forward voltage across the diode for a given current decrease with increase in temperature. The variation of cut-in voltage and the forward voltage with temperature is given by

$$\frac{dV}{dT} = -2.5\ \text{m V/°C} \tag{2.3}$$

2.6 Diode Specifications

Diodes have a wide range of performance specifications. These specifications are the basis of the selection criteria when it comes to choosing the right diode for a given application. Some of the major performance specifications for a diode are as follows:

1. ***Forward Voltage* (V_F):** It is the voltage applied across a forward-biased diode. It is not a specification in itself. It is given along with the corresponding forward current value at which it has been measured. It indicates the diode's static resistance. For example, general purpose diode 1N3611 is specified as 1.1 V@1000 mA, which also indicates that its static resistance is 1.1 Ω.
2. ***Forward Current* (I_F):** It is the direct current flowing through the diode when it is forward-biased.
3. ***Reverse Voltage* (V_R):** It is the voltage across a reverse-biased diode. It is specified along with corresponding reverse current value at which it has been measured. It also indicates the diode's reverse-biased resistance.

4. ***Reverse Current* (I_R):** It is the direct current flowing through a reverse-biased diode.
5. ***Reverse Breakdown Voltage or the Peak Inverse Voltage* (V_{BR}, *PIV*):** It is the maximum reverse voltage that a diode can withstand without breaking down. There are usually two different PIV ratings specified in case of diodes. One is the *repetitive peak inverse voltage* (V_R or V_{RRM}) and the other is the *non-repetitive peak inverse voltage* (V_{RSM}). The non-repetitive rating is obviously greater than the repetitive rating. The one that needs to be considered depends upon the intended application of the diode. As an example, in a rectifier application, it is the repetitive peak inverse voltage rating that is to be considered.
6. ***Power Dissipation* (P_D):** The power dissipated in a diode for a given value of diode voltage (V_D) and current (I_D) is given by

$$P_D = V_D \times I_D \tag{2.4}$$

7. ***Maximum Power Dissipation Rating* [$P_{D(max)}$]:** The maximum power that can be safely dissipated in a diode is referred to as the *maximum power dissipation rating* [$P_{D(max)}$]. The value of maximum power dissipation is specified at 25°C. At higher operating temperatures, its value should be derated as per the power–temperature derating curve of the diode. The maximum power rating decreases linearly with the increase in temperature. Figure 2.8 shows the typical power derating curve for a diode.
8. ***Maximum Junction Temperature* (T_j):** It is the maximum allowable junction temperature of the diode. It is significant in the case of power diodes and helps in finding the size of the heat sink to be used for a given diode current.
9. ***Maximum Average Rectified Current* ($I_{F(av)}$):** It is the maximum average forward rectified output current that can be allowed to pass through the diode.
10. ***Peak Repetitive Forward Current*:** This is the maximum instantaneous value of the repetitive forward current.
11. ***Peak Forward Surge Current*:** During turn-on, malfunction, switching, etc., high values of current may flow through the diode for brief time intervals. Surge current ratings define the maximum value and time duration of such surges in the current level. For instance, a surge rating of 10 A for 10 ms implies that the diode can handle a maximum of 10 A of forward current for time duration not exceeding 10 ms. The surge current rating is significantly higher than the peak forward current rating.
12. ***Ampere Square Seconds* (I^2t):** It indicates the sub-cycle current capability of diode when used as a rectifier. It is usually specified for one complete cycle of a 50 Hz operation.

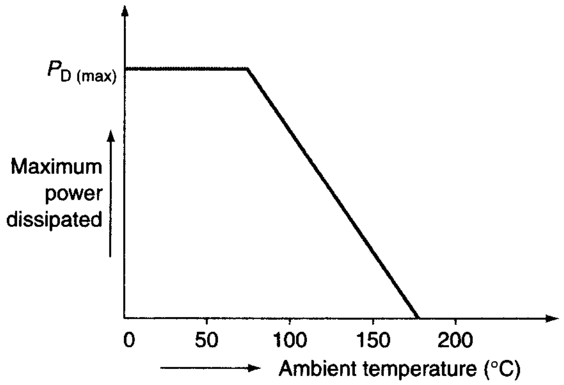


Figure 2.8 Typical power derating curve of a diode.

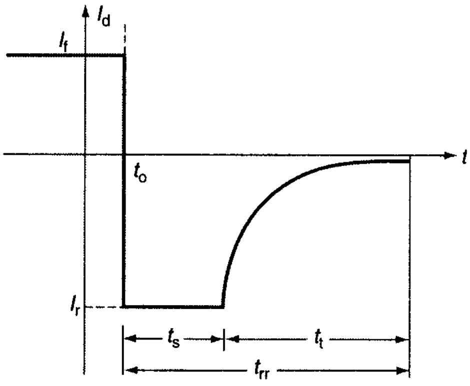


Figure 2.9 Reverse recovery time.

13. ***Reverse Recovery Time* (t_{rr}):** When the diode is switched from the forward-biased condition to the reverse-biased condition abruptly, it is the time required by the reverse current or voltage to reach a specified value. Figure 2.9 shows the current versus time waveform of a diode when the voltage across the diode is abruptly changed to reverse-bias the diode from its forward-bias condition. On application of the reverse voltage, the diode current reverses its direction as shown in the figure and stays at this level for time t_s (storage time). It is the time required for the carriers in the N-region to move to the P-region and the carriers in the P-region to move to the N-region. After this, the current reduces and eventually reaches the reverse saturation value after a certain time called the transition time (t_t). The reverse recovery time is the sum of the storage time and the transition time. The reverse recovery time is a very significant parameter in fast recovery rectifier diodes used in switched mode power supplies and in diodes used for high-frequency switching applications. The reverse recovery time varies from a few nanoseconds (for ultrafast diodes) to about 500 ns (for a typical fast recovery rectifier).
14. ***Forward Recovery Time* (t_{ff}):** It is the time required for the forward current or voltage to reach a specified value after the diode has been abruptly switched from the reverse-biased state to the forward-biased state. This parameter too is significant in switching applications.
15. ***Diode Resistance*:** The diode offers resistance both in the forward- and the reverse-biased conditions. The diode resistance in the forward-biased region varies from few ohms to few hundreds of ohms, while in the reverse-biased region it is in the range of few to few hundreds of MΩ. Diode resistance is discussed in detail in Section 2.7.
16. ***Diode Capacitance*:** It is the inherent capacitance of the diode junction. There are two types of capacitances present, namely, the *transition capacitance* and the *diffusion capacitance*. In the reverse-bias condition, the transition or the depletion capacitance (C_T) is of importance whereas in the forward-bias region diffusion or storage capacitance (C_D) dominates. Diode capacitance plays a very significant role in the functioning of switching diodes. Smaller value of diode capacitance results in faster switching times. Diode capacitances are discussed in detail in Section 2.8.

2.7 Diode Resistance

As the V–I characteristics of a diode are non-linear, the diode resistance varies with change in the applied voltage. Two terms very commonly used to define the resistance of a diode are the *static resistance* and the *dynamic resistance*.

Static Resistance

The static resistance or the DC resistance (R_S) of the diode is the resistance offered by the diode when a steady DC voltage is applied to the diode. This results in the flow of a steady DC current through the diode. Let us consider that application of voltage (V_{D1}) results in current (I_{D1}) through the diode [Figure 2.10(a)]. Then the static resistance of the diode is given by

$$R_S = \frac{V_{D1}}{I_{D1}} \tag{2.5}$$

The static resistance of the diode when forward-biased will be higher near the knee region or below it as compared to the vertical region of the V–I characteristics. In the reverse-biased state, the value of the static resistance will be very high. Typical values of static resistance for silicon diodes vary from few tens to hundreds of ohms in the forward-biased region and from few mega-ohms to few hundreds of mega-ohms in the reverse-biased region.

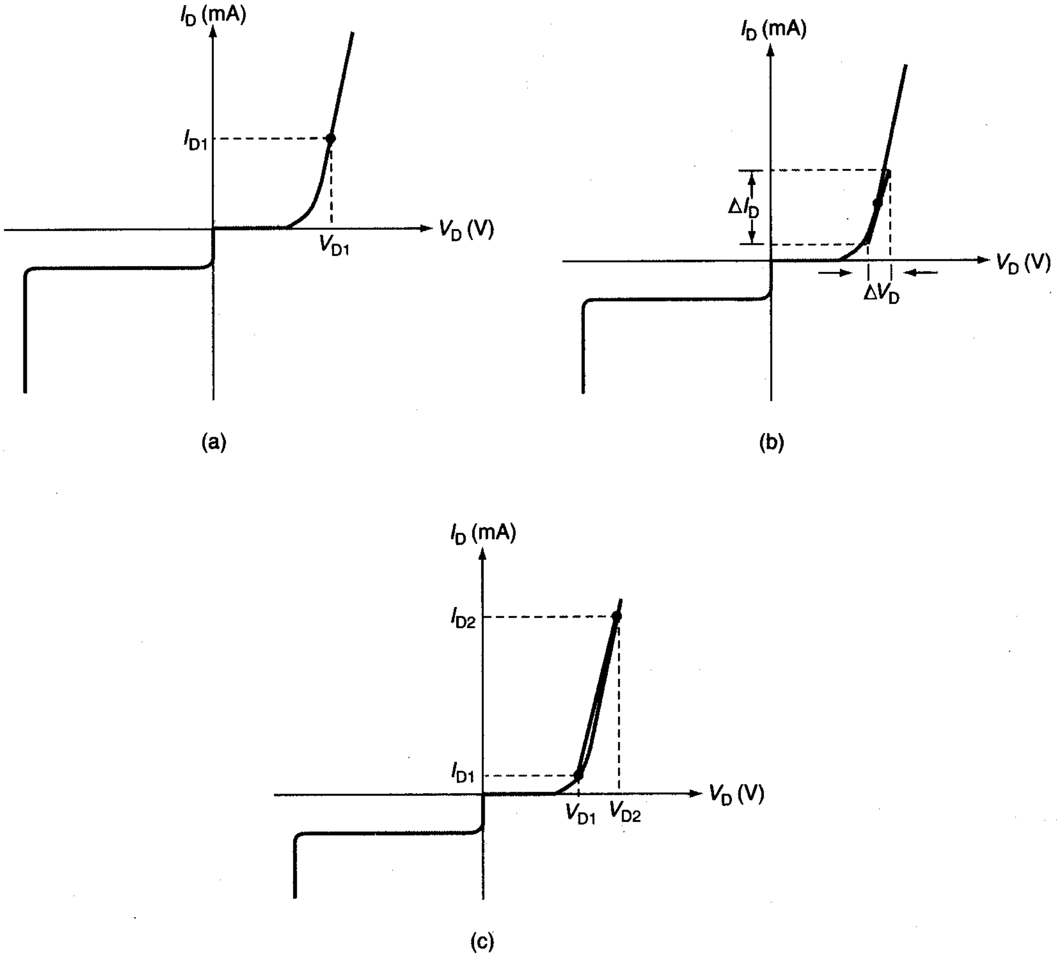


Figure 2.10 (a) Static resistance of a diode; (b) dynamic resistance of a diode; (c) average AC resistance of a diode.

Dynamic Resistance

Dynamic resistance or the AC resistance of a diode is defined as the resistance offered by the diode to a time-varying input signal. The dynamic resistance of the diode having the V–I characteristics shown in Figure 2.10(b) is given by Eq. (2.6). In other words, the dynamic resistance at a particular point in the operating region of the diode is defined by the slope of the tangent drawn at that point.

$$r = \frac{\Delta V_\mathrm{D}}{\Delta I_\mathrm{D}} \tag{2.6}$$

By taking the derivative of universal diode equation given in Eq. (2.1) with respect to applied forward voltage and reversing the result we get the expression for the dynamic resistance of the diode. The derivation is as follows.

Universal diode equation is $I_\mathrm{D} = I_0(e^{V_\mathrm{D}/\eta V_\mathrm{T}} - 1)$. Substituting $V_\mathrm{T} = kT/q$ in the above equation and taking derivative of the equation wrt the forward voltage (V_D) we get

$$\frac{\mathrm{d}I_\mathrm{D}}{\mathrm{d}V_\mathrm{D}} = \frac{qI_0 e^{qV_\mathrm{D}/\eta kT}}{\eta kT}$$

Substituting $I_0 e^{qV_\mathrm{D}/\eta kT} = I_\mathrm{D} + I_0$ in the above equation we get

$$\frac{\mathrm{d}I_\mathrm{D}}{\mathrm{d}V_\mathrm{D}} = \frac{q(I_\mathrm{D} + I_0)}{\eta kT}$$

Now as $I_\mathrm{D} \gg I_0$,

$$\frac{\mathrm{d}I_\mathrm{D}}{\mathrm{d}V_\mathrm{D}} \cong \frac{qI_\mathrm{D}}{\eta kT}$$

Taking the reciprocal of the above equation, we obtain

$$\frac{\mathrm{d}V_\mathrm{D}}{\mathrm{d}I_\mathrm{D}} \cong \frac{\eta kT}{qI_\mathrm{D}}$$

Substituting the value of $k = 8.642 \times 10^{-5}$ eV/K and $T = 300$ K we get

$$\frac{\mathrm{d}V_\mathrm{D}}{\mathrm{d}I_\mathrm{D}} = r \cong \frac{26\eta}{I_\mathrm{D}}$$

Dynamic resistance (r) of a diode in the forward-biased region is given by

$$r \cong \frac{26\eta}{I_\mathrm{D}} \tag{2.7}$$

where $\eta = 1$ for germanium and silicon (for relatively higher values of diode current); $\eta = 2$ for silicon at relatively low levels of diode current, that is below the cut-in-voltage or the knee-point of the diode characteristics; I_D is the forward diode current (in mA). The dynamic ON resistance for the forward -biased diode is also represented as R_D.

