Semiconductor *p-n* and Metal-Semiconductor Junctions

9

9.1 Introduction

Until now, our discussion was based solely on homogeneous semiconductors whose properties are uniform in space. Although a few devices can be made from such semiconductors, the majority of devices and the most important ones utilize nonhomogeneous semiconductor structures. Most of them involve semiconductor *p-n* junctions, in which a *p*-type doped region and an *n*-type doped region are brought into contact. Such a junction actually forms an electrical diode. This is why it is usual to talk about a *p-n* junction as a diode. Another important structure involves a semiconductor in intimate contact with a metal, leading to what is called a metal-semiconductor junction. Under certain circumstances, this configuration can also lead to an electrical diode.

The objective of this chapter will first be to establish an accurate model for the *p-n* junction which can be at the same time mathematically described. This model will be the ideal p-n junction diode. The basic properties of this ideal p-n junction at equilibrium will be described in detail. The non-equilibrium properties of this *p-n* junction will then be discussed by deriving the diode equation which relates the current and voltage across the diode. Deviations from the ideal diode case will also be described. Finally, this chapter will also discuss the properties of metal-semiconductor junctions and compare them with those of *p-n* junctions.

9.2 Ideal *p-n* Junction at Equilibrium

9.2.1 Ideal *p-n* Junction

The ideal *p-n* junction model is also called the abrupt junction or step junction model. This is an idealized model for which we assume that the material is uniformly doped *p*-type with a total acceptor concentration N_A on one side of the junction (e.g.,

M. Razeghi, *Fundamentals of Solid State Engineering*,
https://doi.org/10.1007/978-3-319-75708-7_9

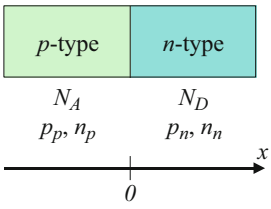

Fig. 9.1 Ideal *p-n* junction model, in which one side of the junction is a purely *p*-type semiconductor and the other a purely *n*-type semiconductor. Both materials are uniformly doped

$x < 0$), and the material is uniformly doped *n*-type with a total donor concentration N_D on the other side (e.g., $x > 0$). For further simplicity, we will consider a homojunction, i.e., both doped regions are of the same semiconductor material. We will restrict our analysis to the one-dimensional case, as illustrated in Fig. 9.1.

In the *p*-type doped region far from the junction area, the equilibrium hole and electron concentrations are denoted p_p and n_p, respectively. In the *n*-type doped region far from the junction area, the hole and electron concentrations are denoted p_n and n_n, respectively. These carrier concentrations satisfy the mass action law in Eq. (7.31):

$$p_p n_p = p_n n_n = n_i^2 \tag{9.1}$$

where n_i is the intrinsic carrier concentration in the semiconductor material considered. We further assume that all the dopants are ionized, which leads to the following carrier concentrations for the *p*- and *n*-type regions, respectively:

$$\begin{cases} p_p = N_A \left(10^{16}\text{cm}^{-3}\right) \\ n_p = \dfrac{n_i^2}{N_A}\left(10^{5}\text{cm}^{-3}\right) \end{cases} \text{and} \begin{cases} n_n = N_D \left(10^{17}\text{cm}^{-3}\right) \\ p_n = \dfrac{n_i^2}{N_D}\left(10^{4}\text{cm}^{-3}\right) \end{cases} \tag{9.2}$$

A few typical values for these concentrations are given in parenthesis. It is important to remember that both a *p*-type, and an *n*-type, isolated semiconductors are electrically neutral.

9.2.2 Depletion Approximation

However, when bringing a *p*-type semiconductor into contact with an *n*-type semiconductor, the material is not electrically neutral everywhere anymore. Indeed, on one side of the junction area, for $x < 0$, there is a high concentration of holes, whereas on the other side there is a low concentration of holes. This asymmetry in carrier density results in the diffusion of holes across the junction as shown in Fig. 9.1. By doing so, the holes leave behind uncompensated acceptors ($x < 0$) which are negatively charged. A similar analysis can be carried out for electrons as there is also an asymmetry in the density of electrons on either side of the *p-n* junction. This leads to their diffusion and makes the material positively charged for $x > 0$ as the electrons leave behind uncompensated donors, as shown in Fig. 9.2.

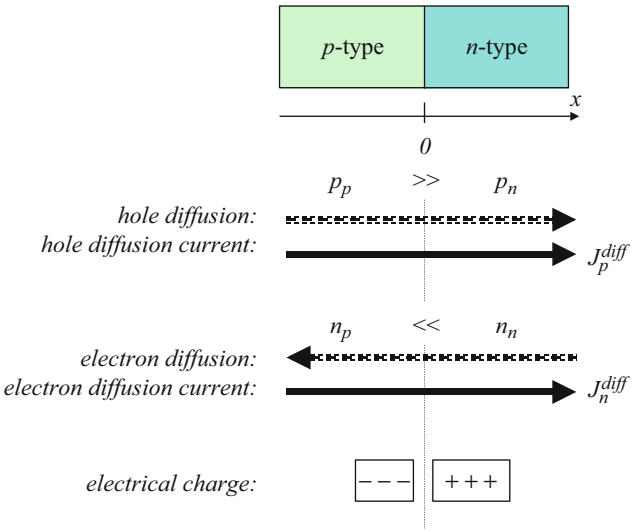

Fig. 9.2 Hole and electron diffusion across a *p-n* junction. The holes diffuse from the left to the right, which leads to a diffusion electrical current from the left to the right as well. By contrast, the electrons diffuse from the right to the left, but this leads to a diffusion electrical current from the left to the right because of the negative charge of electrons. The diffusion process leaves uncompensated acceptors in the *p*-type region and donors in the *n*-type regions, i.e., a net negative charge in the *p*-type region and a net positive charge in the *n*-type region. The presence of these charges results in a built-in electric field

This redistribution of electrical charge does not endure indefinitely. Indeed, as positive and negative charges appear on the $x > 0$ and $x < 0$ sides of the junction, respectively, an electric field strength $E(x)$, called the built-in electric field, will result and is shown in Fig. 9.3 As discussed in Chap. 8, this electric field will generate the drift of the positively charged holes and the negatively charged electrons. By comparing Figs. 9.2 and 9.3, we can see that the drift of these charge carriers counteracts the previous diffusion process. An equilibrium state is reached when the diffusion currents $J^{\text{diffusion}}$ and drift currents J^{drift} are exactly balanced for each type of carrier, i.e., holes and electrons taken independently:

$$\begin{cases} J_{\text{h}}^{\text{diff}} + J_{\text{h}}^{\text{drift}} = 0 \\ J_{\text{e}}^{\text{diff}} + J_{\text{e}}^{\text{drift}} = 0 \end{cases} \tag{9.3}$$

There is a transition region around the *p-n* junction area with a width W_0 in which the electrical charges are present. This region is called the space charge region and is schematically shown in Fig. 9.4a. The charge distribution within this region is modeled as follows: we consider that there is a uniform concentration of negative charges for $-x_{\text{p0}} < x < 0$ equal to $Q(x) = -qN_{\text{A}}$ (where N_{A} is the total concentration of acceptors in the *p*-type region) and a uniform concentration of positive charges for $0 < x < x_{\text{n0}}$ and equal to $Q(x) = +qN_{\text{D}}$ (where N_{D} is the total concentration of donors

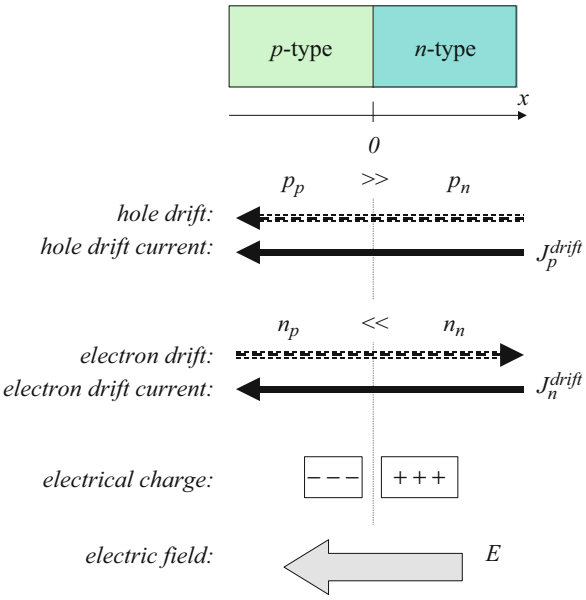

Fig. 9.3 Hole and electron drift across a *p-n* junction. Under the influence of the built-in electric field, the holes drift from the right to the left, which leads to a drift electrical current from the right to the left as well. By contrast, the electrons drift from the left to the right, but this leads to a drift electrical current from the right to the left because of the negative charge of electrons. The drift process counterbalances the diffusion of charge carriers in order to bring the system into equilibrium

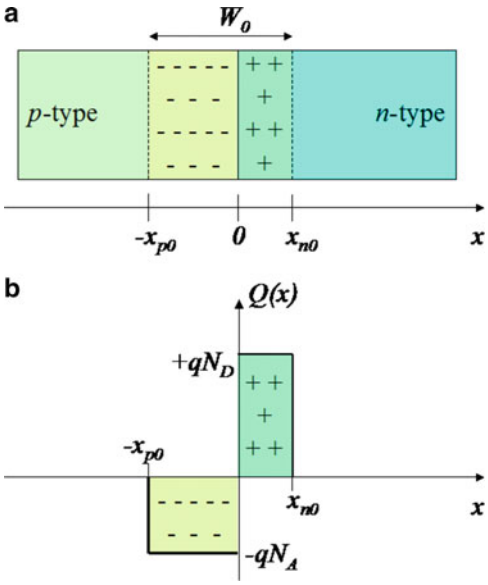

Fig. 9.4 (**a**) Space charge region in a *p-n* junction. Near the junction area, the *p*-type region is negatively charged as a result of the diffusion of charge carriers. (**b**) Electrical charge density in a *p-n* junction. To keep the overall charge neutrality, the total number of negative charges in the *p*-type region is equal to the total number of positive charges in the *n*-type region. In the depletion approximation, the charges are assumed uniformly distributed in space, within the depletion region delimited by $-x_{p0}$ and x_{n0}

in the *n*-type region). The quantities x_{p0} and x_{n0} are positive and express how much the space charge region extends on each side of the junction, as illustrated in Fig. 9.4b. The width of the space charge region, also called depletion width, is then given by:

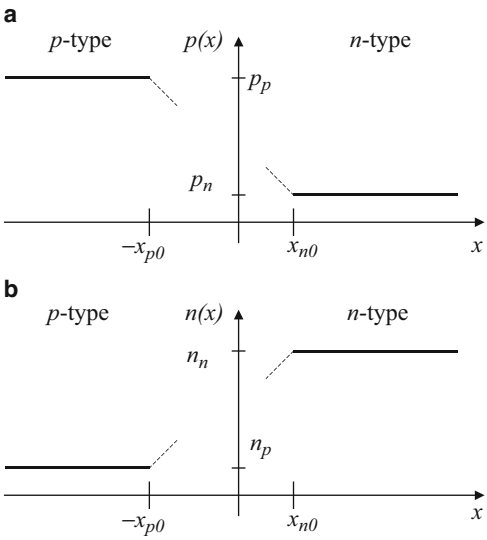

Fig. 9.5 (**a**) Hole and (**b**) electron concentrations in a *p-n* junction. In the depletion approximation, the hole and electron concentrations are assumed to be constant and equal to their equilibrium values outside of the depletion region

$$W_0 = x_{\mathrm{n0}} + x_{\mathrm{p0}} \tag{9.4}$$

Outside of this space charge region, we assume that the semiconductor is electrically neutral without any charge depletion and that the hole and electron concentrations are given by Eq. (9.2). These regions will be called the bulk *p*-type and bulk *n*-type region. The carrier concentrations must therefore somehow go from a high value on one side of the junction to a low value on the other side, and this occurs within the space charge region, as illustrated in Fig. 9.4 In particular, we have (Fig. 9.5):

$$\begin{cases} p\left(-x_{\mathrm{p0}}\right) = p_{\mathrm{p}} & \text{and} \quad p(x_{\mathrm{n0}}) = p_{\mathrm{n}} \\ n\left(-x_{\mathrm{p0}}\right) = n_{\mathrm{p}} & \text{and} \quad n(x_{\mathrm{n0}}) = n_{\mathrm{n}} \end{cases} \tag{9.5}$$

This model is called the depletion approximation. In this model, there are no free holes or electrons in the space charge region: the depletion of carriers is complete. The electric field exists only within this space charge region.

Because the entire *p-n* structure must globally remain electrically neutral, and therefore the space charge region must be neutral as a whole, we must equate the total number of negative charges on one side of the junction to the total number of positive charges on the other side, i.e.:

$$qAN_{\mathrm{A}}x_{\mathrm{p0}} = qAN_{\mathrm{D}}x_{\mathrm{n0}}$$

where A is the cross-section area of the junction, and after simplification:

$$N_{\mathrm{A}}x_{\mathrm{p0}} = N_{\mathrm{D}}x_{\mathrm{n0}} \tag{9.6}$$

Combining Eqs. (9.4) and (9.6), we can express the quantities x_{p0} and x_{n0} as a function of the depletion width W_0:

$$\begin{cases} x_{p0} = \dfrac{N_D}{N_A + N_D} W_0 \\ x_{n0} = \dfrac{N_A}{N_A + N_D} W_0 \end{cases} \tag{9.7}$$

These show that the space charge region extends more in the *p*-type region than in the *n*-type region when $N_D > N_A$ and reciprocally.

Example

Q Estimate the thickness ratio of the depletion region in the *p*-type side ($N_A = 10^{18}$ cm^{-3}) and the *n*-type side ($N_D = 10^{17}$ cm^{-3}) for an abrupt *p-n* junction in the depletion approximation.

