4

PN and Metal–Semiconductor Junctions

CHAPTER OBJECTIVES

This chapter introduces several devices that are formed by joining two different materials together. PN junction and metal–semiconductor junction are analyzed in the forward-bias and reverse-bias conditions. Of particular importance are the concepts of the depletion region and minority carrier injection. Solar cells and light-emitting diode are presented in some detail because of their rising importance for renewable energy generation and for energy conservation through solid-state lighting, respectively. The metal–semiconductor junction can be a rectifying junction or an ohmic contact. The latter is of growing importance to the design of high-performance transistors.

PART I: PN JUNCTION

As illustrated in Fig. 4–1, a PN junction can be fabricated by implanting or diffusing (see Section 3.5) donors into a P-type substrate such that a layer of semiconductor is converted into N type. Converting a layer of an N-type semiconductor into P type with acceptors would also create a PN junction.

A PN junction has rectifying current–voltage (*I–V* or *IV*) characteristics as shown in Fig. 4–2. As a device, it is called a **rectifier** or a **diode**. The PN junction is the basic structure of solar cell, light-emitting diode, and diode laser, and is present in all types of transistors. *In addition, PN junction is a vehicle for studying the theory*

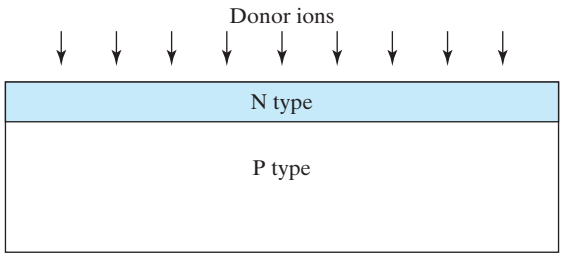


FIGURE 4–1 A PN junction can be fabricated by converting a layer of P-type semiconductor into N-type with donor implantation or diffusion.

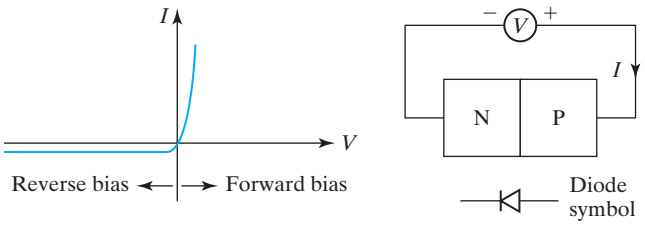


FIGURE 4–2 The rectifying *IV* characteristics of a PN junction.

of the depletion layer, the quasi-equilibrium boundary condition, the continuity equation, and other tools and concepts that are important to the understanding of transistors.

4.1 • BUILDING BLOCKS OF THE PN JUNCTION THEORY •

For simplicity, it is usually assumed that the P and N layers are uniformly doped at acceptor density N_a, and donor density N_d, respectively.[1] This idealized PN junction is known as a **step junction** or an **abrupt junction**.

4.1.1 Energy Band Diagram and Depletion Layer of a PN Junction

Let us construct a rough energy band diagram for a PN junction at equilibrium or zero bias voltage. We first draw a horizontal line for E_F in Fig. 4–3a because there is only one Fermi level at equilibrium (see Sec. 1.7.2). Figure 4–3b shows that far from the junction, we simply have an N-type semiconductor on one side (with E_c close to E_F), and a P-type semiconductor on the other side (with E_v close to E_F). Finally, in Fig. 4–3c we draw an arbitrary (for now) smooth curve to link the E_c from the N layer to the P layer. E_v of course follows E_c, being below E_c by a constant E_g.

QUESTION • *Can you tell which region (P or N) in Fig. 4–3 is more heavily doped? (If you need a review, see Section 1.8.2).*

Figure 4–3d shows that a PN junction can be divided into three layers: the neutral N layer, the neutral P layer, and a **depletion layer** in the middle. In the middle layer, E_F is close to neither E_v nor E_c. Therefore, both the electron and hole concentrations are quite small. For mathematical simplicity, it is assumed that

$$\boxed{n \approx 0 \text{ and } p \approx 0 \qquad \text{in the depletion layer}} \tag{4.1.1}$$

The term *depletion layer* means that the layer is depleted of electrons and holes.

[1] N_d and N_a are usually understood to represent the *compensated* (see end of Section 1.9), or the net, dopant densities. For example, in the N-type layer, there may be significant donor *and* acceptor concentrations, and N_d is the former minus the latter.

