Chapter 2
Junctions and Diodes

"... the totality is not, as it were, a mere heap, but the whole is something besides the parts ..."

Aristotle, *Metaphysics*

A junction between two dissimilar materials is shown schematically in Fig. 2.1 below. Such a junction can take many forms; however, its defining characteristic is the asymmetry that exists upon crossing the junction from one side to the other. This is the essence of the original term used to describe such two-terminal junction devices, viz., *diode*, literally meaning two different paths or roads. Historically, the earliest solid-state junctions were formed in the latter half of the nineteenth century by mechanically pressing a sharp metallic object (e.g., wire) against a semiconducting material; the so-called cat's whisker or point-contact diode.[1] Similar techniques were also used to create junctions between two different semiconductor crystals and between thin layers of metals and semiconductors. Vacuum tube diodes on the other hand, developed around the same time, obtain their asymmetry via the heating of one electrode (the cathode) relative to the other (the anode).[2] The cathode material is chosen such that it readily emits electrons inside the vacuum tube when heated (thermionic emission).

The asymmetry of a junction leads to currents that depend on the polarity of the bias applied to the junction; in other words they possess an asymmetric *I*–*V* characteristic. Such junctions can be engineered to create a wide range of solid-state electronic devices. It is also often necessary, and very important in practice, to mitigate the effects of junctions when making electrical contacts or interconnecting

[1] Often referred to simply as a crystal diode (detector).

[2] Mercury-arc valves were another early type of diode based on a liquid mercury cathode and carbon anode. In this case, electrons are preferentially emitted from the cathode upon formation of a flame or arc discharge (through the contained mercury vapor).

C. Papadopoulos, *Solid-State Electronic Devices: An Introduction*,
Undergraduate Lecture Notes in Physics, DOI 10.1007/978-1-4614-8836-1_2,

Fig. 2.1 General schematic of two different materials in intimate contact. A junction is formed at the interface between A and B

multiple devices. In this chapter, we consider the main types of solid-state junctions and their applications based on diodes. Such junctions form the basic building blocks of all solid-state electronic devices and thus their properties will play a critical role throughout the subsequent chapters.

2.1 *pn* Junctions

The first type of junction we consider is between two regions in a semiconductor having different doping type. The earliest experiments on *pn* junctions are attributed to Ohl working at Bell labs in 1940 on the properties of silicon crystals.[3] Following this discovery, the theoretical understanding of such junctions was largely developed by Shockley.[4]

2.1.1 Thermal Equilibrium and the Built-In Potential

We use the fact that the Fermi level must be constant in thermal equilibrium to construct the band edge diagram for junctions between different materials. In the case of a *pn* junction, before the two materials are in contact we have the situation depicted in Fig. 2.2a: E_0 is the vacuum level and represents the energy of free electrons (in other words, carriers that are no longer bound to the material). The difference between the Fermi level and E_0 is called the *work function*, $q\Phi$, of the material. We also define the difference between the conduction band edge, E_c, and the vacuum level as the *electron affinity*, qX, which is constant for a given semiconductor.

If the two regions now come into contact and combine to form a junction, the large carrier concentration gradients that exist at the interface will cause a transfer of carriers in order to align the Fermi levels and achieve thermal equilibrium. This leads to a *depletion layer* or *space-charge* region near the interface, caused by uncompensated (fixed) impurity ions.[5] The thermal equilibrium band edge

[3] The discovery of an unintentional junction in a rod of crystalline silicon by Ohl led to the terms "n-type" and "p-type" being coined to describe the different doping on either side of the junction. Ohl's experiments also resulted in the demonstration of the first *pn* junction solar cell.

[4] W. Shockley, Bell Syst. Tech. J. **28**, 435 (1949). This was a seminal paper in the history of semiconductor electronics and forms much of the foundation on which Chapters 2 and 3 are based.

[5] To equalize the Fermi levels there will be a net transfer of electrons toward the p-type region and recombination with holes results in the space-charge layer at equilibrium.

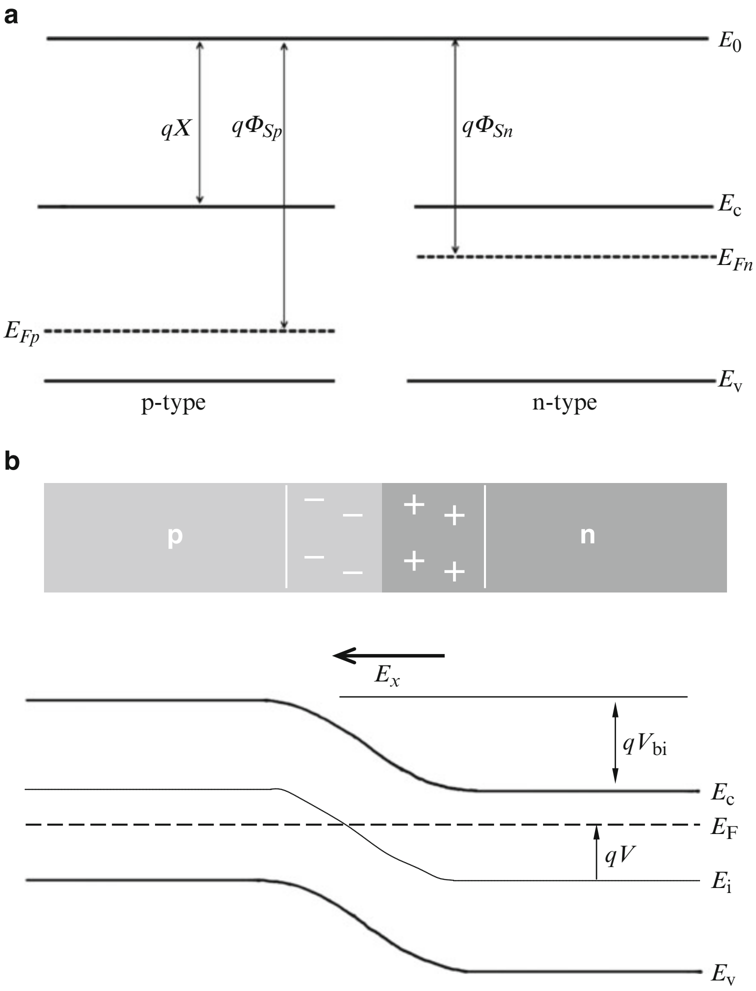


Fig. 2.2 (**a**) Band edge diagrams for isolated p-type and n-type semiconductors, including the vacuum reference energy level, E_0. (**b**) Thermal equilibrium *pn* junction band edge diagram, characterized by the built-in potential barrier V_{bi}. The potential at any point in the junction can also be defined by V as shown. (The electric field in the x-direction (E_x) always points "uphill" with respect to the conduction band edge.) A schematic of the built-in charge of the depletion layer appears above the band edge diagram. The interface should be visualized (in both diagrams) as a two-dimensional sheet or plane (coming out of the page) where the three-dimensional crystals meet. Note: To qualitatively sketch junction band edge diagrams one should start by drawing the thermal equilibrium Fermi level as a horizontal line, followed by sketching the band edge diagrams away from the interface on either side (i.e., in the bulk regions, where the bands are flat). Lastly, the conduction and valance band edges should be smoothly joined together in the junction region where the two materials meet

diagram for the junction thus has the form depicted in Fig. 2.2b. The resulting electric field is characterized by the *built-in potential*, V_{bi}, at equilibrium. In thermal equilibrium the built-in field exactly balances the tendency of carriers to diffuse across the junction so that there is no net current flow.

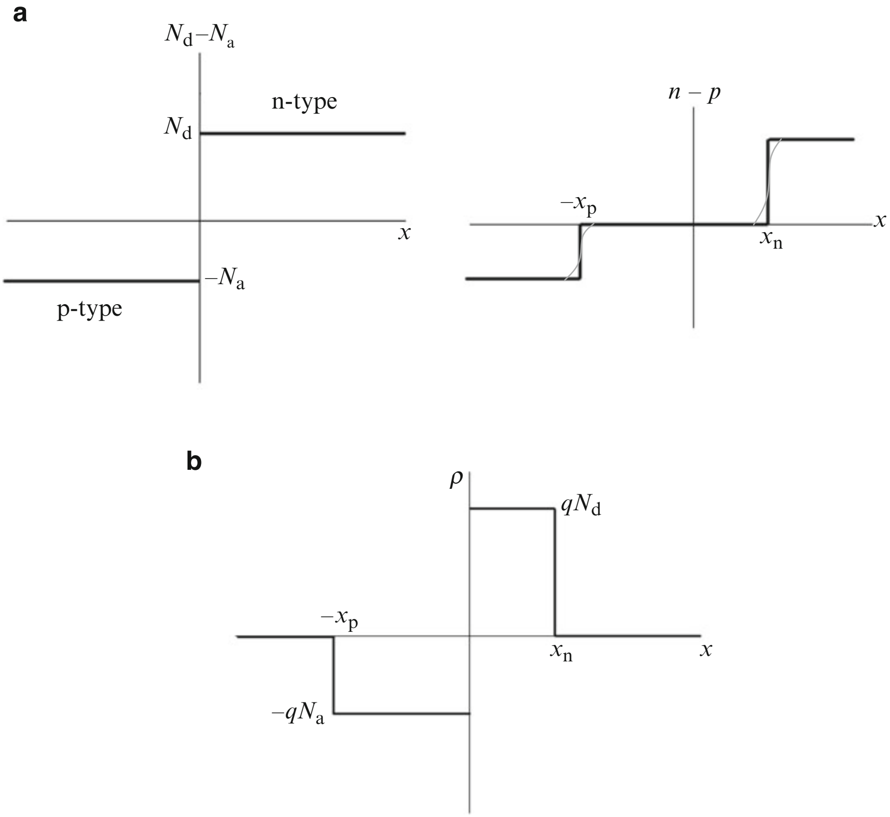


Fig. 2.3 (**a**) Depletion approximation for an abrupt *pn* junction: Doping levels for the step junction (*left*) and resulting mobile carrier concentrations within the depletion approximation (*right*; *thin line* shows behavior of exact solution). (**b**) Charge density for the step *pn* junction under the depletion approximation. (**c**) Electric field across step *pn* junction for charge distribution of Fig. 2.3b. (The field is negative because it is pointing the negative *x*-direction.) (**d**) Electric potential variation across *pn* junction for field shown in Fig. 2.3c (Note that the reference point for the potentials can be shifted without loss of generality)

To solve for the potential barrier V_{bi} one usually employs the *depletion approximation*, which ignores any contribution to the space charge from free electrons and holes, as depicted in Fig. 2.3a. This is a very good approximation due to the exponential dependence of free carrier concentration with Fermi-level position; in the vicinity of the junction the magnitude of $(E_F - E_i)$ becomes small (Fig. 2.2b) and hence the free carrier concentration rapidly decreases. Note that we are also assuming (as in Fig. 2.2) what is known as an abrupt or step junction, in that the doping levels in the p- and n-type materials are constant away from the junction and abruptly change at the junction interfacial region.[6]

[6] The interface or plane where the doping type changes is known as the *metallurgical junction*. The same terminology is used to describe the physical or material transition interface in any type of solid-state junction.

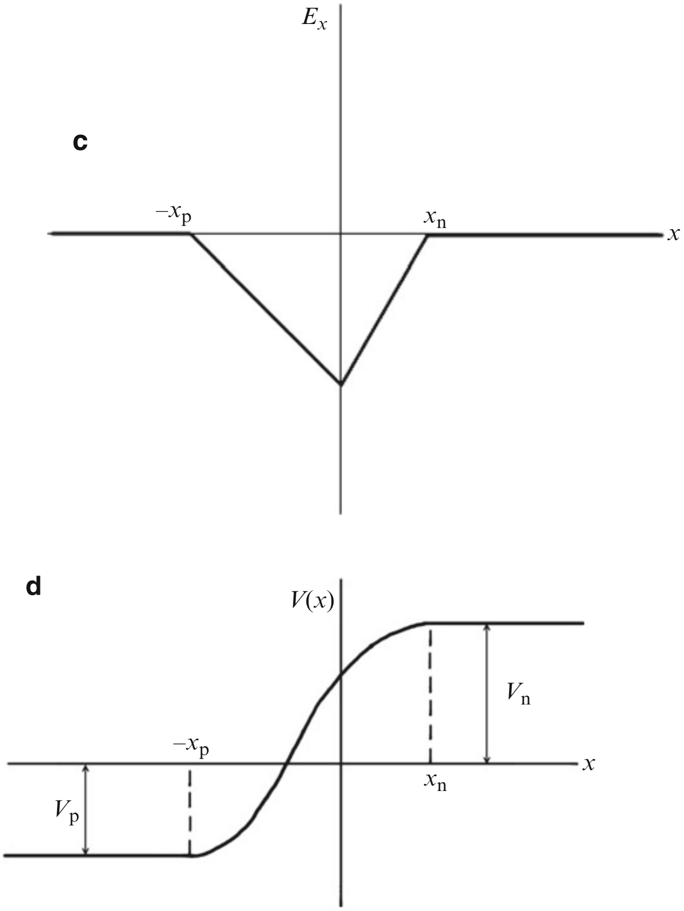


Fig. 2.3 (continued)

This allows us to write Poisson's equation in the space-charge region as[7]

$$\frac{d^2V}{dx^2} = \frac{-q}{\varepsilon_s}(N_d - N_a) \tag{2.1}$$

where the charge density has the form shown in Fig. 2.3b. Thus in the n-type material ($x > 0$) Poisson's equation becomes

[7] Throughout this book the device analysis and descriptions are mainly constrained to one spatial dimension (e.g., x) in order to emphasize and develop an understanding of the basic principles of solid-state electronic devices. The 1D picture will be extended to include additional dimensions/effects when necessary.

$$\frac{d^2V}{dx^2} = -\frac{dE_x}{dx} = \frac{-qN_\mathrm{d}}{\varepsilon_\mathrm{s}} \tag{2.2}$$

which can be integrated to the edge of the depletion region at x_n, where the material becomes neutral and the field vanishes:

$$E_x = -\frac{qN_\mathrm{d}}{\varepsilon_\mathrm{s}}(x_\mathrm{n} - x), \quad 0 \le x \le x_\mathrm{n} \tag{2.3}$$

Similarly, in the p-type region we can find

$$E_x = -\frac{qN_\mathrm{a}}{\varepsilon_s}\left(x + x_\mathrm{p}\right), \quad -x_\mathrm{p} \le x \le 0 \tag{2.4}$$

Thus the field is negative throughout the depletion region and varies linearly with x, reaching a maximum at $x = 0$ (the metallurgical junction) as shown in Fig. 2.3c. The electric field must also be continuous at the interface, $x = 0$, so that

$$N_\mathrm{a}x_\mathrm{p} = N_\mathrm{d}x_\mathrm{n} \tag{2.5}$$

This equation tells us that the width of the depletion region on either side of the junction varies inversely with doping concentration. In other words, the depletion region extends primarily into the side which has the *lightest* doping level. Note that Eq. (2.5) is also another statement of (global) space-charge neutrality (see Appendix A, Eq. (A.31)).

Since

$$E_x = -\frac{dV(x)}{dx}$$

we can integrate the above expressions for the field to obtain the potential variation across the junction. This gives

$$\begin{aligned} V(x) &= V_\mathrm{n} - \frac{qN_\mathrm{d}}{2\varepsilon_\mathrm{s}}(x_\mathrm{n} - x)^2, \quad 0 \le x \le x_\mathrm{n} \\ V(x) &= V_\mathrm{p} + \frac{qN_\mathrm{a}}{2\varepsilon_\mathrm{s}}\left(x + x_\mathrm{p}\right)^2, \quad -x_\mathrm{p} \le x \le 0 \end{aligned} \tag{2.6}$$

where V_n and V_p are the potentials at the neutral edges of the depletion region on either side of the junction as shown in Fig. 2.3d. The parabolic dependence of the potential is to be expected upon integrating Poisson's equation twice for a constant charge density and also describes the curvature of the band edges across the space-charge region.[8]

[8] Electron energies (and hence the energy band edges) are given by $E = -qV$. See Fig. A.13a in Appendix A for the general relation between the band edges and electric field.

The potentials at the boundaries of the junction can be found by noting

$$n = n_{\mathrm{i}}\ \exp\left(\frac{qV_{\mathrm{n}}}{k_{\mathrm{B}}T}\right);\ \ p = n_{\mathrm{i}}\ \exp\left(\frac{-qV_{\mathrm{p}}}{k_{\mathrm{B}}T}\right) \tag{2.7a}$$[9]

which gives

$$\begin{aligned} V_{\mathrm{n}} &= \frac{k_{\mathrm{B}}T}{q}\ln\frac{N_{\mathrm{d}}}{n_{\mathrm{i}}} \\ V_{\mathrm{p}} &= \frac{-k_{\mathrm{B}}T}{q}\ln\frac{N_{\mathrm{a}}}{n_{\mathrm{i}}} \end{aligned} \tag{2.7b}$$

The total built-in potential across the junction, V_{bi}, can be found by integrating the field expressions,

$$V_{\mathrm{bi}} = -\int_{-x_{\mathrm{p}}}^{x_{\mathrm{n}}} E_x dx \tag{2.8}$$

However, this just equals the difference in potentials at the junction edges, which allows us to finally obtain

$$V_{\mathrm{bi}} = V_{\mathrm{n}} - V_{\mathrm{p}} = \frac{k_{\mathrm{B}}T}{q}\ln\frac{N_{\mathrm{d}}N_{\mathrm{a}}}{n_{\mathrm{i}}^2} \tag{2.9}$$

Equation (2.9) can also be obtained in an essentially equivalent manner by noting that the difference between the Fermi levels[10] of the two materials before they are brought together must also equal the potential energy barrier, qV_{bi}. This can be related to the band edge diagram of the junction (Fig. 2.2b) by noting that the difference in conduction band edges on either side of the junction is also equal to qV_{bi}:

$$qV_{\mathrm{bi}} = E_{\mathrm{cp}} - E_{\mathrm{cn}} \tag{2.10}$$

Now, using the equations relating the conduction band edge to electron concentration in a semiconductor (see footnote 9), i.e.,

$$\begin{aligned} n_{\mathrm{n}} &= N_{\mathrm{c}}e^{-(E_{\mathrm{cn}}-E_{\mathrm{F}})/k_{\mathrm{B}}T} \\ n_{\mathrm{p}} &= N_{\mathrm{c}}e^{-\left(E_{\mathrm{cp}}-E_{\mathrm{F}}\right)/k_{\mathrm{B}}T} \end{aligned} \tag{2.11}$$

we can write

$$\frac{n_{\mathrm{n}}}{n_{\mathrm{p}}} = \exp\left(\frac{E_{\mathrm{cp}} - E_{\mathrm{cn}}}{k_{\mathrm{B}}T}\right) \tag{2.12}$$

[9] See Appendix A, Sect. A.2.

[10] This is also equivalent to the difference in work functions of the two separated semiconductors (cf. Fig. 2.2a).

Using Eq. (2.10) now gives

$$V_{\mathrm{bi}} = \frac{k_{\mathrm{B}}T}{q}\ln\left(\frac{n_{\mathrm{n}}}{n_{\mathrm{p}}}\right) = \frac{k_{\mathrm{B}}T}{q}\ln\left(\frac{N_{\mathrm{d}}N_{\mathrm{a}}}{n_{\mathrm{i}}^{2}}\right)$$

as before.

Finally, by using the continuity of $V(x)$ at $x = 0$ in the equations for the potential found above we can write

$$V_{\mathrm{bi}} = V_{\mathrm{n}} - V_{\mathrm{p}} = \frac{q}{2\varepsilon_{\mathrm{s}}}\left(N_{\mathrm{d}}x_{\mathrm{n}}^{2} + N_{\mathrm{a}}x_{\mathrm{p}}^{2}\right) \tag{2.13}$$

which allows us to solve[11] for the depletion widths in terms of the built-in potential:

$$\begin{aligned} x_{\mathrm{n}} &= \left\{\frac{2\varepsilon_{\mathrm{s}}V_{\mathrm{bi}}}{q}\left[\frac{N_{\mathrm{a}}}{N_{\mathrm{d}}(N_{\mathrm{a}}+N_{\mathrm{d}})}\right]\right\}^{1/2} \\ x_{\mathrm{p}} &= \left\{\frac{2\varepsilon_{\mathrm{s}}V_{\mathrm{bi}}}{q}\left[\frac{N_{\mathrm{d}}}{N_{\mathrm{a}}(N_{\mathrm{a}}+N_{\mathrm{d}})}\right]\right\}^{1/2} \end{aligned} \tag{2.14}$$

and thus the total width of the depletion region is given by

$$x_{\mathrm{d}} = x_{\mathrm{n}} + x_{\mathrm{p}} = \left[\frac{2\varepsilon_{\mathrm{s}}}{q}V_{\mathrm{bi}}\left(\frac{1}{N_{\mathrm{a}}} + \frac{1}{N_{\mathrm{d}}}\right)\right]^{1/2} \tag{2.15}$$

Once again we see that the depletion width of a *pn* junction depends most strongly on the material with the lighter doping.

Example 2.1: pn Junction Built-In Potential Calculation A region of n-type silicon with $\rho = 4$ Ω-cm is used to make a *pn* junction with a p-region that has $\rho = 0.2$ Ω cm. Find V_{bi} for the junction.

From the resistivity data for silicon given in Appendix B we can look up that $N_{\mathrm{d}} = 10^{15}$ cm^{-3} and $N_{\mathrm{a}} = 10^{17}$ cm^{-3}. The built-in potential is therefore given by

$$V_{\mathrm{bi}} = \frac{k_{\mathrm{B}}T}{q}\ln\left(\frac{N_{\mathrm{d}}N_{\mathrm{a}}}{n_{\mathrm{i}}^{2}}\right) \approx 0.7\ \mathrm{V}$$

Panel 2.1: Built-In Potential for Heavily Doped Junctions At very high carrier concentrations due to heavily doped semiconductors the equation derived for V_{bi} above is no longer valid because it is based on the exponential approximation

[11] Using $N_{\mathrm{a}}x_{\mathrm{p}} = N_{\mathrm{d}}x_{\mathrm{n}}$.