A The thicknesses of the depletion region in the *p*-type side and the *n*-type side are denoted x_{p0} and x_{n0}, respectively. Their ratio is such that:

$$\frac{x_{p0}}{x_{n0}} = \frac{N_D}{N_A} = \frac{10^{17}}{10^{18}} = 0.1 \cdot$$

9.2.3 Built-In Electric Field

The built-in electric field strength can be calculated using Gauss's law which can be written in our one-dimensional model as:

$$\frac{dE(x)}{dx} = \frac{Q(x)}{\varepsilon} \tag{9.8}$$

where ε is the permittivity of the semiconductor material and $Q(x)$ is the total charge concentration. This relation can be rewritten for either sides of the junction:

$$\begin{cases} \dfrac{dE(x)}{dx} = -\dfrac{qN_A}{\varepsilon} & \text{for} - x_{p0} < x < 0 \\ \dfrac{dE(x)}{dx} = \dfrac{qN_E}{\varepsilon} & \text{for} \, 0 < x < x_{n0} \end{cases} \tag{9.9}$$

From these relations we see that the electric field strength varies linearly on either side of the junction. By integrating Eq. (9.8) using the boundary conditions assumed in the depletion approximation:

$$E(-x_{p0}) = E(x_{n0}) = 0 \tag{9.10}$$

that the electric field strength is equal to zero at the limits of the space charge region ($x = -x_{p0}$ and $x = x_{n0}$), we obtain successively:

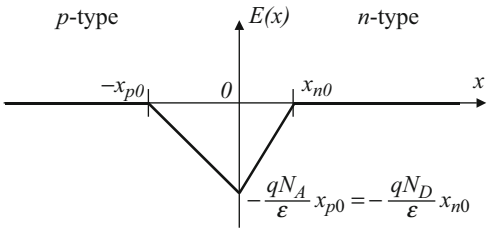

Fig. 9.6 Built-in electric field strength profile across a *p-n* junction. In the depletion approximation, the electric field strength is zero outside the depletion region because there is no net electrical charge. Within the depletion region, the electric field strength varies linearly with distance

$$\begin{cases} E(x) = \int_{-x_{p0}}^{x} dEdx = \int_{-x_{p0}}^{x} -\frac{qN_A}{\varepsilon}dx & \text{for } -x_{p0} < x < 0 \\ E(x) = \int_{x_{n0}}^{x} dEdx = \int_{x_{n0}}^{x} \frac{qN_D}{\varepsilon}dx & \text{for } 0 < x < x_{n0} \end{cases}$$

$$\begin{cases} E(x) = -\frac{qN_A}{\varepsilon}(x + x_{p0}) & \text{for } -x_{p0} < x < 0 \\ E(x) = \frac{qN_D}{\varepsilon}(x - x_{n0}) & \text{for } 0 < x < x_{n0} \end{cases} \tag{9.11}$$

For $x = 0$, we obtain two expressions for the electric field strength from the two previous expressions for $E(x)$:

$$\begin{cases} E(0) = -\frac{qN_A}{\varepsilon}(x_{p0}) \\ E(0) = \frac{qN_D}{\varepsilon}(-x_{n0}) \end{cases} \tag{9.12}$$

And these expressions are equal, according to Eq. (9.6). Therefore, the global electrical neutrality of the *p-n* structure ensures the continuity of the built-in electric field strength. A plot of $E(x)$ is shown in Fig. 9.6.

9.2.4 Built-In Potential

As a result of the presence of an electric field, an electrical potential $V(x)$ also exists and is related to the electric field strength through:

$$E(x) = -\frac{dV(x)}{dx} \tag{9.13}$$

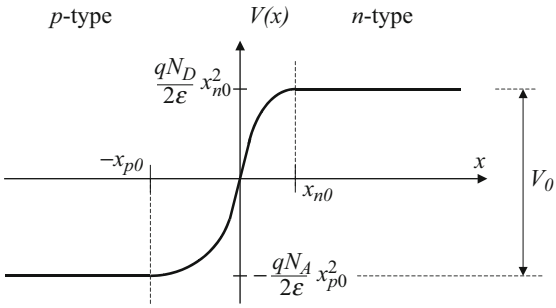

Fig. 9.7 Built-in potential profile across a *p-n* junction. In the depletion approximation, there is no variation of the potential outside the depletion region

The potential is constant outside the space charge region because the electric field strength is equal to zero there. An analytical expression for the electrical potential can be obtained by integrating Eq. (9.11):

$$\begin{cases} V(x) = \dfrac{qN_{\mathrm{A}}}{\varepsilon}\left(\dfrac{x^2}{2} + x_{\mathrm{p0}}x\right) & \text{for} - x_{\mathrm{p0}} < x < 0 \\ V(x) = -\dfrac{qN_{\mathrm{D}}}{\varepsilon}\left(\dfrac{x^2}{2} - x_{\mathrm{n0}}x\right) & \text{for } 0 < x < x_{\mathrm{n0}} \end{cases} \tag{9.14}$$

where we chose the origin of the potential at $x = 0$ and applied the continuity condition of the potential at $x = 0$. This potential is plotted in Figs. 9.6 and 9.7.

The total potential difference across the *p-n* junction is called the built-in potential and is conventionally denoted V_{bi} or V_0. It can be obtained by evaluating the potential difference between $x = -x_{\mathrm{p0}}$ and $x = x_{\mathrm{n0}}$:

$$V_0 = V(x_{\mathrm{n0}}) - V\left(-x_{\mathrm{p0}}\right) \tag{9.15}$$

This can be rewritten as:

$$V_0 = \frac{qN_{\mathrm{D}}}{\varepsilon}\frac{x_{\mathrm{n}}^2 0}{2} + \frac{qN_{\mathrm{A}}}{\varepsilon}\frac{x_{\mathrm{p}}^2 0}{2} \tag{9.16}$$

Expressing $-x_{\mathrm{p0}}$ and x_{n0} as a function of the depletion width given in Eq. (9.7), we obtain:

$$V_0 = \frac{q}{2\varepsilon}\frac{N_{\mathrm{A}}N_{\mathrm{D}}}{(N_{\mathrm{A}} + N_{\mathrm{D}})}W_0^2 \tag{9.17}$$

Another independent expression of the built-in potential can be obtained by expressing the balancing of the diffusion and drift currents. In Chap. 8 we determined analytical expressions for these currents in Eqs. (8.12) and (8.38) for holes and Eqs. (8.12) and (8.36) for electrons. The total current from the motion of holes and that from the motion of electrons are given by:

$$\begin{cases} J_{\mathrm{h}}^{\mathrm{diff}} + J_{\mathrm{h}}^{\mathrm{drift}} = -qD_{\mathrm{p}} \dfrac{dp(x)}{dx} + q\mu_{\mathrm{h}} p(x) E(x) \\ J_{\mathrm{e}}^{\mathrm{diff}} + J_{\mathrm{e}}^{\mathrm{drift}} = qD_{\mathrm{n}} \dfrac{dn(x)}{dx} + q\mu_{\mathrm{e}} n(x) E(x) \end{cases} \tag{9.18}$$

In these expressions, $p(x)$ and $n(x)$ represent the hole and electron concentrations at a position x. Taking into account the condition of Eq. (9.3) stating the exact balancing of the diffusion and drift currents for holes and electrons, we can write:

$$\begin{cases} D_{\mathrm{p}} \dfrac{dp(x)}{dx} = \mu_{\mathrm{h}} p(x) E(x) \\ D_{\mathrm{n}} \dfrac{dn(x)}{dx} = -\mu_{\mathrm{e}} n(e) E(x) \end{cases} \tag{9.19}$$

which can be rewritten using Eq. (9.19) as:

$$\begin{cases} \dfrac{D_{\mathrm{p}}}{\mu_{\mathrm{h}}} \dfrac{1}{p(x)} \dfrac{dp(x)}{dx} = -\dfrac{dV(x)}{dx} \\ \dfrac{D_{\mathrm{n}}}{\mu_{\mathrm{e}}} \dfrac{1}{n(x)} \dfrac{dn(x)}{dx} = \dfrac{dV(x)}{dx} \end{cases}$$

By integrating these equations, we get successively:

$$\begin{cases} \dfrac{D_{\mathrm{p}}}{\mu_{\mathrm{h}}} \displaystyle\int_{-x_{\mathrm{p0}}}^{x_{\mathrm{n0}}} \frac{1}{p(x)} \frac{dp(x)}{dx} dx = -\int_{-x_{\mathrm{p0}}}^{x_{\mathrm{n0}}} \frac{dV(x)}{dx} dx \\ \dfrac{D_{\mathrm{n}}}{\mu_{\mathrm{e}}} \displaystyle\int_{-x_{\mathrm{p0}}}^{x_{\mathrm{n0}}} \frac{1}{n(x)} \frac{dn(x)}{dx} dx = \int_{-x_{\mathrm{p0}}}^{x_{\mathrm{n0}}} \frac{dV(x)}{dx} dx \end{cases}$$

Using Eqs. (9.5) and (9.15), and by taking into account the Einstein relations $\frac{D_{\mathrm{p}}}{\mu_{\mathrm{h}}} = \frac{D_{\mathrm{n}}}{\mu_{\mathrm{e}}} = \frac{k_{\mathrm{b}}T}{q}$ obtained from Eqs. (8.44) and (8.45), we get:

$$\begin{cases} \dfrac{k_{\mathrm{b}}T}{q} \displaystyle\int_{p_{\mathrm{p}}}^{p_{\mathrm{n}}} \frac{dp}{p} = -\int_{0}^{V_0} dV \\ \dfrac{k_{\mathrm{b}}T}{q} \displaystyle\int_{n_{\mathrm{p}}}^{n_{\mathrm{n}}} \frac{dn}{n} = \int_{0}^{V_0} dV \end{cases}$$

which integrates easily into:

$$\begin{cases} \dfrac{k_b T}{q} \ln\left(\dfrac{p_n}{p_p}\right) = -V_0 \\ \dfrac{k_b T}{q} \ln\left(\dfrac{n_n}{n_p}\right) = V_0 \end{cases}$$

i.e.:

$$V_0 = \frac{k_b T}{q} \ln\left(\frac{p_p}{p_n}\right) = \frac{k_b T}{q} \ln\left(\frac{n_n}{n_p}\right) \tag{9.20}$$

This can be rewritten into the form:

$$\frac{p_p}{p_n} = \frac{n_n}{n_p} = \exp\left(\frac{qV_0}{k_b T}\right) \tag{9.21}$$

Using the expressions in Eq. (9.21), we can write the built-in potential as a function of the doping concentrations:

$$V_0 = \frac{k_b T}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{9.22}$$

This potential exists at equilibrium and is a direct consequence of the junction between dissimilarly doped materials. However, it cannot be directly measured using a voltmeter because voltmeters measure the chemical potential difference, and the chemical potential is the same throughout the device since it is at thermal equilibrium with balanced drift and diffusion currents everywhere.

9.2.5 Depletion Width

It is now possible to relate the width W_0 of the space charge region, as well as its extent on either side of the *p-n* junction, with the built-in potential. From the expression of the built-in potential in (Eq. 9.22), we can express the depletion width as:

$$W_0 = \sqrt{\frac{2\varepsilon}{q}\left(\frac{N_A + N_D}{N_A N_D}\right) V_0} \tag{9.23}$$

which becomes, after considering Eq. (9.22):

$$W_0 = \sqrt{\frac{2\varepsilon k_b T}{q^2}\left(\frac{N_A + N_D}{N_A N_D}\right) \ln\left(\frac{N_A N_D}{n_i^2}\right)} \tag{9.24}$$

The extent of the depletion width into each side of the *p-n* junction can then be determined by replacing W_0 from Eq. (9.23) into Eq. (9.7):

$$\begin{cases} x_{p0} = \sqrt{\dfrac{2\varepsilon}{q}\left(\dfrac{N_D}{N_A(N_A+N_D)}\right)V_0} \\ x_{n0} = \sqrt{\dfrac{2\varepsilon}{q}\left(\dfrac{N_A}{N_D(N_A+N_D)}\right)V_0} \end{cases} \tag{9.25}$$

These last two expressions show that the space charge region extends more into the region of lower doping, in accordance with Subsect. 9.2.2.

Example

Q Consider a GaAs abrupt *p-n* junction with a doping level on the *p*-type side of $N_A = 2 \times 10^{17}$ cm^{-3} and a doping level on the *n*-type side of $N_D = 1 \times 10^{17}$ cm^{-3}. Estimate the depletion region widths on the *p*-type side and the *n*-type side at 300 K.

A The depletion region widths sought are given by the following expressions:

$$\begin{cases} x_{p0} = \sqrt{\dfrac{2\varepsilon}{q}\left(\dfrac{N_D}{N_A(N_A+N_D)}\right)V_0} \\ x_{n0} = \sqrt{\dfrac{2\varepsilon}{q}\left(\dfrac{N_A}{N_D(N_A+N_D)}\right)V_0} \end{cases}$$

where ε is the dielectric constant of GaAs ($\varepsilon = 13.1\varepsilon_0$) and V_0 is the built-in potential. The latter is calculated from:

$$\begin{aligned} V_0 &= \frac{k_b T}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \\ &= \frac{\left(1.38066 \times 10^{-23}\right) \times 300}{1.60218 \times 10^{-19}} \ln\left(\frac{\left(2 \times 10^{17}\right)\left(1 \times 10^{17}\right)}{\left(1.79 \times 10^{6}\right)^2}\right) \\ &= 1.297\ V \end{aligned}$$

because the intrinsic carrier concentration in GaAs at 300 K is $n_i = 1.79 \times 10^6$ cm^{-3}. The widths can then be calculated as:

$$\begin{aligned} x_{p0} &= \sqrt{\frac{2\varepsilon}{q}\left(\frac{N_D}{N_A(N_A+N_D)}\right)V_0} \\ &= \sqrt{\begin{array}{l}\dfrac{2 \times \left(13.1 \times 8.85418 \times 10^{-14}\right)}{1.60218 \times 10^{-19}} \\ \times\left(\dfrac{1 \times 10^{17}}{\left(2 \times 10^{17}\right)\left(2 \times 10^{17} + 1 \times 10^{17}\right)}\right) \times 1.297\end{array}} \\ x_{p0} &= 5.6 \times 10^{-6}\text{cm} \\ x_{p0} &= 56\ \text{nm} \end{aligned}$$

and

$$\begin{aligned} x_{n0} &= \sqrt{\frac{2\varepsilon}{q}\left(\frac{N_A}{N_D(N_A+N_D)}\right)V_0} \\ &= \frac{N_A}{N_D}x_{p0} \\ &= \frac{2\times 10^{17}}{1\times 10^{17}}5.6\times 10^{-6} \\ &= 11.2\times 10^{-6}\text{cm} \\ x_{n0} &= 112\text{ nm} \end{aligned}$$

9.2.6 Energy Band Profile and Fermi Energy

Because of the presence of a built-in potential, the allowed energy bands in the semiconductor, e.g., the conduction and the valence bands in particular, are shifted too. The resulting energy band profile is obtained by multiplying the potential by the charge of an electron ($-q$). This is shown in Fig. 9.10e, where it is conventional to plot the bottom of the conduction band (E_C) and the top of the valence band (E_V) across the *p-n* structure.

The reason why we must multiply by the negative charge of an electron is because the resulting band diagram corresponds to the allowed energy states for electrons. This is intuitively understandable because the electrons are more likely to be where there is a higher positive electrical potential; thus the energy band for electrons will be lower there.

We therefore see that the conduction and valence bands are "bent" from the *p*-type to the *n*-type regions. Moreover, the amount of band bending is directly related to the built-in potential:

$$E_{Vp} - E_{Vn} = E_{Cp} - E_{Cn} = qV_0 \tag{9.26}$$

Example

Q Estimate the energy band bending from the *p*-type side to the *n*-type side in a GaAs abrupt *p-n* junction with a doping level on the *p*-type side of $N_A = 2\times 10^{17}$ cm^{-3} and a doping level on the *n*-type side of $N_D = 1\times 10^{17}$ cm^{-3} at 300 K.

