5

Applications of Semiconductors

5.1. INTRODUCTION

One of the principal utilities of semiconductors in a wide range of *electronic devices* (i.e., devices that employ the transport properties of carriers in the material) and *optoelectronic devices* (i.e., devices for the generation and detection of light) is related to their capability to form various electrical junctions and resulting electrostatic inhomogeneities and built-in electric fields. These junctions include (i) a *p–n homojunction* (between two semiconductor regions of opposite doping types), (ii) a *metal–semiconductor junction* (or *Schottky barrier*), and (iii) a *heterojunction* (formed between two dissimilar semiconductors that are joined together by deposition or epitaxial growth). Electronic devices based on such junctions are typically employed as rectifying elements (*p–n diodes*), or as parts in various transistors, or in a wide range of light emitting and light detecting devices. Among these, a *p–n* junction is the most important one for both microelectronic and optoelectronic device applications. This chapter will outline the basic characteristics of a *p–n* junction, as well as of other types of junction and semiconductor devices (for extensive discussion on various semiconductor junctions and their applications in various electronic devices, see books in Bibliography Section B2).

5.2. DIODES

5.2.1. The p–n Junction

A *p–n* junction is formed between two semiconductor regions of opposite doping types. Such junctions formed within a single semiconductor are referred to as homojunctions. For the purpose of illustration, this can be depicted by bringing two oppositely doped regions together and aligning their conduction and valence band energies (see Fig. 5.1). (Note that typical methods of the actual formation of a *p–n* junction, which include diffusion, ion implantation, or epitaxial growth, are

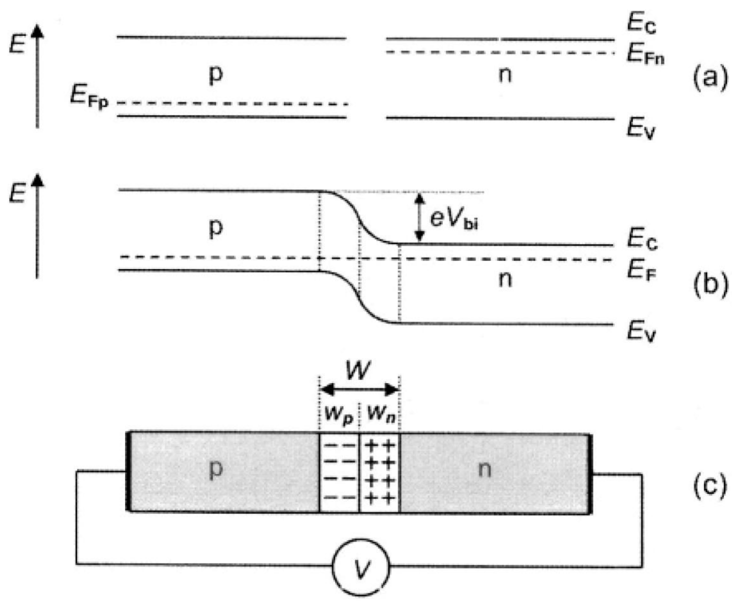

FIGURE 5.1. Schematic illustration of the p–n junction: (a) the energy diagrams of a p- and n-type semiconductors prior to junction formation, (b) the energy diagram after the junction is formed (in thermal equilibrium), and (c) a p–n junction showing the depletion region (or space–charge region).

described briefly further.) It is assumed that (i) the n- and p-doped regions are uniformly doped and (ii) the transition between the two regions is abrupt. (Such a structure is referred to as an *abrupt p–n junction.*) The electrons (in n-type region) and holes (in p-type region) close near the junction diffuse across it. The electrons diffuse into the p-type region, whereas the holes diffuse into the n-type region, and encountering each other, electrons and holes recombine, thus leading to the formation of a region (around the junction) that is depleted of mobile carriers. This region is called the *depletion region* (W). As a result of this diffusion of electrons and holes across the junction, the immobile ionized donors and acceptors in n- and p-type regions are no longer compensated, resulting in the formation of space–charge regions near the junction (see Figs. 5.1 and 5.2). This is associated with the concentrations of acceptor (N_a) and donor (ons, which produce a net negative charge on the p-type side of the junction and a net positive charge on the n-type side, respectively. Such a build-up of oppositely charged regions results in the formation of the junction potential, which effectively prevents further migration of free carriers. Any free carrier, entering the depletion region, is experiencing a force that pushes it back away from the depletion region (i.e., the depletion region is kept free of charge carriers). In other words, the presence of uncompensated charge due to the ionized donors and acceptors produces an electric field that results in a drift of carriers in the opposite direction. The *internal* (*built-in*) *potential* is essentially formed as the result of the Fermi energy difference between the n- and p-type regions. In the thermodynamic equilibrium (no voltage bias is applied), the Fermi level (which is near the valence band edge in a p-type region and near the conduction band in an n-type region) of a p–n junction must be constant across the junction, which necessitates the band bending through