In the reverse-biased region, the value of the dynamic resistance of the diode is given by

$$\text{Dynamic resistance of the diode in the reverse-biased region} \cong \frac{26\eta}{I_0} \tag{2.8}$$

where I_0 is the reverse saturation current. However, the change in the value of reverse saturation current (I_0) is very small with change in the reverse-bias voltage from 0 V to the reverse breakdown voltage resulting in very high value of dynamic resistance. Hence, for all practical purposes the diode can be assumed to be an open circuit in the reverse-bias region. The typical value of dynamic resistance of silicon diodes is of the order of few ohms in the forward-biased region and around few hundreds of mega-ohms in the reverse-biased region.

Average AC Resistance

Another term that is sometimes used to define the resistance of a diode is called the average AC resistance. When a sufficiently large input signal is applied to the diode to produce a broad swing as shown in Figure 2.10(c), the resistance associated with the diode is called the average AC resistance. It is determined by the slope of the straight line formed by joining the two points on the V–I characteristics of the diode corresponding to the maximum and minimum input voltages:

$$\text{Average AC resistance} = \frac{V_{D2} - V_{D1}}{I_{D2} - I_{D1}} \tag{2.9}$$

The static, dynamic and average AC resistances discussed thus far are all contributed by the P–N junction. Other than the junction resistance, the resistance of the semiconductor material (called the body resistance) and the resistance introduced by the connection between the semiconductor material and the external metallic conductor (called the contact resistance) are also present. These resistances together can range from 0.1 Ω to around 2 Ω and in most cases can be ignored.

EXAMPLE 2.1

Refer to Figure 2.11. Determine the static and the dynamic resistances of the diode at points A and B.

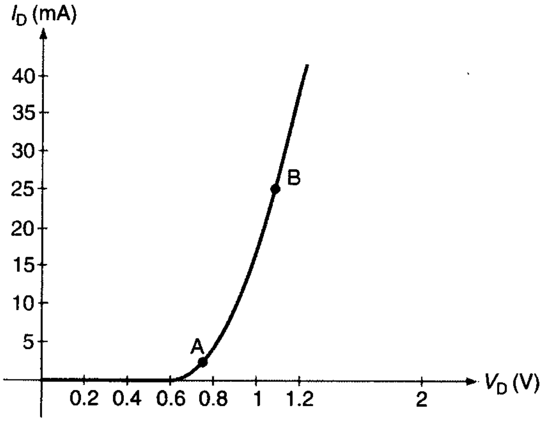
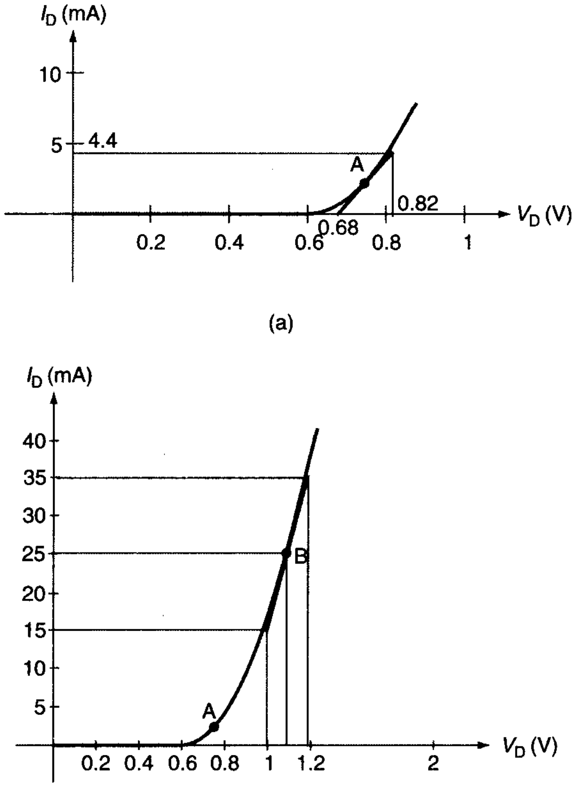


Figure 2.11 | Example 2.1.

Solution

1. Let us first consider point A. The diode current and voltage at point A are 2 mA and 0.75 V, respectively.
2. Static resistance at point A = 0.75/2 × 10^{-3} = 375 Ω.
3. The slope of the tangent line at point A gives the dynamic resistance of the diode at point A. Figure 2.12(a) shows the exploded view of the characteristics near point A.
4. Dynamic resistance = (0.82 – 0.68)/(4.4 × 10^{-3} – 0) = 140/4.4 = 31.8 Ω.
5. Let us now consider point B [Figure 2.12(b)].
6. The diode current and voltage are 25 mA and 1.14 V, respectively.

7. Static resistance of the diode at point B = 1.14/25 × 10^{-3} = 45.6 Ω.
8. The slope of the tangent line at point B gives the dynamic resistance of the diode at point B [Figure 2.12(b)].
9. Dynamic resistance = (1.195 − 1)/(35 − 15) × 10^{-3} = 0.195/20 × 10^{-3} = 9.75 Ω.
10. As is clear from the example, the resistance of a diode in the linear V–I region is much smaller as compared to the resistance near the knee region. Also, the dynamic resistance of a diode is much smaller than the static resistance.



(a)

(b)

Figure 2.12 | Solution to Example 2.1.

2.8 Diode Junction Capacitance

As discussed earlier in Section 2.6 on diode specifications, there are two types of capacitances associated with a junction diode, namely, the ***transition capacitance*** (C_T) and the ***diffusion capacitance*** (C_D). These capacitances in effect come in parallel with the ideal diode as shown in Figure 2.13(a). For low- and mid-frequency low-power applications, the effect of these capacitances on the diode performance is negligible and hence can be ignored. However, in high-frequency and high-power applications, the effect of these capacitances have to be taken into consideration. We will discuss these two diode capacitances in detail in the following sub-sections.

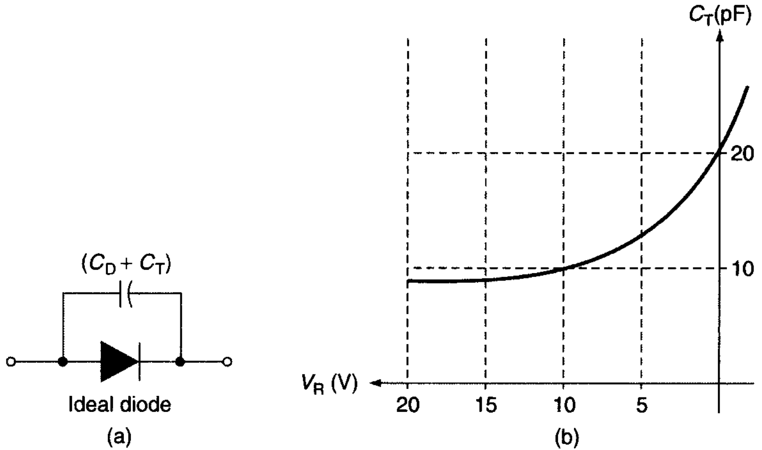


Figure 2.13 Diode capacitance.

Transition Capacitance

The P–N junction acts as a parallel plate capacitor with the P- and the N-regions as the parallel plates and the depletion region as the insulator or the dielectric. As we can recall, the capacitance of a parallel plate capacitor is given by the formula $C_P = (\varepsilon A / d)$, where ε is the permittivity of the dielectric used, A is the area of the plates and d is the separation between the plates. With no applied bias, the width of the depletion region is around 0.5 μm and the associated capacitance is of the order of 20 pF. In the forward-biased state, the width of the depletion region decreases and hence the capacitance increases. In the reverse-biased condition, the depletion region widens with the applied reverse voltage so the corresponding capacitance reduces with increase in applied reverse bias. This capacitance is referred to as the transition capacitance or the space charge capacitance.

Figure 2.13(b) shows the variation of the transition capacitance with the applied reverse voltage. The dependence of the diode capacitance on the applied reverse bias is made use of in a number of electronic devices and systems, for example in variable voltage capacitors known as the varactors. The effect of transition capacitance in the forward-biased state is overshadowed by the presence of diffusion capacitance.

Diffusion Capacitance

In the forward-biased state, the capacitance that is predominant is the diffusion capacitance or the storage capacitance. It is defined by the equation $C_D = dq/dv$, where dq represents the change in the number of minority carriers stored outside the depletion region when a change in voltage (dv) is applied across the diode. In other words, it is dependent on the rate at which the charge is inducted into the P- and the N-regions just outside the depletion region. Its value in the forward-biased region is in the range of 10–20 μF. In the reverse-bias region, its value is much smaller than the transition capacitance and hence the transition capacitance predominates in this region.

Diffusion capacitance affects the switching time of the diode. The switching time constant of the diode is equal to ($r_d \times C_D$), where r_d is the dynamic forward resistance of the diode. The value of switching time constant is very small due to the extremely small value of r_d. Hence, the switching time of the diode is not taken into consideration for normal diode applications and it assumes importance only when the diode is used as a switching device in very high speed applications.

2.9 Diode Equivalent Circuits

An equivalent circuit of a device is a combination of elements suitably connected so as to best represent the actual terminal characteristics of the device. The most accurate equivalent circuit model for a diode is the *piecewise linear equivalent circuit model* in which the diode curves are represented by straight-line segments. The model is shown in Figure 2.14. From the figure it is clear that the assumption has been made that the diode will not conduct until the voltage at the anode exceeds the cathode voltage by the cut-in voltage, which is 0.7 V for silicon diodes and 0.3 V for germanium diodes. Hence, a battery voltage (V_B) has been introduced in the circuit opposite to the conduction direction of the diode. The magnitude of V_B is equal to the cut-in voltage of the diode. When the applied voltage exceeds V_B, the diode starts conducting and the resistance of the diode is expressed as the dynamic-ON resistance (r_d) in the forward-biased condition. A line is drawn on the equivalent model curve with a slope equal to the inverse of the value of the dynamic resistance ($1/r_d$). The ideal diode shown in the circuit is an ideal switch that conducts only in one direction. The piecewise linear equivalent circuit model is the most accurate equivalent model of a diode. However, it does not result in the actual duplication of the diode characteristics, especially in the knee region. Also, the model is equally valid for both DC as well as AC applications.

When the network resistance is much larger than the value of the diode resistance (r_d) then the above model can be simplified as shown in Figure 2.15. Here the diode resistance is assumed to be zero. The model makes an assumption that the diode will not conduct until the cut-in voltage is reached and after that it acts as an ideal closed switch that conducts only in one direction.

Another possible simplification model is shown in Figure 2.16. Here, the curve has been approximated by a straight line through the origin and the slope of the straight line is given by the inverse of the static diode resistance at the point of intersection of the line with the diode V–I characteristics.

Ideal diode is the most simple equivalent diode model. This model is applicable when the applied voltage levels are much larger than the diode's cut-in voltage and the network resistance is of a much larger value

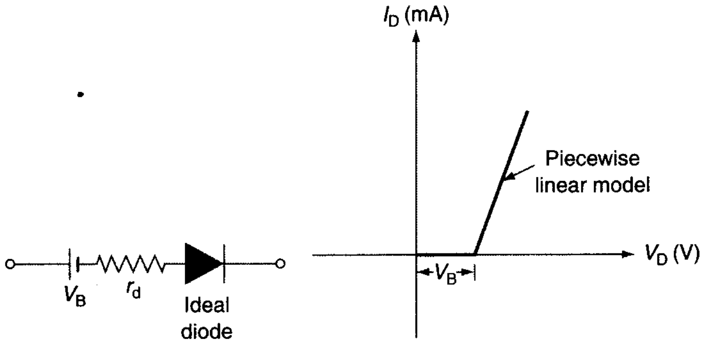


Figure 2.14 | Piecewise linear equivalent model of a diode.

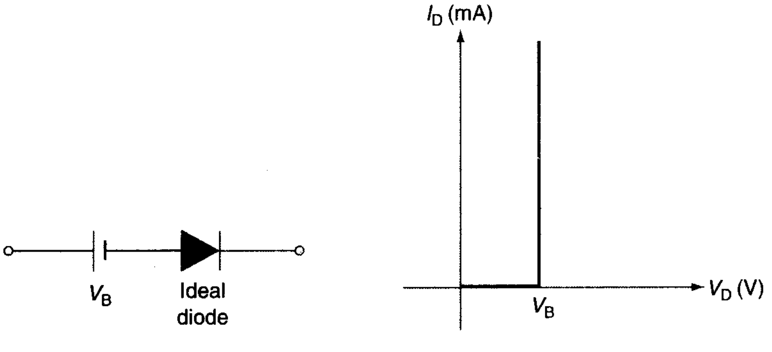


Figure 2.15 | Simplified equivalent diode model.

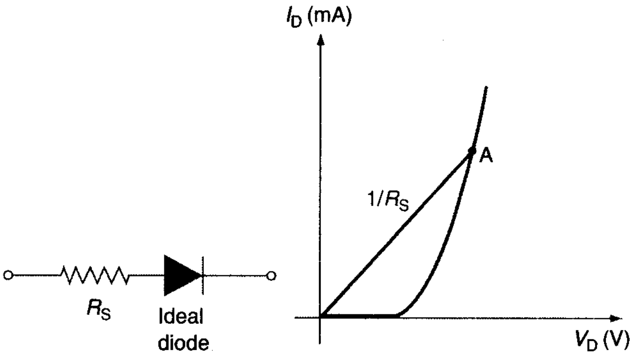


Figure 2.16 Another simplified diode model.