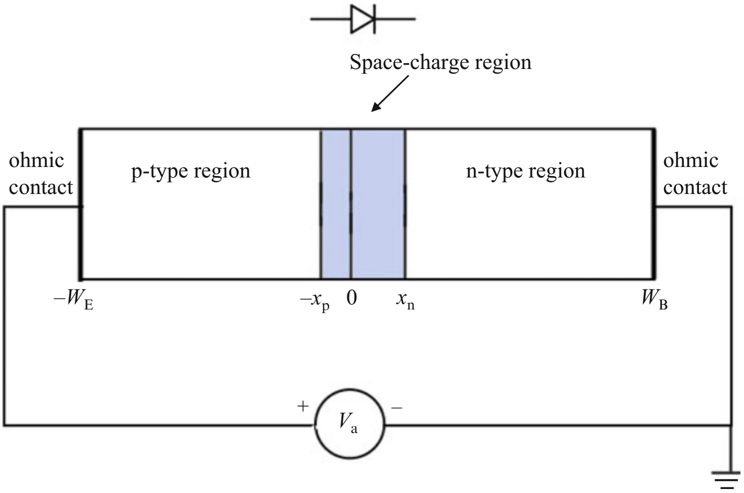


Fig. 2.4 *pn* junction device structure and biasing configuration (Electronic circuit symbol also shown)

of the Fermi–Dirac distribution.[12] When the dopant densities approach N_c or N_v (~ 10^{19} cm^{-3} for Si) the full carrier distribution function should instead be used for derivations.

However, at very high dopant concentrations the Fermi level lies very near the band edges and the potential (V_n or V_p) in the heavily doped side of the junction is approximately one-half of the band gap energy divided by q or about 0.56 V for silicon. For example, the built-in potential for a *pn* junction composed of heavily doped p-type silicon (denoted p^+n)[13] is

$$V_{\mathrm{bi}} = 0.56\ \mathrm{V} + \frac{k_B T}{q}\ln\left(\frac{N_d}{n_i}\right)$$

Similarly, a p^+n^+ junction would have V_{bi} roughly equal to the magnitude of the band gap energy.

2.1.2 *pn Junction I–V Characteristic*

Assume that the *pn* junction is contacted by low-resistance (or ohmic) contacts as illustrated in Fig. 2.4. Since the space-charge region is depleted of mobile carriers, the applied voltage, V_a, will appear almost entirely across this high-resistance potential barrier region of the junction.

If V_a is positive the built-in voltage will be reduced and the junction is said to be *forward biased*, whereas if V_a is negative the built-in voltage is increased and the junction is said to be *reverse biased*. The applied bias will thus alter

[12] See Appendix A, Eq. (A.25).

[13] p^+n and pn^+ junctions are often termed *one-sided pn* junctions.

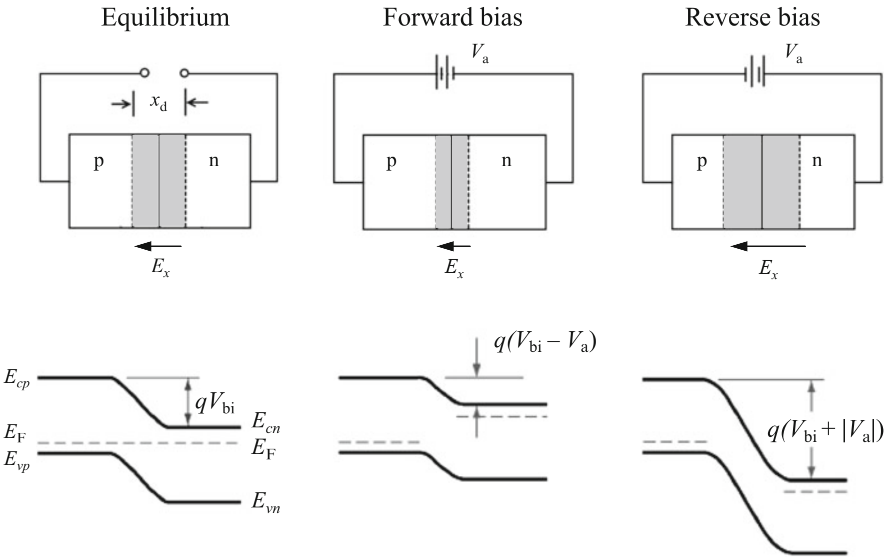


Fig. 2.5 Effect of applied bias on *pn* junction (with similar doping levels on either side). In thermal equilibrium the Fermi level is constant across the entire structure and the built-in electric field opposes the diffusion of carriers across the junction. The applied bias is assumed to drop almost entirely across the depletion layer and thus the Fermi level becomes separated by qV_a. Forward bias reduces the built-in electric field and thus the built-in potential barrier is *reduced* by V_a as shown. Reverse bias on the other hand increases the built-in electric field and the built-in potential is *increased* by V_a

the barrier to diffusion of carriers across the junction that existed in thermal equilibrium as depicted in Fig. 2.5.[14]

Minority carrier densities are crucial to determining current flow through the *pn* junction, whether under forward or reverse bias. We shall see below that the majority carriers act only to supply the minority carrier current “injected” across the junction or to neutralize charge (via recombination) outside of the space-charge region. The reason for this arises from the inherent asymmetry of the *pn* junction; the different doping types lead to few holes in the n-side and conversely few electrons in the p-side. Therefore the majority carriers on either side cannot dominate current flow throughout the *pn* junction structure but must instead facilitate this process from afar, as we now discuss.

In order to derive an expression for the *I–V* characteristic of a *pn* junction we study the dynamics of minority charge carriers in the n- and p-type regions by seeking solutions of the continuity equations for the minority carrier densities on either side of the junction. The most straightforward way of doing this is to assume *low-level injection*, i.e., we assume that the majority carrier concentrations

[14] Note that, more rigorously, the Fermi levels under applied bias should be referred to as *quasi*-Fermi levels since the system is no longer in thermal equilibrium. In this textbook we will not make use of this distinction.

are not appreciably altered by the applied bias. This condition will be seen to be valid for low-to-moderate applied biases, relative to the built-in potential.

Under the assumption of low-level injection the electron density in the n-type region at the boundary with the depletion layer (i.e., at x_{n}) is equal to the dopant density, N_{d}, whether at equilibrium or under bias. Similar comments apply to the hole density on the other side of the junction (at $-x_{\mathrm{p}}$). The equilibrium *minority* carrier concentrations on either side of the junction can now be written as [see Eqs. (2.12) and (2.10)]

$$\begin{aligned} n_{\mathrm{p0}}(-x_{\mathrm{p}}) &= n_{\mathrm{n0}}(x_{\mathrm{n}})\exp\left(\frac{-qV_{\mathrm{bi}}}{k_{\mathrm{B}}T}\right) = N_{\mathrm{d}}(x_{\mathrm{n}})\exp\left(\frac{-qV_{\mathrm{bi}}}{k_{\mathrm{B}}T}\right) \\ p_{\mathrm{n0}}(x_{\mathrm{n}}) &= p_{\mathrm{p0}}(-x_{\mathrm{p}})\exp\left(\frac{-qV_{\mathrm{bi}}}{k_{\mathrm{B}}T}\right) = N_{\mathrm{a}}(-x_{\mathrm{p}})\exp\left(\frac{-qV_{\mathrm{bi}}}{k_{\mathrm{B}}T}\right) \end{aligned} \tag{2.16}$$

Under an applied bias, V_{a}, these become

$$\begin{aligned} n_{\mathrm{p}}(-x_{\mathrm{p}}) &= N_{\mathrm{d}}(x_{\mathrm{n}})\exp\left[\frac{-q(V_{\mathrm{bi}}-V_{\mathrm{a}})}{k_{\mathrm{B}}T}\right] \\ p_{\mathrm{n}}(x_{\mathrm{n}}) &= N_{\mathrm{a}}(-x_{\mathrm{p}})\exp\left[\frac{-q(V_{\mathrm{bi}}-V_{\mathrm{a}})}{k_{\mathrm{B}}T}\right] \end{aligned} \tag{2.17}$$

We can combine the four equations above to express the *excess* minority carrier concentrations at the junction boundaries in terms of their thermal equilibrium values:

$$\begin{aligned} n'_{\mathrm{p}}(-x_{\mathrm{p}}) &= n_{\mathrm{p0}}(-x_{\mathrm{p}})\left[\exp\left(\frac{qV_{\mathrm{a}}}{k_{\mathrm{B}}T}\right)-1\right] \\ p'_{\mathrm{n}}(x_{\mathrm{n}}) &= p_{\mathrm{n0}}(x_{\mathrm{n}})\left[\exp\left(\frac{qV_{\mathrm{a}}}{k_{\mathrm{B}}T}\right)-1\right] \end{aligned} \tag{2.18}$$ [15]

These equations give the important result that the minority carrier density depends exponentially on applied bias while the majority carrier density is assumed insensitive to it (to first order). Note that since the minority carrier densities at thermal equilibrium are typically at least 10 orders of magnitude below the majority carrier densities the assumption of low-level injection will be valid until the exponential factor becomes approximately equal to about 10^{10} for a moderately doped junction and several orders of magnitude larger for a more heavily doped junction.

Having obtained the bias dependence of excess minority carriers at the edges of the depletion region, we can now find solutions of the continuity equations in order to describe how these excess carriers diffuse away from the *pn* junction interface and into the neutral regions, based on an idealized or simplified system, which is typically referred to as *ideal diode analysis*:

[15] Recall form Appendix A, Eq. (A.47), the excess carrier concentration is defined as the difference between the total concentration and the thermal equilibrium concentration.

Since we are interested in the minority carrier diffusion current away from the space-charge region the electric field contribution to this current density is assumed to be negligible.[16] In this case the continuity equations (see Eqs. (A.48) in Appendix A) become

$$\begin{aligned}\frac{\partial n'}{\partial t} &= D_n \frac{\partial^2 n'}{\partial x^2} - \frac{n'}{\tau_n} \\ \frac{\partial p'}{\partial t} &= D_p \frac{\partial^2 p'}{\partial x^2} - \frac{p'}{\tau_p}\end{aligned} \tag{2.19}$$

where it has also been assumed that we have constant doping densities along x (i.e., a step or abrupt junction). Equation (2.19) gives the *diffusion equations*[17] for electrons and holes.

For the case of steady-state (or dc) injection of holes into the n-side of a *pn* junction we can rewrite the diffusion equation as

$$0 = D_p \frac{d^2 p'_{\mathrm{n}}}{dx^2} - \frac{p'_{\mathrm{n}}}{\tau_p} \tag{2.20}$$

This has the exponential solution

$$p'_{\mathrm{n}}(x) = A\ \exp\left(-\frac{x - x_{\mathrm{n}}}{\sqrt{D_p \tau_p}}\right) + B\ \exp\left(\frac{x - x_{\mathrm{n}}}{\sqrt{D_p \tau_p}}\right) \tag{2.21}$$

where A and B are constants and

$$L_p \equiv \sqrt{D_p \tau_p} \tag{2.22}$$

is called the hole *diffusion length.*[18] Similarly, L_n is the corresponding electron diffusion length in the p-type region.

[16] Recall that most of the voltage drop will occur inside the space charge region.

[17] This type of partial differential equation appears in many other fields including heat flow, mass transport in gases and fluids, etc.

[18] The diffusion length represents the average distance a minority carrier travels before recombining.

To evaluate the constants we consider two limiting cases based on the length[19] $x_{\mathrm{B}} = W_{\mathrm{B}} - x_{\mathrm{n}}$ of the neutral n-region from the junction to the ohmic contact:

1. Long-base diode
If x_{B} is much greater than the hole diffusion length, essentially all of the injected holes will recombine before reaching the contact. Thus we can neglect the growing exponential term in the solution and set B equal to zero. The constant A can then be found by using Eq. (2.18) for the excess hole concentration at x_{n} as a function of applied voltage since

$$p'_{\mathrm{n}}(x_{\mathrm{n}}) = A$$

The solution is therefore

$$p'_{\mathrm{n}}(x) = p_{\mathrm{n0}}\left(e^{qV_{\mathrm{a}}/k_{\mathrm{B}}T} - 1\right)\exp\left(-\frac{x - x_{\mathrm{n}}}{L_p}\right) \tag{2.23}$$ [20]

The result is plotted in Fig. 2.6a. The hole diffusion current density can now be found using

$$\begin{aligned} J_p(x) &= -qD_p\frac{dp_{\mathrm{n}}}{dx} \\ &= qD_p\frac{p_{\mathrm{n0}}}{L_p}\left(e^{qV_{\mathrm{a}}/k_{\mathrm{B}}T} - 1\right)\exp\left(-\frac{x - x_{\mathrm{n}}}{L_p}\right) \end{aligned} \tag{2.24}$$

The hole current is therefore greatest at $x = x_{\mathrm{n}}$ (the edge of the space-charge region) and decreases away from the junction because the hole concentration gradient decreases as carriers are lost by recombination. This means that in order for the current to remain constant (as it must in steady state) the *electron current* must *increase* away from the junction as indicated in Fig. 2.6b. This current supplies the electrons with which the holes recombine. On the other hand, at the junction the only electron current flowing is the one injected across the space-charge layer and into the p-region. The electrons injected into the p-region constitute the corresponding minority carrier diffusion current on the other side of the junction.

Thus, the *total* current flowing through the *pn* junction is obtained by summing the two minority carrier injection currents: holes into the n-side plus electrons into the p-side. The injected electron current can be found by a treatment analogous to

[19] The reason for the subscripts "*B*" and "*E*" in the *pn* junction symbols is for historical reasons relating to the bipolar transistor (Base, Emitter) discussed in Chap. 3.

[20] For simplicity $p_{\mathrm{n0}}(x_{\mathrm{n}})$ is written as p_{n0}. Note that for a step junction p_{n0} will be constant throughout the neutral n-type region.

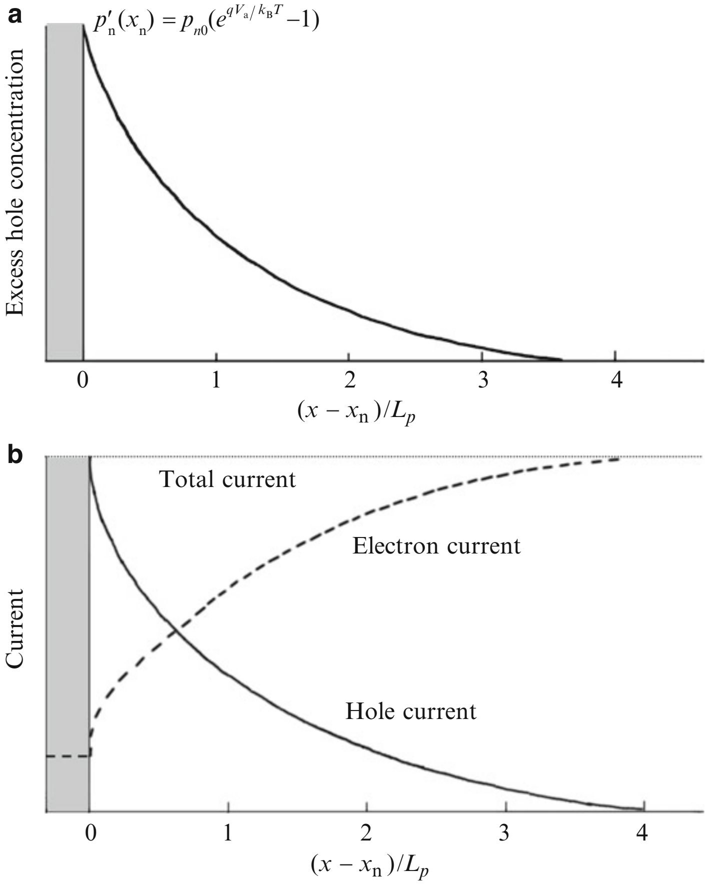


Fig. 2.6 Long-base diode under forward bias. (**a**) Excess minority hole concentration vs. distance in the n-region. (**b**) Electron and hole currents in the n-region

that used above for holes (still under the assumption of a long-“base” diode, so that $x_E = W_E - x_p$ is much greater than the electron diffusion length):

$$J_n(x) = qD_n \frac{n_{p0}}{L_n} \left(e^{qV_a/k_BT} - 1\right) \exp\left(\frac{x + x_p}{L_n}\right) \tag{2.25}$$[21]

To obtain an expression for the total current we now sum the minority carrier current components at $-x_p$ and $+x_n$ to get

[21] Note that the x-coordinate is negative in this expression.

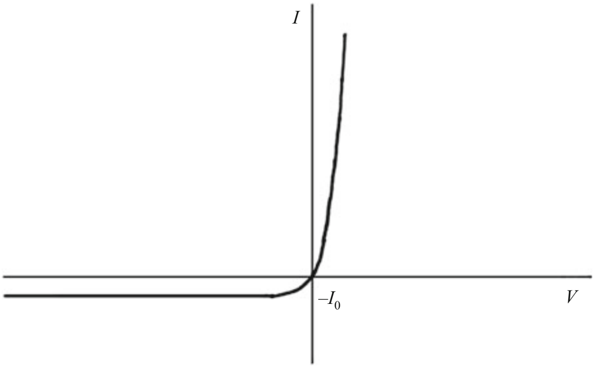


Fig. 2.7 Ideal *pn* junction *I–V* characteristic

$$J = J_p(x_\mathrm{n}) + J_n(-x_\mathrm{p}) = qn_\mathrm{i}^2\left(\frac{D_p}{N_\mathrm{d}L_p} + \frac{D_n}{N_\mathrm{a}L_n}\right)\left(e^{qV_\mathrm{a}/k_\mathrm{B}T} - 1\right)$$
$$= J_0\left(e^{qV_\mathrm{a}/k_\mathrm{B}T} - 1\right) \tag{2.26}$$

where J_0 is the magnitude of the (reverse) *saturation current density* predicted by this model for negative applied bias greater than a few $k_\mathrm{B}T/q$ volts (expressed above for the case of a step *pn* junction). This expression is known as the *ideal diode equation* and is plotted in Fig. 2.7. This form of equation and the exponential dependence is very common in semiconductor electronic devices since it ultimately derives from the fundamental carrier statistics inside these materials, which are governed by the Fermi–Dirac (or Maxwell–Boltzmann) distributions.[22]

2. Short-base diode

In this case the lengths x_B and x_E of the n- and p-type regions are much shorter than the diffusion lengths L_p and L_n. Thus very little recombination occurs in the bulk of the n- and p-type regions. In this limit, almost all the injected minority carriers recombine at the ohmic contacts at either end of the diode structure and thus the excess minority carrier distribution can be approximated as being essentially linear, so that

$$p'_\mathrm{n}(x) = A' + B'\frac{x - x_\mathrm{n}}{L_p} \tag{2.27}$$

To find the constants we once again apply appropriate boundary conditions. The ohmic contact at $x = W_\mathrm{B}$ can be considered a perfect sink for any excess carriers and therefore

[22] The applied bias can be thought of as a small perturbation to the equilibrium properties of a material; hence, even when current flows through electronic devices they are still close to thermal equilibrium under most conditions.

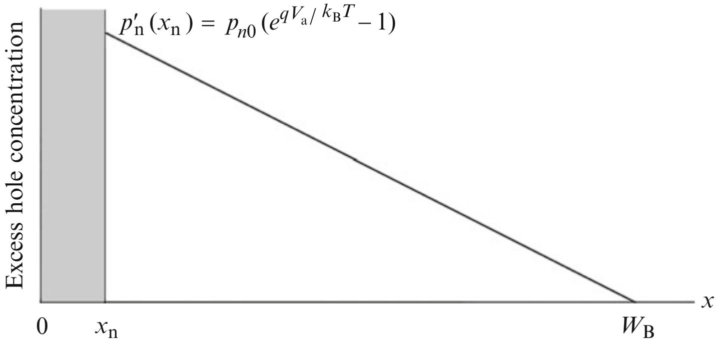


Fig. 2.8 Short-base diode excess hole carrier distribution in n-region under forward bias

$$p'_{\mathrm{n}}(W_{\mathrm{B}}) = 0$$

The other boundary condition is that

$$p'_{\mathrm{n}}(x_{\mathrm{n}}) = A'$$

as in the long-base diode.

The solution for the excess hole density in the n-type region therefore becomes

$$p'_{\mathrm{n}}(x) = p_{\mathrm{n}0}\left(e^{qV_{\mathrm{a}}/k_{\mathrm{B}}T} - 1\right)\left(1 - \frac{x - x_{\mathrm{n}}}{x_{\mathrm{B}}}\right) \tag{2.28}$$

and is plotted in Fig. 2.8. Note the assumption that no recombination occurs in the n-type region is equivalent to letting the lifetime τ_p approach infinity in the diffusion equation. The differential equation that results has a linear solution. A linearly varying concentration indicates that the hole current remains constant throughout the n-type region and that no electron current is needed to compensate for recombining holes.

We can now calculate the hole diffusion current as before

$$J_p(x) = -qD_p\frac{dp_{\mathrm{n}}}{dx} = qD_p\frac{p_{\mathrm{n}0}}{x_{\mathrm{B}}}\left(e^{qV_{\mathrm{a}}/k_{\mathrm{B}}T} - 1\right) \tag{2.29}$$

and also the total current flowing through the thin diode by summing the injected hole and electron diffusion currents to get

$$J = qn_{\mathrm{i}}^2\left(\frac{D_p}{N_{\mathrm{d}}x_{\mathrm{B}}} + \frac{D_n}{N_{\mathrm{a}}x_{\mathrm{E}}}\right)\left(e^{qV_{\mathrm{a}}/k_{\mathrm{B}}T} - 1\right) \tag{2.30}$$

This once again has the form of the ideal diode equation.

We see that the currents through the short-base and long-base diode, Eqs. (2.23) and (2.30), are very similar except for the characteristic length associated with

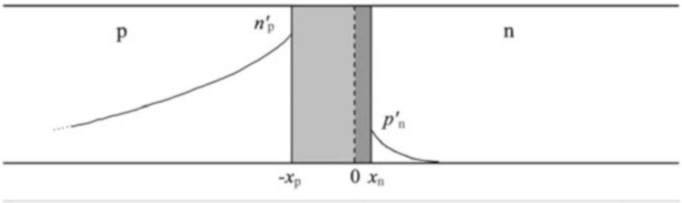


Fig. E2.2 Sketch of excess minority carrier concentrations under forward bias (not to scale). Note the relative magnitudes of the minority carrier concentrations in addition to the difference in diffusion lengths for this problem

each geometry—the minority carrier diffusion length in the long-base diode case and the width of the p- or n-type region in the short-base diode case.