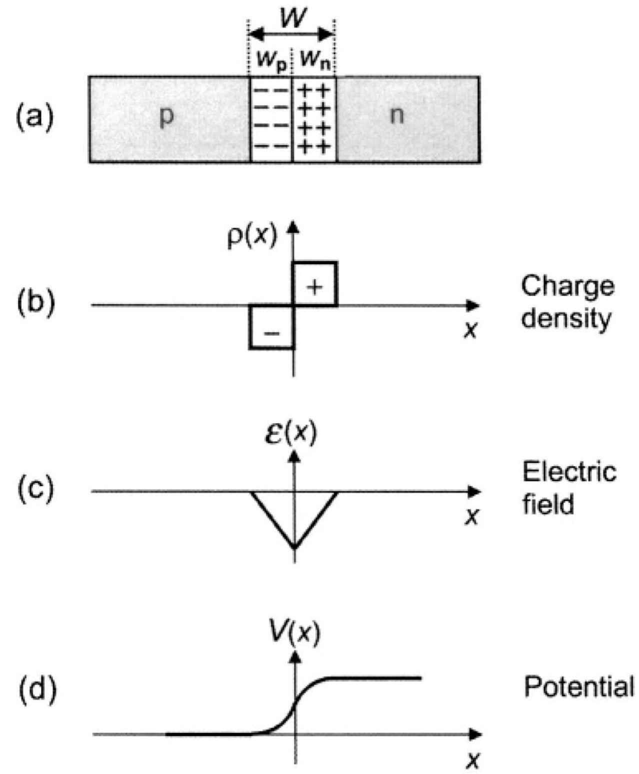

FIGURE 5.2. Schematic illustration of the p–n junction (for $N_a = N_d$) in equilibrium: (a) a p–n junction showing the depletion region (or space–charge region), (b) space–charge distribution, (c) electric field distribution, and (d) potential across junction.

the junction [see Fig. 5.1(b)]. The equilibrium built-in voltage V_{bi} is related to the difference in the Fermi levels (of the two semiconductors) prior to their equalization, i.e.,

$$eV_{bi} = E_{Fn} - E_{Fp} \tag{5.2.1}$$

In order to move across the depletion region, free charge carriers require extra energy to overcome the forces of the space–charge region. In other words, the junction behaves like a barrier for charge flow. Such a barrier is depicted in Fig. 5.1(b) as band bending of the conduction and valence bands in the depletion region. Such a depiction represents the condition of the electrons that now have to "move uphill" in order to traverse across the depletion region from the n-type side to the p-type side. The opposite is true for holes. Thus, free charge carriers require energy to traverse across the depletion region; this can be accomplished by the application of a voltage between the two ends of the p–n junction diode [see Fig. 5.1(c)]. However, depending on polarity, the application of such a voltage may either assist in overcoming the barrier, or vice versa. This effectively results in a rectifying characteristic of a diode, which allows the flow of electrical current in one direction but not in the another. This depends on the *forward-bias* or *reverse-bias* conditions of such a diode. If a voltage is applied to the junction [see Fig. 5.1(c)], it is referred to as *forward biased* when a positive voltage is applied to the p-doped region, and it is *reversed biased* when a negative voltage is applied to the *p*-doped region. Thus, to summarize briefly, the general performance of a *p–n* junction, related to the application of the external bias, can be understood in terms of the *forward-bias* and *reverse-bias* conditions. For forward-bias conditions, the free electrons and holes are pushed towards the junction, thus providing them with additional energy to traverse the junction. Whereas, for reverse-bias conditions, the electrons and holes are pulled away from the junction, making it more difficult for them to traverse the depletion region. This essentially implies that in the forward bias, the potential barrier is lowered, whereas in the reverse bias, the barrier is raised (see Fig. 5.3). Such a performance of the *p–n* junction is employed in many electronic device applications.