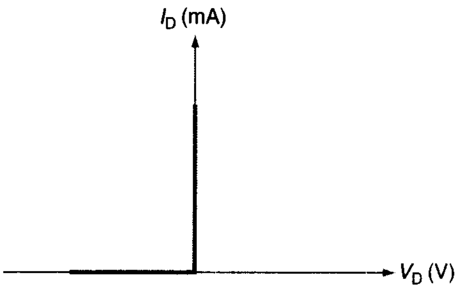
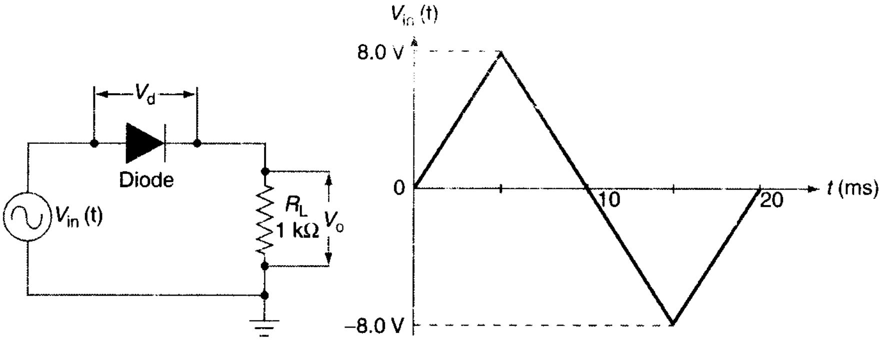


Figure 2.17 Ideal diode model.

than the diode's dynamic-ON resistance. The V–I characteristics of an ideal diode were shown in Figure 2.5. They are reproduced again in Figure 2.17 for reference.

EXAMPLE 2.2

Figure 2.18(a) shows a simple diode circuit. The input waveform applied to the circuit is shown in Figure 2.18(b). Draw the waveforms for the output voltage (V_o) and voltage across the diode (V_d) assuming the diode to be ideal.



(a) (b)

Figure 2.18 Example 2.2.

Solution

1. The ideal diode acts as a short circuit in the forward-biased region and as an open circuit in the reverse-biased region.
2. During the positive half of the input waveform, the diode acts as a short circuit and the whole waveform appears across the load resistance (R_L). Negative half of the input waveform is blocked by the diode and does not appear across R_L.
3. Figure 2.19 shows the output waveform (V_o) and the diode waveform (V_d).

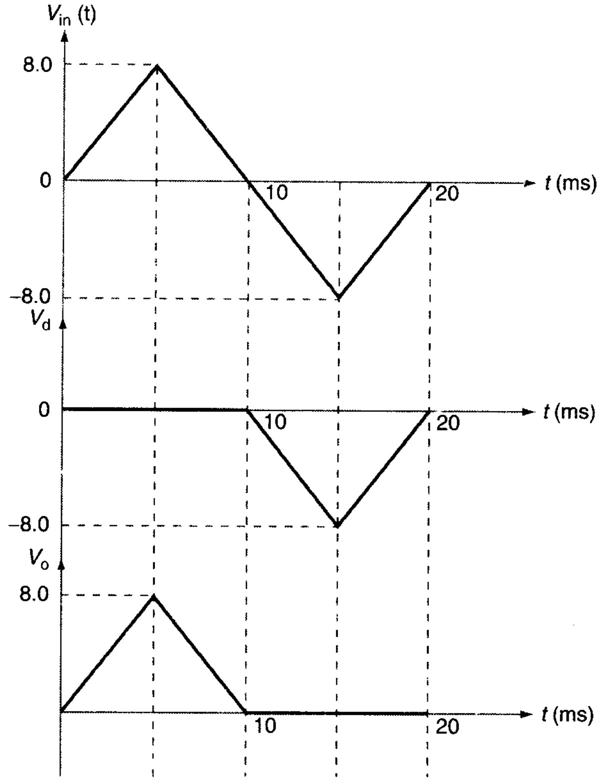
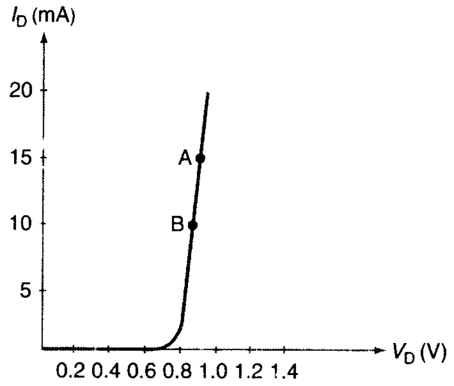


Figure 2.19 | Solution to Example 2.2.

EXAMPLE 2.3 *Draw the piecewise linear equivalent circuit model for the diode shown in Figure 2.20.*

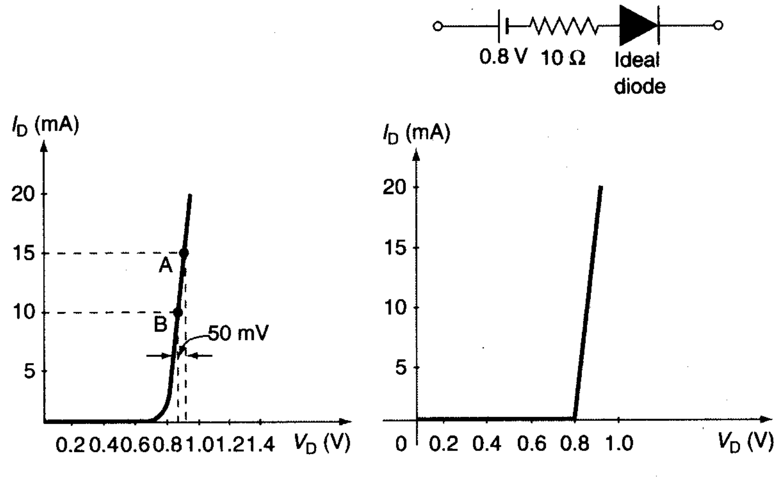


Voltage differential
between points A and B = 50 mV

Figure 2.20 | Example 2.3.

Solution

1. From Figure 2.21(a) we can see that the cut-in voltage is 0.8 V (approx.).
2. The slope of the curve is given by taking two points A and B in the linear region as shown in Figure 2.21(a).
3. Slope = $(V_A - V_B)/(I_A - I_B) = 50\text{ mV}/5\text{ mA} = 10\ \Omega$.
4. The piecewise equivalent model is shown in Figure 2.21(b). It comprises a battery voltage of 0.8 V, a resistance of 10 Ω and an ideal diode. The V–I characteristic curve for the equivalent model is also shown in the figure.



(a) (b)

Figure 2.21 | Solution to Example 2.3.

EXAMPLE 2.4

Using the equivalent model of the diode in Example 2.3, draw the output voltage (V_o) and the diode voltage (V_d) waveforms for the circuit and the input waveform of Example 2.2.

Solution

1. The output voltage (V_o) and diode voltage (V_d) waveforms are shown in Figure 2.22.
2. The diode acts as an open circuit until the input voltage reaches 0.8 V. After that it acts as a resistance with a value of 10 Ω.
3. Therefore the voltage drop across the diode resistance when input voltage is 8 V is given by $[(8 - 0.8)/1 \times 10^3] \times 10 = 72$ mV. Therefore, when the input voltage is 8 V, the effective drop across the diode is 0.872 V.
4. Therefore, the voltage across the load resistance when input voltage is 8 V is $8 - 0.872\text{ V} = 7.128$ V.
5. For the negative portion of the waveform, the diode acts as an open circuit. The whole input voltage appears across the diode and the output voltage (V_o) is zero.

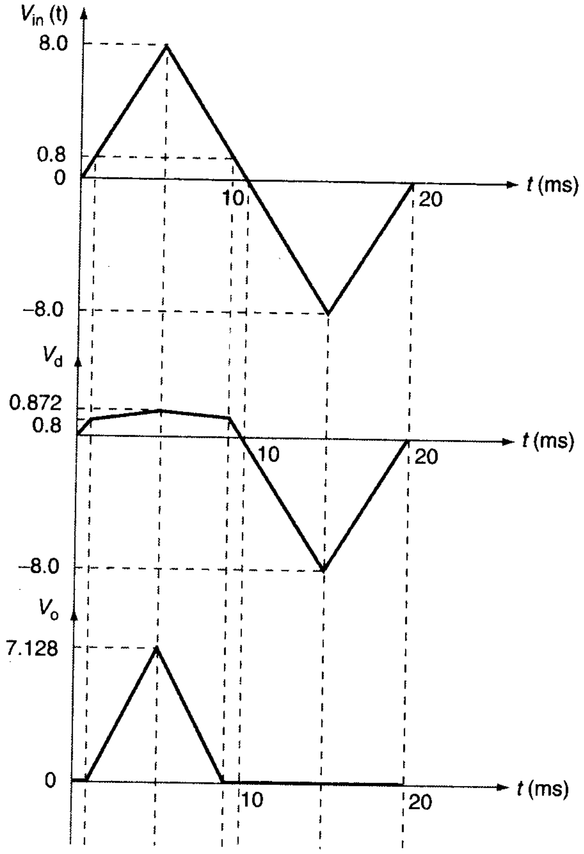


Figure 2.22 | Example 2.4.

2.10 Load-Line Analysis of a Diode Circuit

Load-line analysis is a graphical method of analyzing a circuit. In this method, a load line is drawn on the actual characteristic curve or on the equivalent model curve of the active device used in the circuit. It provides a very accurate method of analyzing the circuit when the actual characteristic curve of the active device is used for analysis. The slope of the load line depends on the applied load. It may be mentioned here that the applied load generally has an important impact on the point or region of operation of the device. The active device of concern in this section is the semiconductor diode.

DC Applied Voltage

Figure 2.23(a) shows the basic diode circuit where a DC input voltage source (V_I) is applied to a series connection of a diode (D) and load resistance (R_L). Applying Kirchhoff's voltage law to the circuit of Figure 2.23(a) we get

$$V_I = V_D + I_D R_L \tag{2.10}$$

where V_D is the diode voltage and I_D the diode current.

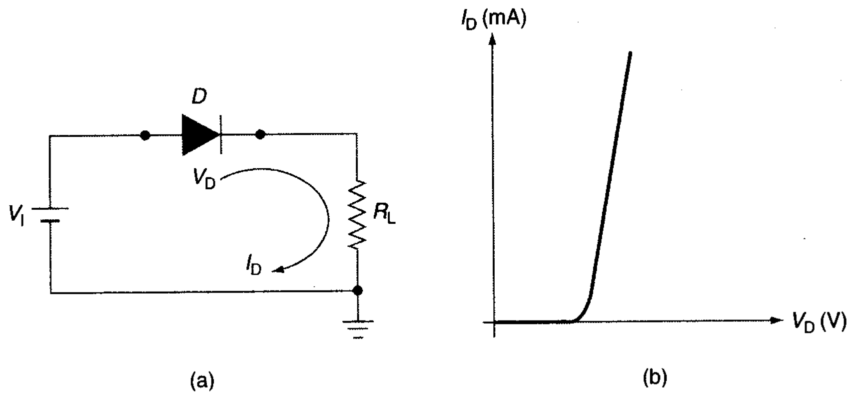


Figure 2.23 | (a) Simple diode circuit; (b) V–I characteristics of a diode.

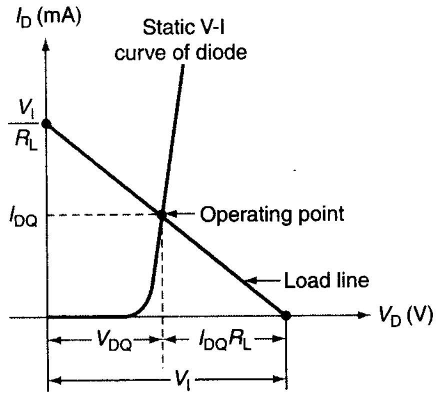


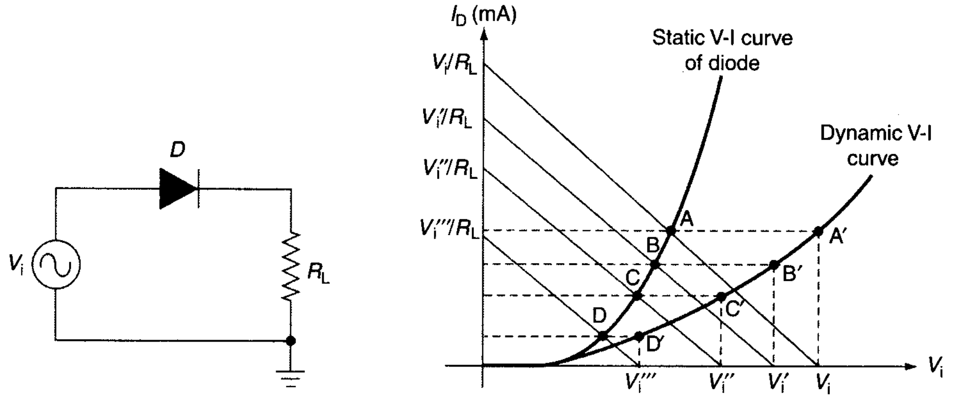
Figure 2.24 | Load-line analysis of a diode circuit for DC input voltage.

The straight line represented by Eq. (2.10) is called the *load line*. This single equation is not sufficient to determine the two unknown variables: diode voltage (V_D) and diode current (I_D). However, these two variables are the same as the diode's V–I characteristic axis variables [Figure 2.23(b)]. Therefore, a second relationship between the two variables is given by the V–I characteristic curve of the diode. The intersection of the load line with the V–I characteristic curve of the diode determines the *operating point* of the circuit also called the *quiescent point* or the *Q-point*.

The load line can be drawn by determining its intercepts on the voltage and the current axis. For $V_D = 0$, $I_D = V_I/R$ and for $I_D = 0$, $V_D = V_I$. The straight line joining these two points is the load line. The slope of the line is dependent on the value of load resistance (R_L) and is given by $-1/R_L$. Thus, for a given input voltage (V_I), lower the value of the load resistance steeper is the load line resulting in a higher value of the current at the Q-point. The process of drawing the load line and determining the Q-point is better illustrated in Figure 2.24. The operating point for the circuit is (V_{DQ}, I_{DQ}), where $V_{DQ} = V_I - I_{DQ} \times R_L$.