A given diode can be approximated by a combination of these two limiting cases; for example, it may be short in the p-type region and long in the n-type region, or vice versa. In these cases we simply take the proper combinations of the minority current densities derived above to come up with the total junction current.[23]

Example 2.2: Ideal Diode Equation Current Components Calculate the hole and electron contributions to the total current density in a forward-biased (ideal, long-base diode) *pn* junction. The diode is made from silicon with n-side doping $N_\mathrm{d} = 10^{18}\ \mathrm{cm}^{-3}$ and p-side doping $N_\mathrm{a} = 5 \times 10^{16}\ \mathrm{cm}^{-3}$. Assume that $\tau_p = 1\ \mu\mathrm{s}$, $\tau_n = 10\ \mu\mathrm{s}$, and an applied bias of 0.7 V. Sketch the excess carrier concentration on either side of the depletion region.

The long-base ideal diode equation currents are given by Eq. (2.26):

$$J_p(x_n) = \frac{qn_i^2 D_p}{N_d L_p}\left(e^{qV_a/k_BT} - 1\right); J_n(-x_p) = \frac{qn_i^2 D_n}{N_a L_n}\left(e^{qV_a/k_BT} - 1\right)$$

Using the mobility data for silicon as a function of impurity concentration (Appendix B) the diffusion coefficients can be found, which leads to diffusion lengths of approximately 20 μm and 150 μm for holes and electrons, respectively. Substituting these values along with the data given in the problem results in

$$J_p(x_\mathrm{n}) \approx 35\ \ \mathrm{mA/cm^2}; J_n(-x_\mathrm{p}) \approx 515\ \ \mathrm{mA/cm^2}$$

A sketch of the excess carrier concentrations is shown in Fig. E2.2. The ratio of hole or electron current to the total current flowing through the junction is known as the *injection efficiency* and will be seen in Chap. 3 to be a key factor in the performance of bipolar transistors.

[23] If one/both of the diode regions have intermediate lengths (i.e., x_B or x_E is comparable to L_p or L_n, respectively), in other words they are neither short nor long, then the diffusion equations lead to solutions for the excess carrier concentrations containing hyperbolic functions [1]. The net result once more is a modification of the saturation current in the ideal diode equation.

In summary, the diode equations predict a large current flow under forward bias and a small constant saturation current under reverse bias. This asymmetry resulted because forward bias aids the injection of carriers from each region across the junction. Under reverse bias the net flow across the junction is instead composed of minority carriers from each region (i.e., electrons on the p-side and holes on the n-side). These are few in number and hence only a small current flows under reverse bias. This implies that the more lightly doped side of the junction determines most of the reverse saturation current, which can also be seen from the diode equations. For example, if the doping on the *n*-side is much less that that on the *p*-side, the hole flow under reverse bias across the junction into the *p*-side will be much greater than the corresponding electron flow into the *n*-side. Since minority carriers on each side of the space-charge region are swept away by the large electric field present in the junction under reverse bias, their concentration is reduced below thermal equilibrium values. This leads to thermal generation of minority carriers in the vicinity of the junction, which is what supplies the carriers for the reverse saturation current in an ideal diode.

Lastly, we note in this section that the depletion width as a function of applied bias can be found in a straightforward manner by modifying Eq. (2.15) for an abrupt *pn* junction by replacing the built-in potential with (V_{bi}–V_{a}):

$$x_{\mathrm{d}} = x_{\mathrm{n}} + x_{\mathrm{p}} = \left[\frac{2\varepsilon_{\mathrm{s}}}{q}\left(\frac{1}{N_{\mathrm{a}}} + \frac{1}{N_{\mathrm{d}}}\right)(V_{\mathrm{bi}} - V_{\mathrm{a}})\right]^{1/2} \tag{2.31}$$

We have seen that if V_{a} is positive, the built-in barrier at the junction will be reduced. In addition, Eq. (2.31) tells us that the depletion region also becomes *narrower* (see Fig. 2.5). Conceptually, we can think of the applied voltage as moving majority carriers toward the edges of the depletion region where they neutralize some of the space charge, which reduces the overall depletion region width. Similarly, if a negative voltage is applied across the junction, the built-in barrier is increased and majority carriers are pulled away from the edges of the depletion region, which therefore *widens*. When the magnitude of V_{a} becomes considerably larger than V_{bi}, the depletion width varies as the square root of reverse bias.

Similarly, the magnitude of the maximum electric field at the junction, $E_{\max}$, and its relationship to applied voltage can be found by noting that the area under the field curve represents the potential across the junction. We saw earlier that for a step junction the field varies linearly with distance (Fig. 2.3c) and therefore in thermal equilibrium we get

$$\frac{1}{2}E_{\max}x_{\mathrm{d}} = V_{\mathrm{bi}} \Rightarrow E_{\max} = \frac{2V_{\mathrm{bi}}}{x_{\mathrm{d}}} \tag{2.32a}$$

which is equivalent to the maximum field obtained using Eqs. (2.3) and (2.4). Under an applied bias this becomes

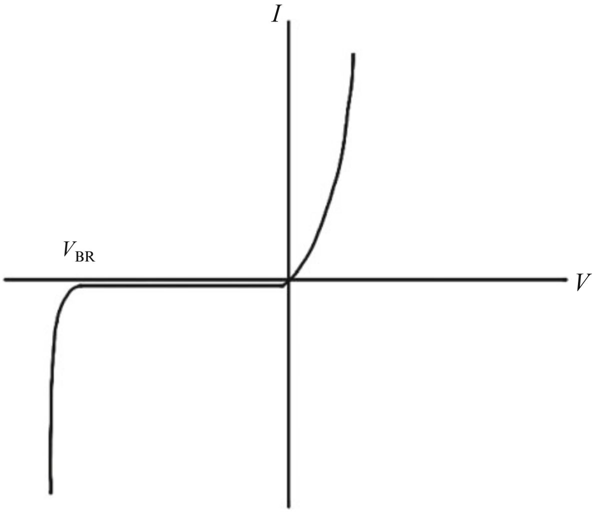


Fig. 2.9 *pn* junction *I–V* characteristic illustrating reverse-bias breakdown

$$E_{\max} = \frac{2(V_{bi} - V_a)}{x_d} \tag{2.32b}$$

and by using Eq. (2.31) the explicit dependence of the maximum electric field on applied bias can be found.[24]

2.1.3 Deviations from Ideal Behavior

1. Reverse-bias breakdown

   An effect not accounted for in the treatment of the ideal diode above is *electrical breakdown* of the junction under large reverse bias, as illustrated schematically in Fig. 2.9. As the reverse bias voltage is increased a critical value is eventually reached where the resulting electric field in the space-charge region causes the magnitude of the current to increase sharply. Generally speaking, such breakdown does not physically damage the junction[25] and is an important process that helps define its stable *I–V* characteristic and potential applications.

   There are two basic mechanisms[26] that cause junction breakdown under high electric fields:

[24] It can be seen that $E_{\max} \propto V_a^{1/2}$ for large reverse biases.

[25] One must take care however to avoid mechanical and/or thermal breakdown (in particular *thermal runaway* processes that cause uncontrollable positive current feedback loops), which can lead to device failure.

[26] The breakdown phenomenon of *punchthrough* may also occur for short-base diodes under reverse bias. This effect is discussed in Sect. 3.4.

- Avalanche breakdown
  Consider an electron traveling in the space-charge region of a reverse-biased *pn* junction: The electron travels an average distance λ, its mean free path, before interacting with an atom in the lattice and losing energy by scattering. The energy gained by the electron between collisions is given by the work done on it by the electric field, which can be written as

$$\Delta \mathrm{E} = q\int_0^{\lambda} \vec{E} \cdot d\vec{x} \tag{2.33}$$

If the electron gains sufficient energy[27] from the field before colliding with an atom in the lattice it can excite an electron out of the silicon–silicon (or other semiconductor) bond so that *three carriers*—the initial electron and an additional electron–hole pair—are created after the collision. This process is illustrated in Fig. 2.10a. Each of the three carriers can then cause similar collisions, etc., which leads to a sudden multiplication or avalanching of the number of charge carriers available to participate in current flow.[28] Hence, current through the junction increases very rapidly.

The multiplication factor for avalanching current compared to the ideal diode saturation current can be written

$$M \equiv \frac{|I|}{I_0} \tag{2.34}$$

and is illustrated in Fig. 2.10b. Empirically, the avalanching multiplication factor for a reverse-biased *pn* junction can be written as

$$M \equiv \frac{1}{1 - (|V_\mathrm{a}|/V_\mathrm{BR})^m} \tag{2.35}$$

where V_BR is the breakdown voltage (taken to be the point at which the current increases rapidly), and m is observed to vary between 2 and 6. Qualitatively, the breakdown voltage will increase with increasing band gap energy (since more energy is required to excite the e–h pairs); decrease with increasing doping level[29] [since the maximum electric field will increase—Eqs (2.3) and (2.4)]; and increase with increasing temperature [due to the increased lattice scattering which limits the mean free path in Eq. (2.33)].

[27] Note that this energy must be at least on the order of the band gap in order to create a new electron–hole pair.

[28] This process is more generally referred to as *impact ionization*.

[29] At very large doping levels the mean free path will also decrease significantly due to impurity scattering; however, in this regime avalanche breakdown is less likely to occur.

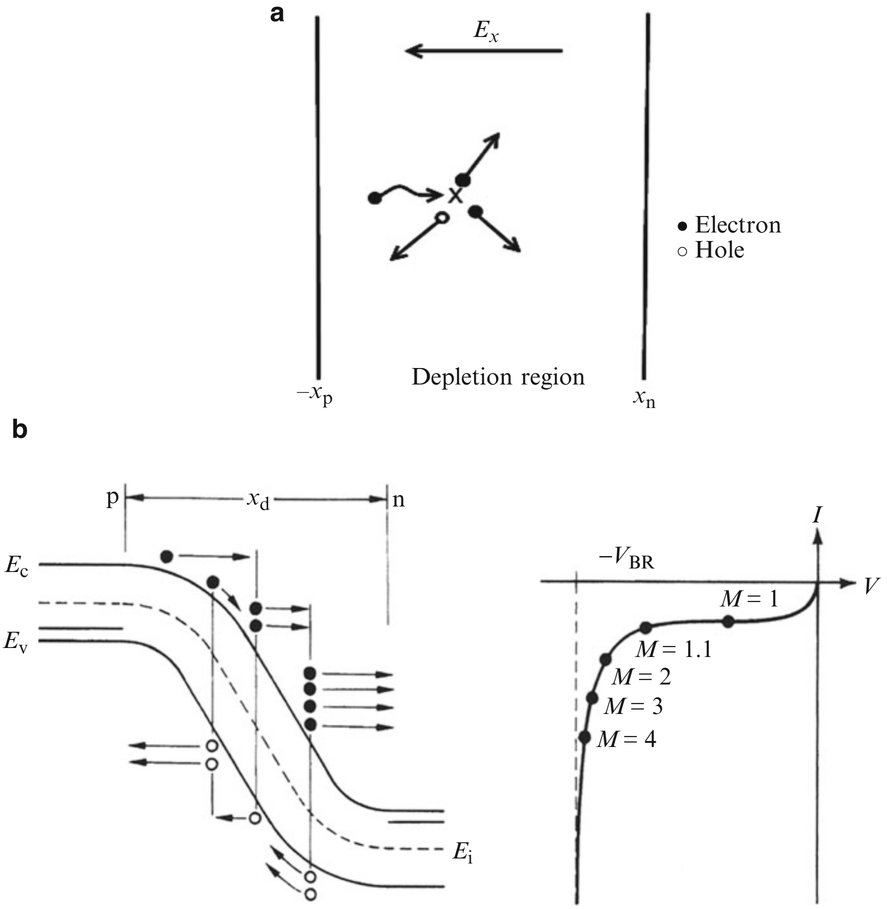


Fig. 2.10 (**a**) Illustration of avalanche breakdown process. The incoming electron (*wavy arrow*) has gained sufficient energy from the field inside the space-charge region such that when it collides with the lattice it is able to free an electron from its bond resulting in the creation of a new electron–hole pair. (**b**) Avalanche breakdown process superimposed on *pn* junction band edge diagram and *I*–*V* curve for reverse bias showing increasing multiplication factor near breakdown voltage (After [7])

- Zener breakdown
  Recall that the width of the depletion region decreases as the dopant concentration increases:

$$x_d = x_n + x_p = \left[\frac{2\varepsilon_s}{q}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)V_{bi}\right]^{1/2} \tag{2.36}$$

Thus the depletion region can be quite narrow for high doping levels resulting in the energy bands across the junction region being bent more steeply. When the width of the depletion layer is small enough, *tunneling*

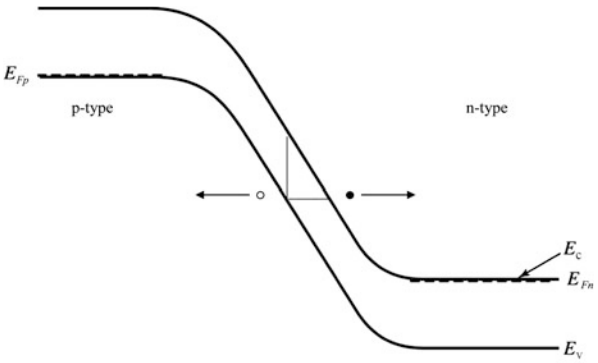


Fig. 2.11 Zener tunneling in a *pn* junction under reverse bias. For heavy doping the narrow depletion width allows an electron from the valence band to tunnel through the band gap leaving behind an empty state (*hole*). The triangular tunneling barrier (*grey lines*) has a height equal to the band gap and width that is inversely proportional to the slope of the band edges (or the electric field strength) (Adapted from [4])

through the band gap can occur (Fig. 2.11). This type of tunneling of electrons from the valence band to the conduction band is called Zener tunneling.[30] Recall[31] that the probability for a particle to tunnel through a barrier is a strong function of the thickness of the barrier and so Zener tunneling is only significant in heavily doped junctions, in which the fields are high and the depletion region is narrow. If we decrease the dopant concentrations, the width of the space-charge region increases and the probability of tunneling decreases rapidly—avalanche breakdown then becomes more likely than Zener breakdown. Zener breakdown also has the opposite temperature dependence compared to avalanching: the tunneling breakdown voltage *decreases* as temperature is increased since the average thermal energy of the electrons will be greater, thus allowing them to tunnel through the barrier more easily, and the band gap energy (or barrier height) will also decrease with temperature.[32]

Devices exhibiting Zener breakdown generally have lower breakdown voltages than those that break down by avalanching. For example, in silicon pure Zener breakdown is usually found in diodes having a breakdown voltage of less than 5 V (Fig. 2.12). As the breakdown voltage increases there will be

[30] Such field-induced *interband tunneling* can occur more generally in solids (e.g., insulators) and is often referred to simply as Zener breakdown; C. M. Zener, Proc. Roy. Soc. (London) **A145**, 523 (1934).

[31] See Appendix A, Example A.4.

[32] See Appendix B for the band gap energy as a function of temperature. Increased thermal energy results in larger crystal lattice spacing, which usually leads to the band structure resembling more of a free particle (i.e., a smaller band gap).

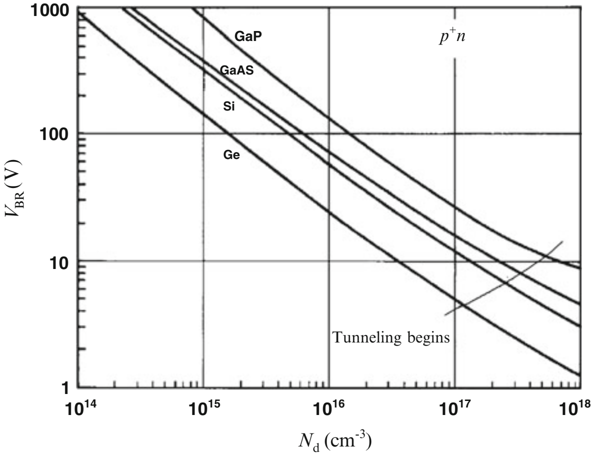


Fig. 2.12 Reverse-bias breakdown data for a one-sided step *pn* junction (Adapted from S. M. Sze, G. Gibbons, Appl. Phys. Lett. **8**, 111 (1966))

a transition region in which both avalanche and tunnel breakdown can occur simultaneously.

Commercially available diodes have well-defined breakdown characteristics and are generally referred to as Zener diodes regardless of their particular breakdown mechanism. Since the voltage across a diode at breakdown is essentially independent of current, Zener diodes are often used as voltage regulators in circuits that require a known value of voltage.

2. Space-charge generation and recombination currents

The analysis used to derive the ideal diode equation ignored processes that occur inside the depletion region, which alter the observed *I–V* characteristic of real *pn* junctions over certain bias ranges compared to the ideal case. Previously we treated the space-charge region simply as a barrier to the diffusion of majority carriers across the junction. However, like the rest of the diode structure, the space-charge region also contains generation–recombination centers.[33] Under forward bias, the injected carriers must pass through space-charge region and a portion of these will recombine before reaching the neutral n- or p-type regions (at x_n or $-x_p$). On the other hand, under reverse bias the opposite process will occur as generation of carriers in the depleted space-charge region also leads to additional current above that predicted by the ideal diode equation analysis. In other words, the ideal diode analysis, in particular its boundary conditions and resulting currents, is still valid; however, the additional current components due to generation/recombination occurring in the space-charge region were not accounted for and must therefore be added to the ideal current we found above.

To find expressions for generation–recombination currents in the space-charge region we may use Shockley–Hall–Read (SHR) (recombination) theory.

[33] Or, more generally, generation/recombination can occur in the space-charge region through a variety of mechanisms.

The essential result of such a treatment is that the space-charge recombination current under an appreciable forward bias can be expressed as

$$J_\mathrm{R} \approx \frac{qx_\mathrm{d}n_\mathrm{i}}{2\tau}\exp\left(\frac{qV_\mathrm{a}}{2k_\mathrm{B}T}\right) \qquad (2.37)^{34}$$

where τ is the carrier lifetime in the space-charge region. If τ is assumed to be approximately equal to the electron and hole lifetimes outside the space-charge region, the ratio between the ideal forward-bias diode current [long-base diode case; Eq. (2.26)] and the space-charge recombination current within this model is

$$\frac{J_\mathrm{ideal}}{J_\mathrm{R}} \approx \frac{2n_\mathrm{i}}{x_\mathrm{d}}\left[\frac{L_n}{N_\mathrm{a}}+\frac{L_p}{N_\mathrm{d}}\right]\exp\left(\frac{qV_\mathrm{a}}{2k_\mathrm{B}T}\right) \qquad (2.38)$$

The space-charge recombination current will therefore become less significant relative to the ideal diode current as forward bias increases. However, the different exponential behavior of J_R can be observed in real *pn* diodes, especially at low currents. For typical silicon diodes, J_ideal exceeds J_R for applied biases greater than about 0.3–0.4 V. In addition, since the ratio is proportional to the intrinsic carrier concentration, *pn* junctions based on larger band gap semiconductors (smaller n_i) will be affected more strongly by recombination processes inside the space-charge region.

Under reverse bias, the net generation of carriers in the space-charge region results in a generation current, J_G, that is proportional to the width of the depletion layer, x_d, and analogous to Eq. (2.38) we can write the ratio as

$$\frac{J_\mathrm{ideal}}{J_\mathrm{G}} \approx \frac{2n_\mathrm{i}}{x_\mathrm{d}}\left[\frac{L_n}{N_\mathrm{a}}+\frac{L_p}{N_\mathrm{d}}\right] \qquad (2.39)$$

Practical values of the parameters for silicon *pn* diodes are such that $J_\mathrm{ideal}/J_\mathrm{G}$ is usually much less than unity and thus the current in reverse-biased silicon *pn* diodes is generated primarily inside the space-charge region as opposed to outside (as in the ideal diode). This behavior once again is more pronounced as the semiconductor band gap energy increases. Since the width of the depletion layer will vary as the square root of applied reverse bias (Eq. 2.31), the generation current in the space-charge region is also a weak function of reverse bias and will gradually increase instead of being the constant reverse saturation

[34] The recombination current density in the space-charge region can be found by evaluating the expression

$$J_\mathrm{R} = q\int_{-x_\mathrm{p}}^{x_\mathrm{n}} U dx,$$

where U is the recombination rate (in units of $\mathrm{s}^{-1}\ \mathrm{cm}^{-3}$). For SHR theory U can be approximated by Eq. (A.49c) in Appendix A, and by using Eq. (2.44) the integral above can be evaluated giving the result of Eq. (2.37). Note that in more detailed recombination models (and in practice) the factor multiplying $k_\mathrm{B}T$ in the exponential can differ from 2.

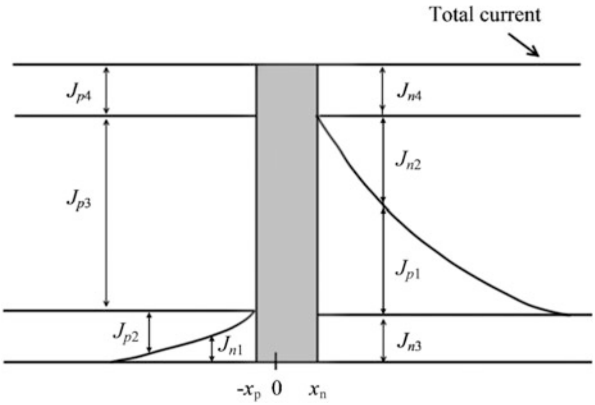


Fig. 2.13 *pn* junction forward-bias current components including space-charge region contributions. Components 1–3 are based on the ideal diode analysis, while the remaining contribution (4) comes from electrons and holes recombining in the space-charge layer shown in gray (Adapted from [4])

current predicted by the ideal diode equation. Lastly, note that ideal diode behavior will become more dominant in both forward and reverse biases as temperature is increased due to the increase in intrinsic carrier concentration.

The observed steady-state dependence of total *pn* junction current on voltage is obtained by summing the ideal diode current and the space-charge generation–recombination current components:

$$I = I_0\left[\exp\left(\frac{qV_a}{k_B T}\right) - 1\right] + I_{GR0}\left[\exp\left(\frac{qV_a}{2k_B T}\right) - 1\right] \qquad (2.40)^{35}$$

We now have a more complete picture of the currents flowing across a *pn* junction as shown in Fig. 2.13.