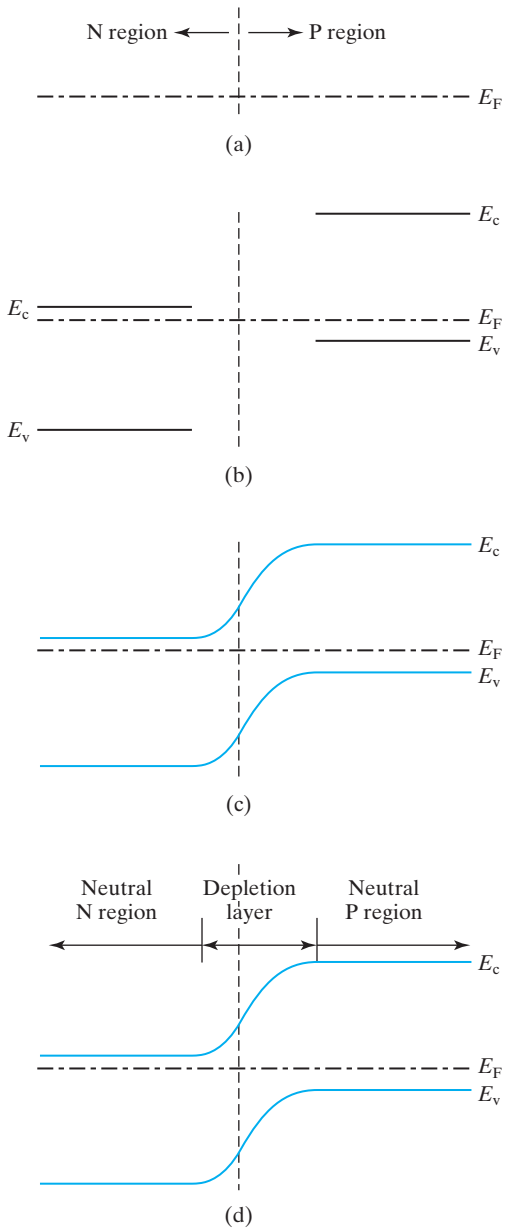


FIGURE 4–3 (a) and (b) Intermediate steps of constructing the energy band diagram of a PN junction. (c) and (d) The complete band diagram.

4.1.2 Built-In Potential

Let us examine the band diagram of a PN junction in Fig. 4–4 in greater detail. Figure 4–4b shows that E_c and E_v are not flat. This indicates the presence of a voltage differential. The voltage differential, ϕ_{bi}, is called the **built-in potential**. A built-in potential is present at the interface of any two dissimilar materials. We are usually

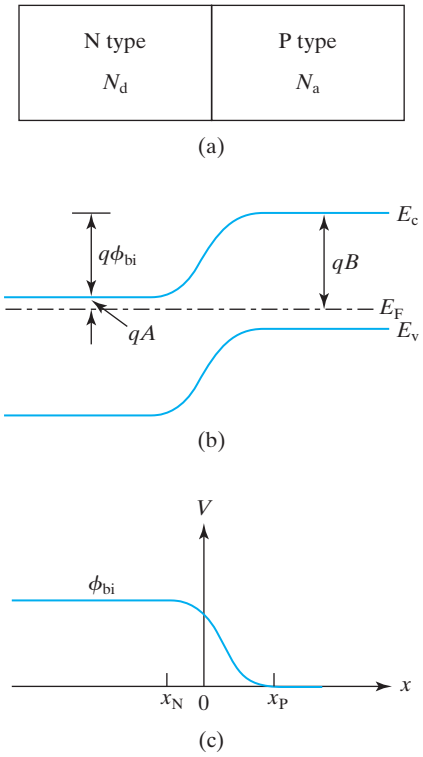


FIGURE 4–4 (a) A PN junction. The built-in potential in the energy band diagram (b) shows up as an upside down mirror image in the potential plot (c).

unaware of them because they are difficult to detect directly. For example, if one tries to measure the built-in potential, ϕ_{bi}, by connecting the PN junction to a voltmeter, no voltage will be registered because the net sum of the built-in potentials at the PN junction, the semiconductor–metal contacts, the metal to wire contacts, etc., in any closed loop is zero (see the sidebar, "Hot-Point Probe, Thermoelectric Generator and Cooler," in Sec. 2.1). However, the built-in voltage and field are as real as the voltage and field that one may apply by connecting a battery to a bar of semiconductor. For example, electrons and holes are accelerated by the built-in electric field exactly as was discussed in Chapter 2. Applying Eq. (1.8.5) to the N and P regions, one obtains

$$\text{N-region} \qquad n = N_d = N_c e^{-qA/kT} \Rightarrow A = \frac{kT}{q} \ln \frac{N_c}{N_d}$$

$$\text{P-region} \qquad n = \frac{n_i^2}{N_a} = N_c e^{-qB/kT} \Rightarrow B = \frac{kT}{q} \ln \frac{N_c N_a}{n_i^2}$$

$$\phi_{bi} = B - A = \frac{kT}{q}\left(\ln \frac{N_c N_a}{n_i^2} - \ln \frac{N_c}{N_d} \right)$$

$$\phi_{bi} = \frac{kT}{q}\ln\frac{N_d N_a}{n_i^2} \tag{4.1.2}$$

The built-in potential is determined by N_a and N_d through Eq. (4.1.2). The larger the N_a or N_d is, the larger the ϕ_{bi} is. Typically, ϕ_{bi} is about 0.9 V for a silicon PN junction.

Since a lower E_c means a higher voltage (see Section 2.4), the N side is at a higher voltage or electrical potential than the P side. This is illustrated in Fig. 4–4c, which arbitrarily picks the neutral P region as the voltage reference. In the next section, we will derive $V(x)$ and $E_c(x)$.

4.1.3 Poisson's Equation

Poisson's equation is useful for finding the electric potential distribution when the charge density is known. In case you are not familiar with the equation, it will be derived from Gauss's Law here. Applying Gauss's Law to the volume shown in Fig. 4–5, we obtain

$$\varepsilon_s \mathscr{E}(x+\Delta x)A - \varepsilon_s \mathscr{E}(x)A = \rho \Delta x A \tag{4.1.3}$$

where ε_s is the semiconductor permittivity and, for silicon, is equal to 12 times the permittivity of free space. ρ is the charge density (C/cm^3) and $\mathscr{E}$ is the electric field.

$$\frac{\mathscr{E}(x+\Delta x) - \mathscr{E}(x)}{\Delta x} = \frac{\rho}{\varepsilon_s} \tag{4.1.4}$$

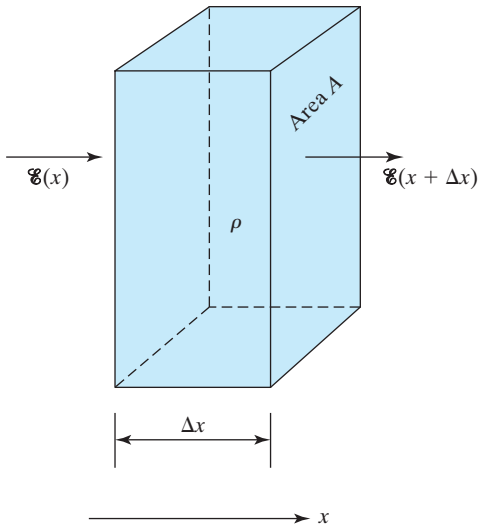


FIGURE 4–5 A small volume in a semiconductor, used to derive the Poisson's equation.