AC Applied Voltage

Let us consider the case when a time-varying input signal is applied to the circuit shown in Figure 2.25(a). Since the voltage applied is time-variant, separate load lines need to be drawn for the instantaneous values of the input voltage. The various load lines are parallel to each other as the value of load resistance (R_L) is

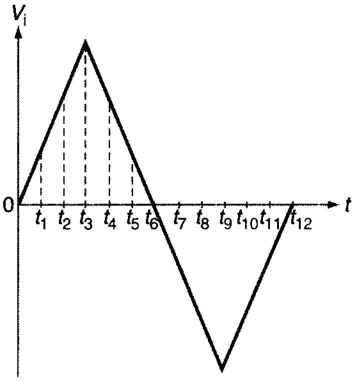


(a) (b)

Figure 2.25 | Dynamic characteristic curve.

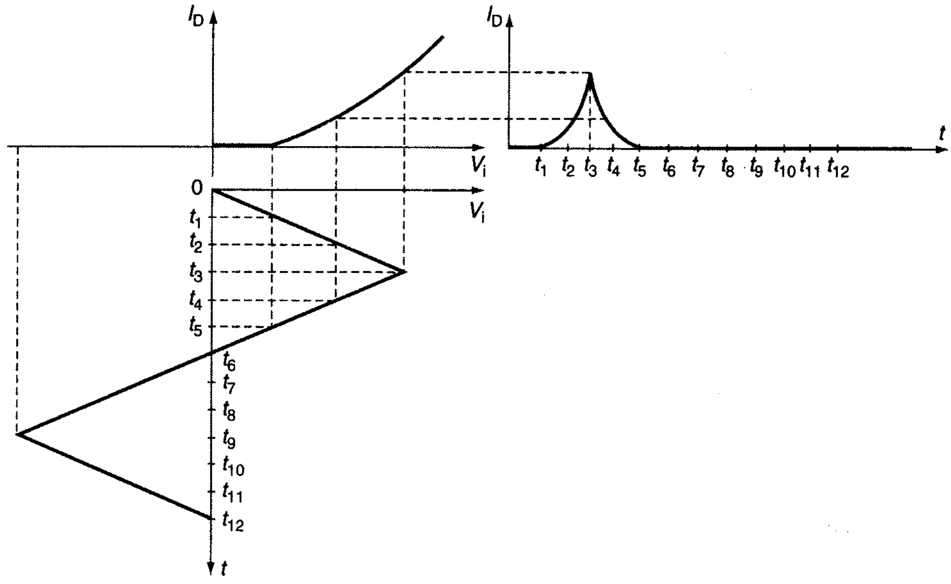
fixed. The intersection of these lines with the static V–I characteristic curve of the diode gives the value of the current in the circuit corresponding to different instantaneous values of the input signal. A better method to determine the current is to draw the dynamic characteristic curve of the circuit which is a plot between the diode current and the input voltage. Figure 2.25(b) shows the procedure for drawing the dynamic characteristic curve. The load line for the maximum value of the input signal is drawn. From the Q-point a horizontal line is drawn. The point where this line intersects with the vertical line drawn from the X-axis corresponding to that input voltage gives a point on the dynamic curve. The process is repeated for a few other values of input voltage to yield sufficient points to construct the dynamic curve.

Let us assume that the waveform shown in Figure 2.26(a) is applied to the circuit shown in Figure 2.25(a). The dynamic curve can be used to draw the output current waveform as shown in Figure 2.26(b). The figure is self-explanatory. It may be mentioned here that the dynamic curve applies only to the circuit containing the same value of load resistance for which it is drawn. Also, in the discussion we have assumed the diode to be an open circuit in the reverse-bias region. However, the dynamic curve for the diode in the reverse-bias region can be drawn on similar lines as drawn for the forward-bias region.



(a)

Figure 2.26 | (a) Input waveform; (b) output waveform construction of a diode circuit for AC input voltage.

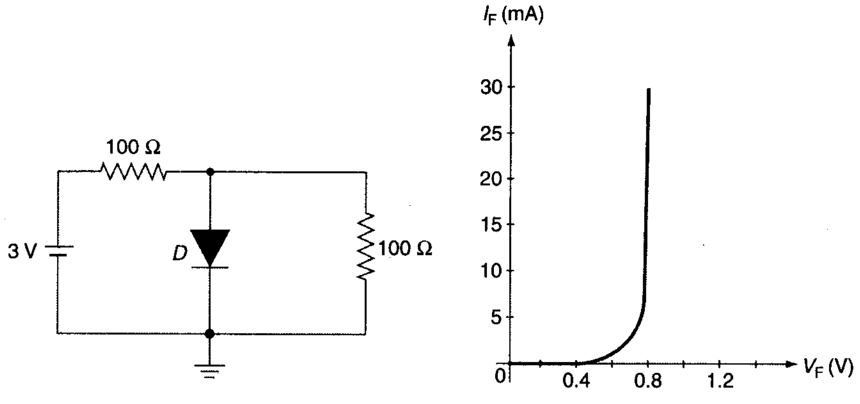


(b)

Figure 2.26 | *Continued.*

EXAMPLE 2.5

Draw the load line for the diode circuit of Figure 2.27(a). The forward characteristics of the diode are shown in Figure 2.27(b). Also determine the operating point.



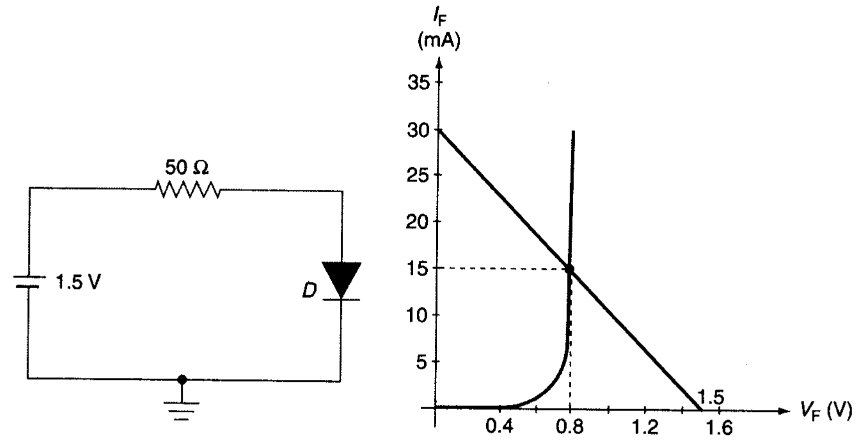
(a) (b)

Figure 2.27 | Example 2.5.

Solution

1. The circuit of Figure 2.27(a) can be simplified using the Thevenin's theorem. Thevenin's equivalent voltage $V_{TH} = (3 \times 100/200) = 1.5$ V. Thevenin's equivalent resistance $R_{TH} = (100 \times 100/200) = 50\ \Omega$. The simplified equivalent circuit is shown in Figure 2.28(a).

2. The load line is given by the equation $V_{TH} = V_D + I_D \times R_{TH}$ or $1.5 = V_D + I_D \times 50$.
3. The co-ordinates of the load line on the X- and the Y-axis are (1.5 V, 0) and (0, 30 mA), respectively. The load line superimposed on the V–I characteristics of the diode is shown in Figure 2.28(b).
4. The operating point is given by the intersection of the load line with the V–I characteristics of the diode. From Figure 2.28(b), we can see that the point is (0.8 V, 15 mA).

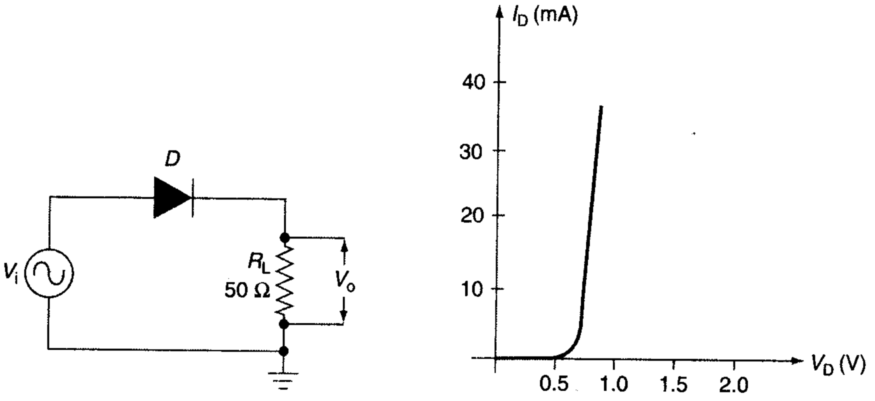


(a) (b)

Figure 2.28 | Solution to Example 2.5.

EXAMPLE 2.6

For the circuit shown in Figure 2.29(a), draw the output waveform of the circuit when a 50 Hz sinusoidal input voltage with a root mean square (RMS) voltage of 1.44 V is applied to the circuit. The diode characteristics are shown in Figure 2.29(b).



(a) (b)

Figure 2.29 | Example 2.6.

Solution

1. The RMS voltage of 1.44 V implies a peak value of 2 V and a frequency of 50 Hz implies a time period of 20 ms.
2. Load line is drawn for the peak value of the input waveform. The load line equation is $2 = V_D + I_D \times 50$. The coordinates of the line on the voltage and the current axes are (2 V, 0) and (0, 40 mA), respectively.
3. The procedure is repeated for other values of the input waveform and the dynamic curve is drawn. The relevant waveforms are shown in Figure 2.30.
4. From the figure, we see that the maximum output current is 24 mA. The peak output voltage across the load resistor is 1.2 V.

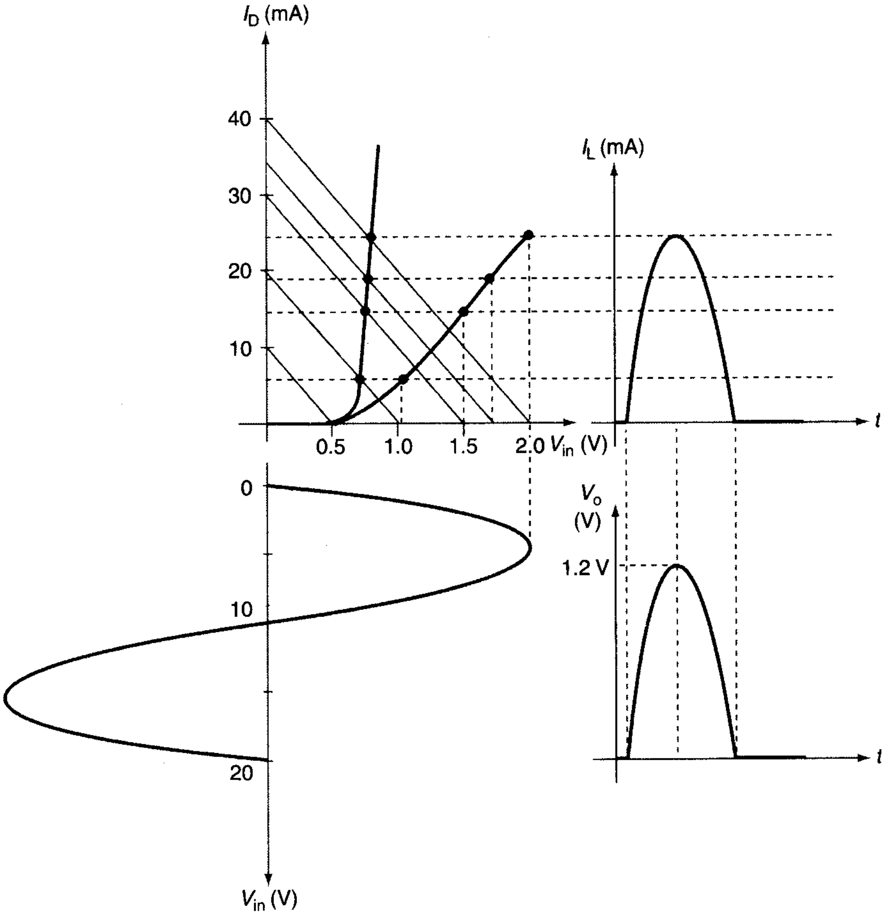
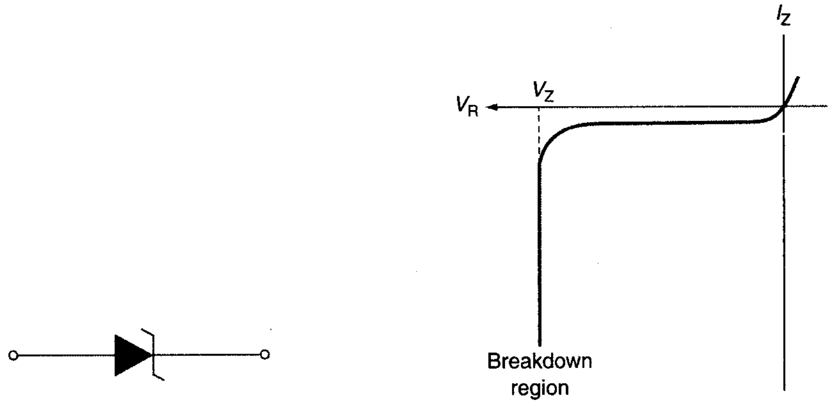


Figure 2.30 Solution to Example 6.

2.11 Breakdown Diodes

As discussed earlier in Section 2.4, when the voltage applied across the diode in the reverse-biased region exceeds the breakdown voltage of the diode, there is a sharp increase in the current flowing through the diode. This region is known as the breakdown region. Breakdown diodes are designed with sufficient power dissipation capabilities to operate in the breakdown region. They are generally employed as



(a) (b)

Figure 2.31 (a) Circuit symbol of breakdown diodes; (b) V–I characteristic of breakdown diodes.

constant-voltage devices or as voltage references. Depending upon the mechanism which leads to breakdown they can be further classified as *Zener diodes* and *avalanche diodes.*

The symbol and the V–I characteristics of breakdown diodes are shown in Figures 2.31(a) and (b), respectively. As is clear from Figure 2.31(b), for reverse voltages less than the breakdown voltage (V_Z), the diode acts as an open circuit, and for voltages greater than the breakdown voltage it acts as a constant voltage reference with the voltage across it equal to the breakdown voltage. It may be mentioned here that the shape of the V–I characteristics is the same for both the Zener and avalanche diodes. The parameters of interest for the breakdown diodes are the breakdown voltage, the dynamic impedance and the power dissipation capability.