3. Effect of series resistance

An initial assumption we made in deriving Eqs. (2.26) and (2.30) was that the applied bias, V_a, is dropped entirely across the junction space-charge region. By examining some typical diode structures (see, e.g., Problem 2) it is possible to get a feel for the accuracy of this assumption. In most cases, it is reasonable to conclude that we can safely assume that the entire applied voltage changes the height of the potential barrier at the *pn* junction for *low-to-moderate* current densities. However, if the applied forward bias is nearly as large as the built-in potential, the barrier preventing carriers from diffusing across the junction is substantially reduced and large currents can flow. Essentially, the high-resistance depletion layer is almost eliminated in this case (the band edges are flattened) and a significant fraction of the applied voltage is dropped across the neutral regions in series with the junction. (This is the reason that the actual forward voltage applied across the *pn* junction is never as large as the built-in voltage.)

[35] I_{GR0} is found from Eq. (2.37).

The effect of series resistance can be included by modifying the ideal diode equation as follows:

$$I = I_0\left[\exp\left(\frac{q(V_\mathrm{a} - IR_\mathrm{s})}{k_\mathrm{B}T}\right) - 1\right] \tag{2.41}$$

where R_s is the effective (parasitic) series resistance.[36]

In general, the diode *I–V* relationship may be written in an empirical parametric form as

$$I = I'_0\left(e^{qV_\mathrm{a}/\eta k_\mathrm{B}T} - 1\right) \qquad (2.42)^{37}$$

where I'_0 is now the effective saturation current and η is called the *ideality factor*. We have seen that the ideality factor will approach 2, when recombination processes in the space-charge region are important, and generally will have a value between 1 (ideal diode) and 2. For applied voltages greater than a few $k_\mathrm{B}T/q$ the -1 term can be neglected and taking the natural log of both sides yields

$$\ln I = \ln I'_0 + \frac{qV_\mathrm{a}}{\eta k_\mathrm{B}T} \tag{2.43}$$

Using a semi-log scale we can now describe the overall diode forward *I–V* characteristic including non-idealities as shown in Fig. 2.14.

An additional non-ideality at large forward bias indicated in Fig. 2.14 is the effect of *high-level injection*. As mentioned earlier, our assumption of low-level injection used to derive the ideal diode equation will begin to breakdown once the forward bias reaches a level that causes the excess minority carrier concentrations to be comparable to the majority carrier or doping levels of the *pn* junction neutral regions. In this case, excess *majority* carriers that act to maintain space-charge neutrality become significant and we can no longer assume that the majority carrier concentration is equal to the dopant density. In order to get a feel for the underlying behavior in the condition of high-level injection, let us examine the hole–electron product at the edges of the depletion region using Eqs. (2.17):

$$p_\mathrm{n}(x_\mathrm{n})n_\mathrm{n}(x_\mathrm{n}) = p_\mathrm{p}(-x_\mathrm{p})n_\mathrm{p}(-x_\mathrm{p}) = n_\mathrm{i}^2\exp\left(\frac{qV_\mathrm{a}}{k_\mathrm{B}T}\right) \tag{2.44}$$

This equation is similar in form to the mass-action law for semiconductors in thermal equilibrium, except that the constant product has a greater value. If we

[36] Contact resistances may also be lumped into this parameter.

[37] The effects of series resistance may also be explicitly included here as in Eq. (2.41).

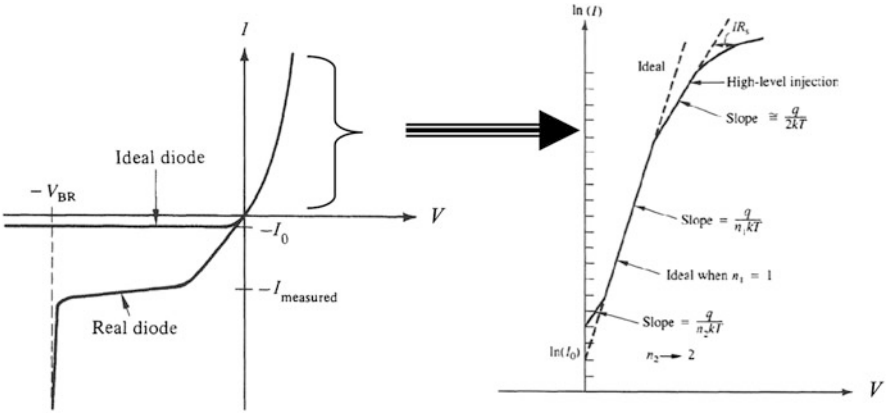


Fig. 2.14 *pn* junction *I–V* characteristic including non-idealities (After [7])

now make the conjecture that Eq. (2.44) holds generally,[38] then this *pn* product will be obeyed regardless of low- or high-level injection conditions. In the case of high-level injection this implies that if the concentration of minority carriers approaches the majority carrier concentration, i.e., *p* and *n* become similar in magnitude, then

$$p_n(x_n) \approx n_i \exp\left(\frac{qV_a}{2k_BT}\right) \tag{2.45}$$

with a similar result for electrons in the p-side. This is the basis for the ideality factor approaching 2 at large biases in Fig. 2.14. Note that the effects of series resistance and high-level injection may occur simultaneously in a given *pn* diode device and care must be taken when analyzing the large forward bias portion of the *I–V* characteristic.

The next step in analyzing *pn* junctions would be to consider non-uniform dopant densities instead of the abrupt or step junctions we have been considering thus far. For example, in a *linearly graded junction*[39] the dopants have a linear concentration profile from the p- to n-type material instead of being constant. Using the depletion approximation and electrostatics allows one to find a linear charge dependence in the resulting space-charge region and thus a field and potential that vary quadratically and to the third power, respectively.

[38] In other words, we are assuming that *quasi-equilibrium* holds. It can be shown more generally that Eq. (2.44), often referred to as the “law of the junction”, is also valid inside the space-charge region and that $pn \leq n_i^2 e^{qV_a/k_BT}$ throughout the *pn* junction (H. K. Gummel, Solid-State Electron. **10**, 209 (1967)).

[39] This can sometimes be used to approximately model *pn* junctions formed via gas phase diffusion processes. (An exponential distribution may also be used.)

Most junctions encountered in practice can be usually be approximated by either a step junction or other analytical profiles (sometimes dependent on the bias applied), which allows important physical insight into their properties. Arbitrary doping profiles are usually handled numerically if a very precise result is required. Similar comments apply to the currents flowing through such junctions. In Chap. 3 it will be seen that approximations can be made that, not surprisingly, show that the ideal diode law is still obtained in practical cases of nonuniform doping although the expression for reverse saturation current will in general be different.

2.1.4 Small-Signal Parameters[40]

Small amplitude time-varying signals are often applied to electronic devices that are operating at a given dc bias point.[41] The response of a device to these small perturbations is characterized by a set of small-signal parameters. This type of approach is particularly useful for applications in amplification, communications, and signal processing where the overall nonlinear device behavior can be approximated by linear components (the small-signal parameters) that are valid for small excursions from the dc biasing point.

2.1.4.1 Conductance

The small-signal or differential conductance of the *pn* junction is given by

$$G = \frac{dI}{dV_{\mathrm{a}}} = \frac{q}{k_{\mathrm{B}}T} I_0 \left(e^{qV_{\mathrm{a}}/k_{\mathrm{B}}T} \right) = \frac{q}{k_{\mathrm{B}}T}(I + I_0) \tag{2.46}$$[42]

The small-signal conductance of a *pn* junction is thus seen to be very sensitive and directly proportional to the current under appreciable forward bias, while it becomes zero under appreciable reverse bias for the idealized junction.

2.1.4.2 Diffusion Capacitance

The variation of minority carrier charges stored in the neutral regions of the *pn* junction under forward bias contributes a small-signal capacitance known as

[40] The results of this section are based on an ideal step *pn* junction diode unless otherwise noted. Non-idealities and/or other junction doping profiles can be included in a straightforward manner.

[41] This includes a device that is unbiased (i.e., zero bias).

[42] Not to be confused with the same symbol also used for carrier generation rate.

the diffusion capacitance, C_d, which characterizes how quickly the minority carrier charge distribution can change in response to a time-varying voltage. The excess minority carrier charge per unit area, Q_n or Q_p, can be used to find the small-signal diffusion capacitance (per unit area) using $C_\mathrm{d} = |dQ_{n,p}/dV_\mathrm{a}|$. For the short-base diode the minority carrier charge can be found using Eq. (2.28)[43] giving

$$C_\mathrm{d} = \frac{q}{k_\mathrm{B}T}\left(\frac{qx_\mathrm{B}p_\mathrm{n0}}{2} + \frac{qx_\mathrm{E}n_\mathrm{p0}}{2}\right)e^{qV_\mathrm{a}/k_\mathrm{B}T} \tag{2.47}$$

where the contribution from both minority electrons and holes on either side of the junction has been included. In the case of a p^+n diode the diffusion capacitance can be expressed as

$$C_\mathrm{d} \approx G\tau_\mathrm{t} \tag{2.48}$$

where a characteristic time, τ_t, multiplies the conductance. To see the significance of this term note that

$$\tau_\mathrm{t} = \frac{x_\mathrm{B}^2}{2D_p} = \frac{Q_p}{J_p} \tag{2.49}$$

The amount of stored charge divided by the rate at which the charge enters or leaves the n-type region is equal to the average time a carrier spends in this region and we call this the average *transit time* of a hole moving through the n-region of the short diode. Similar comments apply to electrons moving through the p-region of the diode.

The long-base diode diffusion capacitance has to take into account the recombination of the minority charges, which normally requires solving the time-dependent continuity or diffusion equations. However, we noted earlier that the long- and short-base diode equations are identical if the diffusion lengths are interchanged with the widths of the neutral regions. Therefore Eq. (2.47) can be modified to give the long-base diode diffusion capacitance as follows:

$$C_\mathrm{d} = \frac{q}{k_\mathrm{B}T}\left(\frac{qL_p p_\mathrm{n0}}{2} + \frac{qL_n n_\mathrm{p0}}{2}\right)e^{qV_\mathrm{a}/k_\mathrm{B}T} \tag{2.50}$$

[43] This is simply the area of the triangle representing the excess minority carrier distribution *vs.* distance.

In the case of a p^+n diode this simplifies to

$$C_\mathrm{d} \approx \frac{G\tau_p}{2} \tag{2.51}$$[44]

and vice versa for a pn^+ diode.

As expected, the diffusion capacitance is negligible under reverse bias because the minority carrier storage is small.

2.1.4.3 Junction Capacitance

In addition to the diffusion capacitance due to excess minority carriers, the fixed charges inside the depletion layer will contribute what is known as the depletion or junction capacitance, C_j, which characterizes the expansion or contraction of majority carrier distributions near the edges of the depletion region in response to a time-varying bias. The space charge per unit area on either side of the (metallurgical) junction is given by

$$Q_s = qN_\mathrm{d}x_\mathrm{n} = qN_\mathrm{a}x_\mathrm{p} \tag{2.52}$$

which can be used to find the small-signal capacitance C_j (per unit area) of the junction:

$$C_\mathrm{j} = \left|\frac{dQ_s}{dV_\mathrm{a}}\right| = qN_\mathrm{d}\frac{dx_\mathrm{n}}{dV_\mathrm{a}} = qN_\mathrm{a}\frac{dx_\mathrm{p}}{dV_\mathrm{a}} \tag{2.53}$$

Since $x_\mathrm{p} = (N_\mathrm{d}/N_\mathrm{a})x_\mathrm{n}$ and $x_\mathrm{d} = x_\mathrm{n} + x_\mathrm{p}$ we can write

$$x_\mathrm{d} = x_\mathrm{n}(1 + N_\mathrm{d}/N_\mathrm{a}) = \left[\frac{2\varepsilon_\mathrm{s}}{q}\left(\frac{1}{N_\mathrm{a}} + \frac{1}{N_\mathrm{d}}\right)(V_\mathrm{bi} - V_\mathrm{a})\right]^{1/2} \tag{2.54}$$

This gives us an expression for x_n in terms of the applied voltage, which can be used to calculate C_j. Taking the derivative with respect to applied voltage and rearranging terms gives

$$\frac{dx_\mathrm{n}}{dV_\mathrm{a}} = \frac{1}{N_\mathrm{d}}\left[\frac{\varepsilon_\mathrm{s}}{2q(1/N_\mathrm{a} + 1/N_\mathrm{d})(V_\mathrm{bi} - V_\mathrm{a})}\right]^{1/2} \tag{2.55}$$

and the junction capacitance is therefore

[44] Equations (2.48) and (2.51) show that the time delays associated with the minority carriers can be thought of in terms of an equivalent *RC* time constant. This implies that a short diode typically responds more quickly than a long diode (since x_B, $x_\mathrm{E} \ll L_p$, L_n).

$$C_\mathrm{j} = \left[\frac{q\varepsilon_\mathrm{s}}{2(1/N_\mathrm{a} + 1/N_\mathrm{d})(V_\mathrm{bi} - V_\mathrm{a})}\right]^{1/2} = \frac{\varepsilon_\mathrm{s}}{x_\mathrm{d}} \qquad (2.56)^{45}$$

Thus for $|V_\mathrm{a}|$ much greater than V_bi, the capacitance of a *pn* step junction varies approximately inversely with the square root of the reverse bias. Devices which employ this voltage variable capacitance are called *varactor* diodes (variable reactor). Varactors find application in filters, oscillators, tuning circuits, etc. Note that the dependence of capacitance on reverse bias will be determined by the doping profile near the junction (since this will determine the distribution of charge in the space-charge region) and can be tailored for specific applications. Generally, the *pn* junction capacitance can be written

$$C_\mathrm{j} = \frac{C_\mathrm{j0}}{\left[1 - \frac{V_\mathrm{a}}{V_\mathrm{bi}}\right]^m} \qquad (2.57)$$

where C_j0 is the zero-bias junction capacitance and m depends on the doping profile of the junction. We have seen that $m = 1/2$ for a step junction. For a linearly graded junction $m = 1/3$, etc. (see [2] for further details).

2.1.4.4 Total Junction Charge Storage

In general, we can see that the relative importance of C_j and C_d depends strongly on the applied junction voltage: Under reverse bias the junction capacitance dominates, but under forward bias the exponential factor in C_d makes diffusion capacitance important as well. For low-to-moderate forward bias it is necessary to consider both types of charge storage in order to obtain an accurate value for the total *pn* junction capacitance as illustrated in Fig. 2.15. For applications where there is a low-voltage sinusoidal excitation of the diode that is biased at a dc operating point, the above analysis thus results in the *pn* diode small-signal equivalent circuit shown in Fig. 2.16.

It is important to discuss some of the limitations of the small-signal parameters presented above: First of all, it is clear that the exponential dependence of the diffusion capacitance cannot be expected to grow indefinitely as the applied bias approaches the built-in potential, just as the ideal diode current (and hence conductance) did not (see Fig. 2.14). In addition, at large forward bias the depletion layer is effectively eliminated and thus the basis for junction capacitance derived above is

[45] This equation has the same form as the capacitance of a parallel plate capacitor with the plates separated by the depletion width. It can be shown that this correspondence holds for arbitrary dopant profiles.

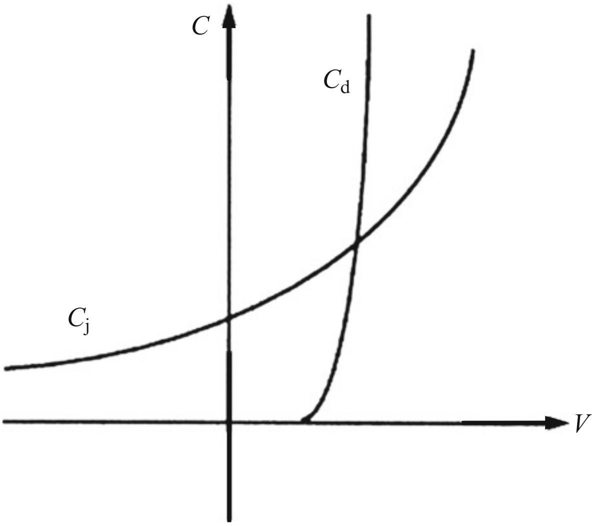


Fig. 2.15 *pn* junction capacitance vs. voltage showing contributions of space-charge (junction capacitance) and excess minority carriers (diffusion capacitance)

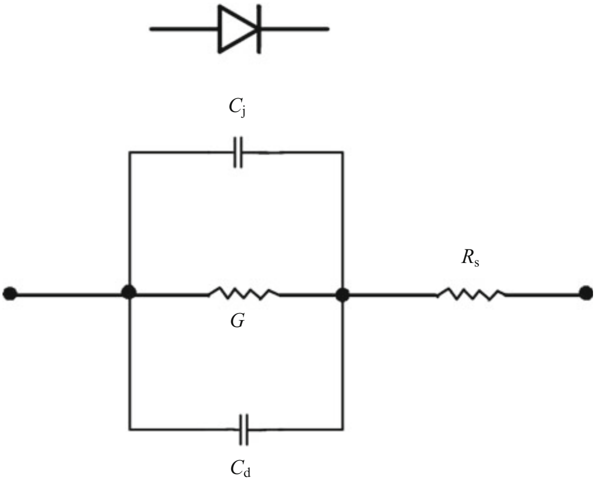


Fig. 2.16 *pn* junction small-signal equivalent circuit. A parasitic series resistance R_s has also been included

no longer valid. For large currents, much of the applied bias will be dropped in the neutral regions of the *pn* junction and this ultimately limits the total junction capacitance at large biases as shown in Fig. 2.17.

A further limitation arises in relation to the frequency of the small-signal excitation applied to the junction. As frequency is increased, a parasitic inductance must also be added to the small-signal equivalent circuit of the diode. The inductive element becomes particularly important at larger forward biases when the capacitive and resistive components of the junction impedance decrease. In addition, the equivalent circuit of Fig. 2.16 should be considered valid only for excitation

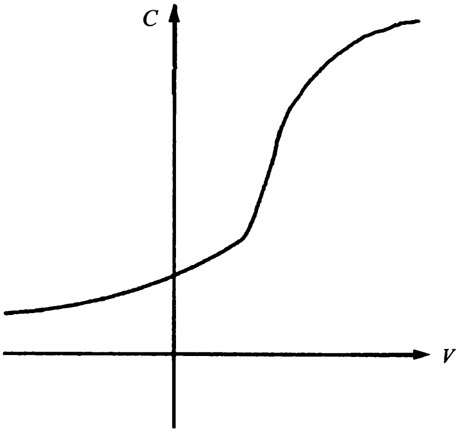


Fig. 2.17 Total *pn* junction capacitance vs. applied bias showing high voltage roll-off

frequencies whose period of oscillation is large compared to the minority carrier lifetime or transit time under forward bias for long- or short-base diodes, respectively. Beyond such frequencies[46] lumped element equivalent circuit models may no longer be applicable and one should employ more detailed distributed calculations that model the entire device by breaking it up into small segments to define a computational grid.[47]

2.1.5 Transient Behavior and (Large Signal) Diode Switching

Switching of a *pn* junction from forward bias (a large current or "on" state) to reverse bias (a very small current, or "off" state) and vice versa is a very common process in applications involving diodes. We can gain insight into the on/off switching behavior of a *pn* junction by considering the buildup and decay of Q_p (the hole charge in the n-region).

Consider the distribution of holes in the n-region of an initially unbiased long-base *pn* diode to which a positive constant-current source is suddenly applied: Current can start to flow across the junction very quickly, but before the steady-state

[46] It is possible to define a *diode cutoff frequency* as $f_T = (2\pi\, RC)^{-1}$, where R and C denote the overall (including parasitics, except for inductance) resistance and capacitance components that are dominant at a particular bias, respectively. f_T can be considered an upper limit to how quickly the diode can respond to small-signal excitations.

[47] Ultimately, this can lead to so-called first-principles calculations that consider the individual atoms making up the device. Such calculations typically require very large computational times, but fortunately this kind of precision is not usually required for most devices at present.

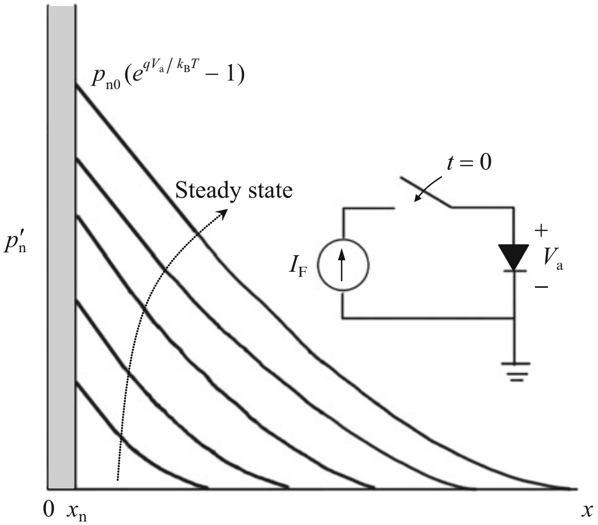


Fig. 2.18 *pn* junction turn-on transient. The slope at the edge of the space-charge region remains constant as the charge increases towards its steady-state distribution (which depends on the current flow and voltage across the junction according to the ideal diode analysis). Current begins to flow very quickly after the current source is connected since only a small number of carriers need to be injected. The long-base diode case is shown (the short-base diode is similar but with a linear triangular distribution instead of the decaying exponential)

charge distribution can be reached holes must be transported into the neutral n-region. The stored charge Q_p increases with time as holes are supplied and the voltage across the junction increases toward its steady-state value. A sketch of the transient increase of stored holes in this situation would have the form shown in Fig. 2.18.

The time required to reach steady state is given by the ratio of the steady-state stored hole charge to the size of the current source (neglecting electrons in the p-side[48]):

$$Q_p(\infty)/I_F = \tau_t\,(\text{short-base}) \Rightarrow \tau_p/2\,(\text{long-base}) \qquad (2.58)^{49}$$

On the other hand, the turnoff time of the diode is limited by the speed at which stored carriers can be removed from the neutral regions: When a reverse bias is

[48] In other words, we are considering a p^+n diode. Non-idealities (recombination in the space-charge region, etc.) are also ignored in this treatment.

[49] This result can be obtained by noting that both the stored charge and the current flowing through the junction can be expressed in terms of the bias appearing across the junction (via exponential factors) using the results of the ideal diode analysis carried out earlier.

suddenly applied across the forward-biased junction the current can reverse direction quickly because the gradient near the edge of the space-charge region can change with only a small change in the number of stored carriers as illustrated in Fig. 2.19a.