Typical methods of the formation of a *p–n* junction include diffusion, ion implantation, or epitaxial growth. In the case of diffusion, the suitable dopant (in sufficient concentration) is diffused (using heat) in the appropriate region, and this results in the formation of a junction. One can also employ ion implantation of, e.g., *n*-type semiconductor with acceptor ions, which results in sharp junctions. (Note that the bombardment with high-energy ions during the implantation process also induces damage to the crystal structure, i.e., it results in the generation of various defects.) It is also possible to employ epitaxial deposition techniques, which allow the formation of various layers of semiconductors together with required dopants included during the growth. Using such techniques allows fabrication of very abrupt junctions (i.e., with no counter-doping in corresponding regions).

As shown in Fig. 5.2, the space–charge density [see Fig. 5.2(b)] changes sharply near the edges of the depletion region. Such a depiction is referred to as the *depletion approximation*, which essentially defines two separate regions, i.e., (i) a depletion region with negligibly low carrier densities and (ii) quasi-neutral (homogeneous)

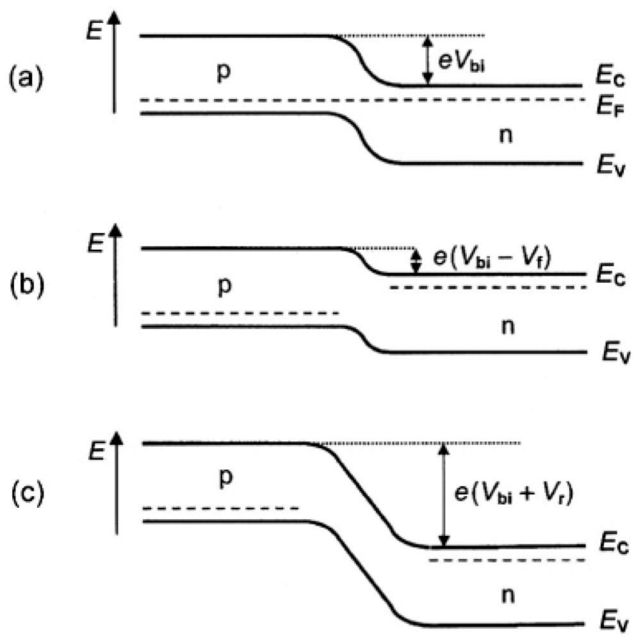

FIGURE 5.3. Schematic illustration of the energy band diagrams of a *p–n* junction (a) in equilibrium, (b) under forward bias, and (c) under reverse bias.

region within which the charge density is assumed to be zero [see Fig. 5.2(b)]. (Note, however, that in practical cases, the carrier densities do not abruptly fall to zero at the edges of the depletion region on both sides, resulting in both the carriers and the space charge having a distribution at the depletion region edge.)

From the analysis of the flow of charge carriers across the *p–n* junction and from the charge neutrality condition for the abrupt junction, one can derive analytical relationships for the diode in equilibrium. In this case, the total current is zero (i.e., the net current across the junction due to both electrons and holes is zero), as the diffusion current and drift current are equal and opposite and, thus, they cancel each other out. Note that (for both types of carrier) the diffusion current is from the *p*-side to the *n*-side, and the drift current is from the *n*-side to the *p*-side. The conditions of the diffusion and drift components cancelling out at equilibrium can be expressed for the *x*-direction as (see Section 4.11)

$$e\mu_e n(x)\mathcal{E}(x) + eD_e \frac{dn(x)}{dx} = 0 \tag{5.2.2}$$

$$e\mu_h p(x)\mathcal{E}(x) - eD_h \frac{dp(x)}{dx} = 0 \tag{5.2.3}$$

The electric field $\mathcal{E}(x) = -dV(x)/dx$, where $V(x)$ is the potential and $n(x)$ and $p(x)$ are the carrier densities at a distance x from the junction. Thus, these equations can be written as

$$-e\mu_e n(x)\frac{dV(x)}{dx} + eD_e \frac{dn(x)}{dx} = 0 \tag{5.2.4}$$

$$e\mu_h p(x)\frac{dV(x)}{dx} + eD_h \frac{dp(x)}{dx} = 0 \tag{5.2.5}$$

For electrons (as an example), Eq. (5.2.4.) can be expressed as

$$\mu_e \frac{dV(x)}{dx} = D_e \frac{1}{n(x)} \frac{dn(x)}{dx} \tag{5.2.6}$$

By integrating this equation over the proper limits (related to the depletion region widths from both sides of an abrupt junction) and by using the Einstein relation ($D = \mu k_B T/e$), one can obtain the following equation:

$$V_n - V_p = \frac{k_B T}{e} \ln \frac{n_n}{n_p} \tag{5.2.7}$$

where V_n and V_p correspond to the potential on each side of the junction, and n_n and n_p correspond to the electron concentration at the each edge of the depletion region. Noting that the potential difference $V_n - V_p = V_{bi}$, and to a good approximation $n_n = N_d$, and [using Eq. (4.4.9): $np = n_i^2$] $n_p = n_i^2/N_a$, the built-in voltage (or barrier voltage) can be related to doping concentrations as

$$V_{bi} = \frac{k_B T}{e} \ln \frac{N_a N_d}{n_i^2} \tag{5.2.8}$$

Thus, Eq. (5.2.8) relates V_{bi} to the given semiconductor parameters, i.e., it depends on the doping of the p- and n-regions, on temperature, and the energy gap E_g [see Eq. (4.4.11): $n_i = (N_c N_v)^{1/2} \exp(-E_g/2k_B T)$ that relates n_i and E_g].

Using Poisson's equation (see Section 4.11), which describes the dependence of the electric field on the space–charge density ρ, i.e., $\nabla \cdot \mathcal{E} = \rho/\varepsilon$, where $\rho = e(N_d^+ - N_a^- - n + p)$, one can derive the electric field distribution within the depletion region. For any point along the x-direction, the electric field gradient can be related to the local space charge as

$$\frac{d\mathcal{E}(x)}{dx} = \frac{e}{\varepsilon}(N_d^+ - N_a^- - n + p) \tag{5.2.9}$$

By neglecting the contributions from n and p within the depletion region, this equation is simplified to (assuming complete ionization)

$$\frac{d\mathcal{E}(x)}{dx} = \frac{e}{\varepsilon} N_d \tag{5.2.10}$$

$$\frac{d\mathcal{E}(x)}{dx} = -\frac{e}{\varepsilon} N_a \tag{5.2.11}$$

for two depletion regions w_n and w_p, respectively (see Fig. 5.1). These equations essentially indicate that within the depletion region the electric field distribution $\mathcal{E}(x)$, which is directed from the n-side to the p-side, has a positive slope on the n-side and a negative slope on the p-side (as shown in Fig. 5.2) with a maximum value $\mathcal{E}_{max}$ of the field at the junction (i.e., $x = 0$) and reaching zero at the boundaries of depletion region (i.e., x_n and $-x_p$ for n- and p-sides, respectively). By integrating Eq. (5.2.10) or (5.2.11) over the proper limits (related to the depletion region boundaries of an abrupt junction), $\mathcal{E}_{max}$ can be related to the built-in voltage V_{bi}. Thus, for the n-side (i.e., for $0 < x < x_n$), one can write

$$\int_{\mathcal{E}_{max}}^{0} d\mathcal{E} = \frac{eN_d}{\varepsilon} \int_0^{x_n} dx \tag{5.2.12}$$

and for the p-side (i.e., for $-x_p < x < 0$), one can write

$$\int_0^{\mathcal{E}_{max}} d\mathcal{E} = -\frac{eN_a}{\varepsilon} \int_{-x_p}^{0} dx \tag{5.2.13}$$

From these equations, one can derive $\mathcal{E}_{\max}$:

$$\mathcal{E}_{\max} = -\frac{eN_d w_n}{\varepsilon} = -\frac{eN_a w_p}{\varepsilon} \tag{5.2.14}$$

Since the electric field $\mathcal{E} = -dV(x)/dx$, where $V(x)$ is the potential, one can relate the electric field with the built-in potential V_{bi} from

$$-V_{bi} = \int_{-x_p}^{x_n} \mathcal{E}(x)\, dx \tag{5.2.15}$$

and,thus

$$-V_{bi} = \frac{\mathcal{E}_{\max}(x_n + x_p)}{2} \tag{5.2.16}$$

or using Eq. (5.2.14) and the fact that $x_n + x_p = W$ (note that essentially x_n and x_p correspond to w_n and w_p), one can write

$$V_{bi} = \frac{eN_d w_n W}{2\varepsilon} = \frac{eN_a w_p W}{2\varepsilon} \tag{5.2.17}$$