Taking the limit of $\Delta x \to 0$,

$$\frac{d\mathscr{E}}{dx} = \frac{\rho}{\varepsilon_s} \tag{4.1.5}$$

$$\boxed{\frac{d^2V}{dx^2} = -\frac{d\mathscr{E}}{dx} = -\frac{\rho}{\varepsilon_s}} \tag{4.1.6}$$

Equation (4.1.5) or its equivalent, Eq. (4.1.6), is *Poisson's equation*. It will be the starting point of the next section.

4.2 • DEPLETION-LAYER MODEL •

We will now solve Eq. (4.1.5) for the step junction shown in Fig. 4–6. Let's divide the PN junction into three regions—the neutral regions at $x > x_P$ and $x < -x_N$, and the **depletion layer** or **depletion region** in between, where p = n = 0 as shown in Fig. 4–6b. The charge density is zero everywhere except in the depletion layer where it takes the value of the dopant ion charge density as shown in Fig. 4–6c.

4.2.1 Field and Potential in the Depletion Layer

On the P side of the depletion layer ($0 \le x \le x_P$)

$$\rho = -qN_a \tag{4.2.1}$$

Eq. (4.1.5) becomes

$$\frac{d\mathscr{E}}{dx} = -\frac{qN_a}{\varepsilon_s} \tag{4.2.2}$$

Equation (4.2.2) may be integrated once to yield

$$\mathscr{E}(x) = -\frac{qN_a}{\varepsilon_s}x + C_1 = \frac{qN_a}{\varepsilon_s}(x_P - x) \qquad 0 \le x \le x_P \tag{4.2.3}$$

C_1 is a constant of integration and is determined with the boundary condition $\mathscr{E} = 0$ at $x = x_P$. You may verify that Eq. (4.2.3) satisfies this boundary condition. The field increases linearly with x, having its maximum magnitude at $x = 0$ (see Fig. 4–6d).

On the N-side of the depletion layer, the field is similarly found to be

$$\mathscr{E}(x) = -\frac{qN_d}{\varepsilon_s}(x - x_N) \qquad x_N \le x \le 0 \tag{4.2.4}$$

x_N is a negative number. The field must be continuous, and equating Eq. (4.2.3) and Eq. (4.2.4) at $x = 0$ yields

$$\boxed{N_a|x_P| = N_d|x_N|} \tag{4.2.5}$$

$|x_N|$ and $|x_P|$ are the widths of the depletion layers on the two sides of the junction. They are inversely proportional to the dopant concentration; the more heavily doped side holds a smaller portion of the depletion layer. PN junctions are usually highly asymmetrical in doping concentration. A highly asymmetrical junction