Avalanche Diodes

In the case of avalanche diodes, on the application of reverse-bias voltage the thermally generated carriers have sufficient energies to disrupt covalent bonds, thereby resulting in free electrons. These free electrons knock out more electrons from the adjacent bonds. The process is regenerative and is referred to as *avalanche multiplication.* Avalanche breakdown mechanism is predominant in lightly doped diodes with broad depletion region and low field intensity. Generally, the avalanche diodes have breakdown voltages greater than 6 V and their breakdown voltage increases with increase in temperature, that is, they have positive temperature coefficient of breakdown voltage. As the temperature increases, the vibrational displacement of atoms in the crystal grows which increases the probability of collision of carriers with the lattice atoms as they cross the depletion region. Hence, they do not have sufficient energy to start the avalanche process resulting in an increase in the breakdown voltage. Silicon diodes with avalanche breakdown phenomenon are available with breakdown voltages ranging from several volts to several hundreds of volts and with power ratings up to 50 W.

Zener Diodes

The breakdown phenomenon in the case of a Zener diode is the result of electrons breaking their covalent bonds due to the existence of a strong electric field at the junction. The new hole–electron pair created increases the reverse current. It does not involve collisions of carriers with the lattice atoms. A Zener breakdown phenomenon occurs for heavily doped diodes having a narrow depletion-region width and high-field intensity. They have breakdown voltages below 6 V. With increase in temperature, the energy of the valence electrons increases, making it easier for these electrons to break the covalent bonds and hence the breakdown

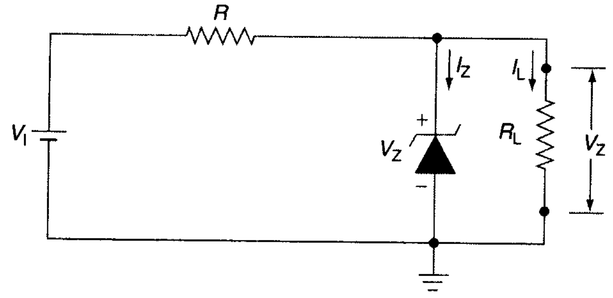


Figure 2.32 | Simple voltage regulator circuit using breakdown diode.

voltage decreases. Hence, these diodes have a negative temperature coefficient of breakdown voltage. Diodes with breakdown voltages between 5 V and 6 V have almost zero temperature coefficient of breakdown voltage. It may be mentioned here that the term Zener diode is generally used for breakdown diodes even with avalanche breakdown phenomenon.

Both Zener and avalanche diodes are used in voltage regulators to regulate the load voltage against variations in load current and input voltage. They are used in these applications because in the breakdown region large change in the diode current produces only a small change in the diode voltage. Figure 2.32 shows a simple voltage regulator circuit employing a Zener diode. The voltage across the load resistor is the same as the Zener breakdown voltage. The topic of voltage regulators is covered in the chapter on linear power supplies (Chapter 14).

2.12 Varactor Diodes

Varactor diodes are used as variable voltage capacitors. They are also referred to as *varicaps* or *variable voltage capacitance diodes* or *tunable diodes*. Their mode of operation depends on the transition capacitance that exists at the P–N junction when the diode is reverse-biased. Junction capacitances were discussed in detail in Section 2.8. Figure 2.33 shows the characteristics of a typical commercially available varactor diode. As shown in the figure, there is a sharp decrease in the transition capacitance initially with increase in reverse-bias voltage. As the reverse-bias voltage increases further, the rate of change of capacitance with voltage decreases. Varactor diodes are normally operated with reverse voltages upto 20–30 V. The relationship between the transition capacitance and the applied reverse bias is expressed by

$$C_T = \frac{K}{(V_\gamma + V_R)^n} \tag{2.11}$$

where K is a constant (depends on the semiconductor material and the diode construction technique); V_γ the knee potential of the diode; V_R the magnitude of the applied reverse bias; $n = 1/2$ for alloy junction and 1/3 for diffused junction.

The circuit symbol and the equivalent circuit of varactor diodes are shown in Figures 2.34(a) and (b), respectively. The variable R_R is the resistance of the diode in the reverse-bias region and is of the order of greater than equal to 1 MΩ; R_S is the geometric resistance of the diode and is of the order of few ohms. The magnitude of C_T varies from few picofarads to around hundred picofarads. In different varactor diode types, the values of minimum and maximum capacitances may vary; however, the ratio of maximum to minimum capacitance is typically 2.5 to 3. Typical application areas of varactor diodes include FM modulators, automatic frequency control devices, adjustable bandpass filters and parametric amplifiers.

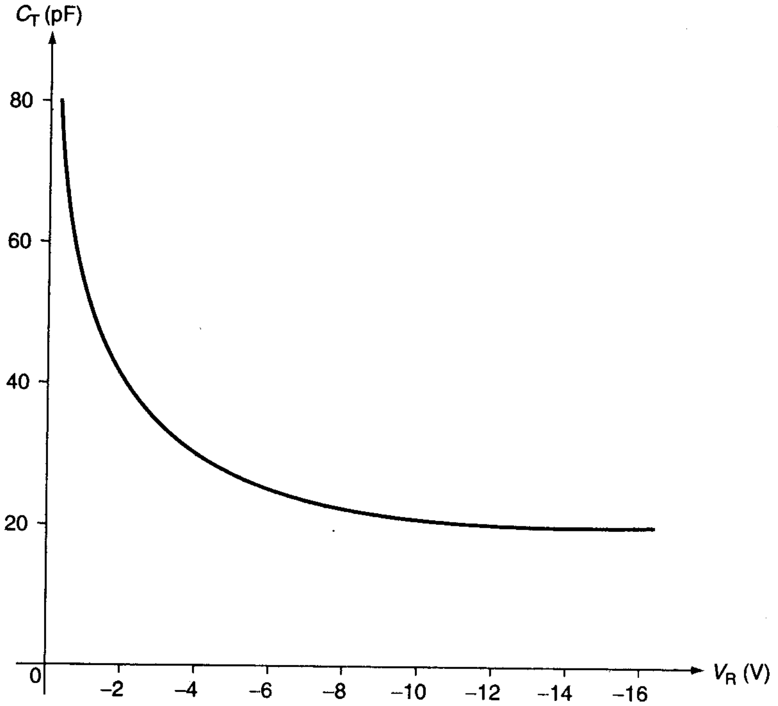


Figure 2.33 | Characteristics of a varactor diode.

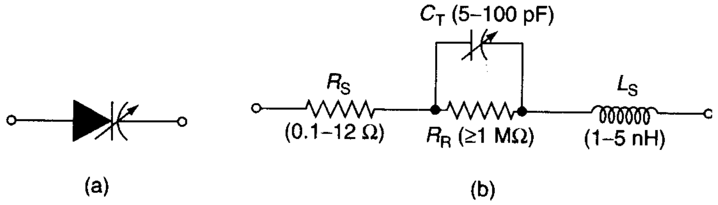


Figure 2.34 | (a) Circuit symbol of a varactor diode; (b) equivalent circuit of a varactor diode.

2.13 Tunnel Diodes

Tunnel diodes have heavily doped P- and N-regions, about 100–1000 times dopant concentration than that of a typical semiconductor diode. Heavy doping results in narrowing of the depletion region. The width of the depletion region of a tunnel diode is about 100–1000 times less than that of a typical semiconductor diode. Owing to this narrow depletion region, the charge carriers instead of climbing up the potential barrier may pierce through the potential barrier resulting in "tunneling" of carriers both in the forward-bias and the reverse-bias regions, hence rendering the diode bi-directional conduction property. In the forward direction, the current reaches the maximum value I_P (called the peak current) at a voltage V_P (peak voltage). At this point, referred to as *peak point*, slope of the V–I curve is zero, that is, $di/dv = 0$. Beyond the peak point, the current starts to decrease with increase in voltage as there are no more carriers available for tunneling. The current approaches zero for a forward voltage of 0.4 V to 0.5 V but then the normal P–N junction effect starts. The forward current of P–N junction diode adds to the current due to the tunneling effect. The current decreases beyond the peak point until a point, referred to as the *valley point*. The region between the peak point and the valley point has negative resistance characteristics as the voltage decreases with increase in current. At the valley point also, the slope of the V–I curve is zero ($di/dv = 0$). Beyond the valley point, the current starts increasing again with increase in voltage and the current reaches the peak value I_P again at a voltage V_F. This is further illustrated in Figure 2.35, showing the V–I characteristics of the tunnel diode. These characteristics may be considered to be composed of two characteristics, one due to the P–N junction and other due to the tunneling phenomenon. The value of voltage swing ($V_F - V_P$) is of the order of 1 V for gallium arsenide tunnel diodes and 0.45 V for germanium tunnel diodes.

The symbol of the tunnel diode and its equivalent circuit in the negative resistance region are shown in Figures 2.36(a) and (b), respectively. The semiconductor material used in construction of tunnel diodes is either germanium or gallium arsenide. Silicon is not used for constructing tunnel diodes. This is because the ratio of the peak current to the valley current (I_P/I_V) in gallium arsenide and

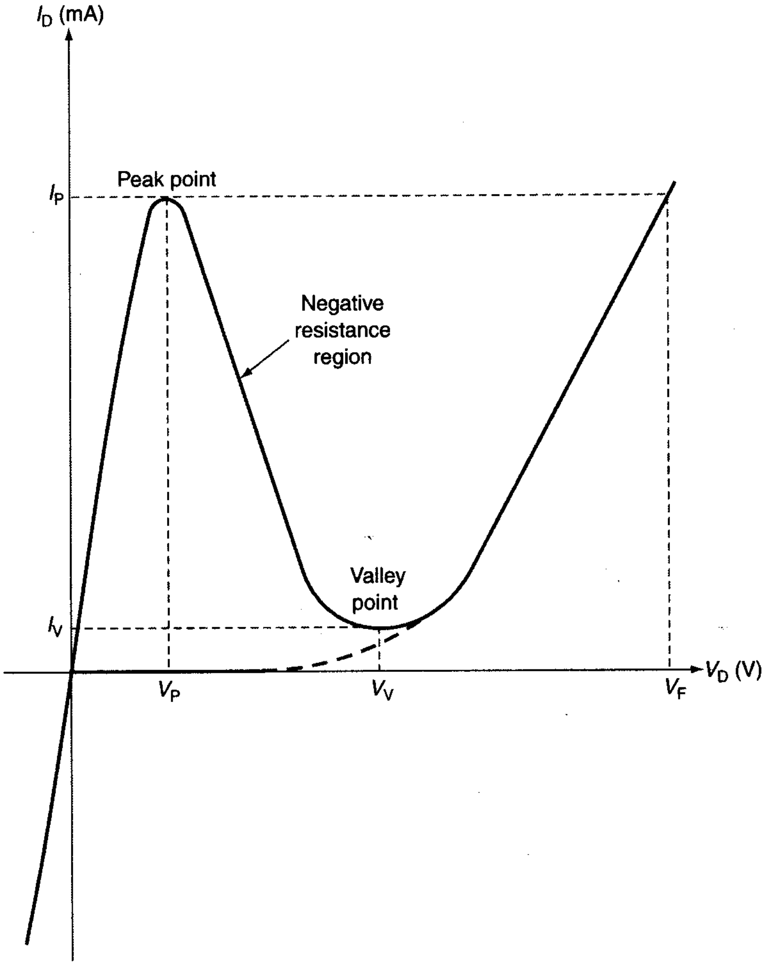


Figure 2.35 V–I characteristics of a tunnel diode.

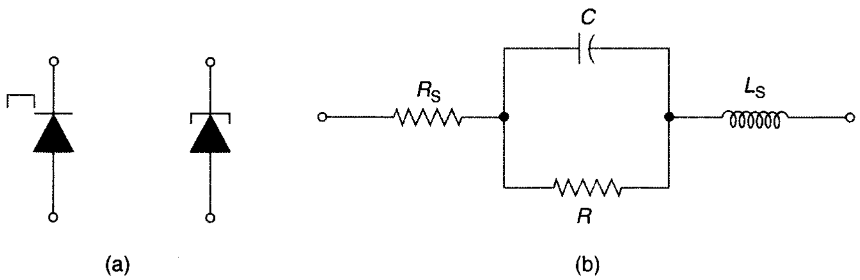


Figure 2.36 (a) Circuit symbol of a tunnel diode; (b) equivalent circuit of the tunnel diode in the negative resistance region.

germanium is quite high, approximately 15 for gallium arsenide and 8 for germanium. The value in the case of silicon is very small (approximately 3).

Tunnel diodes are used in high-speed applications such as in computers with switching times of the order of few nanoseconds to several picoseconds. Owing to their negative resistance characteristics these diodes were earlier used as microwave oscillators. However, now they have been replaced by other devices that have surpassed them in performance.

2.14 Schottky Diodes

Schottky diodes, also known as *hot-carrier diodes*, have a metal–semiconductor junction instead of a semiconductor–semiconductor junction (P–N junction) of a conventional P–N junction diode. Normally N-type silicon is used as the semiconductor while the metal used can be aluminum, platinum, tungsten or molybdenum. This different construction technique renders these diodes some special characteristics as compared to P–N junction diodes such as lower cut-in voltage, increased frequency of operation, etc.

Schottky barrier diodes are majority carrier conduction devices. In both the materials (metal and semiconductor) electrons are the majority carriers. The circuit symbol and the equivalent circuit model for a Schottky diode are shown in Figures 2.37(a) and (b), respectively. The equivalent circuit is an ideal diode in parallel with a capacitor which is equivalent to the junction capacitance. The V–I characteristics of a Schottky diode as compared to a conventional P–N junction diode are shown in Figure 2.38.