A From the previous example, we know that the built-in potential is $V_0 = 1.297$ V. The band bending is therefore equal to $qV_0 = 1.297$ eV.

Away from the space charge region, the Fermi energies in the *p*-type and *n*-type regions are denoted E_{Fp} and E_{Fn}, respectively, as shown in Fig. 9.8 . At equilibrium, these quantities must be equal. Indeed, the hole density in the *p*-type and *n*-type regions is given by Eq. (7.29) in the nondegenerate case:

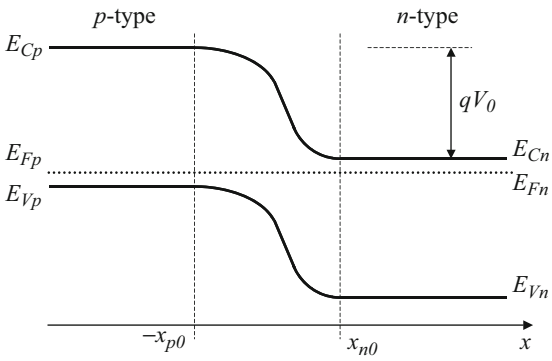

Fig. 9.8 Energy band profile across a *p-n* junction. This profile is obtained by multiplying the potential in Fig. 9.6 by −q, the electrical charge of electrons

$$\begin{cases} p_{\mathrm{p}} = N_{\mathrm{v}} \exp\left(\dfrac{E_{\mathrm{Vp}} - E_{\mathrm{Fp}}}{k_{\mathrm{b}}T}\right) \\ p_{\mathrm{n}} = N_{\mathrm{v}} \exp\left(\dfrac{E_{\mathrm{Vn}} - E_{\mathrm{Fn}}}{k_{\mathrm{b}}T}\right) \end{cases} \tag{9.27}$$

Utilizing Eq. (9.25), we get:

$$\exp\left(\frac{qV_0}{k_{\mathrm{b}}T}\right) = \frac{p_{\mathrm{p}}}{p_{\mathrm{n}}} = \frac{\exp\left(\frac{E_{\mathrm{Vp}} - E_{\mathrm{Fp}}}{k_{\mathrm{b}}T}\right)}{\exp\left(\frac{E_{\mathrm{Vn}} - E_{\mathrm{Fn}}}{k_{\mathrm{b}}T}\right)}$$

$$\exp\left(\frac{qV_0}{k_{\mathrm{b}}T}\right) = \exp\left(\frac{E_{\mathrm{Vp}} - E_{\mathrm{Vn}}}{k_{\mathrm{b}}T}\right) \exp\left(\frac{E_{\mathrm{Fn}} - E_{\mathrm{Fp}}}{k_{\mathrm{b}}T}\right) \tag{9.28}$$

In addition, by using Eq. (9.26) in this expression, we get:

$$1 = \exp\left(\frac{E_{\mathrm{Fn}} - E_{\mathrm{Fp}}}{k_{\mathrm{b}}T}\right)$$

which means that $E_{\mathrm{Fn}} = E_{\mathrm{Fp}}$, i.e., the Fermi energies in the *p*-type and *n*-type regions are equal, and this has already been anticipated in Fig. 9.8 . In fact, this is a general and important property that, at thermal equilibrium, the Fermi energies of dissimilar materials must be equal. This physically means that there must not be a net flow of holes or electrons across the structure at equilibrium.

9.3 Non-equilibrium Properties of *p-n* Junctions

The most interesting and practical properties of a *p-n* junction are observed under non-equilibrium conditions, such as when a voltage is applied across it and/or when it is illuminated. Because of its nonsymmetrical nature, a *p-n* junction will exhibit different properties depending on the polarity of the external voltage or bias applied.

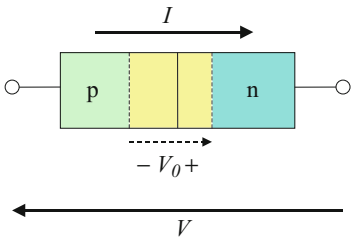

Fig. 9.9 Convention for the polarity of the external voltage and current

The sign convention used for the external voltage and the current in a *p-n* junction is shown in Fig. 9.9: the voltage will be positive if the applied potential on the *p*-type side is higher than that applied on the *n*-type. Note that the built-in voltage V_0 has been taken to be positive.

When an external bias is applied, the diffusion and drift currents do not balance each other anymore. This imbalance results in a net flow of electrical current in one or the other direction. In addition, the internal electric field and voltage across the *p-n* junction, the depletion width, and the energy band profile will all be changed. In this section, we will review how these parameters are modified.

9.3.1 Forward Bias: A Qualitative Description

When an external bias V is applied to the *p-n* structure depicted in Fig., there is usually some voltage drop across both the neutral bulk *p*-type and the *n*-type regions (i.e., outside the space charge region) due to Ohm's law (Sect. 8.2). In other words, the entire external bias is not applied across the transition region because part of it would be "lost" across the neutral regions due to their electrical resistance.

However, in most semiconductor devices which use *p-n* junctions, the length of these neutral regions which the electrical current would have to flow through is small, and any voltage drop would thus be negligible compared to the voltage change across the transition region. In our discussion, for now we will assume that the external bias is applied directly to the limits of the space charge region.

According to the sign convention in Fig. 9.10d, the total voltage across the transition region is now given by V_0-V. There are typically two regimes which need to be considered for the non-equilibrium conditions of a *p-n* junction: forward bias and reverse bias.

In the forward bias regime, corresponding to $V > 0$, the total voltage or potential barrier across the transition region is actually reduced from V_0 to $V_0 - V$, which has a number of consequences. First, the strength of the internal electric field associated with the lower potential barrier is reduced as well, as shown in Fig. 9.10c. This in turn means that the width of the space charge region is reduced because fewer electrical charges are needed to maintain this electric field, as shown in Fig. 9.10b. In other words, W_0 is reduced and is now denoted W, x_{n0} becomes x_n, and x_{p0} becomes x_p, as illustrated in Fig. 9.10a. As the internal voltage is reduced from its

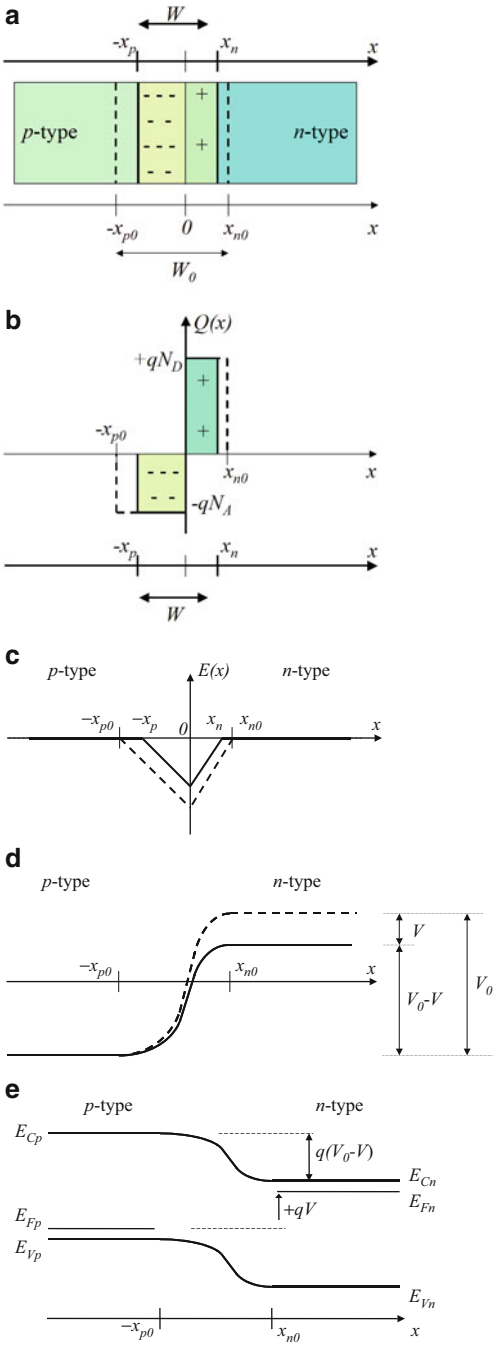

Fig. 9.10 (**a**) Space charge region width, (**b**) electrical charge density, (**c**) electric field strength, (**d**) potential profile, and (**e**) energy band profile of a *p-n* junction under forward bias ($V > 0$). The thick dashed curves represent the equilibrium case for comparison

equilibrium value by an amount equal to V, the energy band profile is changed, and the amount of band bending is reduced by qV, as depicted in Fig. 9.10e. This means that:

$$E_{\mathrm{Vp}} - E_{\mathrm{Vn}} = E_{\mathrm{Cp}} - E_{\mathrm{Cn}} = q(V_0 - V) \tag{9.29}$$

instead of Eq. (9.26). Furthermore, we can still consider that the Fermi energy levels outside the space charge region, i.e., in the neutral bulk p-type (E_{Fp}) and n-type (E_{Fn}) regions, are located at their equilibrium positions because we assumed no voltage drop in these regions. Therefore, because the band bending has been reduced by qV, according to Fig. (e), we must have:

$$E_{\mathrm{Fp}} - E_{\mathrm{Fn}} = -qV \tag{9.30}$$

This means that the Fermi energy is not constant throughout the p-n junction structure, but the Fermi energy levels in the neutral p-type and the n-type regions are separated by qV, where V is the applied external bias. This is a direct consequence of a non-equilibrium condition.

Let us now qualitatively examine the effects of a forward bias on the diffusion and drift currents across the space charge region of a p-n junction. As we saw in the previous section, the diffusion current arises from the difference between the densities of charge carriers on either side of the junction area. It corresponds to the motion of electrons from the n-type region toward the p-type region, and conversely for holes. This means that, at its origin, the diffusion current is related to the motion of majority carriers (e.g., electrons in the n-type region). However, as soon as these carriers reach the other side of the junction, they become minority carriers. Therefore, the diffusion current acts as if it injects minority carriers into one side of the junction by pulling them from the other side of the junction where they are majority carriers.

At equilibrium, the diffusion process is stabilized when the built-in electric field exerts a force that exactly counterbalances the diffusion of these charge carriers. Under a forward bias, as we just saw in Fig. 9.10c, this electric field strength is reduced. Therefore, each type of charge carriers can diffuse more easily, which means that the diffusion currents for both types of carrier increase under a forward bias.

This can also be understood by examining the energy band profile. For example, when the electrons in the n-type region, on the right-hand side of Fig. 9.10c where they are more concentrated, diffuse toward the p-type region where they are less concentrated, the allowed energy states are located at higher energies. This means that the diffusion electrons have to cross a high-energy barrier. Under a forward bias, this energy barrier is reduced, as shown in Fig. 9.10e, and more electrons can thus participate in the diffusion toward the p-type region. A similar argument is valid for holes. As a result, *the diffusion currents for both types of carrier increase under a forward bias*.

By contrast, the *drift current does not change with an external bias*, although this may seem contradictory with the fact that the internal electric field is weaker. This can be understood by examining the drift current in more detail. We saw in Sect. 9.2

that the drift current counterbalanced the diffusion of charge carriers and thus consisted of electrons moving toward the *n*-type region and holes moving toward the *p*-type region. This means that, at its origin, the drift current is related to the motion of minority carriers, such as electrons in the *p*-type region which drift toward the *n*-type region under the influence of the electric field. The drift current thus plays the converse role of the diffusion current. The drift current acts as if it extracts minority carriers from one side of the junction to send them to the other side of the junction where they are majority carriers. Because the concentrations of minority carriers are very small (see Eq. (9.2)), the drift currents are mostly limited by the number of minority carriers available for drift (i.e., electrons on the *p*-type region and holes on the *n*-type region) rather than by the speed at which they would drift (i.e., the strength of the electric field). We then understand why the drift current does not change significantly when an external bias is applied, in comparison to the diffusion current.

9.3.2 Reverse Bias: A Qualitative Description

By contrast, in the reverse bias regime, corresponding to $V < 0$, the total voltage or potential barrier across the transition region is actually increased from V_0 to $V_0 - V$, which also has the opposite effects of a forward bias. The strength of the internal electric field is increased, as shown in Fig. (c). This enlarges the width of the space charge region from W_0 to W (with x_{n0} becoming x_n, and x_{p0} becoming x_p, as illustrated in Fig. 9.11a) because more electrical charges are needed to maintain this electric field, as shown in Fig. 9.11b. As the internal voltage is increased from its equilibrium value by an amount equal to $-V$, the energy band profile is changed, and the amount of band bending is increased by $-qV$, as depicted in Fig. 9.11e. The total amount of band bending is still given by the expression in Eq. (9.29). The difference between the Fermi energy levels outside the space charge region is also still given by Eq. (9.30).

In addition, by contrast with the forward bias case, *the diffusion currents for both types of carrier decrease under a reverse bias*. However *the drift current still does not change significantly* in comparison to the diffusion current when a reverse bias is applied, for the same reason as discussed previously.

9.3.3 A Quantitative Description

In the previous subsections, we have expressed quantitatively the amount of band bending and the difference between the Fermi energy levels of the neutral *p*-type and *n*-type regions as a function of the applied external bias (Eqs. (9.29) and (9.30), respectively).