The dc forward current just before switching ($t = 0–$) is given by

$$I_{\mathrm{F}} = \frac{V_F - V_{\mathrm{a}}}{R_F} \approx \frac{V_F}{R_F} \tag{2.59}$$

where the last expression assumes the voltage source is much greater than the voltage appearing across the diode. Similarly, the magnitude of the reverse current immediately after switching ($t = 0+$) is given by

$$I_{\mathrm{R}} = \frac{V_R + V_{\mathrm{a}}(t)}{R_R} \approx \frac{V_R}{R_R} \tag{2.60}$$

once again assuming a large source voltage in the latter expression.

Thus the diode is able to conduct a large amount of current in the reverse direction and the junction will remain forward biased until the injected minority carrier charge near the edge of the depletion region is removed. A plot of current versus time is initially nearly constant until the excess minority carrier concentration at the edge of the depletion layer gets close to zero (Fig. 2.19b), which is termed the diode *storage time*, t_{s}.

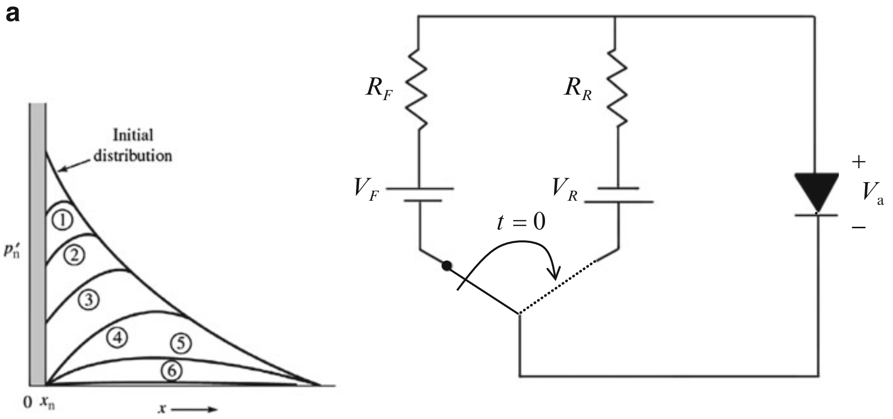


Fig. 2.19 (**a**) *pn* junction behavior during switch off. The initial forward-bias steady minority hole carrier distribution (sketched here for a long-base diode) decreases vs. time; however, the junction can still conduct current in the reverse direction immediately after switching since the slope at the edge of the depletion region can change direction. After a certain storage time, t_{s}, the current begins to decay strongly. (After [4]) (**b**) Current vs. time during switching off of a *pn* junction, illustrating storage time, t_{s}, and recovery time t_{r}, as the voltage across the junction changes from positive to negative. The corresponding decay of the excess minority carrier distributions for a long-base diode is also shown (After [7])

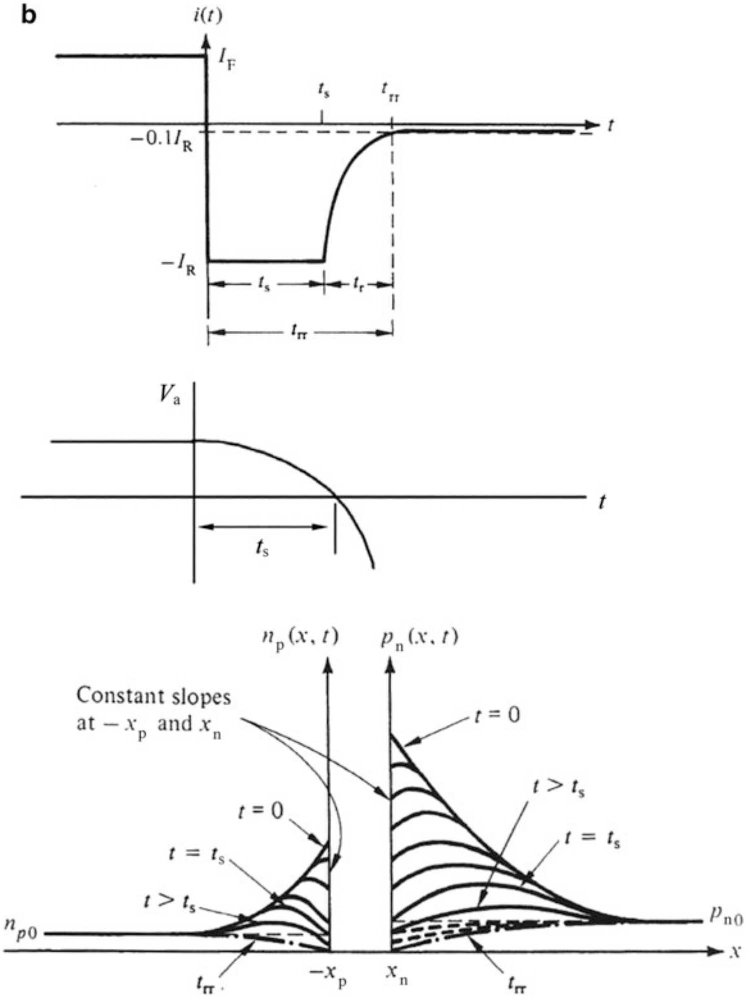


Fig. 2.19 (continued)

The current then decays on a timescale (denoted by the *recovery time*,[50] t_r) mainly determined by carrier lifetime in the neutral region.

A rough estimate of the storage time for a long-base diode can be obtained by noting

$$t_s \approx \frac{\delta \cdot Q_p}{I_R} = \frac{\tau_p}{2} \frac{I_F}{I_R} \cdot \delta \tag{2.61}$$

[50] The recovery time is usually taken as the point at which the reverse current has dropped to 10 % of its initial value.

where Eq. (2.58) has been employed and here δ represents the fraction of the initial charge Q_p that is removed before the reverse current begins to decay.[51] Accurate expressions for the storage and recovery times can be determined by solving the time-dependent continuity equation [3]. For a long-base p^+n junction the storage time turns out to be determined by

$$\operatorname{erf}\sqrt{\frac{t_\mathrm{s}}{\tau_p}} = \frac{I_\mathrm{F}}{I_\mathrm{F} + I_\mathrm{R}} \tag{2.62a}$$

where erf (x) is the error function. An approximate analytical solution can also be found as

$$t_s \approx \tau_p \ln\left[1 + \frac{I_F}{I_R}\right] \tag{2.62b}$$ [52]

The continuity equation solution for the recovery time t_r, results in

$$\operatorname{erf}\sqrt{\frac{t_\mathrm{r}}{\tau_p}} + \frac{\exp(-t_\mathrm{r}/\tau_p)}{\sqrt{\pi t_\mathrm{r}/\tau_p}} = 1 + 0.1\left(\frac{I_\mathrm{R}}{I_\mathrm{F}}\right) \tag{2.62c}$$

In summary, even without any complex mathematical models we can see that in order to switch a *pn* diode quickly we need to be able to produce a large reverse current as well as have a small minority carrier lifetime in the neutral region. In general, a fast *pn* diode should minimize charge storage under forward bias. Another way to accomplish this is by employing a short-base diode whose switching response will instead be *transit time limited*.[53]

Note that we neglected the majority carriers in the above discussion because they respond to changes in electric fields much more quickly than minority carriers since they do not need to recombine. Majority carriers respond within the so-called *dielectric relaxation time*, which is usually on the order of less than a picosecond in silicon. This is also the reason that the small-signal junction capacitance derived earlier is less sensitive to frequency than the forward-bias diffusion capacitance.

[51] Although an analytical expression for this parameter based on the diode material properties is possible, empirically a value of $\delta \sim 0.2$ agrees quite well with data over a fairly broad range of currents.

[52] This equation overestimates the storage time by a progressively larger amount as $I_\mathrm{R}/I_\mathrm{F}$ increases.

[53] Equations (2.59), (2.60), (2.61), (2.62a), (2.62b), and (2.62c) also apply to a short-base diode if τ_p is replaced by $2\tau_\mathrm{t}$.

2.1.5.1 Equivalent *pn* Diode Circuits for Transient Problems

The nature of charge storage at the *pn* junction complicates the use of equivalent circuits for hand calculations involving transient problems. For example, switching diodes that are sequentially forward and reverse biased during circuit operation will have the dominant contribution to charge storage shift between diffusion and depletion layer charges during the transient itself, as we saw above.

If an accurate picture of the current and voltage as functions of time is required the diode can be *piecewise-linearly* approximated, i.e., the nonlinear charge storage and conductance effects can be approximated to first order by linear elements over a small voltage range. If the voltage increment is made small enough, this approximation is accurate and arbitrary precision can be obtained by joining together a sufficient number of piecewise-linear approximations to represent the entire voltage variation in a given transient problem, typically using a computer. The small-signal parameters we found earlier are the piecewise-linear approximations we need in order to represent charge storage and conductance in the *pn* junction for this type of calculation.

Panel 2.2: *pn* Junction Fabrication and Integrated Circuits To form a *pn* junction diode using planar silicon technology (see Appendix A, (sect. A.3)) it is only necessary to diffuse a p-type region into an n-type wafer and make electrical contact to the front and back of the wafer as shown in Fig. 2.20a. However, in a modern integrated circuit (IC) there will in general be many diodes and other devices on the same wafer and we must be able to electrically isolate these devices from each other for proper circuit operation. This is accomplished by using a combination of insulating oxide regions and *pn* junctions that are kept under reverse bias at all times.[54] For example, to form an isolated array of diodes the sequence of steps shown in Fig. 2.20b can be performed.

One problem with IC diodes and other devices formed in a similar manner is so-called *current crowding*, which reduces the effective area of the diode and limits its current carrying capacity (Fig. 2.21a). To get around this difficulty and allow current to flow vertically through the diode junction in a uniform manner, a heavily doped, low-resistance *buried layer* is often diffused onto the substrate. The heavily doped n-type (or p-type) region is formed before growing the epitaxial[55] layer and can be included beneath each junction region as illustrated in Fig. 2.21b whenever vertical currents (i.e., perpendicular to the substrate) are important to integrated

[54] These critical IC interconnection techniques were first developed by Noyce and Lehovec working independently in 1959.

[55] Epitaxy is a type of thin film crystal growth. See Appendix A, Sect. A.3, for further details.

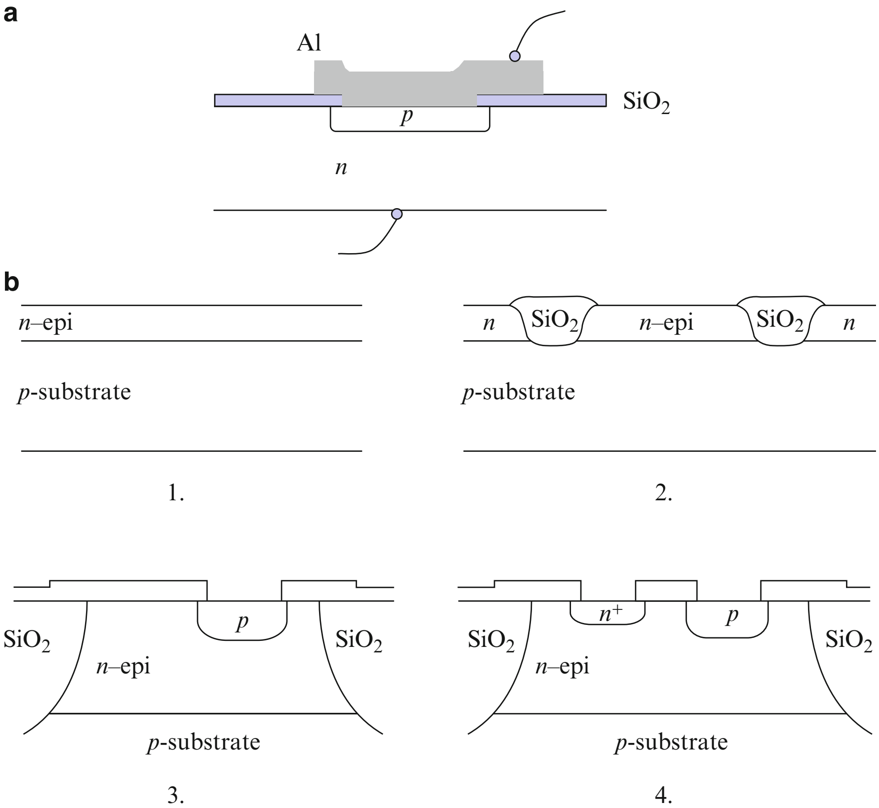


Fig. 2.20 (**a**) Typical planar *pn* junction discrete device structure. (**b**) IC isolated diode array fabrication sequence using an n-type epitaxial layer grown on a p-type substrate. The n^+ layer in image 4 is for making a low-resistance contact to the epitaxial layer as described in Sect. 2.2 and 2.3. Note that the oxide formed above the epitaxial layer is thicker in order to prevent the unintentional formation of field-induced conducting surface channels (see Chap. 4) and is referred to as a *field oxide* (Adapted from [4])

circuit device operation. A modern IC *pn* junction (Fig. 2.21c) employs these and many other advances to create the very precisely controlled structures required for state-of-the-art electronics.

2.2 Metal–Semiconductor Junctions

Most electronic devices are interconnected using metallic wiring that forms metal–semiconductor contacts. For example in an integrated circuit there are typically many millions of metal contacts to silicon. The properties of these contacts can

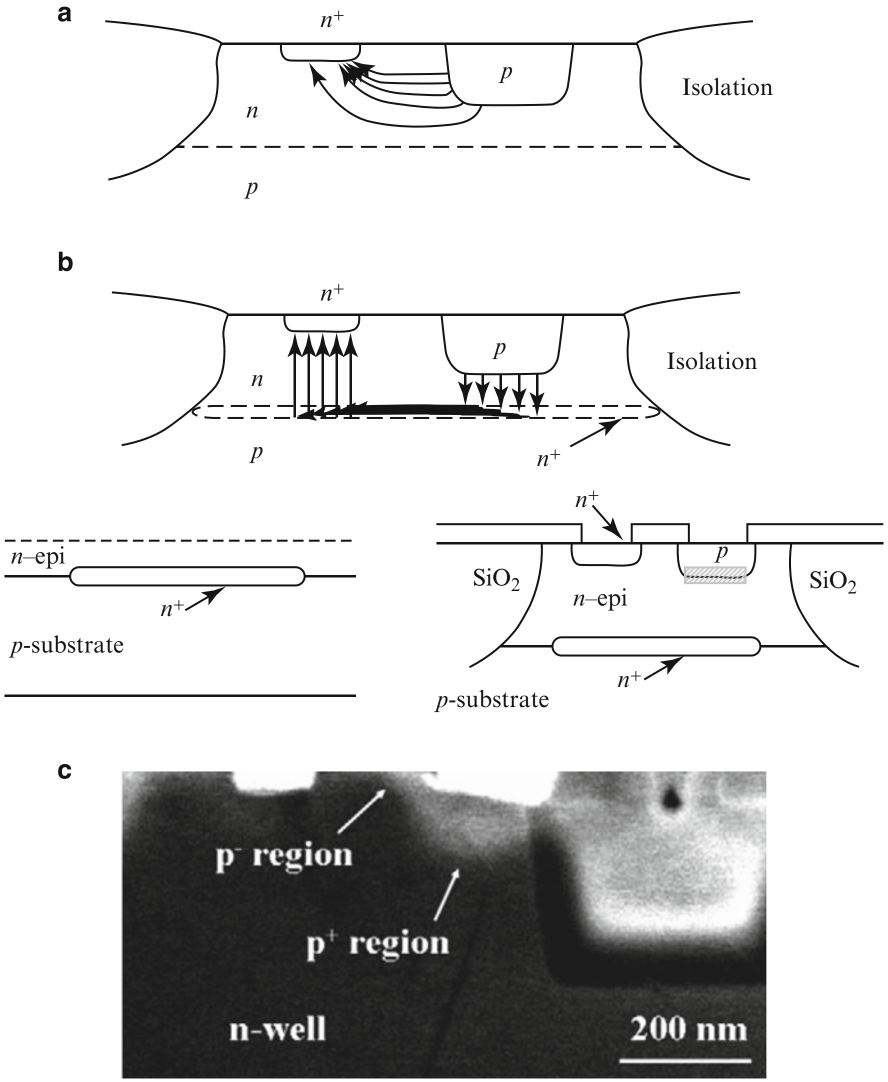


Fig. 2.21 (**a**) IC *pn* junction current crowding phenomenon. (**b**) IC *pn* junction heavily doped buried layer fabrication process to prevent current crowding. The active *pn* junction device interface is highlighted by hatched region (Adapted from [4]). (**c**) Cross section of a typical *pn* junction structure used in integrated circuits (After J.-H. Lee, P.-T. Liu, Microelectronic Engineering **95**, 5 (2012))

vary considerably and we need to understand the nature of thermal equilibrium that is established when a metal–semiconductor junction is formed. The concepts and analysis of metal–semiconductor junctions are similar to what we used to study *pn* junctions and often only need minor modification.

The first metal–semiconductor junctions studied were in the form of point-contact structures: In 1874 Braun was able to realize both rectifying and low-resistance contacts and around the same time Schuster observed rectification in a contact made using similar techniques. The basic theory of metal–semiconductor junctions was developed around 1938 by Schottky along with related work by Mott.[56]

2.2.1 *Metal–Semiconductor Barriers (Blocking Contacts)*

2.2.1.1 Thermal Equilibrium

We once again use the fact that the Fermi level must be constant in thermal equilibrium to construct the band edge diagram. In the case of a copper contact to n-type silicon, before the two materials are brought together we have the situation depicted in Fig. 2.22a. The relative values of the work functions (and hence Fermi levels) of the isolated materials determine the band edge diagram of the metal–semiconductor junction in thermal equilibrium just as they did for the *pn* junction. For the case shown in the figure electrons from the semiconductor (higher Fermi level) will be transferred into the metal (lower Fermi level) in order to achieve equilibrium. Figure 2.22b shows the resulting thermal equilibrium band edge diagram from which we see that the height of the barrier step[57] at the interface is given by

$$q\phi_{\mathrm{B}} = q(\Phi_{\mathrm{M}} - \mathrm{X}) \tag{2.63a}$$[58]

where the built-in potential of the semiconductor is found from

$$\phi_{\mathrm{bi}} = \Phi_{\mathrm{M}} - \Phi_{\mathrm{S}} \tag{2.63b}$$

The transfer of electrons from the semiconductor into the metal results in a built-in electric field at the junction due to the uncompensated donor ions near the interface, which creates a space-charge region as shown in Fig. 2.23a (cf. Fig. 2.3b for a *pn* junction). We are once again ignoring any contribution to the charge density from free electrons and holes in the semiconductor, i.e., the depletion approximation is assumed to hold. In the metal there exists a thin layer of negative interfacial or surface charge[59] that is equal in magnitude to the space charge in the semiconductor.

[56] Many of the important features of metal–semiconductor junctions (and *pn* junctions) were also developed theoretically by Davydov between 1938 and 1939.

[57] Known as the Schottky barrier.

[58] The analogous barrier height for a p-type semiconductor is given by $q\phi_{\mathrm{B}} = E_{\mathrm{g}} - q(\Phi_{\mathrm{M}} - X)$.

[59] Recall from electrostatics that free extra charge cannot exist in the interior of a (metallic) conductor.