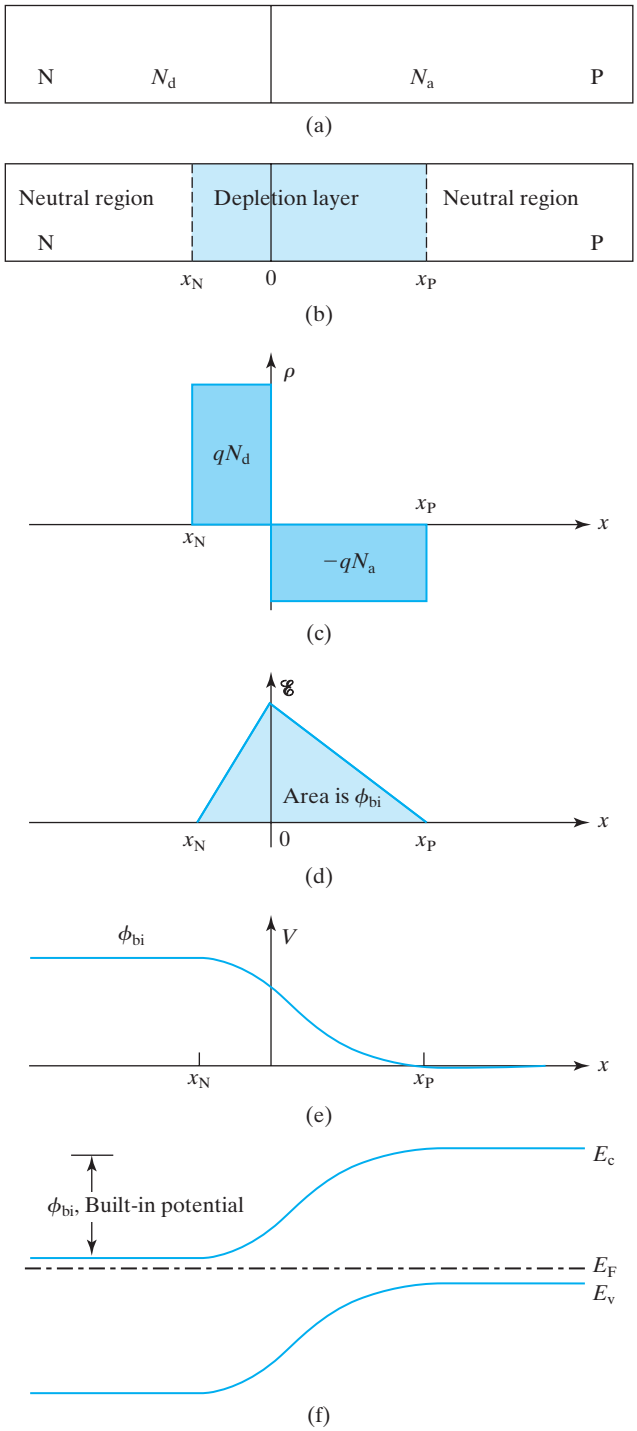


FIGURE 4–6 (a) Step PN junction; (b) depletion approximation; (c) space charge profile; (d) electric field from integration of ρ/ε_s (Poisson's equation); (e) electric potential from integrating $-\mathscr{E}$; and (f) energy band diagram.

is called a **one-sided junction,** either an **N^+P junction** or a **P^+N junction**, where N^+ and P^+ denote the heavily doped sides. *The depletion layer penetrates primarily into the lighter doping side, and the width of the depletion layer in the heavily doped material can often be neglected*. It may be helpful to think that a heavily doped semiconductor is similar to metal (and there is no depletion layer in metal).

Equation (4.2.5) tells us that the area density of the negative charge, $N_a|x_P|$ (C/cm^2), and that of the positive charge, $N_d|x_N|$ (C/cm^2), are equal (i.e., the net charge in the depletion layer is zero). In other words, the two rectangles in Fig. 4–6c are of equal size.

Using $\mathscr{E} = -dV/dx$, and integrating Eq. (4.2.3) yields

$$V(x) = -\frac{qN_a}{2\varepsilon_s}(x_P - x)^2 \qquad 0 \le x \le x_P \tag{4.2.6}$$

We arbitrarily choose the voltage at $x = x_P$ as the reference point for $V = 0$. Similarly, on the N-side, we integrate Eq. (4.2.4) once more to obtain

$$\begin{aligned} V(x) &= D - \frac{qN_a}{2\varepsilon_s}(x - x_N)^2 \\ &= \phi_{bi} - \frac{qN_d}{2\varepsilon_s}(x - x_N)^2 \qquad x_N \le x \le 0 \end{aligned} \tag{4.2.7}$$

where D is determined by $V(x_N) = \phi_{bi}$ (see Fig. 4–6e and Eq. (4.1.2)). $V(x)$ is plotted in Fig. 4–6e. The curve consists of two parabolas (Eqs. (4.2.6) and (4.2.7)). Finally, we can quantitatively draw the energy band diagram, Fig. 4–6f. $E_c(x)$ and $E_v(x)$ are identical to $V(x)$, but inverted as explained in Section 2.4.

4.2.2 Depletion-Layer Width

Equating Eqs. (4.2.6) and (4.2.7) at $x = 0$ (because V is continuous at $x = 0$), and using Eq. (4.2.5), we obtain

$$x_P - x_N = W_{dep} = \sqrt{\frac{2\varepsilon_s\phi_{bi}}{q}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)} \tag{4.2.8}$$

$x_N + x_P$ is the total **depletion-layer width**, represented by W_{dep}.

If $N_a >> N_d$, as in a P^+N junction,

$$\boxed{W_{dep} \approx \sqrt{\frac{2\varepsilon_s\phi_{bi}}{qN_d}}} \approx |x_N| \tag{4.2.9}$$

If $N_d >> N_a$, as in an N^+P junction,

$$\boxed{W_{dep} \approx \sqrt{\frac{2\varepsilon_s\phi_{bi}}{qN_a}}} \approx |x_P| \tag{4.2.10}$$

$$|x_N| = |x_P|N_a/N_d \cong 0$$

EXAMPLE 4–1 A P^+N junction has $N_a = 10^{20}cm^3$ and $N_d = 10^{17}cm^{-3}$. What is (a) the built-in potential, (b) W_{dep}, (c) x_N, and (d) x_P?

SOLUTION:

a. Using Eq. (4.1.2),

$$\phi_{bi} = \frac{kT}{q}\ln\frac{N_d N_a}{n_i^2} \approx 0.026\ \text{V}\ln\frac{10^{20}\times 10^{17}\text{cm}^{-6}}{10^{20}\text{cm}^{-6}} \approx 1\ \text{V}$$

b. Using Eq. (4.2.9),

$$W_{dep} \approx \sqrt{\frac{2\varepsilon_s\phi_{bi}}{qN_d}} = \left(\frac{2\times 12\times 8.85\times 10^{-14}\times 1}{1.6\times 10^{-19}\times 10^{17}}\right)^{1/2}$$

$$= 1.2\times 10^{-5}\text{cm} = 0.12\ \mu\text{m} = 120\ \text{nm} = 1200\ \text{Å}$$

c. In a P^+N junction, nearly the entire depletion layer exists on the N-side.

$$|x_N| \approx W_{dep} = 0.12\,\mu\text{m}$$

d. Using Eq. (4.2.5),

$$|x_P| = |x_N|N_d/N_a = 0.12\ \mu\text{m}\times 10^{17}\text{cm}^{-3}/10^{20}\text{cm}^{-3} = 1.2\times 10^{-4}\ \mu\text{m}$$
$$= 1.2\ \text{Å} \approx 0$$

The point is that the heavily doped side is often hardly depleted at all. *It is useful to remember that $W_{dep} \approx 0.1\ \mu m$ for $N = 10^{17}\ cm^{-3}$.* For more examples of the PN junction, see http://jas.eng.buffalo.edu/education/pn/pnformation2/pnformation2.html.