In fact, most of the relations that were derived in Sect. 9.2 for the equilibrium case are valid when an external bias voltage V is applied, provided we make the following transformations:

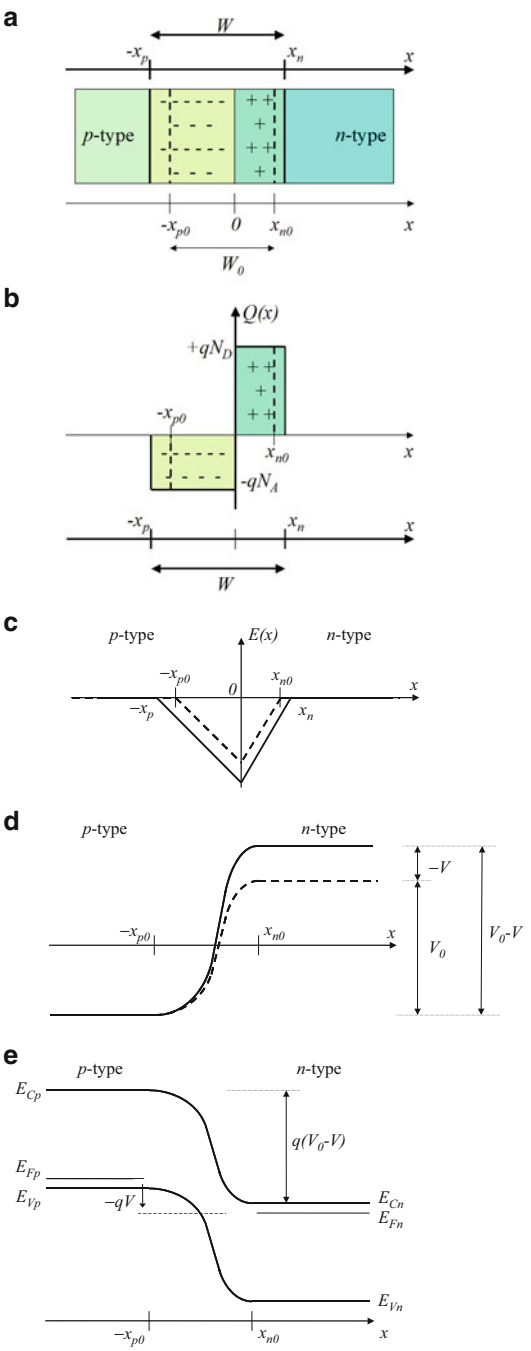

Fig. 9.11 (**a**) Space charge region width, (**b**) electrical charge density, (**c**) electric field strength, (**d**) potential profile, and (**e**) energy band profile of a *p-n* junction under reverse bias ($V < 0$). The thick dashed curves represent the equilibrium case for comparison

$$\begin{cases} W_0 & \rightarrow & W \\ x_{p0} & \rightarrow & x_p \\ x_{n0} & \rightarrow & x_n \\ V_0 & \rightarrow & V_0 - V \end{cases} \tag{9.31}$$

This statement is justified by the fact that most of the expressions in Sect. 9.2 have been obtained without invoking the equilibrium condition of Eq. (9.3) but by using the electrical charge neutrality principle and Gauss's law instead which are valid at all times.

The following few relations will be important for future discussions. The depletion width can be obtained from Eq. (9.23) by using Eq. (9.31):

$$W = \sqrt{\frac{2\varepsilon}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)(V_0 - V)} \tag{9.32}$$

for $V < V_0$. We clearly see that the depletion width shrinks when a forward bias is applied ($V > 0$), whereas it expands when a reverse bias is applied ($V < 0$). This confirms the qualitative discussion of the previous subsection.

Example

Q Calculate the ratio of the depletion region width W under a forward bias of 0.3 V to the equilibrium width W_0, for a GaAs abrupt *p-n* junction with a doping level on the *p*-type side of $N_A = 2 \times 10^{17}$ cm^{-3} and a doping level on the *n*-type side of $N_D = 1 \times 10^{17}$ cm^{-3} at 300 K.

A The depletion width W under a bias V is given by the expression: $W = \sqrt{\frac{2\varepsilon}{q}\left(\frac{N_A+N_D}{N_A N_D}\right)(V_0 - V)}$, where the built-in potential is $V_0 = 1.297$ V, as determined in earlier examples. The ratio sought is therefore:

$$\frac{W}{W_0} = \sqrt{\frac{(V_0 - V)}{V_0}} = \sqrt{\frac{(2.297 - 0.3)}{1.297}} = 0.877$$

The depletion width is then:

$$\begin{aligned} W &= 0.877 W_0 = 0.877\left(x_{p0} + x_{n0}\right) \\ &= 0.877(56 + 112) \\ &= 147 \text{ nm} \end{aligned}$$

The extent of the space charge region inside the *p*-type and *n*-type regions, as shown in Figs. 9.9a and 9.10a, can be obtained from Eq. (9.25):

$$\begin{cases} x_p = \sqrt{\dfrac{2\varepsilon}{q}\left(\dfrac{N_D}{N_A(N_A+N_D)}\right)(V_0 - V)} \\ x_n = \sqrt{\dfrac{2\varepsilon}{q}\left(\dfrac{N_A}{N_D(N_A+N_D)}\right)(V_0 - V)} \end{cases} \tag{9.33}$$

Similarly, the non-equilibrium hole and electron concentrations at the edges of the space charge region, denoted $p(-x_p)$, $p(x_n)$, $n(-x_p)$, and $n(x_n)$, can be obtained by considering Eq. (9.21):

$$\frac{p(-x_p)}{p(x_n)} = \frac{n(x_n)}{n(-x_p)} = \exp\left(\frac{q(V_0 - V)}{k_b T}\right) \tag{9.34}$$

In addition, following our previous discussion, we realize that the majority carrier concentrations are little changed under a moderate forward or a reverse bias, i.e., $p(-x_p) = p_p$ and $n(-x_n) = n_n$, which after replacing in Eq. (9.34) to:

$$\frac{p_p}{p(x_n)} = \frac{n_n}{n(-x_p)} = \exp\left(\frac{q(V_0 - V)}{k_b T}\right)$$

and by using Eq. (9.21) to eliminate p_p and n_n from this latest equation:

$$\frac{p(x_n)}{p_n} = \frac{n(-x_p)}{n_p} = \exp\left(\frac{qV}{k_b T}\right) \tag{9.35}$$

These expressions are important as they show that, when an external bias voltage is applied, the *minority carrier concentrations* at the boundary of the space charge region, $p(x_n)$ and $n(x_p)$, are directly and simply related to the equilibrium minority carrier concentrations p_n and n_p, and the applied bias voltage V. All these relations will prove important in the derivation of the diode equation for an ideal *p-n* junction which will be the topic of the next subsection.

Example

Q Calculate the minority carrier concentrations at x_n and $-x_p$ for the GaAs *p-n* junction described in the previous example.

A The minority carrier concentrations at x_n and $-x_p$ are given by: $\frac{p(x_n)}{p_n} = \frac{n(-x_p)}{n_p} = \exp\left(\frac{qV}{k_b T}\right)$, where p_n and n_p are the minority carrier concentrations in the neutral *n*-type side and *p*-type side, respectively, at equilibrium. These are given by the action mass law:

$$p_n = \frac{n_i^2}{N_D} = \frac{\left(1.79 \times 10^6\right)^2}{1 \times 10^{17}} = 3.20 \times 10^{-5}\text{cm}^{-3}\text{and}$$

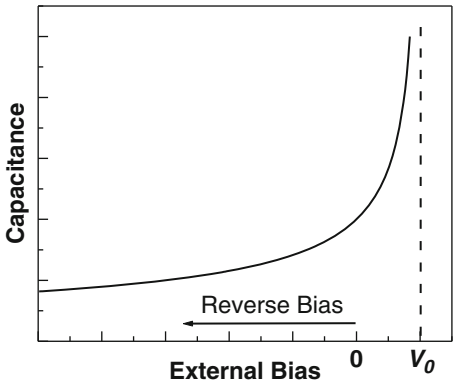

Fig. 9.12 Depletion layer capacitance as a function of bias voltage, showing the increase in capacitance with forward bias and the decrease with reverse bias

$$n_{\mathrm{p}} = \frac{n_{\mathrm{i}}^2}{N_{\mathrm{A}}} = \frac{\left(1.79 \times 10^6\right)^2}{2 \times 10^{17}} = 1.60 \times 10^{-5}\mathrm{cm}^{-3}..$$

In addition, the exponential is numerically equal to:

$\exp\left(\frac{qV}{k_{\mathrm{b}}T}\right) = \exp\left(\frac{\left(1.60218\times10^{-19}\right)\times0.3}{\left(1.38066\times10^{-23}\right)\times300}\right) = 1.1 \times 10^5$. Thus, we get:

$$p(x_{\mathrm{n}}) = \left(3.20 \times 10^{-5}\right)\left(1.1 \times 10^5\right) \approx 3.5\ \mathrm{cm}^{-3} \text{ and}$$

$$n(x_{\mathrm{p}}) = \left(1.60 \times 10^{-5}\right)\left(1.1 \times 10^5\right) \approx 1.76\,\mathrm{cm}^{-3}$$

9.3.4 Depletion Layer Capacitance

The depletion layer is relatively devoid of mobile carriers and can therefore be thought of as somewhat similar to the dielectric in a capacitor. Positive and negative charges are separated by this depletion layer, and this leads to a capacitance associated with the *p-n* junction. This capacitance can be thought of as like that of a parallel plate capacitor and expressed as:

$$C_{\mathrm{dep}} = \frac{\varepsilon A}{W} \tag{9.36}$$

However rather than being constant, the capacitance of a *p-n* junction varies with the reverse bias via the voltage dependence of the depletion width as shown in Fig. 9.12.

More formally, the capacitance of the *p-n* junction can be derived starting from the definition of capacitance:

$$C_{\mathrm{dep}} = \left|\frac{dQ}{dV}\right| \tag{9.37}$$

where dQ is the incremental change in charge stored on either side of the junction for an incremental increase in voltage of dV. For the abrupt junction, the charge stored on either side of the junction can be expressed as:

$$Q_{\text{dep}} = qAN_{\text{D}}x_{\text{n}} = qAN_{\text{A}}x_{\text{p}} \tag{9.38}$$

where x_{n} and x_{p} are given by Eq. (9.33). Substituting in Eq. (9.38) for either term gives the equation:

$$Q_{\text{dep}} = A\sqrt{2q\varepsilon\frac{N_{\text{A}}N_{\text{D}}}{(N_{\text{A}}+N_{\text{D}})}(V_0 - V)}$$

which can then be differentiated with respect to V to yield:

$$C_{\text{dep}} = \frac{A}{\sqrt{(V_0 - V)}}\sqrt{\frac{q\varepsilon}{2}\frac{N_{\text{A}}N_{\text{D}}}{(N_{\text{A}}+N_{\text{D}})}} \tag{9.39}$$

which we can see reduces to Eq. (9.36) above when $V = 0$.

The voltage dependence of the *p-n* junction capacitance is used in varactor diodes or varicaps, in tuning circuits where the diode is reverse-biased to prevent forward conduction, and a small DC tuning voltage is applied to vary the capacitance. Additionally, measuring the capacitance of a diode as a function of bias can be used to extract information about the built-in voltage and the doping profile. This can be done by plotting $1/C_{\text{dep}}$ vs. applied voltage:

$$V = A^2\left[\frac{q\varepsilon(N_{\text{A}}N_{\text{D}})}{2(N_{\text{A}}+N_{\text{D}})}\right]\frac{1}{C_{\text{dep}}^2} - V_o \tag{9.40}$$

In the case of an abrupt one-sided junction (such as a p^+n^- or a metal-semiconductor Schottky diode (see Sect. 9.5)), this equation reduces further, and the carrier concentrations can be extracted more directly:

$$\begin{aligned} V &= \frac{A^2q\varepsilon}{2}N_{\text{A}}\frac{1}{C_{\text{dep}}^2} - V_o, \quad (N_{\text{D}} >> N_{\text{A}}) \\ V &= \frac{A^2q\varepsilon}{2}N_{\text{D}}\frac{1}{C_{\text{dep}}^2} - V_o, \quad (N_{\text{A}} >> N_{\text{D}}) \end{aligned} \tag{9.41}$$

9.3.5 Ideal p-n Junction Diode Equation

The diode equation refers to the mathematical expression which relates the total electrical current I through an ideal *p-n* junction to the applied external bias voltage V. It is also referred as the current-voltage or *I-V* characteristic of the diode. To determine it, we must focus our analysis on the minority carriers, i.e., holes in the *n*-type region and electrons in the *p*-type region.

In addition to the depletion approximation model considered so far, a few more assumptions need to be considered:

(i) First, we assume that there are no external sources of carrier generation.
(ii) No recombination of charge carriers occurs within the space charge region.
(iii) We assume that the applied biases are moderate enough to ensure that the minority carriers remain much less numerous than the majority carriers in the neutral regions.
(iv) Finally, we assume that the change in minority carrier concentrations in the neutral regions does not result in a non-negligible electric field.

In virtue of assumptions (i) and (ii), any hole or electron that has diffused across the space charge region must be present at its boundaries, i.e., at $-x_p$ and x_n, respectively. When a bias V is applied, the concentrations of these holes and electrons, which are in excess of their equilibrium concentrations, are given by:

$$\begin{cases} \Delta p_n = p(x_n) - p_n \\ \Delta n_p = n(-x_p) - n_p \end{cases}$$

This becomes after using Eq. (9.35):

$$\begin{cases} \Delta p_n = p_n\left(e^{\frac{qV}{k_b T}} - 1\right) \\ \Delta n_p = n_p\left(e^{\frac{qV}{k_b T}} - 1\right) \end{cases} \tag{9.42}$$

Here, and in the rest of the text, we will use the extended meaning of the term "excess carrier." For example, if Δp_n and Δn_p are positive, i.e., $V > 0$ or forward bias, then there are net real excesses of holes and electrons at the space charge boundaries, and we talk about minority carrier injection. This is shown in Fig. 9.13.

But if Δp_n and Δn_p are negative, i.e., $V < 0$ or reverse bias, then there are net real deficiencies of holes and electrons, and we talk about minority carrier extraction. In this case, the minority carriers at the boundaries of the space charge region are less numerous than in the bulk neutral material; therefore there is a diffusion of minority carriers from the bulk neutral region toward the edges of the space charge region. This is illustrated in Fig. 9.14.

Returning to the forward bias case, the excess holes, present at $x = x_n$ with a concentration Δp_n, will be diffusing deeper into the neutral n-type region where their equilibrium concentration is only p_n. As they diffuse, they will experience recombination as discussed in Chap. 8, with a characteristic diffusion length L_p in the steady-state regime. The excess hole concentration is therefore reduced as we advance deeper in the material. This situation has already been encountered in Chap. 8 and the analytical expression for $\delta p_n(x_1)$, the excess hole concentration at a position x_1, is obtained for Eq. (8.55):

$$\delta p_n(x_1) = \Delta p_n e^{-\frac{x_1}{L_p}} \tag{9.43}$$

where L_p is the hole diffusion length in the n-type region. In this expression, we chose another axis, denoted x_1, oriented in the same direction as the original axis x and with its origin at $x = x_n$. It is important to remember that the excess

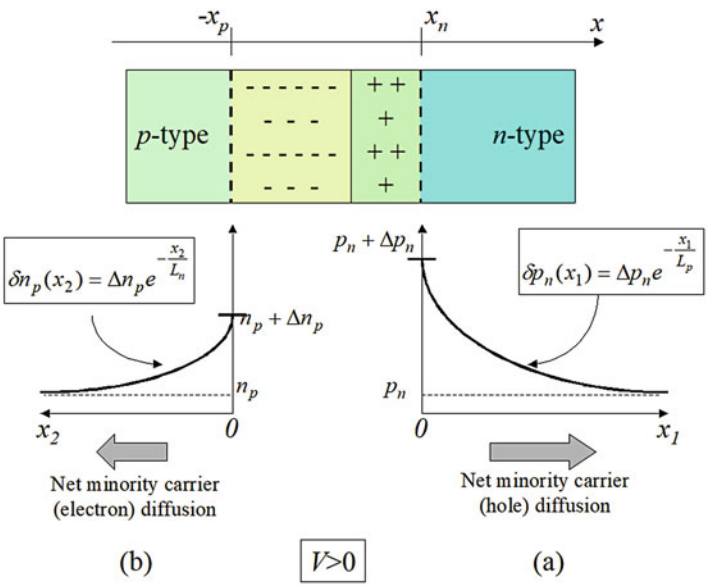

Fig. 9.13 (**a**) Excess hole concentration profile in the *n*-type region, and (**b**) excess electron concentration profile in the *p*-type region, under a forward bias. The excess carrier concentrations decrease, following an exponential decay, as they go further from the edges of the depletion region

concentration of holes at $x = x_n$ remains constant at Δp_n given by Eq. (9.42) because holes are continuously injected or extracted through the space charge region into or from the *n*-type region due to the application of the external bias voltage. We can make use of Fig. 8.7 to plot the spatial profile of the excess hole concentration in Fig. 9.13a for the forward bias case and Fig. 9.14a for the reverse bias case.