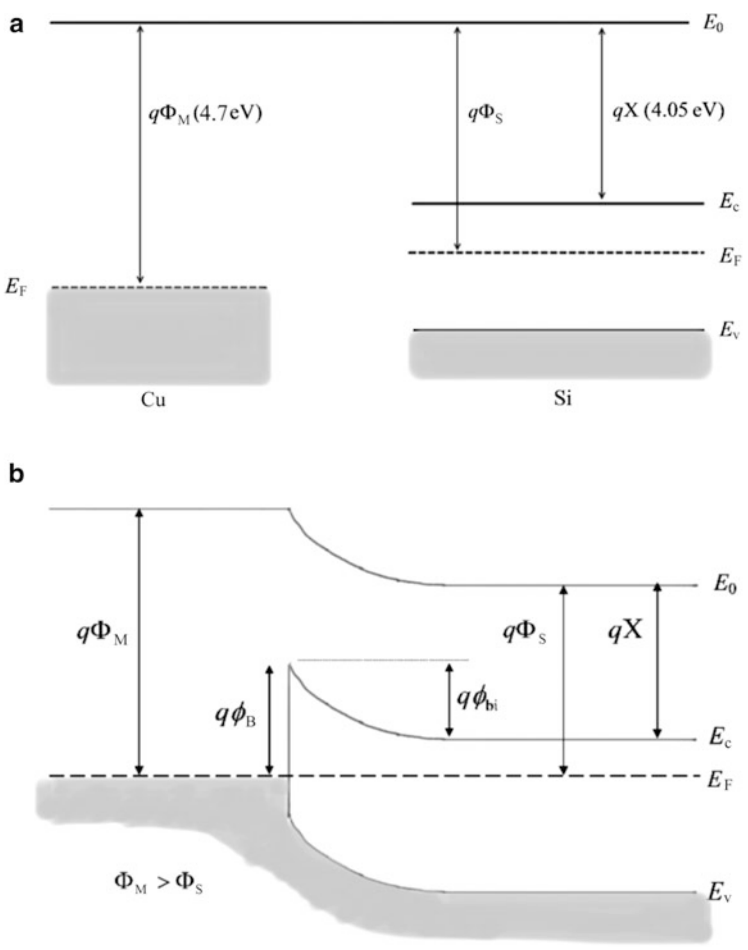


Fig. 2.22 (**a**) Copper and n-type silicon band edge diagrams in isolation. (**b**) Copper–silicon junction thermal equilibrium band edge diagram. Note the constancy of the electron affinity in the semiconductor regardless of position. In this example electrons were transferred from the semiconductor and therefore the bands bend upward at the interface. For the corresponding junction to a p-type semiconductor electrons are instead transferred into the semiconductor and the bands bend downward

We can now analyze the metal–semiconductor interface using electrostatics as we did for the *pn* junction. The electric field is thus given by

$$E_x = -\frac{qN_\mathrm{d}}{\varepsilon_\mathrm{s}}(x_\mathrm{d} - x), \quad 0 \le x \le x_\mathrm{d} \tag{2.64}$$

and plotted in Fig. 2.23b. The maximum field, located at the metallurgical junction, has the value

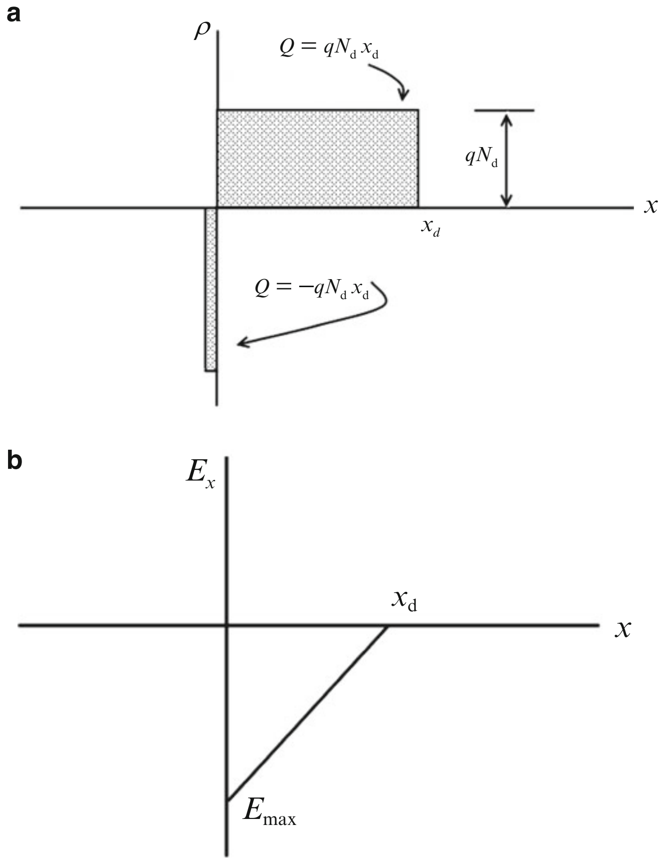


Fig. 2.23 (**a**) Space-charge density for metal–semiconductor junction in Fig. 2.22 (charge per unit area Q is also shown). (**b**) Corresponding electric field in depletion region of metal–semiconductor junction

$$E_{\mathrm{max}} = \frac{-qN_{\mathrm{d}}x_{\mathrm{d}}}{\varepsilon_{\mathrm{s}}} \tag{2.65}$$

The potential variation in the semiconductor is given by

$$V(x) = \phi_{\mathrm{bi}} - \frac{qN_{\mathrm{d}}}{2\varepsilon_{\mathrm{s}}}(x_{\mathrm{d}} - x)^2, \quad 0 \le x \le x_{\mathrm{d}} \tag{2.66}$$

where the built-in potential, expressed as the negative of the area under the field curve, is

$$\phi_{\mathrm{bi}} = -\frac{1}{2}E_{\mathrm{max}}x_{\mathrm{d}} = \frac{1}{2}\frac{qN_{\mathrm{d}}x_{\mathrm{d}}^2}{\varepsilon_{\mathrm{s}}} \tag{2.67}$$

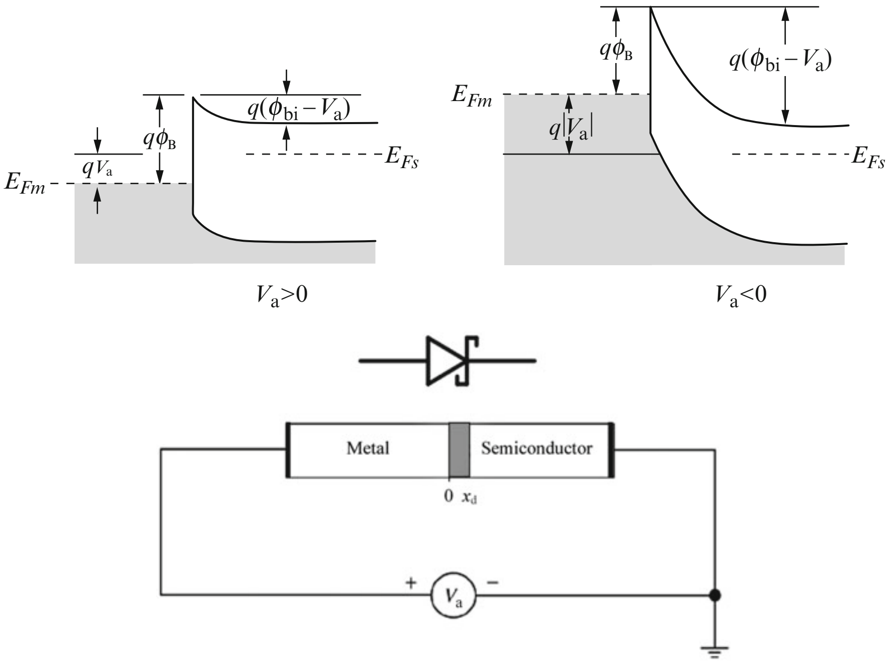


Fig. 2.24 Effect of applied bias on metal–semiconductor (n-type) junction potential barrier in forward and reverse directions. The interface barrier height ϕ_B is assumed unaffected. Low-resistance contacts are made to both the metal and semiconductor regions and the applied bias is assumed to drop entirely across the semiconductor depletion layer (For a Schottky barrier to a p-type semiconductor, the forward- and reverse-bias polarities are interchanged) (Electrical symbol also shown)

Using this expression we can write the depletion width as

$$x_\mathrm{d} = \sqrt{2\phi_\mathrm{bi}\varepsilon_\mathrm{s}/qN_\mathrm{d}} \tag{2.68}$$

Note that the results derived above in Eqs. (2.64), (2.65), (2.66), (2.67), and (2.68) are identical to those for a $p^{+}n$ junction. Similarly, a metal–semiconductor barrier based on p-type material would be analogous to a $n^{+}p$ junction.

2.2.1.2 *I–V* Characteristic

The potential barrier at the interface, ϕ_B, makes it difficult for free electrons to travel from the metal to the semiconductor. To first order this barrier is independent of applied bias since it is determined by the material properties of the metal/semiconductor and not any built-in charges. On the other hand, the built-in potential, ϕ_bi, impeding electrons from traveling from the semiconductor to the metal can be altered from its thermal equilibrium value by an applied bias. Analogous to the *pn* junction, the built-in potential is reduced when the metal is biased positively with respect to the semiconductor, and it is increased when the metal is made more negative as illustrated in Fig. 2.24.

The applied bias changes the depletion width to

$$x_{\mathrm{d}} = \sqrt{2(\phi_{\mathrm{bi}} - V_{\mathrm{a}})\varepsilon_{\mathrm{s}}/qN_{\mathrm{d}}} \tag{2.68}$$

Similarly, the magnitude of the maximum electric field in the junction becomes

$$E_{\max} = \frac{2(\phi_{\mathrm{bi}} - V_{\mathrm{a}})}{x_{\mathrm{d}}} \tag{2.69}$$

which is equivalent to Eq. (2.32b).

To obtain an expression for the metal–semiconductor barrier *I–V* characteristic we make use of the fact that in thermal equilibrium the net current across the metal–semiconductor interface must be zero. In other words, the rate at which electrons cross over the barrier into the semiconductor from the metal is balanced by the rate at which electrons cross the barrier into the metal from the semiconductor. We can apply these arguments at the boundary plane of the band edge diagram in thermal equilibrium:

The electron currents in either direction across the junction boundary due to thermal motion of carriers are proportional to the density of electrons at the boundary.[60] In the semiconductor this density is (referring to Fig. 2.22b)

$$n_{\mathrm{s}} = N_{\mathrm{c}}\exp\left(-\frac{q\phi_{\mathrm{B}}}{k_{\mathrm{B}}T}\right) = N_{\mathrm{d}}\exp\left(-\frac{q\phi_{\mathrm{bi}}}{k_{\mathrm{B}}T}\right) \tag{2.70}$$

Thus the condition for thermal equilibrium at the junction corresponds to

$$|J_{\text{M-S}}| = |J_{\text{S-M}}| = KN_{\mathrm{d}}\exp\left(-\frac{q\phi_{\mathrm{bi}}}{k_{\mathrm{B}}T}\right) \tag{2.71}$$

where $J_{\text{M-S}}$ and $J_{\text{S-M}}$ are the thermally induced current densities directed from the metal toward the semiconductor and vice versa, and K is a constant of proportionality. When a bias V_{a} is applied the built-in potential of the semiconductor is changed and we can expect the flux of electrons from the semiconductor toward the metal to be modified by a factor

$$n_{\mathrm{s}} = N_{\mathrm{d}}\exp\left(-\frac{q(\phi_{\mathrm{bi}} - V_{\mathrm{a}})}{k_{\mathrm{B}}T}\right) \tag{2.72}$$

The flux of electrons from the metal to the semiconductor, however, is not affected by the applied bias because the barrier at the interface is assumed to remain fixed at its equilibrium value. We can subtract these two components to

[60] For a system obeying Maxwell–Boltzmann statistics, it can be shown that the electron (or hole) current across a plane due to carriers in thermal motion can be expressed as $J = qn\overline{v}_{\mathrm{th}}/4$, where $\overline{v}_{\mathrm{th}}$ is the mean or average thermal velocity, $\overline{v}_{\mathrm{th}} = \sqrt{\frac{8k_{\mathrm{B}}T}{\pi m^*}}$.

obtain an expression for the net current density from the metal into the semiconductor under applied bias:

$$\begin{aligned} J &= J_{\text{M-S}} - J_{\text{S-M}} \\ &= KN_{\text{d}}\text{exp}\left(-\frac{q(\phi_{\text{bi}} - V_{\text{a}})}{k_{\text{B}}T}\right) - KN_{\text{d}}\text{exp}\left(-\frac{q\phi_{\text{bi}}}{k_{\text{B}}T}\right) \end{aligned} \tag{2.73}$$

which can be written

$$J = J_0[\exp(qV_{\text{a}}/k_{\text{B}}T) - 1] \tag{2.74a}$$

where $J_0 = KN_{\text{d}}\text{exp}(-q\phi_{\text{bi}}/k_{\text{B}}T)$. Equation (2.74a) is in the form of the ideal diode equation, once again showing the strong asymmetry or rectification of current flow across the junction. Thus, this type of metal–semiconductor junction is referred to as a *Schottky diode*.

Given the resemblance to a *pn* junction it is not surprising that the Schottky diode has a similar current–voltage dependence although the presence of the metal–semiconductor barrier causes the expression for saturation current density to be quite different. If we repeat the above analysis starting from Eq. (2.70) but instead work with the expression involving N_{c} and rewrite Eq. (2.71) in terms of $J = qn\overline{v}_{\text{th}}/4$, the saturation current density can be found explicitly as

$$J_0 = \frac{4\pi q m_n k_{\text{B}}^2 T^2}{h^3}\exp\left(-\frac{q\phi_B}{k_{\text{B}}T}\right) \tag{2.74b}$$

The model leading to Eqs. (2.74) captures the essential physics of transport across a metal–semiconductor barrier.

The principles of electron transport over the barrier and into the metal were first derived by Bethe in 1942.[61] If scattering inside the depletion region is significant this model needs to be modified to include drift–diffusion processes. Additional transport processes that can occur are tunneling through the barrier (discussed further below), generation–recombination in the depletion region (similar to a *pn* junction), and minority carrier contributions. The relative importance of these effects will depend on the particular metal/semiconductor combination as well as the applied bias. In addition, if the doping profile in the semiconductor is not constant[62] the saturation current density will also be modified. In more detailed treatments the saturation current density is also not completely independent of applied voltage. However, the dependence is weak compared to the exponential term and, as we saw for the *pn* diode, the current–voltage relationship for a metal–semiconductor diode can be approximated quite accurately using Eq. (2.42):

[61] This is known as the thermionic emission model.

[62] For example, in a *Mott barrier* a thin, lightly doped semiconductor region contacts the metal, which transitions to a highly doped bulk region a short distance from the junction.

$$I = I'_0 \left[\exp\left(\frac{qV_\mathrm{a}}{\eta k_\mathrm{B} T} \right) - 1 \right]$$

where I'_0 is independent of voltage and η is usually found experimentally to be slightly greater than 1 for a Schottky diode.[63]

2.2.2 Metal–Semiconductor Ohmic Contacts (Non-blocking)

In the metal–semiconductor barrier junction examined in the previous section, the applied voltage was dropped mainly across the high-resistance junction region and currents are therefore limited by the metal–semiconductor contact. The opposite case, in which the contact offers negligible resistance to current flow, defines an *ohmic contact*. Being able to make low-resistance metal–semiconductor junctions is very important for electrically contacting devices without modifying their inherent *I*–*V* characteristics. This is even more important in modern integrated circuit devices which are smaller and subsequently have larger current densities and regions with increased doping levels.

Two standard approaches are used to make ohmic contacts:

1. Tunnel contacts

   The metal–semiconductor junction barrier structure we considered previously, for example, can be made ohmic if the effect of the barrier on current flow is made negligible. This can be achieved by heavily doping the semiconductor so that the barrier width, given by the depletion width in Eq. (2.68), namely,

   $$x_\mathrm{d} = \sqrt{\frac{2\varepsilon_\mathrm{s}\phi_\mathrm{bi}}{qN_\mathrm{d}}}$$

   is made very small. When the barrier width approaches a few nanometers, *tunneling* transport through the barrier can take place when a bias is applied as depicted in Fig. 2.25a. Many electrons are available to take part in tunneling and currents rise very rapidly with applied bias, regardless of polarity. Hence, a metal–semiconductor contact at which tunneling is the significant transport process has a very small resistance and it is virtually always an ohmic contact. To ensure a very thin barrier, the semiconductor is often doped until it is *degenerate* (i.e., until the Fermi level is very close to or enters either the valence or the conduction band). Tunnel contacts to heavily doped semiconductors are widely used in integrated circuits.

[63] At higher doping levels (~10^{18} cm^{-3}) and/or lower temperature η begins to deviate from unity [2] as tunneling becomes more important.

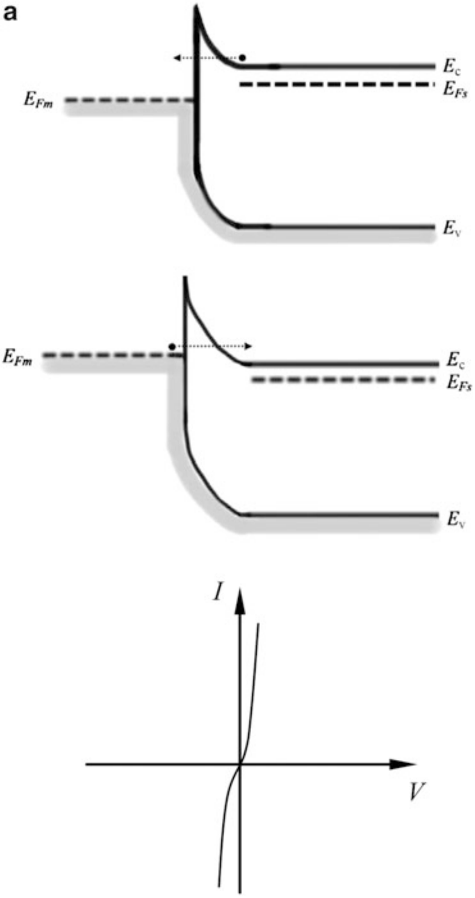


Fig. 2.25 (**a**) Ohmic tunnel contact and corresponding *I–V* characteristic showing rapid current increase with voltage for both polarities due to electron tunneling. (**b**) Schottky ohmic contact. Electrons are transferred from the metal into the semiconductor and therefore the bands bend downward at the interface because of the enhanced electron concentration. For the corresponding ohmic contact to a p-type semiconductor electrons are instead transferred into the metal and the bands bend upward due to the increased hole concentration

2. Schottky ohmic contacts
   Another way to achieve ohmic behavior is to choose materials with the proper work function difference so majority carriers become more numerous near the contact than they are in the bulk of the semiconductor. In our rectifying contact example, the metal had a greater work function than the n-type semiconductor and electrons were therefore transferred from the semiconductor to the metal to reach thermal equilibrium. If we instead choose a metal with a smaller work function, electrons will be transferred from the metal to the semiconductor. In this case the semiconductor surface is not depleted when it comes into equilibrium with the metal, but rather has an enhanced majority carrier concentration

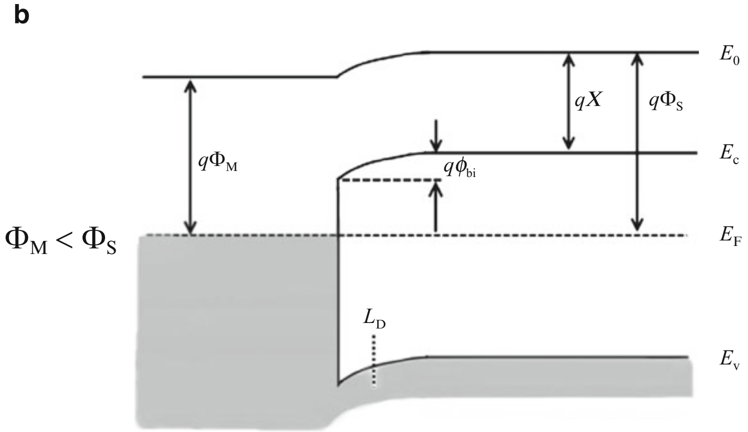


Fig. 2.25 (continued)

(Fig. 2.25b). Note in this diagram the *Debye length*, L_D, is a measure of the spatial extent of the additional charge carriers near the semiconductor interface.[64]

We can summarize Schottky contacts thusly; metal contacts to n-type material can be classified as being *ohmic* when the bands bend *down* toward the interface and rectifying or *blocking* when the bands bend *up*. The inverse conditions apply for contacts to p-type material[65] (see Figs. 2.22b and 2.25b).

The essential condition for ohmic behavior is an unimpeded transfer of charge carriers between the two materials forming the contact: At contacts there are generally built-in potentials and unless very thin barriers are present (i.e., tunneling is significant) majority carriers must be more numerous than they are in the bulk in order to achieve an ohmic contact.

[64] This may also be viewed as a measure of the *screening length* or the distance over which the free charges in the semiconductor rearrange themselves in response to the additional electrons in order to cancel out the electric field inside the bulk of the semiconductor. The Debye length is proportional to the density of free charges in a material; for most metals L_D is typically well below 1 nm, while it can range from 10 to 100 nm or more in a semiconductor depending on the doping level, with a maximum value for the intrinsic case.

[65] Note that the equations given for a metal contact to n-type material apply equally for contacts to p-type material with appropriate substitutions for the doping type (N_a for N_d), effective mass (m_p for m_n), etc.

2.2.3 Deviations from Ideal Behavior

2.2.3.1 Surface States and Barrier Height

The metal–semiconductor junctions considered thus far were ideal in the sense that we assumed the allowed energy states at the junction interface surface were the same as those in the bulk semiconductor material. In general, any surface of a crystal will introduce extra allowed states for electrons, often referred to as Shockley–Tamm sates. These states arise because the electrons on the surface are bonded only from the side directed toward the bulk as shown in Fig. 2.26a. In addition, impurities and crystal defects are sources of other types of surface states.

In real metal–semiconductor junctions the surface states almost always modify the height of the junction barrier from the idealized case. To account for these interfacial effects the metal–semiconductor junction can be treated as containing a very thin intermediate region sandwiched between the two materials: In Fig. 2.26b the n-type semiconductor is depleted of electrons near its surface by acceptor surface states. If a rectifying contact is now formed with a metal having a larger work function, electrons will be transferred from the surface states to the metal. However, since the density of surface states is usually very large (proportional to the density of atoms at the interface), a negligible movement of the Fermi level at the semiconductor surface transfers sufficient charge to equalize the Fermi levels. This is referred to as *Fermi-level pinning*.[66]

When the Fermi level is pinned the barrier height of a metal–semiconductor junction becomes

$$q\phi_{\mathrm{B}} = \left(E_{\mathrm{g}} - q\phi_0\right) \tag{2.75}$$

where $q\phi_0 = E_{\mathrm{F}} - E_{\mathrm{v}}$, compared to the ideal case (no surface states) we saw earlier (Eq. 2.63a):

$$q\phi_{\mathrm{B}} = q(\Phi_{\mathrm{M}} - X)$$

In general the barrier height will have an intermediate value that depends on the magnitude of surface states near the Fermi level of the semiconductor. However, in practice, Schottky barrier heights for most standard semiconductors based on the diamond lattice (Si, Ge, GaAs, etc.) are usually more accurately described by Fermi-level pinning with a smaller dependence on metal work function. Experimentally, it is found for these semiconductors,

[66] The importance of surface states on metal–semiconductor interfaces was pointed out by Bardeen in 1947. The effect of Fermi-level pinning is somewhat analogous to adding a very thin heavily doped layer between the metal and semiconductor.

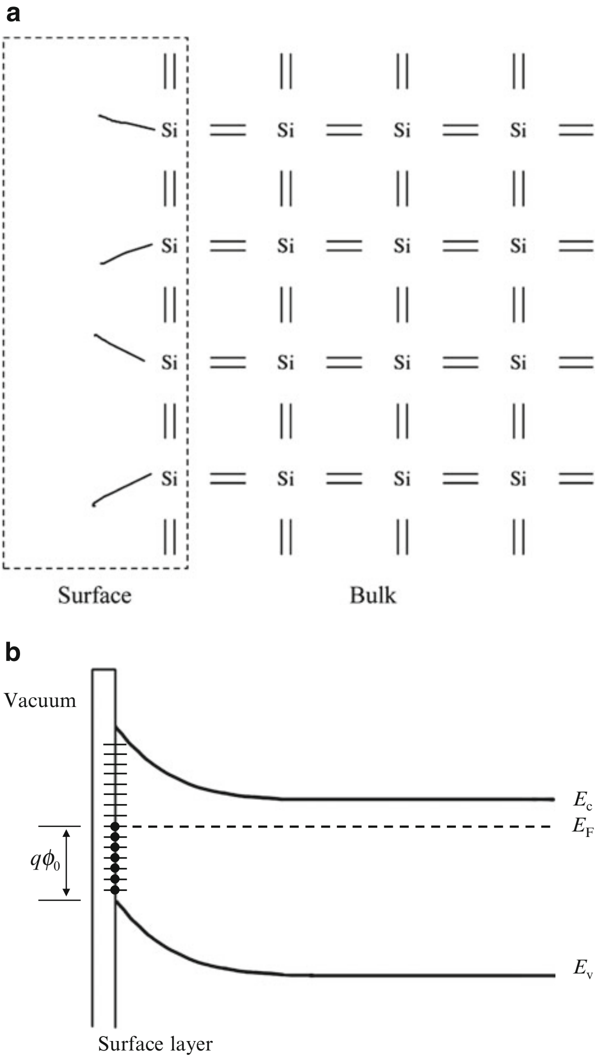


Fig. 2.26 (**a**) Illustration of incomplete or dangling bonds at the surface of a semiconductor that lead to a large density of localized surface interface states. (**b**) Surface transition layer representing semiconductor interface states leading to a space-charge region near the surface of the semiconductor at thermal equilibrium

$$q\phi_0 \approx \frac{1}{3}E_g \tag{2.76}$$

so that the barrier height $q\phi_B$ is roughly 2/3 of the band gap energy. Although surface states and other non-idealities will modify the exact Schottky barrier height, the ideas and properties of idealized metal–semiconductor junctions discussed above are still valid[67] as long as an accurate barrier height is used, obtained from experimental data or other means.