From Eqs. (4.2.9) and (4.2.10), we learn that *the depletion-layer width is determined by the lighter doping concentration.* Those two equations can be combined into

$$W_{dep} = \sqrt{2\varepsilon_s\phi_{bi}/qN} \tag{4.2.11}$$

$$1/N = 1/N_d + 1/N_a \approx 1/\textit{lighter dopant density} \tag{4.2.12}$$

4.3 • REVERSE-BIASED PN JUNCTION •

When a positive voltage is applied to the N region relative to the P region, the PN junction is said to be **reverse-biased**. The zero-biased and reverse-biased PN junction energy diagrams are shown in Fig. 4–7. Under reverse bias, there is very little current since the bias polarity allows the flow of electrons from the P side to the N side and holes from the N side to the P side, but there are few electrons (minority carriers) on the P side and few holes on the N side. Therefore, the current is negligibly small. Since the current is small, the IR drop in the neutral regions is also negligible. All the reverse-bias voltage appears across the depletion layer. The potential barrier increases from $q\phi_{bi}$ in Fig. 4–7b to $q\phi_{bi} + qV_r$ in Fig. 4–7c.

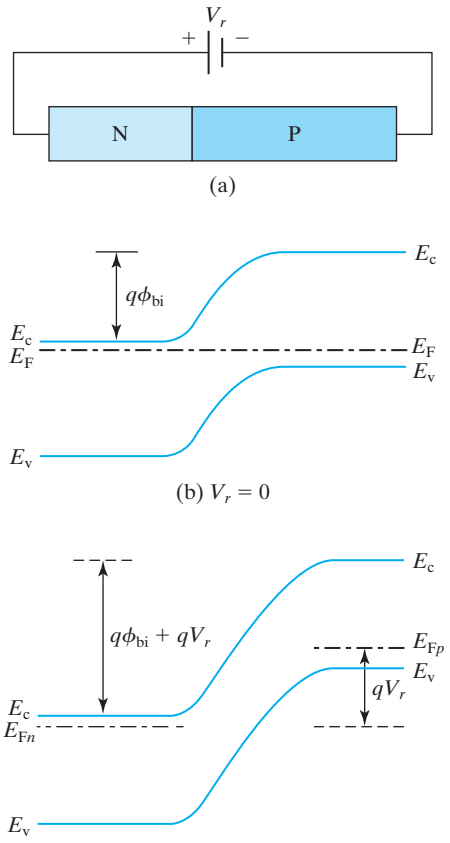


FIGURE 4–7 Reverse-biased PN junction (a) polarity of reverse bias; (b) energy band diagram without bias; and (c) energy band diagram under reverse bias.

The equations derived in the previous section for $V_r = 0$ are also valid under reverse bias if the ϕ_{bi} term is replaced with $\phi_{bi} + V_r$. The depletion layer width becomes

$$W_{dep} = \sqrt{\frac{2\varepsilon_s(\phi_{bi} + V_r)}{qN}} = \sqrt{\frac{2\varepsilon_s \times \text{potential barrier}}{qN}} \tag{4.3.1}$$

The depletion layer widens as the junction is more reverse biased. Under reverse bias, the depletion layer needs to widen in order to dissipate the larger voltage drop across it.

4.4 • CAPACITANCE-VOLTAGE CHARACTERISTICS •

The depletion layer and the neutral N and P regions in Fig. 4–8 may be viewed as an insulator and two conductors. Therefore, the PN junction may be modeled as a parallel-plate capacitor with capacitance

$$C_{dep} = A\frac{\varepsilon_s}{W_{dep}} \tag{4.4.1}$$

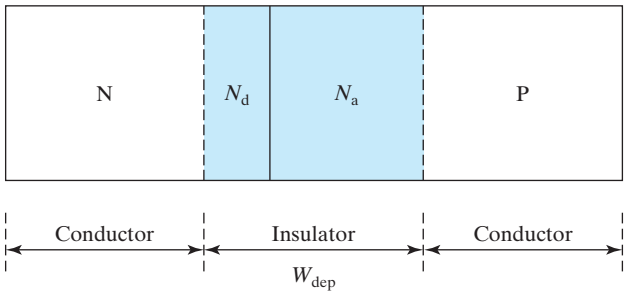


FIGURE 4–8 The PN junction as a parallel-plate capacitor.

where C_{dep} is the **depletion-layer capacitance** and A is the area. PN junction is prevalent in semiconductor devices and its capacitance is an unwelcome capacitive load to the devices and the circuits. C_{dep} can be lowered by reducing the junction area and increasing W_{dep} by reducing the doping concentration(s) and/or applying a reverse bias. *Numerically, $C \approx 1fF/\mu m^2$ when $W_{dep} = 0.1\mu m$.*

Using Eq. (4.4.1) together with Eq. (4.3.1), we obtain

$$\frac{1}{C_{dep}^2} = \frac{W_{dep}^2}{A^2 \varepsilon_s^2} = \frac{2(\phi_{bi} + V_r)}{qN\varepsilon_s A^2} \tag{4.4.2}$$

Equation (4.4.2) suggests a linear relationship between $1/C_{dep}{}^2$ and V_r. Figure 4–9 illustrates the most common way of plotting the C–V data of a PN junction. From the slope of the line in this figure, one can determine N (or the lighter dopant concentration of a one-sided junction; see Eq. (4.2.12)). From the intercept with the horizontal axis, one can determine the built-in potential, ϕ_{bi}.