Conversely, the excess electrons present at $x = -x_p$ with a concentration Δn_p will diffuse deeper into the neutral *p*-type region, with a diffusion length L_n. This leads to the spatial profile $\delta n_p(x_2)$ shown in Fig. 9.13b for the forward bias case and Fig. 9.14b for the reverse bias case, and it is analytically given by:

$$\delta n_p(x_2) = \Delta n_p e^{-\frac{x_2}{L_n}} \tag{9.44}$$

where L_n is the electron diffusion length in the *p*-type region. It is important to note that, here, we chose the sign convention for the axis x_2 in the opposite direction of the original axis x because the electrons diffuse in this opposite direction.

There are essentially two methods to compute the diode equation. The first one consists of analyzing the diffusion currents in the *p-n* junction. From our discussion in Subsect. 9.3.1, we understand that, when an external bias is applied, the drift currents across the space charge region do not vary, whereas the diffusion currents change. The sum of the increments in the hole and the electron diffusion currents across the space charge region is thus a direct measure of the net electrical current through the *p-n* junction since no net current is originally present at equilibrium,

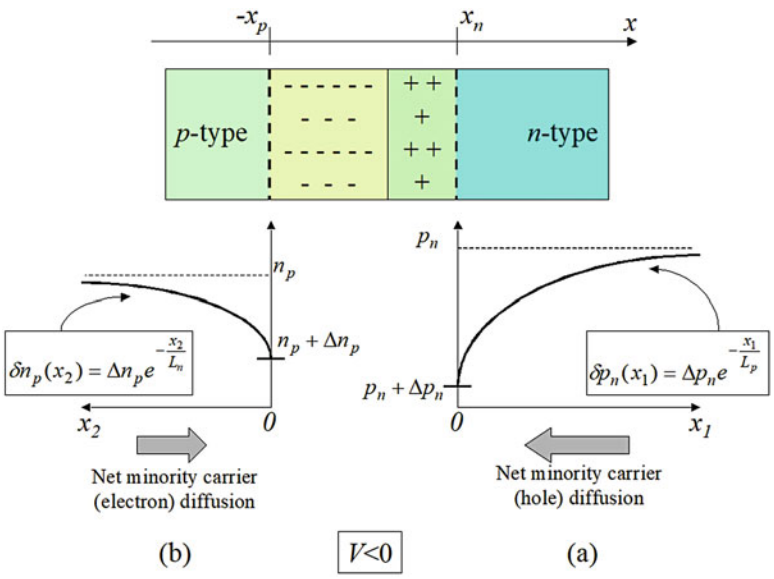

Fig. 9.14 (**a**) "Excess" hole concentration profile in the *n*-type region, and (**b**) "excess" electron concentration profile in the *p*-type region, under a reverse bias. These carrier concentrations change following an exponential dependence as they go further away from the edges of depletion region

because we have assumed there are no external sources of carrier generation and because the total electrical current is constant throughout a two-terminal device, such as the *p-n* junction earlier shown in Fig. 9.8.

The incremental diffusion currents are the diffusion currents which result from the excess carriers in the material. The diffusion current densities for electrons and holes can be obtained from Eqs. (8.36) and (8.38) and are given by:

$$\begin{cases} J_{\mathrm{h}}^{\mathrm{diff}}(x_1) = -qD_{\mathrm{p}}\dfrac{d(\delta p_{\mathrm{n}}(x_1))}{dx_1} \\ J_{\mathrm{e}}^{\mathrm{diff}}(x_2) = qD_{\mathrm{n}}\dfrac{d\left(\delta n_{\mathrm{p}}(x_2)\right)}{dx_2} \end{cases} \tag{9.45}$$

Using the expressions of the excess carrier concentrations in Eqs. (9.43) and (9.44), we get:

$$\begin{cases} J_{\mathrm{h}}^{\mathrm{diff}}(x_1) = +q\dfrac{D_{\mathrm{p}}}{L_{\mathrm{p}}}\Delta p_{\mathrm{n}} e^{-\frac{x_1}{L_p}} \\ J_{\mathrm{e}}^{\mathrm{diff}}(x_2) = -q\dfrac{D_{\mathrm{n}}}{L_{\mathrm{n}}}\Delta n_{\mathrm{p}} e^{-\frac{x_2}{L_n}} \end{cases} \tag{9.46}$$

In order to obtain the total current through the *p-n* junction, we must evaluate the diffusion current densities for holes and electrons at the limits of the space charge region at $x = x_n$ and $x = -x_p$, respectively, or equivalently at $x_1 = x_2 = 0$:

$$\begin{cases} J_h^{diff}(0) = +q\dfrac{D_p}{L_p}\Delta p_n \\ J_e^{diff}(0) = -q\dfrac{D_n}{L_n}\Delta n_p \end{cases} \tag{9.47}$$

Example

Q Estimate the ratio of the diffusion current densities of holes and electrons for the GaAs *p-n* junction described in the previous example.

A The ratio of the diffusion currents is given by: $\left|\dfrac{J_h^{diff}(0)}{J_e^{diff}(0)}\right| = \dfrac{D_p}{D_n}\dfrac{L_n}{L_p}\dfrac{\Delta p_n}{\Delta n_p}$, where Δp_n and Δn_p are the excess minority carrier concentrations at the limits of the depletion region. These quantities are given by: $\Delta p_n = p_n\left(e^{\frac{qV}{k_bT}} - 1\right)$ and $\Delta n_p = n_p\left(e^{\frac{qV}{k_bT}} - 1\right)$. Their ratio is then: $\dfrac{\Delta p_n}{\Delta n_p} = \dfrac{p_n}{n_p} = \dfrac{{}^{n_i^2}/_{N_D}}{{}^{n_i^2}/_{N_A}} = \dfrac{N_A}{N_D}$. In addition, the diffusion lengths can be expressed as a function of the minority carrier lifetime on the *n*-type and the *p*-type sides. These lead to the ratio: $\left|\dfrac{J_h^{diff}(0)}{J_e^{diff}(0)}\right| = \dfrac{D_p}{D_n}\dfrac{\sqrt{D_n\tau_n}}{\sqrt{D_p\tau_p}}\dfrac{N_A}{N_D}$. Assuming that the minority carrier lifetimes are the same for holes and electrons, we get: $\left|\dfrac{J_h^{diff}(0)}{J_e^{diff}(0)}\right| = \sqrt{\dfrac{D_p}{D_n}}\dfrac{N_A}{N_D}$. The ratio of the diffusion coefficients can be calculated using the majority carrier mobilities through the Einstein relations and we obtain: $\left|\dfrac{J_h^{diff}(0)}{J_e^{diff}(0)}\right| = \sqrt{\dfrac{\mu_h}{\mu_e}}\dfrac{N_A}{N_D}$ and

$$\left|\frac{J_h^{diff}(0)}{J_e^{diff}(0)}\right| = \sqrt{\frac{400}{8500}}\frac{2\times10^{17}}{1\times10^{17}} \approx 0.43$$

In all these expressions of current densities, it is important to remember that the sign convention for the current density $J_h^{diff}(x_1)$ is the same as the axis x, whereas for $J_e^{diff}(x_2)$ it is opposite that of axis x. The total current density is the sum of the hole and electron diffusion currents, with however a sign difference:

$$J_{total} = J_h^{diff}(0) - J_e^{diff}(0) \tag{9.48}$$

The minus sign for $J_e^{diff}(0)$ accounts for the sign convention chosen for axis x_2. Inserting Eq. (9.47) into this relation, we get:

$$J_{\text{total}} = q\left(\frac{D_\text{p}}{L_\text{p}}\Delta p_\text{n} + \frac{D_\text{n}}{L_\text{n}}\Delta n_\text{p}\right) \tag{9.49}$$

and using Eq. (9.42), we finally obtain:

$$J_{\text{total}} = q\left(\frac{D_\text{p}}{L_\text{p}}p_\text{n} + \frac{D_\text{n}}{L_\text{n}}n_\text{p}\right)\left(e^{\frac{qV}{k_\text{b}T}} - 1\right) \tag{9.50}$$

The total current is given by the total current density multiplied by the area of the *p-n* junction. If we assume a uniform area A, we get:

$$I_{\text{total}} = AJ_{\text{total}} = qA\left(\frac{D_\text{p}}{L_\text{p}}p_\text{n} + \frac{D_\text{n}}{L_\text{n}}n_\text{p}\right)\left(e^{\frac{qV}{k_\text{b}T}} - 1\right) \tag{9.51}$$

By introducing a new term I_0, this can be rewritten as:

$$I_{\text{total}} = I_0\left(e^{\frac{qV}{k_\text{b}T}} - 1\right) \tag{9.52}$$

with:

$$I_0 = qA\left(\frac{D_\text{p}}{L_\text{p}}p_\text{n} + \frac{D_\text{n}}{L_\text{n}}n_\text{p}\right) \tag{9.53}$$

Equations (9.52) and (9.53) represent the diode equation for an ideal *p-n* junction. This function is plotted in Fig. 9.15.

We see that under a forward bias, the current increases exponentially as a function of applied voltage. By contrast, under reverse bias, the current rapidly tends toward $-I_0$. The value of the current I_0 is therefore called the reverse saturation current. The physical meaning of this current can be understood as follows. When a strong reverse bias is applied ($V < 0$), the density of minority carriers at the boundary of the space charge region quickly falls to zero according to Eq. (9.35). This means that, inside the depletion region, there is no diffusion of carriers, but only drift currents are

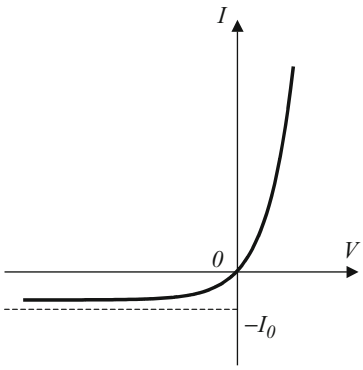

Fig. 9.15 Current-voltage characteristic for an ideal *p-n* junction diode. The dependence of the current on the voltage follows an exponential expression. The current is zero when the voltage is zero, without external excitation

present. Outside the depletion region however, the only charge motion is the diffusion of minority carriers from the neutral regions toward the depletion region, as illustrated by the block arrows in Fig. 9.14. We can therefore say that the saturation current in Eq. (9.53) corresponds to the total drift, across the space charge region, of minority carriers which have been extracted or able to reach the limits of the space charge region through diffusion from the neutral regions.

The *p-n* junction diode acts like a one-way device: when it is forward-biased, current can flow from the *p*-type to the *n*-type region without much resistance, whereas when it is reverse-biased, a very large resistance prevents the current from flowing in the opposite direction from the *n*-type to the *p*-type region.

The second method which can be used to determine the diode equation consists of calculating the total charge accumulated on each side of the junction area. This second method is called the charge control approximation. Let Q_p be the steady-state excess positive charge in the *n*-type region which is given by integrating Eq. (9.43):

$$Q_p = qA \int_0^\infty \delta p_n(x_1) dx_1 = qA\Delta p_n \int_0^\infty e^{-\frac{x_1}{L_p}} dx_1$$

i.e.:

$$Q_p = qAL_p \Delta p_n \tag{9.54}$$

where A is the area of the *p-n* junction. This excess charge is illustrated in Fig. 9.16a, in the forward bias case. The hole diffusion current must then be able to maintain this excess positive charge, even though the holes are recombining. As the average lifetime of holes in the *n*-type region is the recombination lifetime τ_p defined in Subsect. 8.5.3, the hole diffusion current must be able to supply Q_p positive charges during a time equal to τ_p. This current must therefore be $I_p = \frac{Q_p}{\tau_p}$.

Similarly, the excess negative charge in the *p*-type region is given by:

$$Q_n = qAL_n \Delta n_p \tag{9.55}$$

and is shown in Fig. 9.16b. The electron diffusion current into the *p*-type region is $-I_n = -\frac{Q_n}{\tau_n}$. In this last expression, we made use of the same sign convention as for axis x_2. The total current is therefore given by:

$$I_{total} = I_p + I_n = qA\frac{L_p}{\tau_p}\Delta p_n + qA\frac{L_n}{\tau_n}\Delta n_p$$

or:

$$I_{total} = qA\left(\frac{L_p}{\tau_p}\Delta p_n + \frac{L_n}{\tau_n}\Delta n_p\right) \tag{9.56}$$

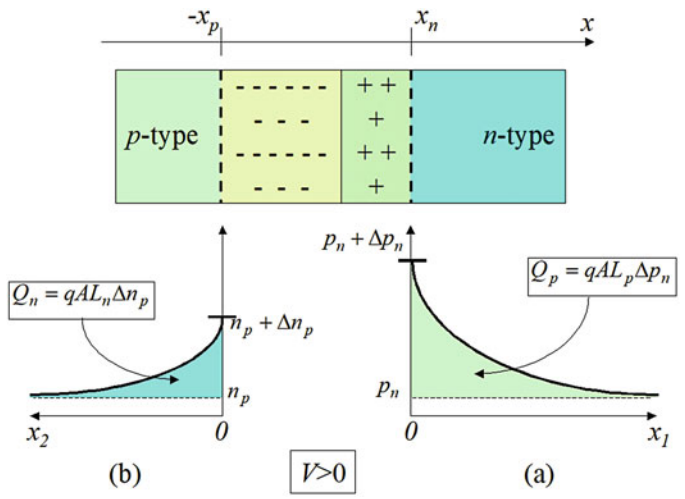

Fig. 9.16 (**a**) Excess positive charge in the *n*-type region and (**b**) excess negative charge in the *p*-type region, under a forward bias. The total excess charges are calculated by integrating the excess carrier concentrations over the volume of the regions outside the depletion region

Using the definition of the diffusion lengths given in Eqs. (8.53) and (8.56), and using Eq. (9.42), we can transform this last expression into:

$$I_{\text{total}} = AJ_{\text{total}} = qA\left(\frac{D_{\text{p}}}{L_{\text{p}}}p_{\text{n}} + \frac{D_{\text{n}}}{L_{\text{n}}}n_{\text{p}}\right)\left(e^{\frac{qV}{k_{\text{b}}T}} - 1\right)$$

and thus get the diode equation obtained in Eq. (9.51).