The junction barrier for a Schottky diode, in both the forward- and reverse-bias regions, is less than that of P–N junction diode. This results in lower cut-in voltage of the order of 0.3 V for silicon-metal Schottky diode as compared to a cut-in voltage of 0.7 V for silicon P–N junction diodes. Lower junction barrier also results in higher currents at the same applied voltage in both the forward- and the reverse-bias conditions. Thus, they dissipate less power than a normal diode. This, however, results in larger reverse saturation current as compared to a conventional P–N junction diode which is highly undesirable. Also, the peak inverse voltage (PIV) rating for a Schottky barrier diode is less than that of a comparable P–N junction diode.

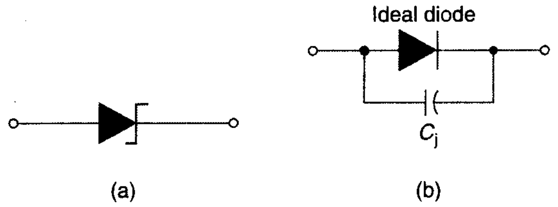


Figure 2.37 (a) Circuit symbol of a Schottky diode; (b) equivalent circuit of a Schottky diode.

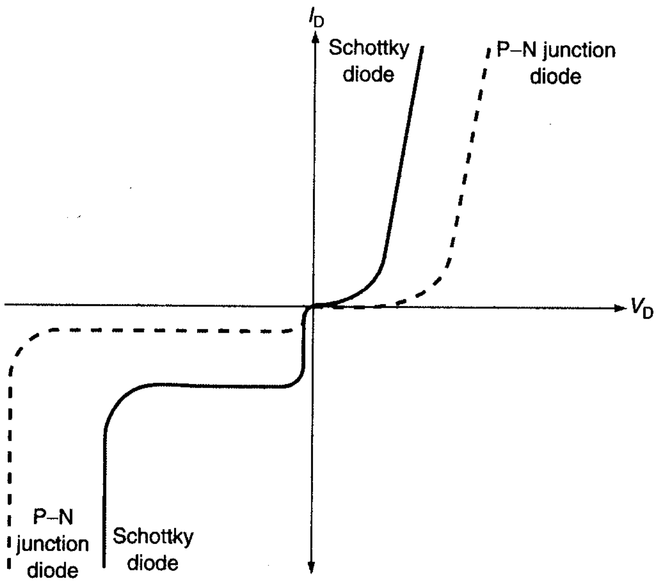


Figure 2.38 V–I characteristics of a Schottky diode.

Schottky diodes are used as high-efficiency rectifiers which are essential in applications such as switched mode power supplies (SMPS), switching regulators, etc. The absence of minority carriers in Schottky diodes results in significantly lower value of reverse recovery time (as low as 20 ns). Thus, they are effective at operating frequencies extending up to several gigahertz. Other application areas include low-voltage/high-current power supplies, AC to DC converters, mixers and detectors in communication systems.

2.15 Point-Contact Diodes and Power Diodes

Point-Contact Diodes

These diodes are intended primarily for radio frequency (RF) applications owing to their extremely small internal capacitance, considerably less than that of a conventional junction diode. They basically have a metal–semiconductor junction and have been replaced by Schottky barrier diodes because Schottky diodes offer lower forward resistance, wide dynamic range and better noise performance as compared to point-contact diodes.

Power Diodes

Power diodes are designed to operate at high-power levels and at high operating temperatures. They are mainly used as rectifiers. They are generally constructed using silicon because silicon offers higher current, temperature and PIV ratings. Such diodes have large junction area to ensure low forward diode resistance so that the I^2R loses can be reduced. The current capability of power diodes is increased by placing two or more diodes in parallel whereas the PIV rating is increased by stacking the diodes in series. Generally, they are mounted in conjunction with heat sinks for thermal management.

2.16 Light-Emitting Diodes

A semiconductor P–N junction diode designed to emit light when forward-biased is called a light-emitting diode (LED). When a P–N junction is forward-biased, the electrons in the N-type material and the holes in the P-type material travel towards the junction. Some of these holes and electrons recombine with each other and in the process radiate energy. The energy will be released either in the form of photons of light or in the form of heat. In silicon and germanium diodes, most of the energy is released as heat and the emitted light is insignificant. However, in some materials such as gallium phosphide (GaP), gallium arsenide (GaAs) and gallium arsenide phosphide (GaAsP) substantial photons of light are emitted. Hence, these materials are used in the construction of LEDs.

The V–I characteristics of LEDs are similar to that of a normal P–N junction diode with the difference that the cut-in voltage in the case of LEDs is around 1.5 V as compared to 0.7 V for silicon diodes and 0.3 V for germanium diodes. Figures 2.39(a) and (b) show the process of light emission in an LED and its circuit symbol, respectively. As can be seen from the figure, the conducting surface connected to the P-type material is smaller in size to allow maximum number of photons to contribute to the output light energy. The wavelength of emitted light is the function of bandgap energy of the semiconductor material and is expressed by the empirical formula

$$\lambda = \frac{1240}{\Delta E} \tag{2.12}$$

Table 2.1 | Commonly used LED materials

Material	Bandgap energy (eV)	Wavelength (nm)
GaAs	1.43	910
GaP	2.24	560
$GaAs_{60}P_{40}$	1.91	650
AlSb	1.60	775
InSb	0.18	6900

where λ is the wavelength (nm); ΔE the bandgap energy (eV). Table 2.1 enlists some of the materials used for making LEDs along with their bandgap energies and wavelengths. LEDs are discussed in detail in the chapter on optoelectronic devices (Chapter 7).

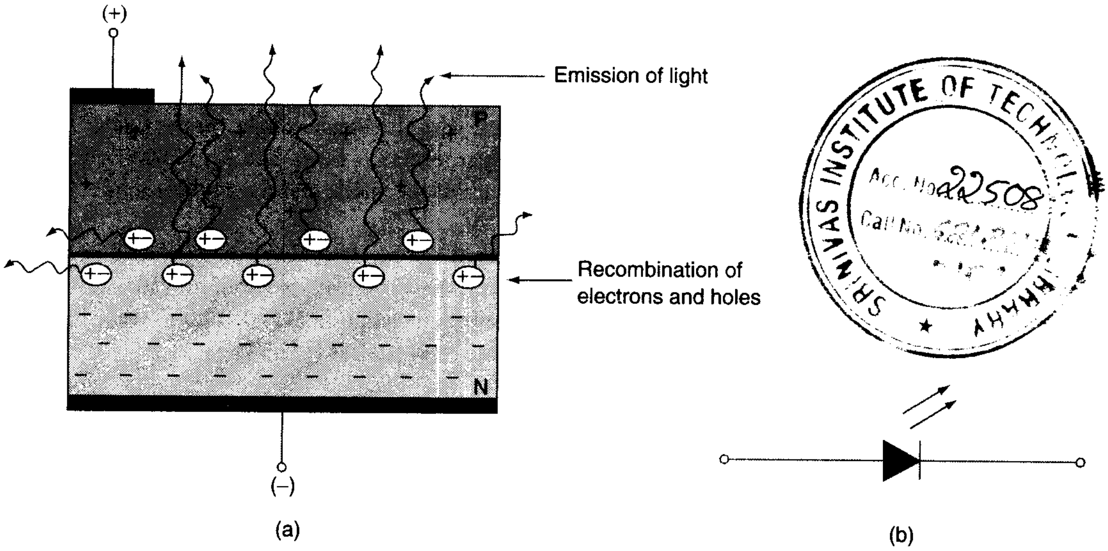
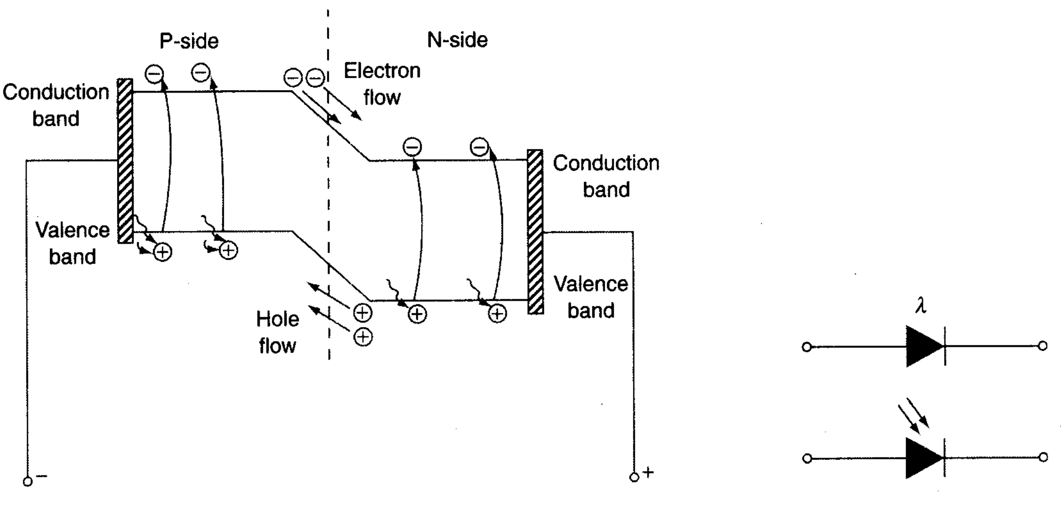


Figure 2.39 | (a) Process of light emission in an LED; (b) circuit symbol of an LED.



(a)

(b)

Figure 2.40 | (a) Photoeffect in a photodiode; (b) circuit symbol of a photodiode.

2.17 Photodiodes

Photodiode is a junction diode through which significant current flows when light falls on it. Photodiodes are operated either in the reverse-bias mode (referred to as the photoconductive mode) or with no external bias (referred to as the photovoltaic mode). When no light is incident on the photodiode, the current flowing through it is the reverse saturation current. This current is also referred to as the *dark current*. When operated in the photoconductive mode, the impinging photons of incident light create electron–hole pairs on both sides of the junction. The number of electron–hole pairs generated is directly proportional to the number of incident photons. The photo-induced electrons in the conduction band of the P-region will move across the junction down the potential hill along with the thermally generated minority carriers. Similarly, holes produced in the valence band of the N-region are available to add to the current flow by moving across to the P-region. Figures 2.40(a) and (b) show the photoeffect in a photodiode and its circuit symbol. Figure 2.41 shows the variation of the photocurrent with the incident light. In the figure, I_{L1}, I_{L2}, I_{L3} and I_{L4} are the photocurrent levels corresponding to light levels L_1, L_2, L_3 and L_4, respectively. When operated in the photovoltaic mode, a voltage is developed across the anode and the cathode terminals. The dark current in the photovoltaic mode is nearly zero.

The spectral response of photodiodes is a function of the energy bandgap of the material used in its construction. Some of the commonly used materials are silicon (200–1100 nm), germanium (500–1900 nm), indium gallium arsenide (700–1700 nm) and mercury cadmium telluride (1900–10,000 nm). Detailed description of photodiodes is given in the chapter on optoelectronic devices (Chapter 7).

2.18 Connecting Diodes in Series and in Parallel

Diodes in Series

Semiconductor junction diodes are connected in series to enhance the peak inverse voltage rating beyond what is available in a single diode. In order to ensure that there is equal division of reverse voltage across the individual diodes, the diodes should have closely matched reverse-bias characteristics. Equal division of

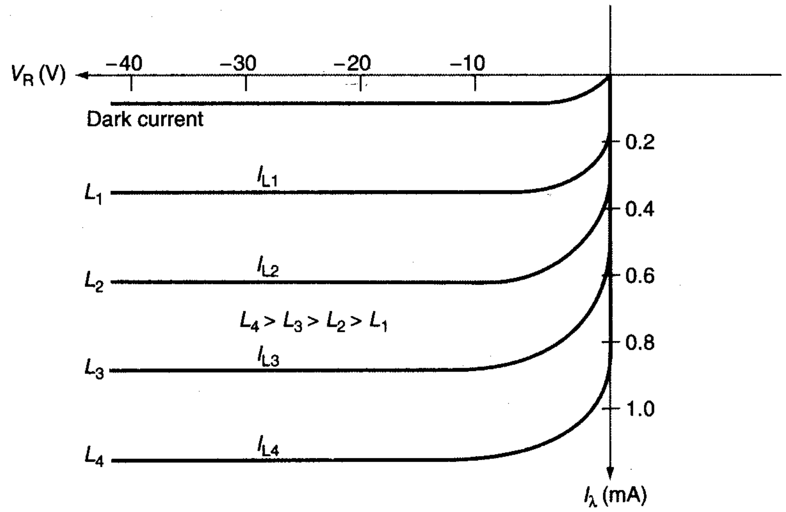


Figure 2.41 Variation of photocurrent with the incident light in a photodiode.

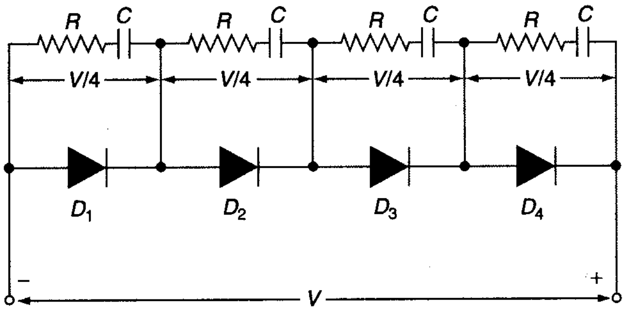


Figure 2.42 Connecting diode in series.

reverse voltage can however be forced by connecting series R-C networks across individual diodes. An arrangement for series connection of four diodes is shown in Figure 2.42. The value of the resistors (R) used should be much smaller than the reverse-bias resistance of the individual diodes.