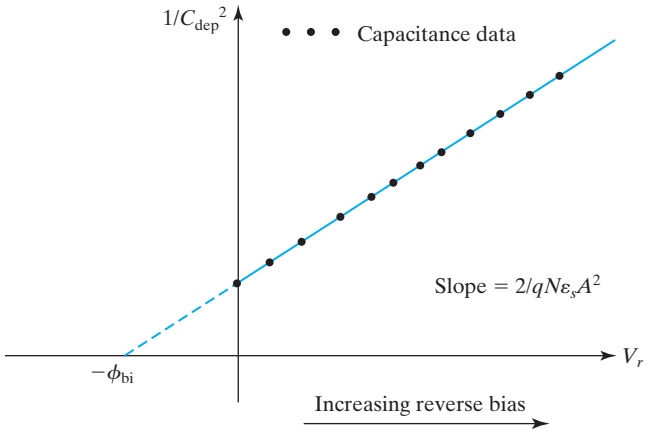


FIGURE 4–9 The common way of plotting the C–V data of a PN junction.

EXAMPLE 4–2 The slope of the line in Fig. 4–9 is 2×10^{23} F^{-2} V^{-1}, the intercept is 0.84 V, and the area of the PN junction is 1 μm^2. Find the lighter doping concentration, N_l, and the heavier doping concentration, N_h.

SOLUTION:

Using Eq. (4.4.2), the lighter doping concentration is

$$\begin{aligned} N_l &= 2/(\text{slope} \times q\varepsilon_s A^2) \\ &= 2/(2 \times 10^{23} \times 1.6 \times 10^{-19} \times 12 \times 8.85 \times 10^{-14} \times 10^{-8}\text{cm}^2) \\ &= 6 \times 10^{15}\text{cm}^{-3} \end{aligned}$$

(There is no way to determine whether the lightly doped side is the N side or the P side.) Now, using Eq. (4.1.2),

$$\phi_{bi} = \frac{kT}{q}\ln\frac{N_h N_l}{n_i^2}$$

$$N_h = \frac{n_i^2}{N_l}e^{q\phi_{bi}/kT} = \frac{10^{20}\text{cm}^{-6}}{6 \times 10^{15}\text{cm}^{-3}}e^{0.84/0.026} = 1.8 \times 10^{18}\text{cm}^{-3}$$

This example presents an accurate way to determine N_l, less so for N_h. If the intercept data has a small experimental error of 60 mV and the correct value is 0.78 V, Eq. (4.1.2) would have yielded $N_h = 1.8 \times 10^{17}$ cm^{-3}. *We should be aware of when a conclusion, though correct, may be sensitive to even small errors in the data.*

4.5 • JUNCTION BREAKDOWN[2] •

We have stated that a reverse-biased PN junction conducts negligibly small current. This is true until a critical reverse bias is reached and **junction breakdown** occurs as shown in Fig. 4–10a. There is nothing inherently destructive about junction breakdown. If the current is limited to a reasonable value by the external circuit so that heat dissipation in the PN junction is not excessive, the PN junction can be operated in reverse breakdown safely. A **Zener diode** is a PN junction diode designed to operate in the breakdown mode with a breakdown voltage that is tightly controlled by the manufacturer. A Zener protection circuit is shown in Fig. 4–10b. If the breakdown voltage of the Zener diode is 3.7 V, the maximum voltage that can appear across the integrated circuit (IC) leads C and D would be 3.7 V (even in the presence of a surge voltage of, say, 50 V, across the lines A and B). The resistance R is chosen to limit the current to a level safe for the Zener diode. Figure 4–10b can also represent a rudimentary voltage-reference circuit. In that case, a voltage supply (battery) with $V >$ 3.7 V is connected between A and B. The voltage that appears at C and D will be maintained at 3.7 V, within a tight range specified by the manufacturer, even if the battery voltage fluctuates with usage and temperature.

[2] This section may be omitted in an accelerated course.

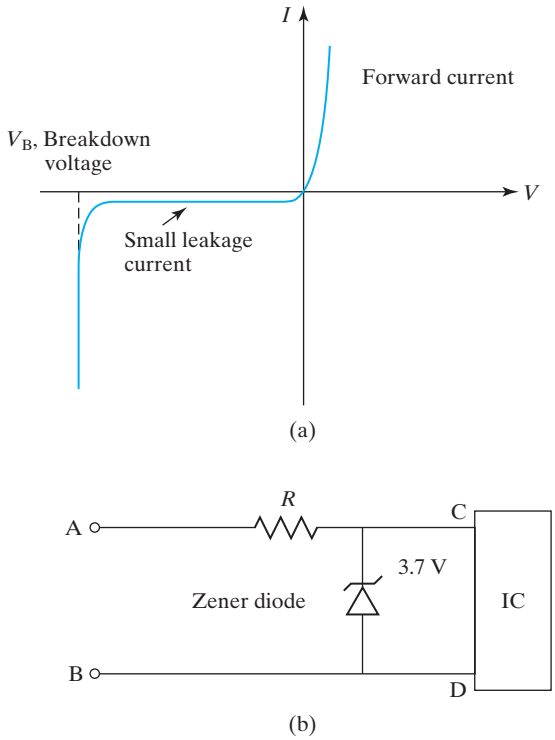


FIGURE 4–10 Reverse breakdown in a PN junction. (a) *IV* characteristics; (b) a Zener protection circuit or voltage-reference circuit.