9.3.6 Minority and Majority Carrier Currents in Neutral Regions

In the previous discussion, we saw that the total electrical current through a *p-n* junction device was determined by the diffusion currents across the space charge region which result in minority carriers being injected into or extracted from the neutral regions under the influence of an applied external bias.

For the sake of clarity, let us consider the example of a forward-biased *p-n* junction, as the one shown in Fig. 9.13. We saw that the excess minority carriers diffuse into the neutral regions following an exponential decay given in Eqs. (9.43) and (9.44). This leads to diffusion currents which also follow an exponential decay, as obtained in Eq. (9.46). However, we know that the total electrical current throughout a two-terminal device is constant. Therefore, the decrease in diffusion current, for example, that of holes in the right-hand side of the figure, as we move away from the space charge region has to be compensated by another current. This is achieved through the drift of majority carriers, for example, electrons in the neutral *n*-type region. Indeed, through their diffusion and recombination, the minority

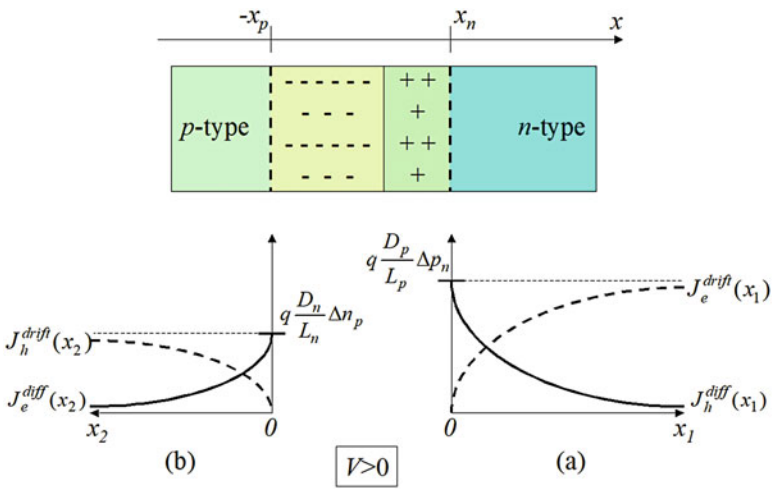

Fig. 9.17 Diffusion current of minority carriers and drift current of majority carriers in the (**a**) *n*-type region and (**b**) *p*-type region, under a forward bias. As the minority carriers diffuse further away from the edges of the depletion region, they recombine with majority carriers. The diffusion current of minority carriers is therefore reduced. But, this process also results in the flow of majority carriers in the opposite direction, which compensates the decrease in diffusion current with a drift current in the same proportion

carriers "consume" majority carriers (e.g., electrons). There thus must be a flow of majority carriers (e.g., electrons) in the opposite direction to resupply those lost in the recombination process. This flow of majority carriers generates a drift current.

Therefore, in the neutral regions, there are two components which make up the total electrical current: the diffusion current of minority carriers and the drift current of majority carriers. These are shown in Fig. 9.17 . This means, in particular, that there must be an electric field present in the neutral regions; otherwise there would not be any drift current. This apparently contradicts our assumption at the beginning of Subsect. 9.3.1 that there was no potential drop within the neutral regions. In fact, the potential drop is very small in comparison with any applied external bias voltage and therefore can be neglected in our model.

An analytical expression for the drift current can be easily determined, on each side of the *p-n* junction. Indeed, the total hole and electron current densities must be constant at the values given by the diode equation in Eq. (9.47). As we know the expression for the diffusion current densities $J_{\mathrm{h}}^{\mathrm{diff}}(x_1)$ and $J_{\mathrm{e}}^{\mathrm{diff}}(x_2)$ from Eq. (9.46), the drift current densities will be the difference:

$$\begin{cases} J_{\mathrm{h}}^{\mathrm{drift}}(x_2) = J_{\mathrm{e}}^{\mathrm{diff}}(0) - J_{\mathrm{e}}^{\mathrm{diff}}(x_2) \\ J_{\mathrm{e}}^{\mathrm{drift}}(x_1) = J_{\mathrm{h}}^{\mathrm{diff}}(0) - J_{\mathrm{h}}^{\mathrm{diff}}(x_1) \end{cases} \tag{9.57}$$

Recalling Eqs. (9.46) and (9.49), we get successively:

$$\begin{cases} J_{\mathrm{h}}^{\mathrm{drift}}(x_2) = -q\dfrac{D_{\mathrm{n}}}{L_{\mathrm{n}}}\Delta n_{\mathrm{p}} + q\dfrac{D_{\mathrm{n}}}{L_{\mathrm{n}}}\Delta n_{\mathrm{p}} e^{-\frac{x_2}{L_{\mathrm{n}}}} \\ J_{\mathrm{e}}^{\mathrm{drift}}(x_1) = q\dfrac{D_{\mathrm{p}}}{L_{\mathrm{p}}}\Delta p_{\mathrm{n}} - q\dfrac{D_{\mathrm{p}}}{L_{\mathrm{p}}}\Delta p_{\mathrm{n}} e^{-\frac{x_1}{L_{\mathrm{p}}}} \end{cases}$$

$$\begin{cases} J_{\mathrm{h}}^{\mathrm{drift}}(x_2) = q\dfrac{D_{\mathrm{n}}}{L_{\mathrm{n}}}\Delta n_{\mathrm{p}}\left(e^{-\frac{x_2}{L_{\mathrm{n}}}} - 1\right) \\ J_{\mathrm{e}}^{\mathrm{drift}}(x_1) = q\dfrac{D_{\mathrm{p}}}{L_{\mathrm{p}}}\Delta p_{\mathrm{n}}\left(1 - e^{-\frac{x_1}{L_{\mathrm{p}}}}\right) \end{cases} \tag{9.58}$$

It is important to remember that the sign convention chosen for $J_{\mathrm{h}}^{\mathrm{drift}}(x_2)$ is opposite that of axis x.

9.4 Deviations from the Ideal *p-n* Diode Case

Before deriving the ideal diode equation in the previous section, it was necessary to make several assumptions. In reality, these assumptions are not necessarily valid, and the ideal diode equation gives only qualitative agreement with actual measurements of the *I-V* characteristics of real *p-n* junction diodes. This deviation from the ideal case is mainly due to (a) generation of carriers in the depletion region, (b) surface leakage effects at the periphery of a real junction, (c) recombination of carriers in the depletion region, (d) the high-injection condition (when the injection of minority carriers exceeds the doping density), and finally (e) all the applied bias not being dropped across the depletion region due to series resistance effects. The above deviations are illustrated in the figure below. The special case of reverse breakdown will be discussed in Subsect. 9.4.3 (Fig. 9.18).

9.4.1 Reverse Bias Deviations from the Ideal Case

Part of the deviation of the leakage current from the ideal reverse saturation current arises from the thermal generation of electron-hole pairs within the space charge region. The built-in electric field separates these carriers and they drift toward the neutral regions of the diode. This drift results in an excess current that is in addition to the diffusion of minority carriers, discussed in the ideal case. Section 8.6 introduced the concept of thermal generation of carriers, and along with it a thermal generation rate per unit volume $G_{\mathrm{t}}(T)$, expressed in $\mathrm{cm}^{-3}\cdot\mathrm{s}^{-1}$. Since the volume of the depletion region is equal to WA, assuming no recombination occurs, the current due to generation in the depletion region can be expressed as:

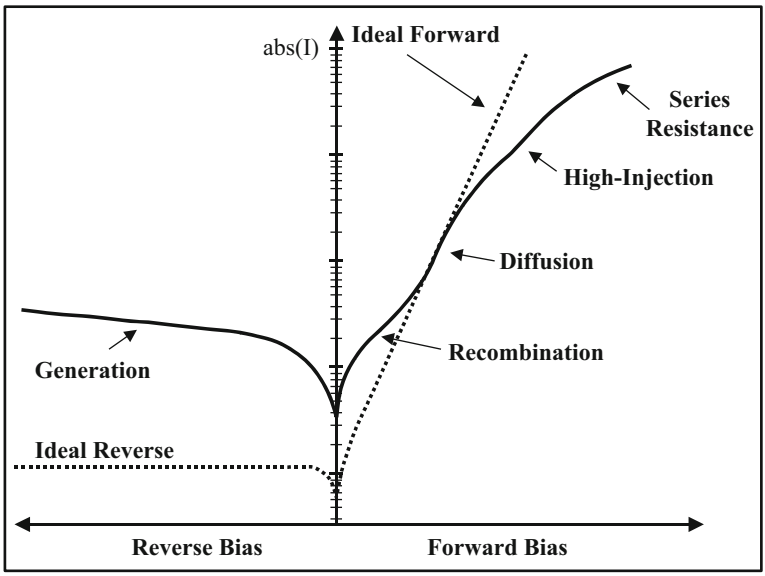

Fig. 9.18 The current-voltage characteristic for a real Si *p-n* junction diode (solid) does not exactly match the behavior of a Si junction diode predicted by the ideal diode model (dotted), both shown above in semilog scale. A real Si diode shows the following deviations from the ideal (diffusion limited) case: reverse leakage current due to thermal generation and surface leakage effects, recombination in the depletion region, high-injection deviation, and series resistance effects

$$I_{\text{gen}} = qWAG_{\text{t}}(T) \tag{9.59}$$

Under reverse bias the current can then be expressed as the sum of the diffusion and generation components:

$$I_{\text{rev}} = qA\left(\frac{D_{\text{p}}}{L_{\text{p}}}p_{\text{n}} + \frac{D_{\text{n}}}{L_{\text{n}}}n_{\text{p}}\right) + qWAG_{\text{t}}(T). \tag{9.60}$$

Since the depletion layer width (W) depends upon the applied bias, the reverse current of the diode now shows a bias dependence: as the reverse bias is increased, the depletion width widens, and hence this increases the generation current leading to a corresponding increase in the reverse leakage current as a function of applied bias. In addition to excess carriers arising from thermal generation, it is possible for external photoexcitation to create carriers in the depletion region – this is the case of a photodiode.

This leakage current is further compounded by the surface leakage. Surface leakage effects are due to the finite extent of the *p-n* junction area and the characteristics of the junctions that occur at the periphery of the diode. This is due primarily to ionic charges on or outside the semiconductor that induce corresponding image charges within the semiconductor. These charges create their own surface

depletion region that acts as a parallel conduction channel that bypasses the *p-n* junction and allows current to flow along the surface of the diode. Typically this leakage current increases with reverse bias.

9.4.2 Forward Bias Deviations from the Ideal Case

Under forward bias recombination dominates over the generation processes. In order to supply the carriers lost to recombination, the net external current flowing through the diode is increased. This current is called the recombination current (I_{rec}). The recombination rate is at its maximum near the center of the depletion region, where nearly equal number of electrons and holes are available to contribute to recombination. Assuming a linear variation of the potential across the depletion region, the potential at the center can be taken as $\frac{V_0-V}{2}$. In this case the carrier concentration at the center of the depletion region depends upon $\exp\left(-\frac{q(V_0-V)}{2k_bT}\right)$ rather than $\exp\left(\frac{q(V_0-V)}{k_bT}\right)$. The rate at which electrons and holes are recombining is then proportional to $\exp\left(\frac{qV}{2k_bT}\right)$. By introducing a material constant (I_{R0}) dependent upon the minority carrier recombination lifetimes in the respective halves of the depletion layer, and the overall depletion layer width, it becomes possible to arrive at an expression for the recombination current (I_R):

$$I_R \approx I_{R0}\exp\left(\frac{qV}{2k_bT}\right) \tag{9.61}$$

Combining this new equation for the recombination current together with the existing minority carrier diffusion current yields a new expression for the total current though the diode:

$$I = I_0\exp\left(\frac{qV}{k_bT}\right) + I_{R0}\exp\left(\frac{qV}{2k_bT}\right) \tag{9.62}$$

In working with real diodes, this equation is generally represented in an empirical form by introducing a new factor n called the ideality factor:

$$I \approx I_0\exp\left(\frac{qV}{nk_bT}\right) \tag{9.63}$$

In this combined equation, the ideality factor n tends toward 2 when recombination current dominates and tends toward 1 when diffusion current dominates and varies from 1 to 2 when both currents are comparable. In the case of silicon diodes operating at room temperature, both processes can be seen to operate as the current injection is increased from low to moderate levels.

Under higher levels of current injection (under forward bias), the diode enters the high-injection regime where the injected minority carrier density becomes

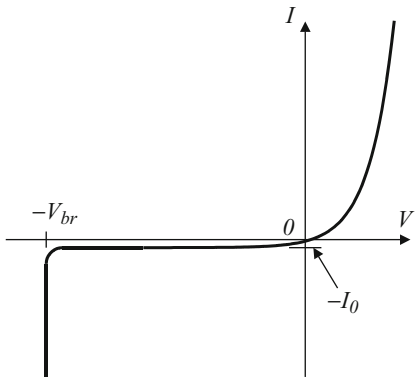

Fig. 9.19 Current-voltage characteristic for an ideal *p-n* junction diode showing a reverse breakdown. When the voltage across the *p-n* junction is equal to the reverse breakdown voltage, the current increases dramatically. If it is not limited, this current can damage the diode through heating

comparable or greater than the majority carrier density. In this case the current becomes proportional to $\exp\left(\frac{qV}{2k_bT}\right)$, as is shown in Fig. 9.19.

Under higher reverse bias, the contact potentials and the potential drop across the bulk regions of the semiconductor cease to be negligible, and the series resistance of the *p-n* diode no longer dominates. At this point the exponential increase in current begins to subside in favor of a more linear increase, limited by the series resistance of the diode. The empirical diode equation introduced above can be modified to take this behavior into account, by introducing a term (R_S) for the series resistance. Thus the equation becomes:

$$I \approx I_0 \exp\left(\frac{q(V - IR_S)}{nk_bT}\right) \tag{9.64}$$

9.4.3 Reverse Breakdown

In the ideal *p-n* junction diode model, we saw that the current through a *p-n* junction diode was limited by the saturation current $-I_0$ when a reverse bias was applied. Even in the non-ideal case the reverse current was seen to increase slowly. In reality, this model holds only up to a certain value of reverse bias $-V_{br}$, called the breakdown voltage. At that point, the current suddenly increases dramatically as shown in Fig. 9.19 . This phenomenon is called reverse breakdown. The peak value for the internal electric field strength (i.e., at $x = 0$) corresponding to this applied reverse bias is called the critical electric field.

This situation is not necessarily a damaging one for the *p-n* junction and is reversible, as long as the current can be limited to prevent too much power from being dissipated inside the device. Otherwise, parts of the device can be physically destroyed (e.g., melted).