2.2.3.2 Effects at High and Low Applied Bias

Other deviations from ideal behavior, similar to those observed in *pn* junctions, can also occur in Schottky barrier diodes:

Under large reverse bias avalanche breakdown is generally observed for Schottky diodes based on moderately doped semiconductors, whereas breakdown via tunneling can occur for more heavily doped diodes. Generation currents arising in the depletion region can also cause the reverse saturation current in a Schottky diode to gradually increase with the magnitude of reverse bias. Similarly, space-charge region recombination currents under forward bias can also play a role. The deviation from ideal diode exponential behavior caused by space-charge generation–recombination in Schottky diodes is usually small compared to *pn* junctions and is typically more evident in junctions with large barriers and at low temperatures. In addition, unlike the ideal case, the barrier height itself will also vary with applied reverse bias due to the effects of the applied field and image forces.[68] This so-called *Schottky barrier lowering* is more important for small barrier heights and causes the saturation current to increase gradually with reverse bias. Finally, under large applied forward bias the series resistance of the semiconductor must also be taken into account.

2.2.4 Small-Signal Parameters

2.2.4.1 Conductance

The small signal conductance of an ideal Schottky diode is equivalent to the *pn* diode expression:

$$G = \frac{dI}{dV_\mathrm{a}} = \frac{q}{k_\mathrm{B}T} I_0 \left(e^{qV_\mathrm{a}/k_\mathrm{B}T} \right) = \frac{q}{k_\mathrm{B}T}(I + I_0)$$

with the only difference being the particular value of the saturation current.

[67] Essentially only Eqs. (2.63a) and (2.63b) will require modification.

[68] Electrons that are emitted from the metal into the semiconductor under reverse bias will induce images charges of the opposite sign in the planar metal surface near the interface, which causes the barrier height to be lowered within a few nanometers from the metallurgical junction.

2.2.4.2 Junction Capacitance

The small-signal junction capacitance per unit area for a Schottky diode can be found using the space charge stored in the semiconductor:

$$Q_{\mathrm{s}} = \sqrt{2q\varepsilon_{\mathrm{s}}N_{\mathrm{d}}(\phi_{\mathrm{bi}} - V_{\mathrm{a}})} \tag{2.77}$$

and therefore,

$$C = \left|\frac{dQ_{\mathrm{s}}}{dV_{\mathrm{a}}}\right| = \sqrt{\frac{q\varepsilon_{\mathrm{s}}N_{\mathrm{d}}}{2(\phi_{\mathrm{bi}} - V_{\mathrm{a}})}} = \frac{\varepsilon_{\mathrm{s}}}{x_{\mathrm{d}}} \tag{2.78}$$

Solving this equation for the total voltage across the junction gives

$$(\phi_{\mathrm{bi}} - V_{\mathrm{a}}) = \frac{q\varepsilon_{\mathrm{s}}N_{\mathrm{d}}}{2C^2} \tag{2.79}$$

This indicates that a plot of $1/C^2$ versus the applied voltage should be a straight line. The slope can be used to obtain the doping level in the semiconductor and the intercept with the voltage axis should equal the built-in voltage.[69] Such plots are often used to study semiconductors. In practice, the built-in voltage obtained is not as accurate as the doping level. Commercial profilers can also use this data to plot dopant concentration as a function of position on a semiconductor wafer.

The small-signal equivalent circuit for the Schottky barrier diode is identical to the *pn* diode circuit shown in Fig. 2.16, except for one important difference: Since the Schottky diode is predominantly a majority carrier device, diffusion capacitance due to minority carrier charge storage is absent. This makes the Schottky diode an intrinsically fast device that can respond to frequencies into the THz regime.[70] Similar comments apply to the transient behavior of Schottky diodes: their characteristic timescale is no longer associated with recombination/diffusion processes but rather the transit time associated with the drift of majority carriers in the depletion region, i.e., the dielectric relaxation time. As mentioned earlier, this is a fast process and thus Schottky diodes can be switched very rapidly.[71]

[69] Equation (2.79) is also valid for a one-sided abrupt *pn* junction (p^+n or pn^+).

[70] As for the *pn* diode, we can define a Schottky diode cutoff frequency as $f_T = (2\pi\ RC)^{-1}$. Typically, the fastest Schottky diodes are made from semiconductors with the highest carrier mobility in order to reduce the effect of series resistance.

[71] The external circuit parameters will largely determine how quickly a Schottky diode can be switched as opposed to intrinsic delays in the device itself.

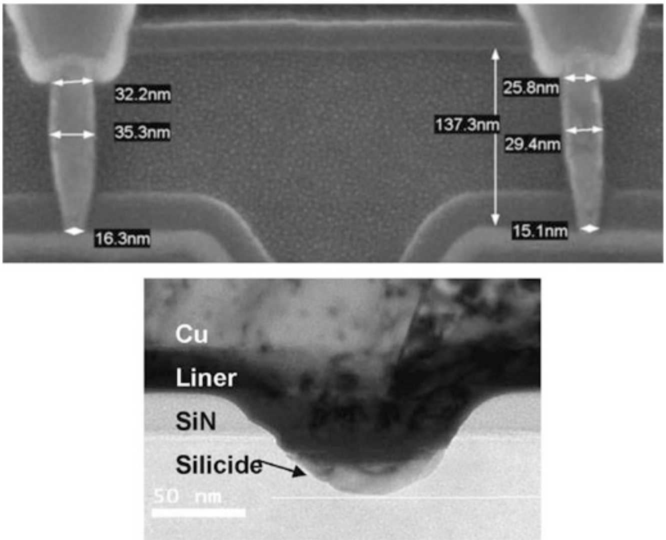


Fig. 2.27 Electron microscope images of IC metal–semiconductor contacts based on copper/silicide conductors (*Source*: K. Ohuchi et al., 8th International workshop on Junction Technology, IWJT, 2008; S.-C. Seo et al., Copper Interconnect Technology Conference, IITC, 2009)

Panel 2.3: Metal–Semiconductor Contact and Device Structures Modern metal–semiconductor contacts and diodes are usually made via planar processing. For creating ohmic tunnel contacts in integrated circuits, a chemical reaction between the metal and underlying silicon contact area is commonly used to form a low-resistance metallic *silicide* (a silicon–metal compound) region, which reduces the lateral resistance of the contact. This process creates a stable, low-resistance electrical connection. Some examples of IC contacts are shown in Fig. 2.27.

Schottky barrier diodes are used extensively in integrated circuits because they are relatively easy to fabricate and combine with other devices on a chip. A typical planar metal–semiconductor diode is shown in Fig. 2.28a. One issue that must often be dealt with in such structures is premature breakdown under moderate levels of reverse bias due to the larger electric fields that exist at the edge of the device.[72] The concentrated electric fields at the edges of the junction can be mitigated in several ways: Fig. 2.28b, c show two common techniques—using an overlapping electrode metal contact or an additional insulating oxide layer to create larger separation at the edges in order to reduce the field. The applications of such structures will be expounded in Sect. 2.4 and discussed further in Chap. 3.

[72] This edge or surface breakdown mechanism can also be important in *pn* diodes but is generally not as severe unless high power devices are required.

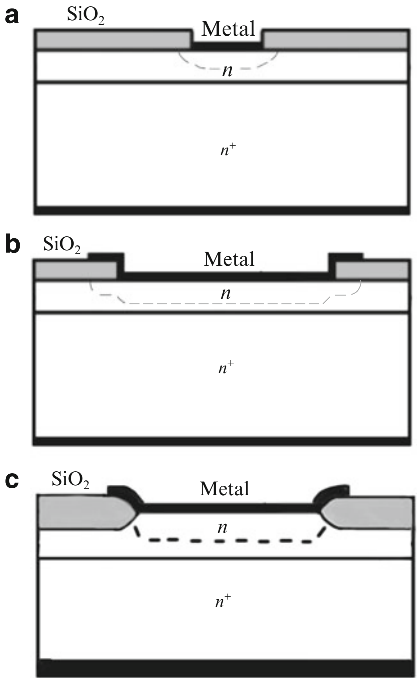


Fig. 2.28 (**a**) Planar Schottky diode device schematic. (**b**, **c**) Approaches for reducing the electric field at the diode edges (*Dotted lines* indicate edge of depletion region) (Adapted from [2])

2.3 Other Types of Junctions

In this section we briefly discuss the properties of two other important junctions that are encountered in practice. Although full details are not provided here, the *pn* and metal–semiconductor results provide sufficient background to understand most other junction types as well.

2.3.1 Isotype Junctions

Figure 2.29 shows the thermal equilibrium band edge diagram for a semiconductor that has two regions with different doping levels that are of the *same type*. Typical examples of these so-called isotype[73] junctions are p^+p and n^+n interfaces. In such junctions, there will be a transfer of carriers (electrons or holes) from the more heavily doped region into the more lightly doped region to achieve thermal equilibrium. The built-in potential of the isotype junction can be found by applying Eq. (2.12).

[73] J. B. Gunn, J. Electron. Control **4**, 17 (1958) is an early study on isotype junctions.

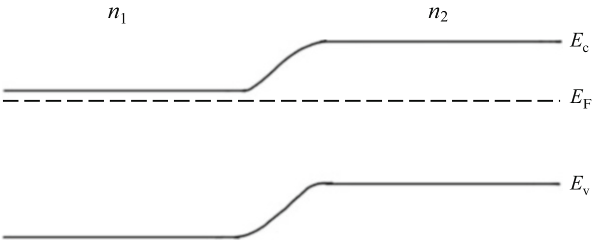


Fig. 2.29 Isotype junction band edge diagram for two n-type regions, $n_1 > n_2$

In contrast to *pn* junctions, however, since the carriers are transferred into a region of the same doping type, there will be an *enhancement* of the carrier concentration near the interface in the more lightly doped side of the junction. This situation is very similar to the ohmic Schottky contact shown in Fig. 2.25b, and indeed an isotype junction may be considered as an ideal ohmic contact in the sense that it consists of the same material (i.e., interface states will not play a role).

2.3.2 Heterojunctions

Thus far when considering the junction between two semiconductor regions we have always assumed that it consisted of the same type of semiconductor, known as a *homojunction*. If we lift this restriction, the resulting interface is now called a *heterojunction*. Such junctions began being studied theoretically in the 1950s followed by important experimental studies in the 1960s.[74]

We can analyze junctions between different types of semiconductors in a manner analogous to those already examined. In particular, an idealized energy-band model (Anderson model[75]) can be used that is adequate for most purposes as illustrated in Fig. 2.30 for the case of a *pn* junction between a narrow and wide band gap semiconductor, respectively. Similar comments apply to isotype heterojunctions,

[74] Shockley proposed using heterojunctions for devices in 1951 while in the same year Gubanov presented studies on their theoretical properties. In 1957, Kroemer provided important extensions to the earlier heterojunction studies and device proposals, followed by pioneering experimental studies by Anderson around 1960. Subsequently, the ability to grow different semiconductor layers with atomic precision and minimal lattice mismatch allowed high-quality heterojunction interfaces to be developed.

[75] R. L. Anderson, IBM J. Res. Dev. 4, 283 (1960). The Anderson model is similar to the idealized Schottky barrier model and can be modified to also include the effects of non-idealities such as interface states, etc.

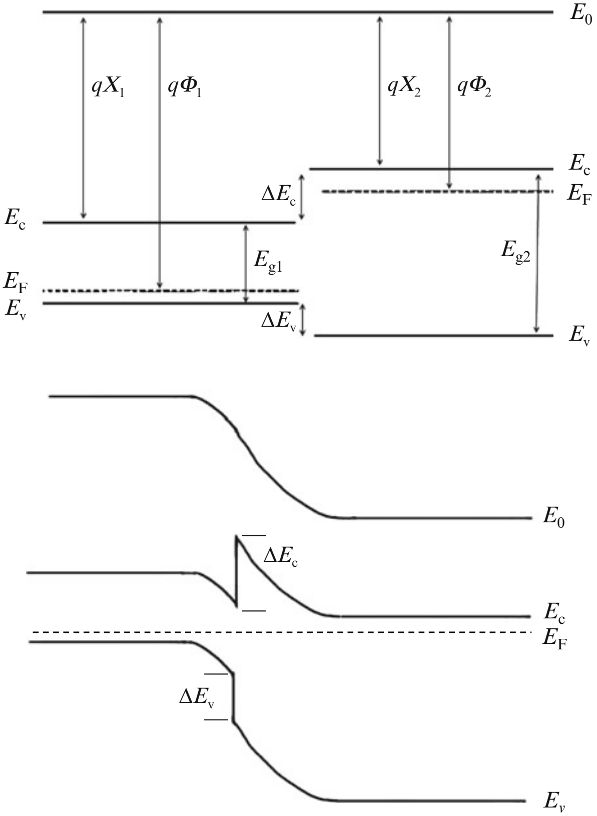


Fig. 2.30 *pn* heterojunction band edge diagram (cf. Figs. 2.2b and 2.22b)

although one must realize that the electrical properties of any type of heterojunction will depend critically on the relative magnitudes of the band gap energies, work functions, electron affinities, and applied bias. As such, heterojunctions can display *I*–*V* behavior that is reminiscent of diffusive transport across a *pn* junction, transport over or through a metal–semiconductor barrier, and/or anywhere in between.

2.4 Applications of Single-Junction Devices or Diodes

The most common application of a two-terminal junction device or diode is current rectification, since it has very little resistance to current flow in one direction and a very high resistance in the other. The effectiveness of a rectifier can be characterized by defining a *rectification ratio* that relates current under forward bias to that under reverse bias.

We have also seen that because of their voltage variable junction capacitance both *pn* and Schottky diodes can be employed as varactors in various applications. Varactors typically operate under reverse bias in order to maximize the voltage range that can be used to adjust the capacitance and avoid excessive currents.

Another common use of a diode is to provide *isolation*, as we have already seen for the case of reverse-biased *pn* junctions in integrated circuits. More generally, diodes are used to provide electrical isolation and protection for a wide variety of applications and systems in the form of electrostatic discharge (ESD) protection, which prevents high voltages by passing large currents through either forward-biased or reverse-biased (Zener) diodes. Other applications where diodes are used include diode logic circuits, battery/mains switching circuits and some matrix or crossbar switches.

The well-defined forward-bias characteristics of junction diodes also enable them to be used as very accurate temperature sensors. A constant current is typically passed through the diode while the forward voltage variation is measured as a function of temperature (see Problem 3). The "freezing out" of dopants at very low temperatures increases the effective resistance of the diode and causes the voltage to increase much more rapidly, which sets a lower limit for diode thermometers of about 1 or 2 K in practice. On the other hand, depending on the type of semiconductor and doping levels diode temperature sensors can operate up to 500 K or more.[76]

2.4.1 *pn* Diodes

pn junctions generally have a smaller reverse saturation current density than typical Schottky diodes. In addition, *pn* junctions are usually preferred for Zener diode applications because of the greater control that can be achieved through dopant concentrations and distributions.

A very important and distinct application of *pn* junctions is in *optoelectronics*, in particular light-emitting devices such as light-emitting diodes (LEDs) and laser diodes that rely on the emission of light via e–h pair recombination in and around the space-charge region of junctions made using direct band gap semiconductors. *pn* junctions are also commonly used for solar cells and photodiodes (the sensing elements in many digital cameras) as discussed further below. Applications in *solid-state lighting* (based on high-efficiency LED light sources) and solar energy production are some of the strongest growth areas for *pn* junction devices.

[76] As temperature increases the forward voltage drop becomes progressively smaller and this ultimately limits sensitivity at high temperatures. The increased intrinsic carrier concentration may also affect the majority carrier levels in the semiconductor at elevated temperatures (see Appendix A) and alter the expected junction behavior.

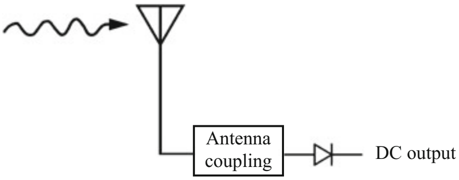


Fig. 2.31 Diode radio detection, general schematic. The incoming radiation (small signal or large signal) is rectified by the diode, which results in a nonzero dc output current component that can be used to produce an audible output or processed further (e.g., amplified, etc.), depending on the application

2.4.2 Schottky Diodes

As discussed earlier, since the Schottky diode is a majority carrier device it can be operated up to THz frequencies, which usually makes it the diode of choice for high-speed applications, such as microwave detectors, mixers, varactors, oscillators, etc. Schottky diodes are also used aboard satellites to detect changes in the Earth's atmospheric chemical species via GHz and THz gas spectroscopy.[77]

Another advantage of Schottky diodes for some applications is that they typically have a lower voltage drop (i.e., lower "turn-on" voltage) in forward bias than *pn* junctions. For example, Schottky diodes fabricated on n-type silicon will typically have a built-in potential about 200–300 mV smaller than a comparable *pn* junction. Thus a Schottky diode placed in parallel with a *pn* junction diode will prevent it from passing significant current in the forward direction or in other words the *pn* junction is "clamped" to a lower forward voltage. Such Schottky-clamped devices can be used to improve the switching speed of digital logic circuits (Chap. 3), in addition to general voltage clamping circuit and other applications where a low forward voltage drop and/or fast switching is advantageous.

Lastly, from a manufacturing point of view, Schottky diodes generally involve simpler processing steps compared to *pn* junctions. In addition, lower temperatures during fabrication may provide an economic and environmental advantage for some applications due to the smaller input energy needed and also increases process compatibility.

Panel 2.4: Microwave Detection and the Development of Semiconductors Historically, one of the most important applications of diode rectifiers has been the detection of radio signals as discussed briefly in Chap. 1 in relation to vacuum tube diodes. A basic radio detection scheme is shown in Fig. 2.31. Early telegraph and

[77] The Earth Observing System (EOS) Microwave Limb Sounder (MLS) is one prominent example, which is part of NASA's Aura satellite and uses GaAs-based Schottky diode mixers to detect radiation via heterodyning. Some images of satellite data collection results are shown at the end of this chapter.

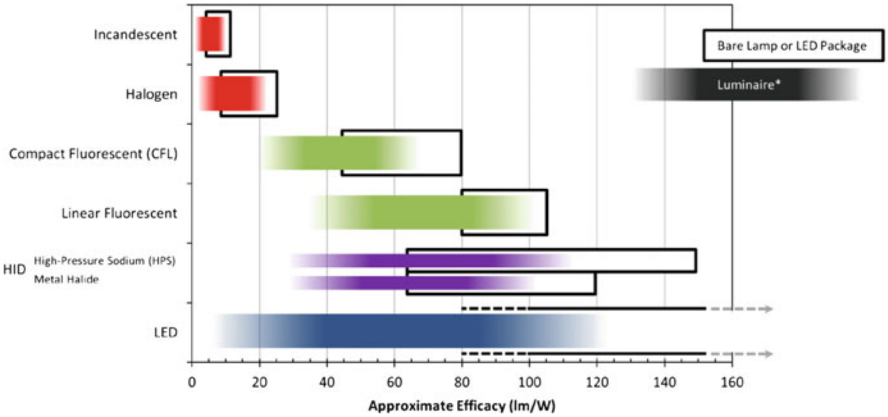


Fig. 2.32 Light output per watt for various modern electric light sources. Solid-state lighting based on LEDs continues to improve in performance and increase its market share (Luminaire refers to the entire lighting package) (Source: US Department of Energy, 2013)

radio receivers relied on point-contact diode detectors. These so-called crystal radio sets were very widespread in the early twentieth century until the development of vacuum tube technology. However, the need for microwave technology during the Second World War, in particular microwave radar, required detectors that operated at higher frequencies than earlier radio. Due to their size, typical vacuum tube devices were not able to respond fast enough to enable microwave detection. Thus, a resurgence in the use of the much faster point-contact diodes occurred. Germanium and silicon crystal rectifiers would play key roles as radar receivers in the war. Very importantly, this led to renewed interest in semiconductors following the war and paved the way for the development of modern electronics in the second half of the twentieth century.

To complete the discussion on applications we mention that isotype junctions are very useful as ohmic contacts (as alluded to above) and this is their most widespread application, particularly for planar integrated circuit technology, and heterojunctions are widely used for optoelectronics including LEDs and diode lasers because they often allow tremendous improvements in performance by enhancing the radiative recombination of carriers. For example, modern LEDs based on different types of heterojunctions have now become among the most efficient sources of light available and these set the standard for solid-state lighting applications at present (see Fig. 2.32). Heterojunctions are also often used in solar cells, which is discussed more generally in the following panel.