4.5.1 Peak Electric Field

Junction breakdown occurs when the **peak electric field** in the PN junction reaches a critical value. Consider the N^+P junction shown in Fig. 4–11. Employing Eqs. (4.2.4) and (4.3.1) and evaluating the peak electric field at $x = 0$, we obtain

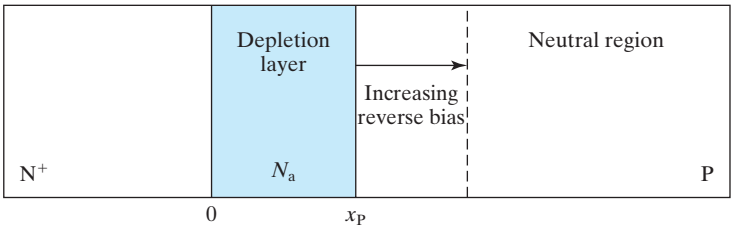
$$\mathscr{E}_p = \mathscr{E}(0) = \left[\frac{2qN}{\varepsilon_s}(\phi_{\text{bi}} + |V_r|)\right]^{1/2} \tag{4.5.1}$$

As the reverse bias voltage V increases in Fig. 4–11, $\mathscr{E}_p$ increases with it. When $\mathscr{E}_p$ reaches some critical value, $\mathscr{E}_{\text{crit}}$, breakdown occurs. Equating $\mathscr{E}_p$ in Eq. (4.5.1) with $\mathscr{E}_{\text{crit}}$ allows us to express the breakdown voltage in terms of the doping concentration. Remember, N is the lighter dopant density in a one-sided junction. In general, N is an average of N_a and N_d [see Eq. (4.2.12)].

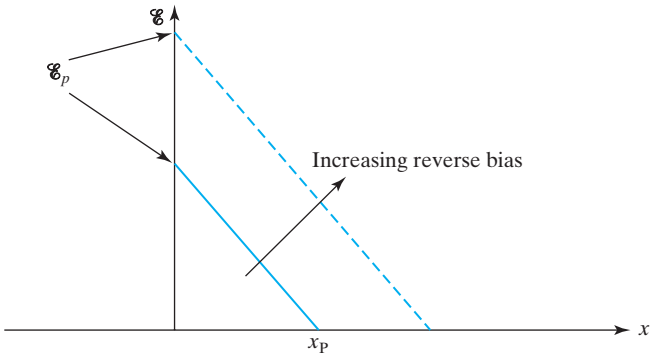
$$V_B = \frac{\varepsilon_s \mathscr{E}_{\text{crit}}^2}{2qN} - \phi_{\text{bi}} \tag{4.5.2}$$

4.5.2 Tunneling Breakdown

When a heavily doped junction is reverse biased, as shown in Fig. 4–12, only a small distance separates the large number of electrons in the P-side valence band and the empty states in the N-side conduction band. Therefore, tunneling of electrons can



(a)



(b)

FIGURE 4–11 The field distribution in a one-side PN junction. (a) N^+ P junction with $x_N \approx 0$ and (b) electric field profile.

occur (see Fig. 4–12b). The tunneling current density has an exponential dependence on $1/\mathscr{E}$ [1]:

$$J = G e^{-H/\mathscr{E}_p} \tag{4.5.3}$$

where G and H are constants for a given semiconductor. The IV characteristics are shown in Fig. 4–12c. This is known as **tunneling breakdown.** The critical electric field for tunneling breakdown is proportional to H, which is proportion to the 3/2 power of E_g and 1/2 power of the effective mass of the tunneling carrier. The critical field is about 10^6 V/cm for Si. V_B is given in Eq. (4.5.2). Tunneling is the dominant breakdown mechanism when N is very high and V_B is quite low (below a few volts). Avalanche breakdown, presented in the next section, is the mechanism of diode breakdown at higher V_B.

4.5.3 Avalanche Breakdown

With increasing electric field, electrons traversing the depletion layer gain higher and higher kinetic energy. Some of them will have enough energy to raise an electron from the valence band into the conduction band, thereby creating an electron–hole pair as shown in Fig. 4–13. This phenomenon is called **impact ionization**. The electrons and holes created by impact ionization are themselves also accelerated by the electric field. Consequently, they and the original carrier can create even more carriers by impact ionization. The result is similar to a snow

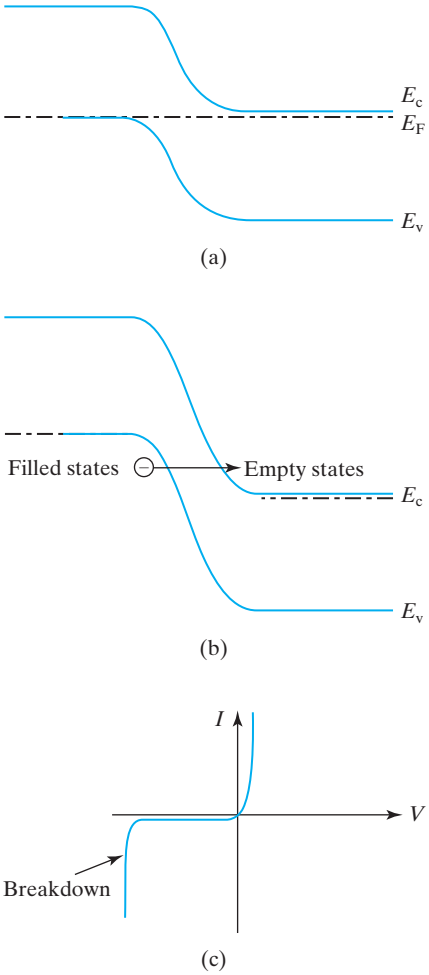


FIGURE 4–12 Tunneling breakdown. (a) Heavily doped junction at zero bias; (b) reverse bias with electron tunneling from valence band to conduction band; and (c) *IV* characteristics.