Diodes in Parallel

Semiconductor diodes are connected in parallel to enhance the forward current capability. Parallel connection of diodes is trickier than the series connection of the same. Parallel connected diodes must have closely matched forward characteristics, lest they will not have equal division of current. The diode with the lowest forward voltage drop draws larger current initially. This heats the diode junction, thus further reducing the forward voltage drop. (The forward voltage drop reduces at a rate of 2.5 mV/°C.) This process is cumulative and ends up with the diode getting damaged. As a result, the other diodes in the parallel connection have to share a larger burden of current. Again the diode with comparatively lower forward voltage becomes the target and is thus prone to damage. The process continues until all the diodes get damaged. An equal division of forward current can be forced by using series resistors (Figure 2.43) or

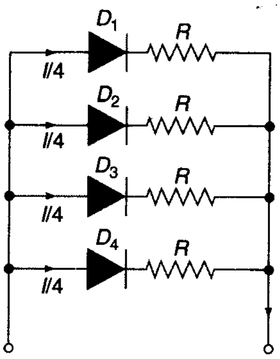


Figure 2.43 | Diodes in parallel.

balancing inductors with each of the parallel connected diodes. The value of resistors (R) used should be much larger than the forward-bias resistance of the individual diodes.

2.19 Diode Numbers and Lead Identification

Semiconductor diodes have a "1" prefixed in their type number identification. However, no distinction is made among the various materials used in diode construction. Majority of semiconductor diodes are fabricated in one of the following package styles, namely DO-7, DO-35 and DO-41. These package styles are shown in Figure 2.44. The anode (P-side) and cathode (N-side) of the diodes made in these and similar package styles are marked in several ways (Figure 2.45). One of the methods is to indicate the anode with a positive sign and the cathode with a negative sign. The most commonly used method is to put a circular band near the cathode. The other terminal without the band is of course the anode. Yet another popular style of marking the diode's leads is to put an arrow along the length of the diode with the arrow pointing towards the cathode or to put a dot near the cathode.

In the case of Zener diodes made in these package styles, a band is put near the cathode. In some Zener diodes, a positive sign is put near one of the leads with the other lead usually unmarked. In this case, a

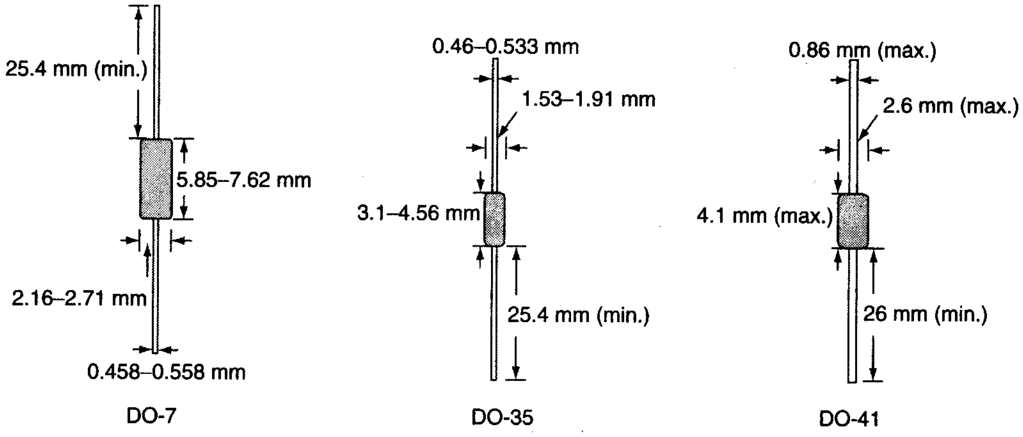


Figure 2.44 | Diode packages.

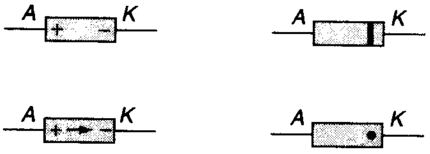
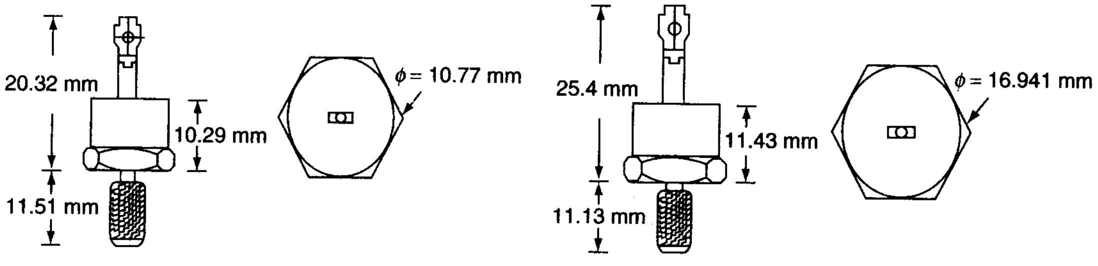


Figure 2.45 Lead identification of diodes.



DO-4 Package

DO-5 Package

Figure 2.46 Stud-mounted metal packages for diodes.

plus (+) sign indicates cathode and not the anode as stated in the case of conventional diodes. Remember that in the usual mode of operation of a Zener diode, cathode is more positive with respect to the anode.

High-current versions of semiconductor diodes are usually made in stud-mounted metal packages such as DO-4, DO-5 and so on. The DO-4 and the DO-5 packages are shown in Figure 2.46. The stud-mounted diodes are made both with the stud acting as the anode and the stud as the cathode. The stud-mounted diodes are marked by either putting the plus (+) and minus (–) signs, respectively, on the anode and the cathode or by showing the diode symbol with the anode pointing towards the anode terminal and the cathode pointing towards the cathode terminal.

Lead identification in diodes can also be done using a multimeter. Multimeter leads are connected to the diode and the multimeter is set to the position showing the diode symbol. If the display shows the cut-in voltage of the diode then the diode terminal connected to the red (positive) lead of the multimeter is the anode and the one connected to the black lead (negative) is the cathode. If the connections are reversed, the multimeter should show an OL indication. The test gives correct results only when the diode used is healthy.

2.20 Diode Testing

Both P–N junction and Zener diodes can be tested using a digital or an analog multimeter, an ohmmeter or a curve tracer. When using a digital multimeter, set the multimeter at a position showing the diode symbol. The multimeter leads are connected to the diode such that the diode is forward-biased, that is, the red (positive) lead of the multimeter is connected to the anode and the black (negative) lead is connected to the cathode. The display provides an indication of its forward-bias voltage. An OL indication in this position indicates an open or a defective diode. Now interchange the multimeter leads to reverse-bias the diode. The multimeter will give an OL indication (open circuit) if the diode under test is healthy. A low resistance or a short circuit in this position indicates a shorted or a defective diode.

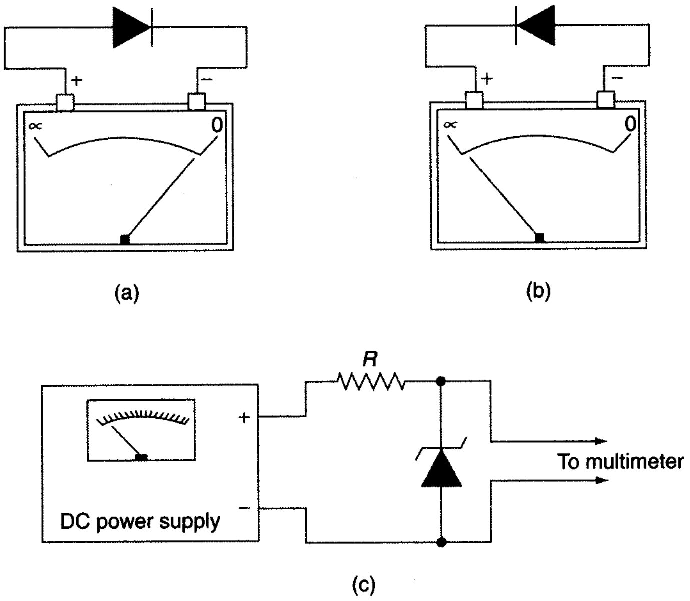


Figure 2.47 Testing diodes using multimeter.

The diode can be tested using an analog multimeter or an ohmmeter in a similar manner. Select the meter in one of the lower resistance ranges. The leads are connected to the diode in such a way that the diode is forward-biased. For a healthy diode the meter shows a very low resistance confirming that the diode is working properly in the forward-bias mode [Figure 2.47(a)]. The resistance shown by the meter is a function of the current established through the diode by the internal battery of the meter. Now, interchange the leads to reverse-bias the diode. The multimeter would show an open circuit if the diode is healthy [Figure 2.47(b)]. An open circuit in both the tests and a low resistance or short in both the tests indicates an open and a shorted diode, respectively.

Zener diodes can also be tested in the same fashion. Another important parameter one would like to check in the case of Zener diodes is the Zener breakdown voltage. The breakdown voltage of a given diode can be ascertained by rigging up a small test circuit as shown in Figure 2.47(c). Resistance *R* is typically 100 Ω. The input DC voltage being fed from a regulated DC power supply is gradually increased while continuously monitoring the voltage across the Zener diode with a multimeter. The voltage across the Zener diode is observed to increase with the increase in the input voltage. Infact, the voltage across the Zener diode is almost equal to the input voltage, until it reaches the breakdown voltage. Beyond that, the output voltage stays put despite changes in the input voltage. After having reached the breakdown voltage, the current through the diode is given by the input–output voltage differential divided by the resistance (*R*). The power dissipated in the Zener diode in this case would be equal to the product of the breakdown voltage and the current flowing through the circuit. So, while carrying out this test, one should remember not to exceed the input voltage to a point that forces the Zener diode under test to dissipate more power than it can safely handle. Typically, the current through the diode should not be allowed to exceed 10 mA while carrying out the test.

KEY TERMS

Ampere square seconds (I^2t)
Avalanche diode
Average AC resistance
Breakdown diode
Cut-in voltage or the knee voltage

Diffusion capacitance or the storage capacitance
Diode capacitance
Dynamic resistance
Equivalent circuit
Forward bias
Forward current (I_F)
Forward recovery time (t_{ff})
Forward voltage (V_F)
Ideal diode
LED
Load-line analysis
Maximum average rectified current ($I_{F(av)}$)
Peak forward surge current
Peak repetitive forward current
Photodiode
Piecewise linear equivalent circuit model
Point-contact diode
Power diode
Power dissipation (P_D)
Reverse bias
Reverse breakdown voltage or the peak inverse voltage
Reverse current (I_R)
Reverse recovery time (t_{rr})
Reverse voltage (V_R)
Schottky diode
Semiconductor diode
Static resistance
Transition capacitance
Tunnel diode
Varactor diode
Zener diode

OBJECTIVE-TYPE EXERCISES

Multiple-Choice Questions

1. P-side of a semiconductor diode is applied a potential of 0.5 V whereas the N-side is applied a potential of −1.0 V. The diode will
   a. conduct.
   b. not conduct.
   c. conduct partially.
   d. breakdown.

2. In a semiconductor diode, V–I relationship is such that
   a. current varies linearly with voltage.
   b. current increases exponentially with voltage.
   c. current varies inversely with voltage.
   d. none of these.

3. The capacitance appearing across a reverse-biased semiconductor junction
   a. increases with increase in bias voltage.
   b. decreases with increase in bias voltage.
   c. is independent of bias voltage.
   d. none of these.

4. There are two semiconductor diodes A and B. One of them is Zener whereas other is avalanche. Their ratings are 5.6 V and 24 V, respectively, then
   a. A is Zener, B is avalanche.
   b. A is avalanche, B is Zener.
   c. both of them are Zener diodes.
   d. both of them are avalanche diodes.

5. The static resistance of a diode is
   a. its opposition to the DC current flow.
   b. its opposition to AC current flow.
   c. resistance of diode when forward-biased.
   d. none of these.

6. The important specifications of a Zener diode are
   a. its breakdown voltage and power dissipation.
   b. breakdown voltage, dynamic impedance and power dissipation.
   c. breakdown voltage and dynamic impedance.
   d. none of these.

7. Typical value of impurity concentration in a tunnel diode is
   a. 1 part in 10^8 parts.
   b. 1 part in 10^6 parts.
   c. 1 part in 10^3 parts.
   d. 1 part in 10 parts.

8. The photodiodes are operated in
   a. reverse-bias condition.
   b. zero-bias condition.
   c. either of the two.
   d. none of the two.

9. The cut-in voltage for a LED is of the order
  a. 1 V.
  b. 0.7 V.
  c. 0.3 V.
  d. 1.5 V.

10. A varactor diode may be advantageous at microwave frequencies (indicate false answer)
  a. for electronic tuning.
  b. as an oscillator.
  c. as a parametric amplifier.
  d. for frequency multiplication.

Fill in the Blanks

1. The ideal diode acts as a _______ switch when forward-biased and acts as a _______ switch when reverse-biased.

2. In an open-circuit P–N junction diode, the current due to majority carrier flow is _______ to/than the current due to the minority carrier flow.

3. The Schottky diodes have a _______ junction.

4. The reverse-bias current of a diode _______ with increase in temperature.

5. _______ is the most accurate diode equivalent model.

REVIEW QUESTIONS

1. Show the basic diode action diagrammatically when it is
  a. Unbiased
  b. Reverse-biased
  c. Forward-biased

2. Sketch V–I characteristics of a semiconductor junction diode for both silicon and germanium diodes. Indicate differences, if any, in the characteristics curve of the two types.

3. Briefly describe the following terms:
  a. Static resistance
  b. Dynamic resistance
  c. Junction capacitance
  d. Tunneling effect

4. Why do we need to connect diodes in series and in parallel and what are the precautions to be observed while doing so?

5. What are breakdown diodes? How are they classified on the basis of breakdown mechanism?

6. Show how a Zener diode may be used to regulate the output voltage.

7. Explain in detail
  a. The effect of temperature on Zener diode.
  b. The depletion region and the N- and the P-regions of a P–N junction diode form a parasitic capacitance. What is the effect of the applied reverse-bias voltage on this capacitance?
  c. How is it possible to determine the polarity of a rectifying diode (which terminal is the anode and which terminal is the cathode) from its physical appearance?
  d. What do you understand by the term "derating"? What is its significance for semiconductor devices?
  e. The effect of temperature on the reverse saturation current of the diode.

8. "*Tunnel diode is a voltage controllable device.*" Comment.