There are two major mechanisms for the reverse breakdown: avalanche breakdown which occurs at higher reverse biases as a result of impact ionization and Zener breakdown which occurs at lower reverse biases as a result of tunneling across the junction.

9.4.4 Avalanche Breakdown

As a stronger reverse bias is applied, the electric field strength across the space charge region increases. The charge carrier particles, holes, and electrons which drift across the depletion region can therefore achieve higher velocities.

When the reverse bias is strong enough, typically higher than $6E_g/q$ and can even go up to 1000 V, the electric field strength can become so large that a hole or an electron can gain sufficient kinetic energy to impact on a semiconductor lattice atom and ionize it, or even break a chemical bond. This phenomenon is called impact ionization. It may seem conceptually difficult to envision a hole impacting on the crystal lattice, but this can be better understood when we realize that when a hole moves in one direction, it in fact corresponds to the motion of an electron in the opposite direction with the same velocity. An accelerated particle must typically acquire energy at least equal to the bandgap energy E_g in order to break a chemical bond, because this corresponds to the energy required to excite an electron from the valence band to the conduction band. Therefore, for wider bandgap semiconductors, higher electric field strength is necessary to ensure impact ionization.

As a result of impact ionization, an electron-hole pair (EHP) is created within the space charge region in addition to the impacting particle. The electron and the hole from the pair will then be spatially separated by the electric field present at that location: the electron drifting toward the *n*-type side and the hole toward the *p*-type side, as illustrated in Fig. 9.20.

The electrons and holes thus generated can themselves be further accelerated by the electric field. If they reach a sufficient high kinetic energy within the space charge region, they can in turn contribute to create additional EHPs through ionizing collisions. This results in a cascade or avalanche effect. One initial charge carrier thus has the potential to create many additional carriers, and a dramatic increase in current is achieved as the one shown in Fig. 9.19.

It is possible to characterize the avalanche breakdown quantitatively by introducing a multiplication factor M such that the reverse current near breakdown is given by MI_0 where I_0 is the saturation current. This factor actually means that an incident electron results in a total of M electron-hole pairs. This factor is empirically given by:

$$M = \frac{1}{1 - \left(\frac{V_r}{V_{br}}\right)^n} \tag{9.65}$$

where V_r is the reverse bias, V_{br} is the breakdown voltage, and n is an exponent in the range 3~6. From this expression, we clearly see that the reverse current, MI_0, increases sharply when V_r nears V_{br} as depicted in Fig. 9.19.

The avalanche process is more likely to occur when a wide enough space charge region can be sustained to ensure sufficient acceleration. This can be more easily achieved by using lightly doped *p-n* junctions because, if heavily doped junctions are used, another phenomenon can more easily occur: the tunneling of charge carriers from one side of the junction to the other.

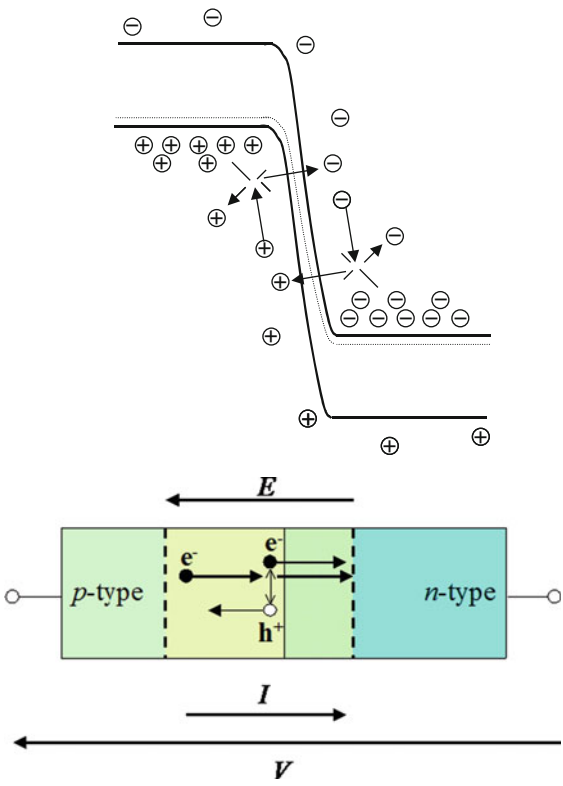

Fig. 9.20 Impact ionization process: under strong reverse bias, electrons and holes are injected into the depletion region; when they gain enough kinetic energy, they impact on the semiconductor lattice to create electron-hole pairs. These newly created carriers can then lead to the same impact ionization process if they can gain enough kinetic energy within the space charge region

Example

Q A voltage-stabilizing diode takes advantage of the steep slope in the breakdown regime to clamp the voltage. For such a kind of diode with $V_{\mathrm{br}} = -14$ V, estimate how many times the current will increase when the reverse bias goes from -13.990 to -13.995 V. Assume $n = 6$.

A The multiplication factor if given by: $M = \frac{1}{1-\left(\frac{V_{\mathrm{r}}}{V_{\mathrm{br}}}\right)^n}$. For the two reverse biases mentioned, we get the ratio of the multiplication factor:

$$\frac{M_1}{M_2} = \frac{1-\left(\frac{V_2}{V_{\mathrm{br}}}\right)^n}{1-\left(\frac{V_1}{V_{\mathrm{br}}}\right)^n} = \frac{1-\left(\frac{13.990}{14}\right)^6}{1-\left(\frac{13.995}{14}\right)^6} = 2$$

The current will thus increase by a factor 2 when the voltage is reduced by 0.005 V.

9.4.5 Zener Breakdown

Under a more moderate reverse bias, typically less than $6E_g/q$, the top of the valence band in the *p*-type side E_{Vp} is already higher than the bottom of the conduction band in the *n*-type side E_{Vc}. This situation is illustrated in Fig. 9.21. This means that the electrons at the top of the valence band in the *p*-type side have the same or higher energy than the empty states available at the bottom of the conduction band in the *n*-type side.

This staggering of the energy bands also results in a reduced spatial separation between the conduction and valence bands, as shown by *d* in Fig. 9.21 .Moreover, in heavily doped *p-n* junctions, the space charge region is already narrow (with a width *W*) and does not expand much under a moderate reverse bias.

The staggered alignment of the energy bands and their spatial proximity favor the tunneling of electrons from the valence band in the *p*-type side into the conduction band in the *n*-type side, as shown in Fig. 9.21. This leads to a negative current. This process is called the Zener effect. As there are many electrons in the valence band and many empty available states in the conduction band, the tunneling current can be substantial.

The Zener tunneling probability T_Z is strongly field dependent on the applied bias *V* and the bandgap E_g. It can be written as:

$$T_Z = \exp\left\{-\frac{4\sqrt{2m^*}}{3qV\hbar}E_g^{3/2}\right\} \tag{9.66}$$

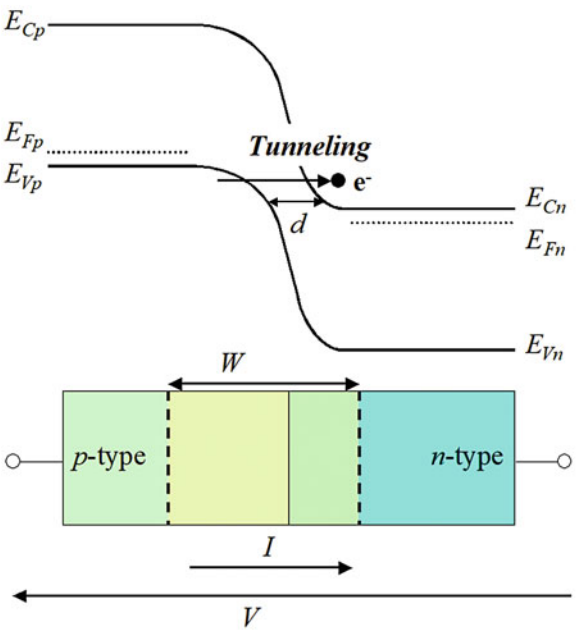

Fig. 9.21 Zener breakdown mechanism involving electrons tunneling from the valence band of the *p*-type side to the conduction band of the *n*-type side

9.5 Metal-Semiconductor Junctions

As we have already mentioned in Subsect. 9.2.6 and illustrated in the case of a *p-n* junction, two dissimilar materials in contact with each other and under thermal equilibrium must have the same value of Fermi energy.

When a metal is brought into contact with a semiconductor, a certain amount of band bending occurs to compensate the difference between the Fermi energies of the metal and that of the semiconductor. In fact, this difference in Fermi energy means that electrons in one material have a higher energy than in the other. These will therefore tend to flow from the former to the later material. There is thus a transfer of electrons across the metal-semiconductor junction in a similar way as the charge transfer in the case of a *p-n* junction. Such a junction is also often called a metallurgic junction or a metal contact because metals are commonly used in semiconductor industry to connect or "contact" a semiconductor material to an external electrical circuit.

The charge transfer can be readily achieved because, as we saw in Fig. 5.11 in Subsect. 5.2.7, the Fermi energy in a metal lies within an energy band, which makes it easy for electrons to be emitted from or received by a metal. This charge redistribution gives rise to a local built-in electric field which counterbalances this redistribution. When sufficiently large electric field strength is established around the metallurgic junction, the redistribution stops.

Since the overall charge neutrality must be maintained, the excess electrical charges inside the semiconductor and that inside the metal must be of an equal amount but with opposite signs. However, because a metal has a much higher charge density than a semiconductor, the width over which these excess charges spread inside the metal is negligibly thin in comparison to the width inside the semiconductor. This is somewhat similar to the case of a *p-n* junction with one side heavily doped. As a result, the built-in electric field and the band bending are primarily present inside the semiconductor as well. The following section aims at giving a quantitative description of the physical properties of a metal-semiconductor junction.

9.5.1 Formalism

The physical parameters which need to be considered in this description are depicted in Fig. 9.22. For the metal, these include its Fermi energy E_{Fm} and work function $\Phi_m > 0$. As we saw when discussing the photoelectric effect in Chap. 4, the work function of a metal is the energy required to extract one electron from the metal surface and pull it into the vacuum. In a more quantitative manner, the work function is the energy difference between the Fermi energy and the vacuum level as shown in Fig. 9.22. For the semiconductor, the parameters of interest also include its Fermi energy E_{Fs}, its work function $\Phi_s > 0$, and also its electron affinity $\chi > 0$. The latter is the energy required to extract one electron from the conduction band of the semiconductor into the vacuum and is given by the energy difference between the bottom

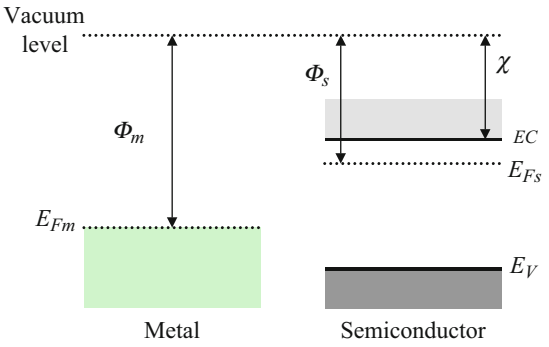

Fig. 9.22 Fermi energies, work functions in a metal and a semiconductor, when considered isolated from each other. The vacuum level is the same for both materials, but the Fermi energies are generally different

of the conduction band and the vacuum level. A few values of electron affinity for elements in the periodic table are given in Fig. A.12 in Appendix A.3.

The amount of band bending and the direction of electron transfer depend on the difference between the work functions of the metal and the semiconductor. When these materials are isolated, their vacuum levels are the same, as illustrated in Fig. 9.22. But, when these materials come into contact, the Fermi energy must be equal on both sides of the junction. The vacuum level is at an energy Φ_m above the top of the metal Fermi energy, while it is Φ_s above the semiconductor Fermi energy. This means that the energy bands in the semiconductor must shift upward by an amount equal to $\Phi_m - \Phi_s$ in order to align the Fermi energy on both sides of the junction.

On the one hand, if $\Phi_m > \Phi_s$, the energy bands of the semiconductor actually shift downward with respect to those of the metal, and electrons are transferred from the semiconductor into the metal, as shown in Fig. 9.23. The signs of the charge carriers which appear on either side of the junction and the direction of the built-in electric field, also shown in Fig. 9.23, are determined from the analysis conducted for a *p-n* junction. On the other hand, if $\Phi_m < \Phi_s$, the energy bands in the semiconductor shift upward with respect to those of the metal, and the electrons are transferred from the metal into the semiconductor.

9.5.2 Schottky and Ohmic Contacts

The electrical properties of a metal-semiconductor junction depend on whether a depletion region is created as a result of the charge redistribution. This phenomenon in turn depends on the difference in work function Φ_m-Φ_s, and on the type of the semiconductor (*n*-type or *p*-type).

Indeed, we know that when $\Phi_m > \Phi_s$, electrons are extracted from the semiconductor into the metal.

If the semiconductor is *n*-type, then this process depletes the semiconductor of its electrons or majority charge carriers. A depletion region thus appears near the junction, and we obtain a diode-like behavior similar to a *p-n* junction when an

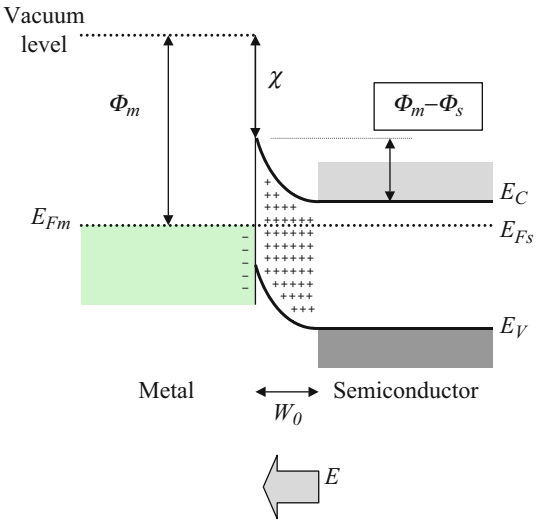

Fig. 9.23 Energy levels, accumulated charge carriers, and built-in electric field in a metal-semiconductor junction. When the metal and the semiconductor are brought into contact, at equilibrium, the energy band profile of the semiconductor near the junction is modified so that the Fermi energies become equal in both materials

external bias is applied. This is shown in Fig. 9.24a. This situation is often called a rectifying contact or Schottky contact.

However, if the semiconductor is *p*-type, the electrons which are extracted from the semiconductor are taken from the *p*-type dopants which then become ionized. This process thus creates more holes or majority charge carriers. In this case, there is no depletion region, but rather majority carriers are accumulated near the junction area, and we do not observe a diode-like behavior. Majority carriers are free to flow in either direction under the influence of an external bias. This is shown in Fig. 9.25a. This situation is often called an ohmic contact and the current-voltage characteristics are linear.