From the condition of charge neutrality (i.e., the total negative charge in the p-side depletion region exactly balances the total positive charge in the n-side depletion region, see Fig. 5.1), one can write

$$w_p N_a = w_n N_d \tag{5.2.18}$$

where w_p and w_n are the widths of the p-side and n-side charged regions, respectively. In addition, one can express the total depletion width W as

$$W = w_p + w_n \tag{5.2.19}$$

Thus, one can write

$$w_p = W \frac{N_d}{N_a + N_d} \tag{5.2.20}$$

$$w_n = W \frac{N_a}{N_a + N_d} \tag{5.2.21}$$

The depletion width W (at equilibrium) can be related to doping concentrations and the built-in voltage as

$$W = \left(\frac{2\varepsilon V_{bi}}{e} \frac{N_a + N_d}{N_a N_d}\right)^{1/2} \tag{5.2.22}$$

The expressions for the widths w_p and w_n are

$$w_p = \left(\frac{2\varepsilon V_{bi}}{e} \frac{N_d}{N_a(N_a + N_d)} \right)^{1/2} \tag{5.2.23}$$

$$w_n = \left(\frac{2\varepsilon V_{bi}}{e} \frac{N_a}{N_d(N_a + N_d)} \right)^{1/2} \tag{5.2.24}$$

From the condition of charge neutrality ($w_p N_a = w_n N_d$), one can also conclude that for a specific case, e.g., for $N_a \gg N_d$, one obtains that $w_n \gg w_p$. This indicates that in such a case, the much greater fraction of the depletion region occurs on the *n*-side of the junction, i.e., in the side of the junction with lower doping. In other words, the actual magnitude of W is largely determined by the doping concentration of the lower-doped side of the junction, which basically determines the *p–n* junction characteristics. Such a structure (i.e., when one side of the junction has significantly greater doping concentration than the another side) is referred to as a *one-sided abrupt p–n junction.*

In general, the knowledge about the depletion width W is vital, since (i) it determines the limit on the dimensions of the diode, and (ii) in practical applications of a reverse-biased diode, the depletion width may also determine its breakdown voltage.

For the case of a biased diode, the expression for the depletion width has to include the applied voltage V (which can be either positive or negative for forward or reverse bias, respectively):

$$W = \left(\frac{2\varepsilon(V_{bi} - V)}{e} \frac{N_a + N_d}{N_a N_d} \right)^{1/2} \tag{5.2.25}$$

From this expression, it follows that W is increased (decreased) under reverse (forward) bias conditions (see Fig. 5.3).

Another fundamental consequence of the analysis of the net flow of charge carriers across the *p–n* junction (i.e., analysis of diffusion and recombination of charge carriers) is its current–voltage characteristic, which can be expressed as

$$I = I_0[\exp(eV/k_B T) - 1] \tag{5.2.26}$$

This equation (referred to as the *diode equation*) describes the characteristic of the *p–n* junction which is referred to as *rectification.* (This characteristic is applied to a wide range of applications.) The current across the diode increases exponentially with the application of a forward-bias voltage, whereas the current is limited to a relatively small saturation value with the application of reverse-bias voltage (see Fig. 5.4). Note that this equation and $I(V)$ characteristic in Fig. 5.4 represent an ideal case (in practical devices, however, deviations from this characteristic may occur). The saturation current I_0 depends on various parameters related to the minority carriers as

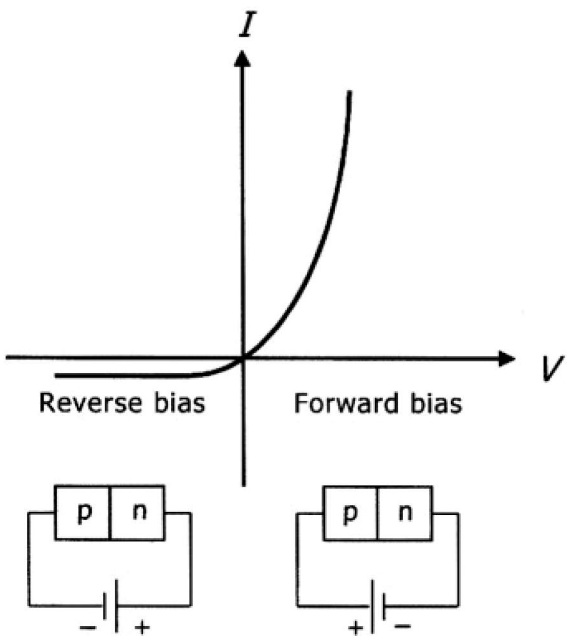

FIGURE 5.4. Schematic representations of the current–voltage curve of a p–n junction diode and of the corresponding circuits.