9. What are Schottky diodes? How are they different from conventional P–N junction diodes? What are their major applications?

10. Explain the principle of operation of a LED.

11. What is meant by the term "reverse-recovery time"? Is it due to the majority carriers or the minority carriers?

12. What is the difference between the static and dynamic resistance of a P–N junction diode? Using the diode equation show that the dynamic resistance of an ideal P–N junction diode under forward-bias condition is inversely proportional to the forward current.

13. For a silicon diode, calculate the amount by which the diode voltage would have to be increased to double the diode current.

14. Mention two instruments which can be used to check the health of a diode. Also explain the checking procedure.

15. Explain in detail the operating principle of a photodiode. Also, draw its V–I characteristics.

PROBLEMS

1. Refer to Figure 2.48. Determine the static resistance of the diode at points A and B.

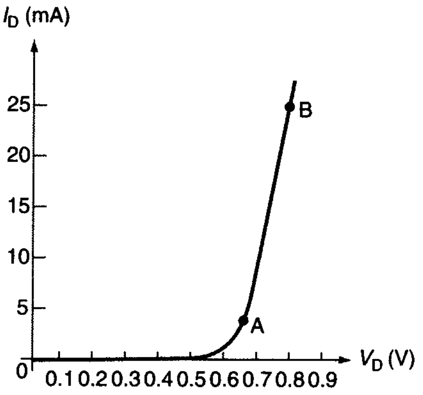


Figure 2.48 | Problem 1.

2. Draw the piecewise linear equivalent circuit model for the diode in Problem 1.

3. A simple diode circuit is shown in Figure 2.49(a). Assuming the diode to be ideal, draw the output waveform (V_o) and waveform across the resistance (V_r) for the input waveform (V_{in}) shown in Figure 2.49(b).

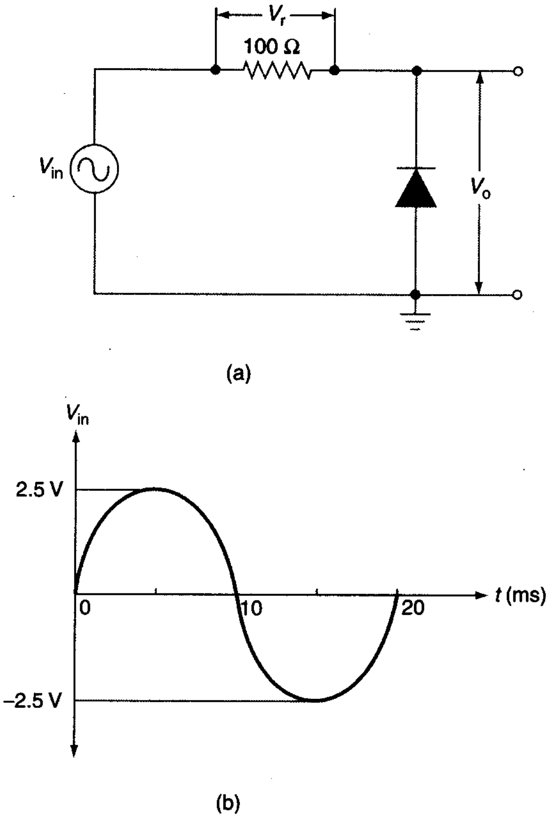
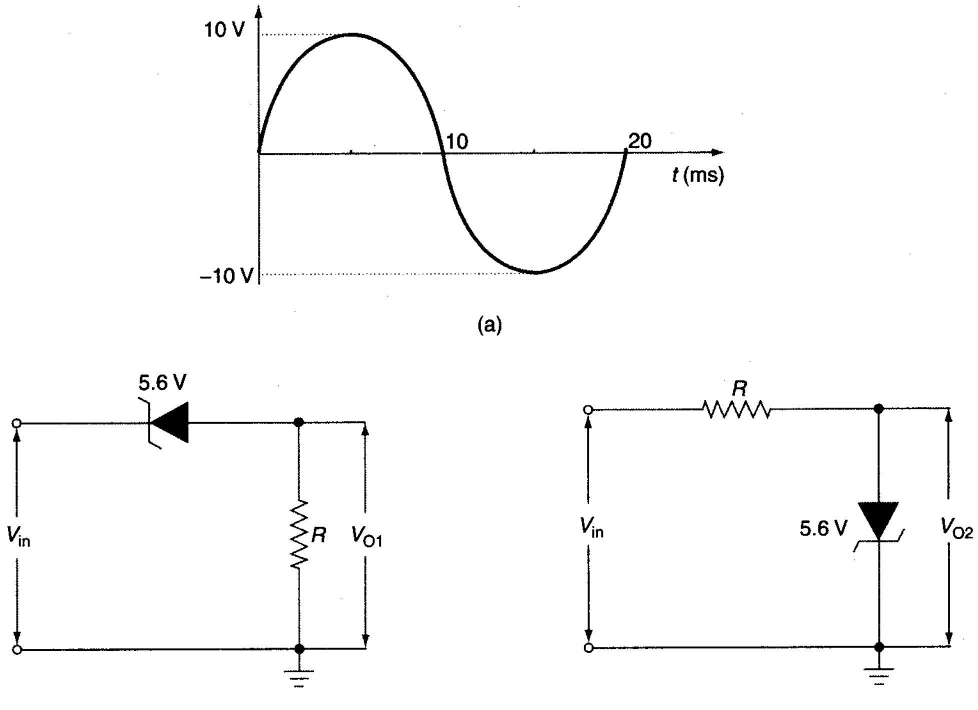


Figure 2.49 | Problem 3.

4. Sketch the output waveform when the input waveform (V_{in}) of Figure 2.50(a) is fed to the circuits of Figures 2.50(b) and (c).



(a)

(b)

(c)

Figure 2.50 | Problem 4.

5. Refer to Figure 2.51. Draw the V_O, V_{O1} and V_{O2} waveforms assuming the diode to be ideal.

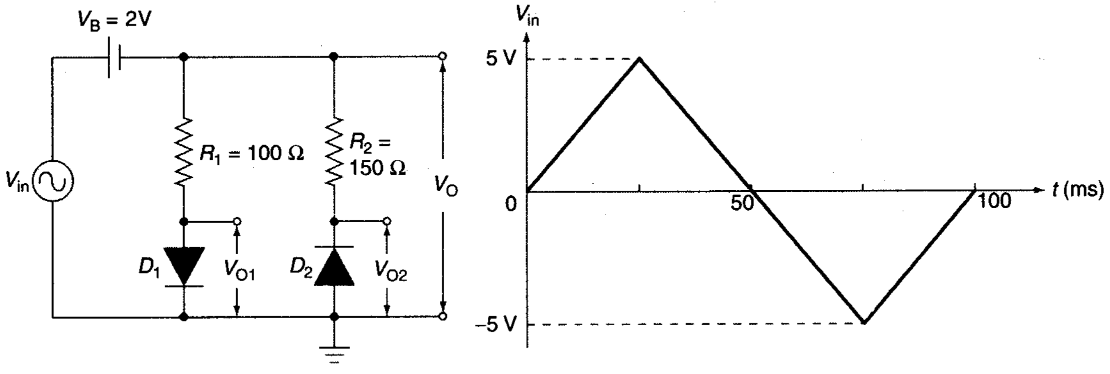


Figure 2.51 | Problem 5.

ANSWERS

Multiple-choice Questions

1. (a)
2. (b)
3. (b)
4. (a)
5. (a)
6. (b)
7. (c)
8. (c)
9. (d)
10. (b)

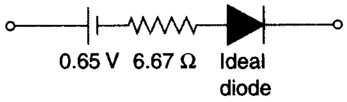
Fill in the Blanks

1. Closed, open
2. Equal
3. Metal-semiconductor
4. Increases
5. Piecewise linear equivalent model

Problems

1. 170 Ω, 32 Ω

2.



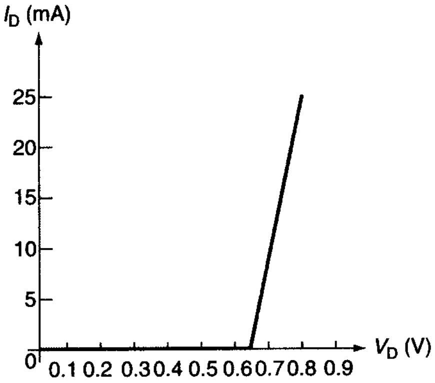


Figure 2.52 Solution to Problem 2.

3.

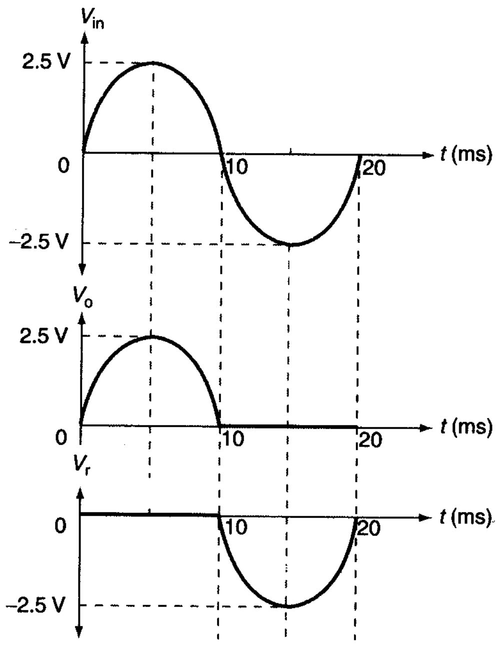


Figure 2.53 | Solution to Problem 3.

4.

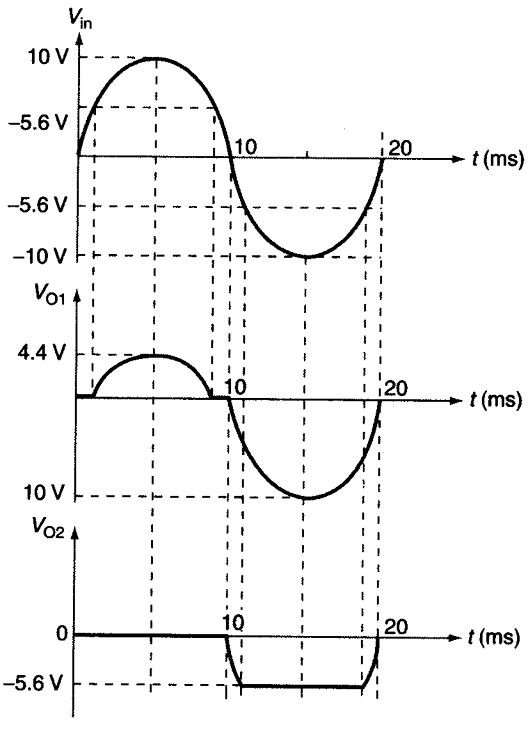


Figure 2.54 | Solution to problem 4.

5.

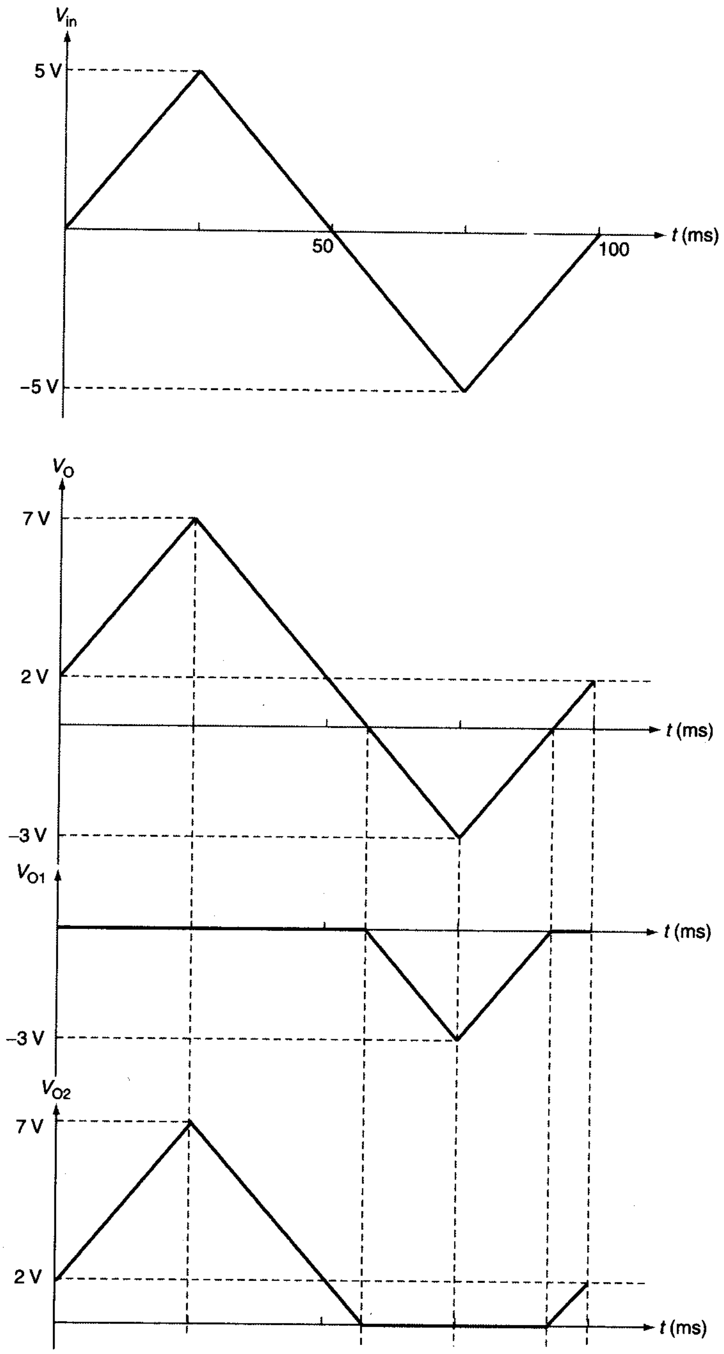


Figure 2.55 | Solution to Problem 5.