If we now consider $\Phi_m < \Phi_s$, electrons are extracted from the metal into the semiconductor. The previous analysis needs to be reversed. In other words, for an *n*-type semiconductor, the junction will be an ohmic contact, while for a *p*-type semiconductor, the junction will be a Schottky contact.

These four configurations are shown in Figs. 9.23 and 9.24 and summarized in Table 9.1.

In the case of a Schottky contact, the existence of the depletion region means that there is a potential barrier across the junction which can be shifted by an amount equal to $-qV$ when an external voltage V is applied between the metal and the semiconductor. This in turn influences the current flow in a similar way as for a *p-n* junction. This is shown in Fig. 9.26 for the case of an *n*-type semiconductor. It is however important to understand that majority carriers are responsible for the current transport in a metal-semiconductor junction, whereas in a *p-n* junction, it is due to the minority carriers.

The sign convention for a metal-semiconductor junction is the same as for a *p-n* junction by considering the type of the semiconductor. Although the current

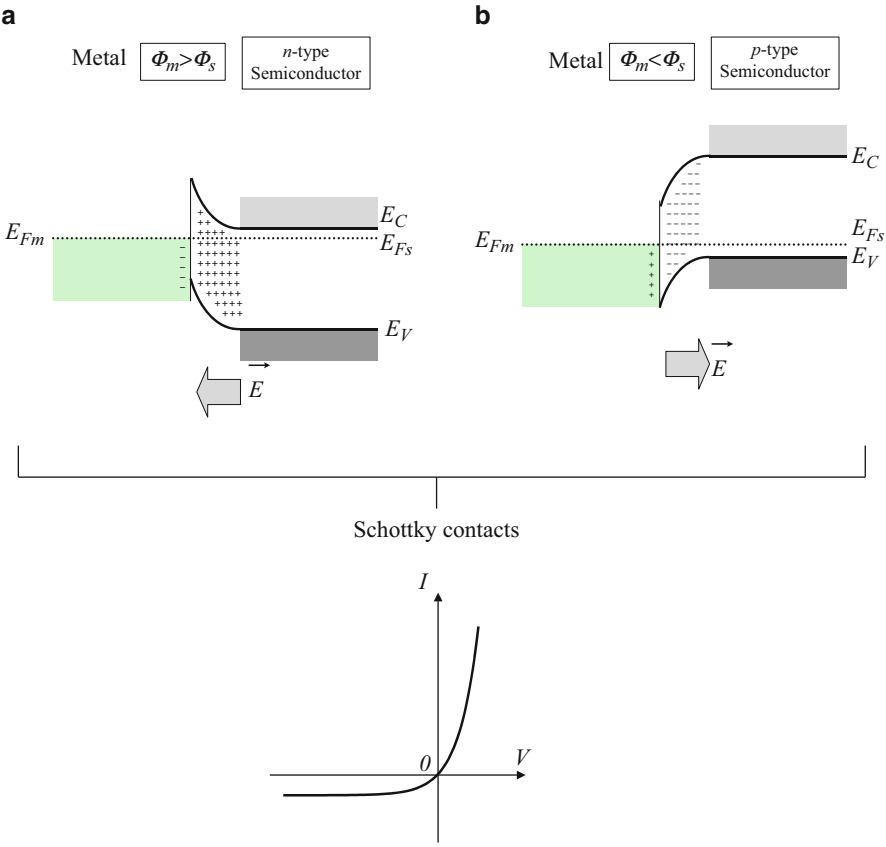

Fig. 9.24 These two of the four possible metal-semiconductor junction configurations lead to a Schottky contact: (**a**) $\Phi_m > \Phi_s$ and *n*-type, (**b**) $\Phi_m < \Phi_s$ and *p*-type. A Schottky contact is obtained in each case because the majority carriers in the semiconductor experience a potential barrier which prevents their free movement across the metal-semiconductor junction, and therefore as shown at the bottom of the figure, the *I-V* characteristic shows rectifying behavior

transport mechanism in a Schottky contact is somewhat different from that in a *p-n* junction, the current-voltage relation for an ideal Schottky contact has a similar expression as for an ideal *p-n* junction:

$$I = I_0\left(e^{\frac{qV}{k_b T}} - 1\right) \tag{9.67}$$

where I_0 is the reverse saturation current and is exponentially proportional to the difference between the metal work function Φ_m and the semiconductor electron affinity χ:

$$I_0 = AB_e T^2 e^{\left(-\frac{(\Phi_m - \chi)}{k_b T}\right)} \tag{9.68}$$

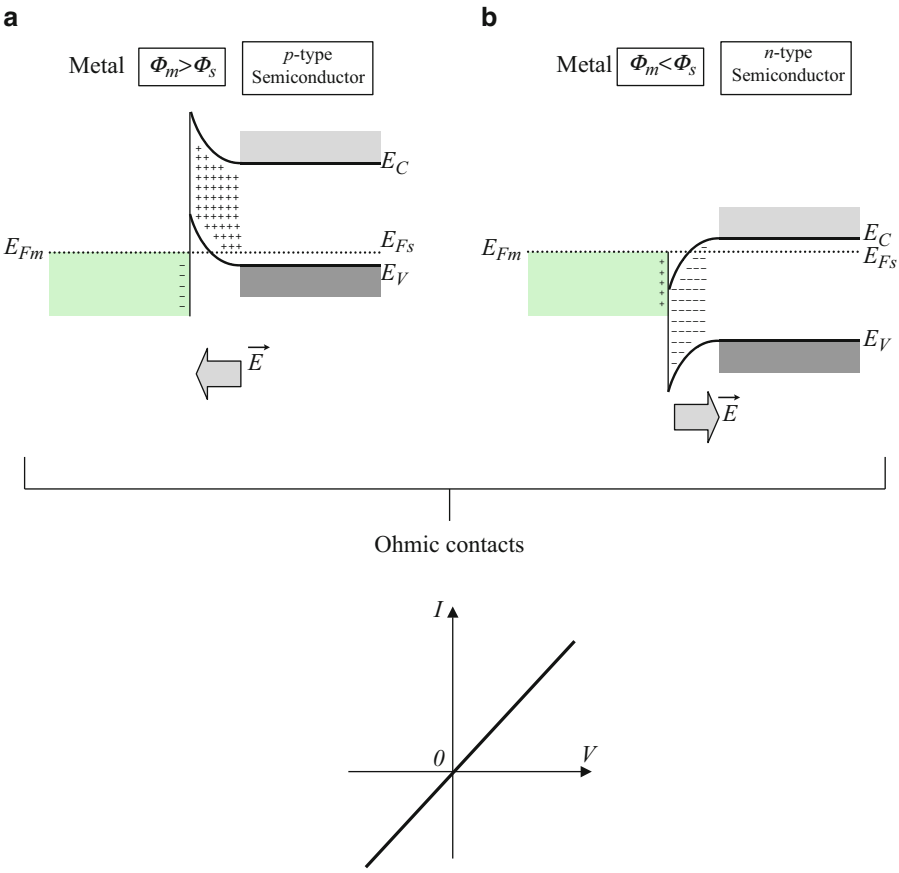

Fig. 9.25 These two of the four possible metal-semiconductor junction configurations lead to an ohmic contact: (**a**) $\Phi_m > \Phi_s$ and *p*-type, (**b**) $\Phi_m < \Phi_s$ and *n*-type. Unlike the configurations shown in Fig. 9.24, the energy band profiles here are such that the majority carriers in the semiconductor can move across the metal-semiconductor junction without experiencing a potential barrier, and therefore as shown at the bottom of the figure, the *I-V* characteristic shows ohmic behavior

Table 9.1 Four possible metal-semiconductor junction configurations and the resulting contact types

	Semiconductor	Junction
$\Phi_m > \Phi_s$	*n*-type	Schottky
$\Phi_m < \Phi_s$	*p*-type	Schottky
$\Phi_m > \Phi_s$	*p*-type	Ohmic
$\Phi_m < \Phi_s$	*n*-type	Ohmic

B_e is the effective Richardson constant, and for most metal-semiconductor Schottky junctions, it varies from 10 to 100 K^{-2} cm^{-2}. The quantity $(\Phi_m - \chi)$ is often denoted $q\Phi_B$, where Φ_B is called the Schottky potential barrier height. For a real Schottky contact, one needs to take into account thermionic emission (Appendix

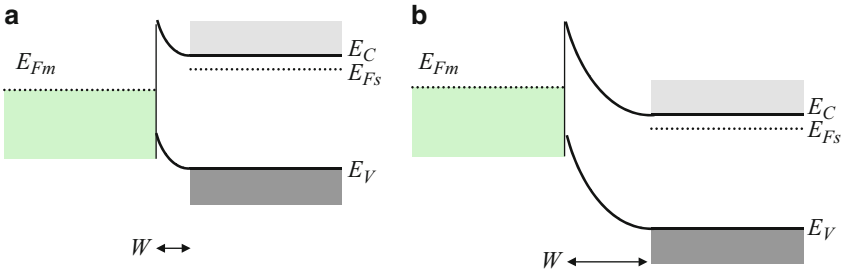

Fig. 9.26 Band alignment in a Schottky metal-(*n*-type) semiconductor contact under (**a**) forward bias where the potential barrier is reduced, and under (**b**) reverse bias where the potential barrier is increased, thus reducing the tunneling of carriers

A.9), as well as impurity and interface states. In this case, the current-voltage relation is given by:

$$I = I_0\left(e^{\frac{qV}{nk_bT}} - 1\right) \tag{9.69}$$

where n is the ideality factor as mentioned before and is typically between 1 and 2.

9.6 Summary

In this chapter, we have presented a complete mathematical model for an ideal *p-n* junction, based on an abrupt homojunction model and the depletion approximation. We introduced the concepts of a space charge region, built-in electric field, built-potential, and depletion width at equilibrium. We have discussed the balance of electrical charges, as well as that of the diffusion and drift currents within the space charge region.

The non-equilibrium properties of *p-n* junctions have also been discussed. The forward bias and reverse bias conditions were examined. We emphasized the importance of minority carrier injection and extraction. We derived the diode equation and understood the nature of the currents outside the space charge region. We have discussed the avalanche and Zener breakdown mechanisms as deviations from the ideal *p-n* junction diode behavior under strong reverse bias conditions.

Finally, we presented the electrical properties of metal-semiconductor junctions and introduced the concepts of Schottky and ohmic contacts.

Problems

1. A *p-n* junction diode has a concentration of $N_A = 10^{17}$ acceptor atoms per cm^3 on the *p*-type side and a concentration of N_D donor atoms per cm^3 on the *n*-type side. Determine the built-in potential V_0 at room temperature for a germanium diode for values of N_D ranging from 10^{14} to 10^{19} cm^{-3}. Also determine the peak

value of the electric field strength for this same range, and plot both of these values as a function of N_D on a semilog scale.

2. Consider a GaAs step junction with $N_A = 10^{17}$ cm^{-3} and $N_D = 5 \times 10^{15}$ cm^{-3}. Calculate the Fermi energy in the *p*-type and *n*-type regions at 300 K. Draw the energy band diagram for this junction. Determine the built-in potential from the diagram and from Eq. (9.22). Compare the results.
3. Consider an asymmetric p^+-*n* junction, which has a heavily doped *p*-type side relative to the *n*-type side, i.e., $N_A >> N_D$. Determine a simplified expression for the width of the space charge region given in Eq. (9.23).
4. Calculate the depletion width for a Si *p-n* junction that has been doped with 10^{18} acceptor atoms per cm^3 on the *p*-type side and 10^{16} donor atoms per cm^3 on the *n*-type side. Compare this depletion width to the width of the depletion region on the n-side (from Eq. (9.33)). What percentage of the width lies within the *n*-type semiconductor. T = 300 K.
5. A silicon *p-n* diode with $N_A = 10^{18}$ cm^{-3} has a built-in voltage of 0.814 eV and capacitance of 10^{-8} F·cm^{-2} at an applied voltage of 0.5 V. Determine the donor density. A = 1 cm^2.
6. Plot the diode equation for an ideal Si *p-n* junction diode with an area 50 μm^2, an acceptor concentration $N_A = 10^{18}$ cm^{-3}, a donor concentration $N_D = 10^{18}$ cm^{-3}, recombination lifetimes equal to $\tau_n = \tau_p = 1$ μs, and diffusion coefficients equal to $D_n = 35$ cm^2·s^{-1} and $D_p = 12.5$ cm^2·s^{-1}.
7. Consider a Si *p-n* step junction with $N_A = 10^{17}$ cm^{-3} and $N_D = 10^{16}$ cm^{-3}, with recombination lifetimes $\tau_p = 0.1$ μs and $\tau_n = 0.01$ μs and carrier mobilities $\mu_h = 450$ cm^2/Vs and $\mu_e = 800$ cm^2/Vs at 300 K.
8. Determine the total reverse saturation current density, the reverse saturation current density due to holes and that due to electrons.
9. Assume a forward bias equal to $V_0/2$ is applied, where V_0, the built-in potential, is equal to 0.7546 V. Calculate the injected minority carrier currents at the edges of the space charge region.
10. Assume a reverse bias equal to $-V_0/2$ is applied. Calculate the minority carrier currents at the edges of the space charge region.
11. A Si *p-n* junction is doped with an acceptor concentration $N_A = 5 \times 10^{18}$ cm^{-3} and a donor concentration $N_D = 5 \times 10^{15}$ cm^{-3}. The critical electric field strength for breakdown is equal to 10^5 V·cm^{-1}. Determine the breakdown voltage and the corresponding depletion width. Do the same for a donor concentration $N_D = 5 \times 10^{17}$ cm^{-3}.
12. Consider an ideal metal-semiconductor junction between *p*-type silicon and polycrystalline aluminum. The Si is doped with $N_A = 5 \times 10^{16}$ cm^{-3}. The metal work function is 4.28 eV and the Si electron affinity is 4.01 eV. Draw the equilibrium band diagram and determine the barrier height ϕ_B.
13. Consider the same silicon-aluminum metal-semiconductor junction. The cross-sectional area of the junction is 10 μm^2. Assume that B_e is 30 AK^{-2} cm^{-2} and the ideality factor *n* is 1. Calculate the reverse saturation current and plot the *I-V* curve as a function of applied bias.

Further Reading

Ashcroft NW, Mermin ND (1976) *Solid state physics*. Holt Rinehart and Winston, New York
Neudeck GW (1989) The PN junction diode. Addison-Wesley, Reading
Pierret RF (1989) Advanced semiconductor fundamentals. Addison-Wesley, Reading
Sapoval B, Hermann C (1995) Physics of semiconductors. Springer, New York
Streetman BG (1990) Solid state electronic devices. Prentice-Hall, Englewood Cliffs
Sze SM (1981) Physics of semiconductor devices. Wiley, New York
Wang S (1989) Fundamentals of semiconductor theory and device physics. Prentice-Hall, Englewood Cliffs