Panel 2.5: Photovoltaics The amount of energy being generated by the sun is immense. Nuclear fusion creates approximately 10^{20} J/s of power; most of it emitted in the form of electromagnetic radiation in the UV and IR regions of the spectrum. This power output is expected to continue for another ten billion years or so and is the source of input energy for the earth. Of the energy produced by the

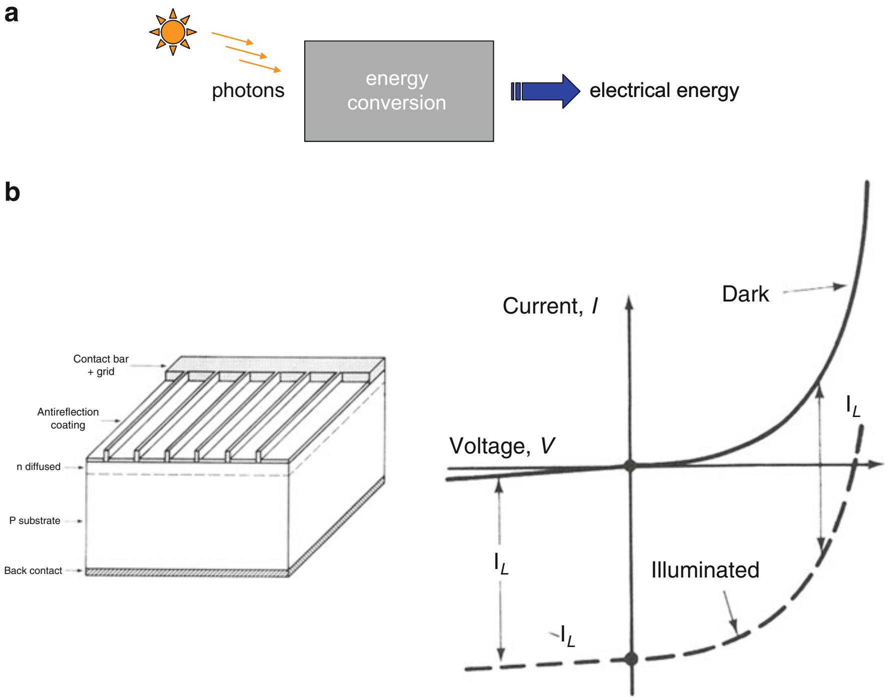


Fig. 2.33 (**a**) General concept for a solar cell that converts the energy present in sunlight into usable electrical energy. The conversion of light to electrical energy can be direct or indirect. (**b**) Illuminated diode *I*–*V* characteristic and simple *pn* junction solar cell device structure. The ideal diode characteristic is shifted downward by the photogenerated current I_L (After [10])

sun approximately 1,367 W/m^2 reaches the earth's atmosphere in the form of solar EM radiation (the solar constant).[78] Note that even with atmospheric losses the total amount of solar energy striking the earth's surface is approximately 10,000 times the total energy consumption of the world's entire population (~ 15 TW). Thus if this energy could be harnessed it could provide much of the world's energy needs. Over the past decade interest and growth in such solar cell technology have increased dramatically due to a combination of social, economic, and environmental factors.

The basic concept of a solar cell is shown in Fig. 2.33a. The development of practical solar cell technology has been based on finding an efficient and inexpensive means to convert the incoming solar radiation (photons) into electrical energy.

[78] This is however attenuated by the atmosphere before reaching the earth and the amount of "air mass" (AM) through which the light passes is used to describe the incident solar energy. Terrestrial solar cell performance is typically specified with respect to the AM1.5 spectrum (48° from vertical) or about 1,000 W/m^2.

In 1839 A. E. Becquerel observed that AgCl-coated Pt electrodes in an electrochemical cell (a semiconductor–liquid junction) produced a voltage when exposed to light. Subsequently, similar behavior was observed in many other materials, particularly solid-state junctions, and this phenomenon is now generally referred to as the *photovoltaic effect*. Virtually all solar cell technology currently in use is based on the photovoltaic effect, and thus these devices are also known as solar photovoltaic cells. Upon illumination solar cells can deliver power to an external circuit in a direct and clean manner (i.e., essentially nonpolluting).

The most common solar cells and those developed first commercially are based on *pn* junctions (see Fig. 2.33b for a basic solar cell structure). When the junction is illuminated with photons near or above the band gap energy the excess electron–hole pairs that are created will be swept out of the depletion region by the built-in electric field and result in current flow, i.e., *photovoltaic energy conversion* will occur. By extending our previous "dark" treatment of the *pn* junction we can show that the illuminated junction *I–V* characteristic is essentially a modification of the ideal diode equation as shown in Fig. 2.33b.

We can proceed in an identical manner to the previous ideal diode analysis; however, an extra term is now added to the diffusion equations that represents the generation of electron–hole pairs due to the incident light:

$$\begin{aligned} \frac{\partial n'}{\partial t} &= D_n \frac{\partial^2 n'}{\partial x^2} - \frac{n'}{\tau_n} + G \\ \frac{\partial p'}{\partial t} &= D_p \frac{\partial^2 p'}{\partial x^2} - \frac{p'}{\tau_p} + G \end{aligned} \tag{2.80}$$

As before, we can now look for the steady-state solutions to these equations. For simplicity, we assume that the junction is uniformly illuminated with photons, i.e., G is a constant. For steady state in the n-side of a *pn* junction we then have

$$0 = D_p \frac{d^2 p'_{\mathrm{n}}}{dx^2} - \frac{p'_{\mathrm{n}}}{\tau_p} + G \tag{2.81}$$

This is similar to the dark junction case, but the addition of the generation term has now resulted in a *nonhomogeneous* (linear) differential equation. Recall that a general solution to such an equation consists of the solution to the homogeneous equation (that we obtained earlier) plus any particular solution. We therefore get the following solution:

$$p'_{\mathrm{n}}(x) = A \exp\left(-\frac{x - x_{\mathrm{n}}}{L_p}\right) + B \exp\left(\frac{x - x_{\mathrm{n}}}{L_p}\right) + G\tau_p \tag{2.82}$$

where once again A and B are constants and L_p is the hole diffusion length (and L_n is the analogous electron diffusion length). To find the constants we can again consider two limiting cases based on the length W_B of the n-region from the junction to the ohmic contact (see Fig. 2.4). For the long-base diode the solution turns out to be[79]

$$p'_\mathrm{n}(x) = \left[p_\mathrm{n0}\left(e^{qV_\mathrm{a}/k_\mathrm{B}T} - 1\right) - G\tau_p\right]\exp\left(-\frac{x - x_\mathrm{n}}{L_p}\right) + G\tau_p \tag{2.83}$$

Note that, as expected, far away from the junction the excess carrier concentration approaches a constant value determined by the generation rate and the carrier lifetime. Using the above expression we can now calculate the diffusion current density due to holes at the edge of the space-charge region:

$$\begin{aligned} J_p(x) &= -qD_p\frac{dp_\mathrm{n}}{dx} \\ &= qD_p\frac{p_\mathrm{n0}}{L_p}\left(e^{qV_\mathrm{a}/k_\mathrm{B}T} - 1\right)\exp\left(-\frac{x - x_\mathrm{n}}{L_p}\right) - qGL_p\exp\left(-\frac{x - x_\mathrm{n}}{L_p}\right) \end{aligned} \tag{2.83}$$

The total current flowing through the *pn* junction is obtained as before by summing the two minority carrier currents flowing across the junction (electrons and holes) giving

$$\begin{aligned} J &= J_p(x_\mathrm{n}) + J_n(-x_\mathrm{p}) \\ &= qn_\mathrm{i}^2\left(\frac{D_p}{N_\mathrm{d}L_p} + \frac{D_n}{N_\mathrm{a}L_n}\right)\left(e^{qV_\mathrm{a}/k_\mathrm{B}T} - 1\right) - qG\left(L_p + L_n\right) \\ &= J_0\left(e^{qV_\mathrm{a}/k_\mathrm{B}T} - 1\right) - qG\left(L_p + L_n\right) \end{aligned} \tag{2.84}$$

That is, the ideal diode equation (dark) current is simply reduced or shifted by the photogenerated current. Thus, the photovoltaic effect creates a current that flows in the reverse direction through the junction.

A contribution that is not included in the above derivation is photo generation of carriers within the depletion region. We can easily include this additional current by adding another term to the photogenerated current proportional to the depletion width, x_d:

$$J = J_0\left(e^{qV_\mathrm{a}/k_\mathrm{B}T} - 1\right) - qG\left(L_p + L_n + x_\mathrm{d}\right) \tag{2.85}$$

[79] Note the excess carrier concentration at the edge of the depletion region will be determined (for the case of low-level injection, as before) by the voltage appearing across the junction, V_a (cf., Eq. (2.23)). (In the case of an illuminated junction V_a is not strictly an applied bias, but we keep the same notation for simplicity.)

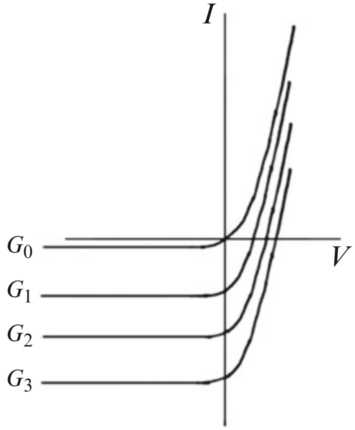


Fig. 2.34 *pn* junction *I–V* characteristics for increasing levels of illumination denoted by optical electron–hole pair generation rate *G*

Note that this extra term can be ignored if it is small compared to the minority carrier diffusion lengths. The above equation also indicates that the photogenerated current is caused by carriers generated within a diffusion length of the depletion region. The diffusion lengths, along with the depletion region itself, define the active "collection regions" of a *pn* junction solar cell.

For the short-base diode, we can repeat the above analysis; however, recall that the net result was simply a modification of the reverse saturation current density term appearing in the ideal diode equation, and other than replacing the diffusion lengths with the physical dimensions of the neutral regions, the functional form of the solution will be identical to that above.

Our (idealized) analysis of the illuminated *pn* junction has essentially shown that the current is given by

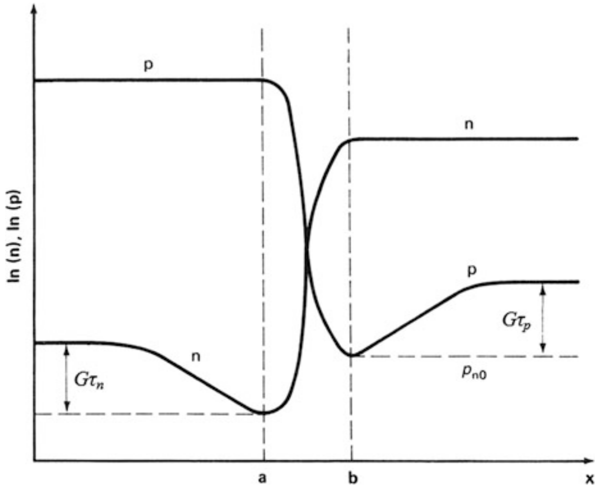
$$I = I_{\text{dark}} - I_L \tag{2.86}$$

where I_{dark} is the usual ideal diode equation current and I_L is the current due to optical generation:

$$I = I_0\left(e^{qV_\text{a}/k_\text{B}T} - 1\right) - qAG\left(L_p + L_n + x_\text{d}\right) \tag{2.87}$$

To first order this means that the illuminated *I–V* characteristics are identical to the dark characteristics except that they are translated downward by I_L as illustrated in Fig. 2.34. The point at which the curves cross the current axis is known as the *short-circuit current*, I_{sc} (i.e., when $V_\text{a} = 0$), which has a magnitude equal to I_L. On the other hand, when there is an open circuit across the junction device, $I = 0$ and the open-circuit voltage is

Fig. 2.35 *pn* junction short-circuit current caused by minority carriers diffusing towards the space-charge region during steady-state illumination (Adapted from [10])



$$V_{\text{oc}} = \frac{k_{\text{B}}T}{q}\ln\left(\frac{I_L}{I_0} + 1\right) \tag{2.88}$$

Thus, V_{oc} depends on the optical generation rate through I_L and the properties of the semiconductor through I_0. Recall that V_{oc} will never be larger than the built-in potential of the junction. Depending on the intended application, an illuminated *pn* junction can be operated in different quadrants of its *I*–*V* characteristic. For example, the third quadrant is typically used for photodetection and the device is then a *photodiode*. For the purpose of power/energy generation we are most interested in the *fourth quadrant* since in this case power can be extracted from the device (similar to a battery).

It is useful to briefly discuss the physical mechanisms of the voltage appearing across an illuminated junction: When light shines on the *pn* junction, one without an external bias voltage, each absorbed photon creates an e–h pair. When these carriers diffuse to the junction (or are created within the depletion layer) the built-in electric field separates them. This separation of charge produces a forward voltage across the barrier since the electric field of the photoexcited carriers is opposite to the built-in field (cf. Fig. 2.2b). As stated previously this is the origin of the photovoltaic effect and causes the appearance of the open-circuit voltage defined above for an illuminated junction.

We can also obtain some insight by examining the distribution of carriers near the illuminated junction when it is short-circuited as shown in Fig. 2.35. The origin of the short-circuit current can now be seen as due to the diffusion of carriers from high concentration (away from the junction) to low concentration (towards the junction) where these are swept away by the built-in electric field. The concentration gradient is maintained by the photogeneration of carriers. In steady state,

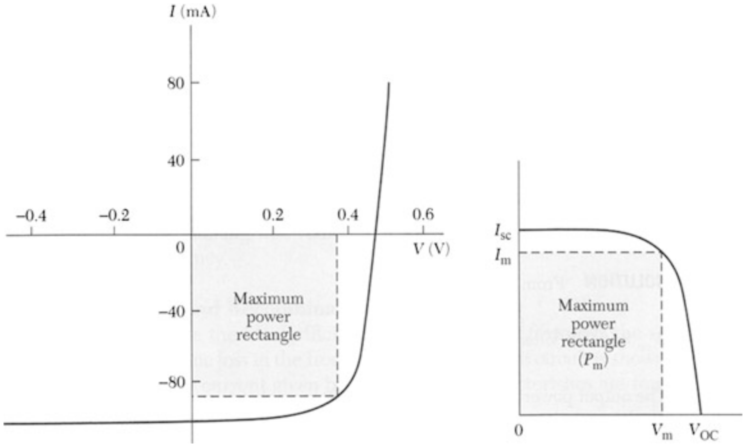


Fig. 2.36 Solar cell maximum power rectangle superimposed on diode *I–V* characteristic. By convention the data is often flipped vertically in order to give positive values as shown by the diagram on the right-hand side (After [2])

the illuminated junction will in general be somewhere in between the two extremes of short- and open-circuit behavior as the system responds to the nonequilibrium condition created by the exposure to light. In a sense, the response of the illuminated junction to a bias is opposite to that of the "dark" junction (first quadrant) since the output current magnitude now decreases with bias instead of increasing.

The output power for any operating point in the fourth quadrant can be found from the usual expression

$$P = IV = I_0 V \left(e^{qV_a/k_B T} - 1 \right) - I_L V \tag{2.89}$$

and the condition for maximum power output is obtained when $dP/dV = 0$ or

$$V_m = \frac{k_B T}{q} \ln \left[\frac{1 + (I_L/I_0)}{1 + (qV_m/k_B T)} \right] = V_{oc} - \frac{k_B T}{q} \ln \left(1 + \frac{qV_m}{k_B T} \right) \tag{2.90}$$

which along with the corresponding current I_m determines the maximum power output P_m. These parameters define the *maximum power rectangle* shown in Fig. 2.36. We can see that the area of the rectangle will always be less than the product of the short-circuit current and open-circuit voltage. The ratio

$$\text{FF} = \frac{I_m V_m}{I_{sc} V_{oc}} \tag{2.91}$$

is known as the *fill factor* and is an important figure of merit for solar cell design. Lastly, the *power conversion efficiency* provides an overall measure of solar cell performance:

$$\eta \equiv \frac{P_{\mathrm{m}}}{P_{\mathrm{in}}} = \frac{I_{\mathrm{m}}V_{\mathrm{m}}}{P_{\mathrm{in}}} = \frac{\mathrm{FF}I_{\mathrm{sc}}V_{\mathrm{oc}}}{P_{\mathrm{in}}} \qquad (2.92)^{80}$$

where P_{in} is the incident optical energy per unit time or the input power. Thus, to maximize efficiency all three solar cell parameters—fill factor, short-circuit current, and open-circuit voltage—should be maximized.[81]

Photovoltaics currently only provide a very small portion of the world's energy: Approximately 100 GW of solar power is installed worldwide. At present most of these installations are based on silicon (~80 %), although other materials are being actively developed as well. The photovoltaic market has been growing annually at a rate of approximately 50 %. New installations currently exceed 20 GW annually. If these types of growth rates continue terawatt levels of power could be reached by 2030. Many alternative materials and technologies are currently being developed to provide efficient solar energy conversion with lower production costs. In addition, these emerging technologies may allow implementations not readily achievable with existing devices (e.g., flexible panels, access to different regions of the electromagnetic spectrum, enhanced stability, integration with other devices and materials, etc.).

Although the material on junctions in this chapter has been somewhat extensive, it is worth reemphasizing the point made at the beginning of the chapter that this is because such junctions are the basis of virtually all modern electronics: Being able to control the distribution of charges in a semiconductor to create junctions has allowed the development of solid-state devices and circuits that can perform complex tasks,[82] which are the foundation of today's information-based society, as will be discussed further in Chaps. 3 and 4.

References

1. Shockley, W.: The Theory of p–n Junctions in Semiconductors and p–n Junction Transistors. Bell Syst. Tech. J. **28**, 435 (1949)
2. Sze, S.M., Ng, K.K.: Physics of Semiconductor Devices, 3rd edn. Wiley Interscience, Hoboken (2007)
3. Kingston, R.H.: Switching Time in Juction Diodes and Junction Transistors. Proc. IRE **42**, 829 (1954)

[80] Not to be confused with the diode ideality factor.

[81] Sources of solar cell losses include the inability to absorb photons with energy less than the band gap, heat generated by large energy photon absorption, reflection losses, carrier recombination, and parasitic resistances.

[82] Previously this was only possible in other areas of technology, for example, using the potential energy of a spring or the pressure/temperature differential of gases or of a chemical reaction to perform a useful task. Solid-state electronics has far surpassed the complexity of any other type of man-made "machine" in terms of the number of working parts operating together, as exemplified by the integrated circuit.

4. Muller, R.S., Kamins, T.I.: Device Electronics for Integrated Circuits, 3rd edn. Wiley, New York (2003)
5. Grove, A.S.: Physics and Technology of Semiconductor Devices. Wiley, New York (1967)
6. Shur, M.S.: Introduction to Electronic Devices. Wiley, New York (1995)
7. Neudeck, G.W.: The PN Junction Diode, 2nd edn. Prentice-Hall, Boston (1989)
8. Smith, R.A.: Semiconductors, 2nd edn. Cambridge University Press, Cambridge (1978)
9. Neamen, D.A.: Semiconductor Physics and Devices, 3rd edn. McGraw-Hill, New York (2003)
10. Green, M.A.: Solar Cells. Prentice-Hall, Upper Saddle River (1982)

Problems

1. *Long- vs. short-base diodes*. An ideal silicon *pn* diode is formed by diffusing a high concentration of phosphorus into a 75-μm-thick boron-doped wafer having a resistivity of 5 Ω-cm and minority carrier lifetime of 5 μs. The junction is formed 10 μm below the surface with area 10^{-4} cm^2. (1) Find the built-in potential. (2) Calculate the current flowing through the diode under an applied forward bias of 0.5 V. (3) Is I_0 for an ideal diode always constant?
2. *Ohmic voltage drops*. Consider a silicon short-base diode with the following parameters:

$$N_\mathrm{d} = 2 \times 10^{17}\mathrm{cm}^{-3} \text{ and } N_\mathrm{a} = 5 \times 10^{18}\mathrm{cm}^{-3}$$
$$x_\mathrm{B} = x_\mathrm{E} = 5 \ \mu m \text{ and } A = 10^{-4}\mathrm{cm}^2$$

   Find the voltage dropped in the neutral regions of the diode for an applied forward bias of 0.7 V.
3. [83] *Diode temperature dependence*. An ideal long-base Si *pn* diode has $N_\mathrm{a} = 10^{17}$ cm$^{-\,3}$, $N_\mathrm{d} = 7 \times 10^{16}cm^{-\,3}$, and a cross-sectional area of $10^{-\,3}$cm^2. (1) If $\tau_n = \tau_p = 1$ μs, calculate the current flowing through the junction under an applied bias of 0.5 V. Repeat your calculation for a temperature of 500 K. (2) Sketch the thermal equilibrium band edge diagram of the *pn* junction for both temperatures. (3) If the junction in this question were required to absorb light, at what wavelength would it begin to absorb strongly?
4. *pn diode storage time*. Compare the accuracy of Eqs. (2.61) and (2.62b) to the full solution of the continuity equation given by Eq. (2.62a), for $I_\mathrm{R}/I_\mathrm{F}$ ranging from 0.01 to 100.

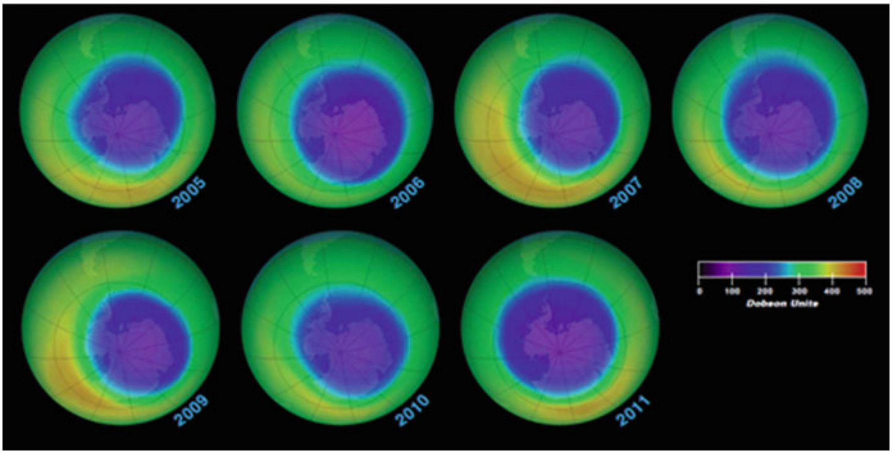
[83] This problem may be somewhat difficult and/or lengthy.

5. *Metal–semiconductor band edge diagrams*. An ideal metal–semiconductor junction is formed between platinum (work function 5.3 eV) and p-type silicon. What type of contact results?

Microwave Limb Sounder onboard the Aura Satellite

The MLS (located near the foreground in the image below) observes microwave emission from gas molecules (e.g., O_3, H_2O, CO, SO_2) from 118 GHz to 2.5 THz using GaAs-based Schottky diodes (solar panels based on silicon *pn* junctions power the satellite as well). The data shown is based on annual MLS Earth ozone concentration measurements taken over the south pole.

Source: JPL/NASA


2005
2006
2007
2008
2009
2010
2011
0
100
200
300
400
500
Dobson Units

http://www.springer.com/978-1-4614-8835-4

Solid-State Electronic Devices
An Introduction
Papadopoulos, C.
2014, XIII, 277 p. 151 illus., 45 illus. in color., Softcover
ISBN: 978-1-4614-8835-4