$$I_0 = eA\left(\frac{D_e n_p}{L_e} + \frac{D_h p_n}{L_h}\right) \tag{5.2.27}$$

where A is the cross-sectional area of the junction, D_e and D_h are the diffusion constants of electrons and holes, L_e and L_h are the diffusion lengths of electrons and holes, and n_p and p_n are the equilibrium minority carrier concentrations of electrons in p-region and holes in n-region, respectively. (Note that the diffusion constant $D = \mu k_B T/e$.) As mentioned earlier, the relationship between the minority carrier diffusion length and the diffusion constant is $L = (D\tau)^{1/2}$, where τ is the minority carrier lifetime. Using Eq. (4.4.9), i.e., $np = n_i^2$, and also recalling that for *shallow impurities* at room temperature almost the entire donor or acceptor sites are ionized and the free carrier density corresponds to the impurity concentration, the above expression for I_0 can be expressed in terms of dopant concentrations. Note that in the case of donors, the electron density n equals the concentration of donors (i.e., $n \cong N_d$), and in the case of acceptors, the density of holes p equals the concentration of acceptors (i.e., $p \cong N_a$). Thus, $n_p = n_i^2/p_p = n_i^2/N_a$, and $p_n = n_i^2/n_n = n_i^2/N_d$. Accordingly, I_0 can be expressed as

$$I_0 = en_i^2 A\left(\frac{D_e}{L_e N_a} + \frac{D_h}{L_h N_d}\right) \tag{5.2.28}$$

or alternatively, by using the expression $L = (D\tau)^{1/2}$,

$$I_0 = en_i^2 A\left[\frac{1}{N_a}\left(\frac{D_e}{\tau_e}\right)^{1/2} + \frac{1}{N_d}\left(\frac{D_h}{\tau_h}\right)^{1/2}\right] \tag{5.2.29}$$

These equations indicate the dependence of the reverse saturation current on the properties of the minority carriers. Thus, I_0 decreases as the carrier lifetimes and/or the doping concentrations increase.

Due to the charge separation in the depletion region (i.e., the presence of two layers of space charge in the depletion region), it behaves like a capacitor, and the (junction) capacitance can be expressed as

$$C = A\left[\frac{e\varepsilon N_a N_d}{2(N_a + N_d)}\right]^{1/2} \frac{1}{(V_{bi} - V)^{1/2}} \tag{5.2.30}$$

Thus, a *p–n* junction capacitance can be varied by the applied voltage; this property can be employed in electronic circuits. Such junction devices that are employed for their voltage-controlled variable capacitance are referred to as *varactor diodes.*

As mentioned earlier, Eq. (5.2.26) for $I(V)$ characteristic (see Fig. 5.4) represents an ideal case. However, in practical devices, various breakdown phenomena and deviations from the ideal characteristic may occur. The basic breakdown mechanisms include *Zener* (or *tunnel*) *breakdown* and *avalanche breakdown.* In the case of Zener breakdown, which typically occurs in heavily doped diodes at low reverse voltages, the heavy doping results in a very narrow barrier width, and the valence band electrons that are at sufficiently short distances from empty states in the conduction band, tunnel through the barrier. In the case of avalanche breakdown, which occurs at high reverse voltages, the carriers, gaining sufficient energy in the junction electric field, cause ionizing collisions that produce additional electron–hole pairs which may result in an uncontrolled current flow and rapid increase in the reverse current.

5.2.2. Schottky Barrier

The semiconductor diode structures can also be obtained by joining two dissimilar materials of high purity. For example, bringing a metal into a contact with a semiconductor (e.g., by depositing a metal onto a semiconductor) can form a metal–semiconductor diode structure (a *Schottky barrier*). Depending on the relative values of the work functions of the metal and a semiconductor and on the type of a semiconductor (i.e., *n*-type or *p*-type), one can obtain either rectifying junctions or ohmic contacts. From the equilibrium criterion of the Fermi level equalization across the junction, it follows that if the work function of the metal $e\Phi_m$ is greater than that of *n*-type semiconductor $e\Phi_s$, the flow of electrons from the semiconductor into a metal occurs. This results in the formation of a depletion layer (with a positive space charge) in a semiconductor near the junction and the accumulation of a negative surface charge on a metal (see Fig. 5.5). This is accompanied by the upward bending (within a depletion region in a semiconductor) of the energy band edges and the formation of a potential barrier having a height of $eV_{bi} = e(\Phi_m - \Phi_s)$, which prevents the electron diffusion from the conduction band of a semiconductor into the metal (see Fig. 5.5). Note that