avalanche on a mountainside. (Furthermore, holes accelerate to the left and generate electrons *upstream*, thus providing positive feedback.) When $\mathscr{E}_p$ reaches $\mathscr{E}_{\text{crit}}$, the carrier creation rate and the reverse current rise abruptly. This is called **avalanche breakdown.** If the ϕ_{bi} term in Eq. (4.5.2) is ignored,

$$V_{\text{B}} = \frac{\varepsilon_s \mathscr{E}_{\text{crit}}^{\,2}}{2qN} \tag{4.5.4}$$

$\mathscr{E}_{\text{crit}}$ is about 5×10^5 V/cm at $N = 10^{17}\text{cm}^{-3}$, and is approximately proportional to $N^{0.2}$ [2]. Therefore Eq. (4.5.2) may be reduced to

$$V_{\text{B}}(V) \approx 15 \times \left(\frac{10^{17}}{N}\right)^{0.6} \propto 1/N^{0.6} \tag{4.5.5}$$

V_{B} is about 15 V at $N = 10^{17}\text{cm}^{-3}$.

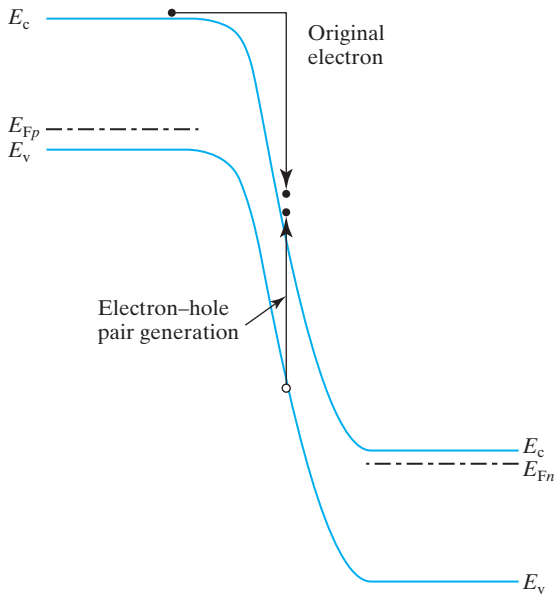


FIGURE 4–13 Electron–hole pair generation by impact ionization. The incoming electron gives up its kinetic energy to generate an electron–hole pair.

In general, it is necessary and effective to reduce the junction doping concentration(s) if a larger breakdown voltage is desired. The energy that a carrier must possess in order to initiate impact ionization increases (and therefore, $\mathscr{E}_{crit}$ also increases) with increasing band gap, E_g. Thus, the breakdown voltage, for given N_a and N_d, progressively increases from Ge to Si to GaAs diodes, due to increasing E_g.

• Applications of High-Voltage Devices •

While IC devices typically have junction breakdown voltages under 20 V, silicon **power devices** can operate at 100–1,000 V because their junction breakdown voltages can be that high. They are used to control gasoline-electric hybrid cars, diesel-electric trains, urban subway trains, and industrial processes. They are even used in the **HVDC** or **high-voltage DC** utility power transmission systems where giga-watt power may be converted from AC to DC at the generator end (e.g., in Northwestern U.S.), transmitted as HVDC power, and converted back to AC for voltage down-transformation and distribution at the user-market end (e.g., Southern California). Compared to AC transmission, HVDC improves the power grid stability and reduces power transmission loss.

QUESTION: Estimate the order of magnitude of the doping density, N, required to achieve a 1,000 V junction breakdown voltage.

SOLUTION: Using Eq. (4.5.5), we estimate that N should be 10^{14} cm^{-3} or smaller.

4.6 • CARRIER INJECTION UNDER FORWARD BIAS— QUASI-EQUILIBRIUM BOUNDARY CONDITION •

Let us now examine the PN junction under forward bias. As shown in Fig. 4–14, a forward bias of V reduces the barrier height from ϕ_{bi} to $\phi_{bi} - V$. This reduces the drift field and upsets the balance between diffusion and drift that exists at zero bias. Electrons can now diffuse from the N side into the P side. This is called **minority-carrier injection**. Similarly, holes are *injected from the P side into the N side*. Figure 4–15 presents a way of visualizing carrier injection. As the barrier is reduced, a larger portion of the "Boltzmann tail" of the electrons on the N side can move into the P side (see Fig. 1–20). More electrons are now present at x_P and more holes appear at $-x_N$ than when the barrier is higher.

On the N side, $E_c - E_{Fn}$ is of course determined by N_d [Eq. (2.8.1)]. Let us assume that E_{Fn} remains constant through x_P because the depletion layer is narrow.[3] Therefore, at the edge of the neutral P region,

$$\begin{aligned} n(x_P) &= N_c e^{-(E_c - E_{Fn})/kT} = N_c e^{-(E_c - E_{Fp})/kT} e^{(E_{Fn} - E_{Fp})/kT} \\ &= n_{P0} e^{(E_{Fn} - E_{Fp})/kT} = n_{P0} e^{qV/kT} \end{aligned} \tag{4.6.1}$$

n_{P0} is the equilibrium (denoted by subscript 0) electron concentration of the P region (denoted by the subscript P), simply n_i^2/N_a. *The minority carrier density has been raised by* $e^{qV/kT}$.

A similar equation may be derived for $p(x_N)$.

$$\boxed{\begin{aligned} n(x_P) &= n_{P0} e^{qV/kT} = \frac{n_i^2}{N_a} e^{qV/kT} \\ p(x_N) &= p_{N0} e^{qV/kT} = \frac{n_i^2}{N_d} e^{qV/kT} \end{aligned}} \tag{4.6.2}$$

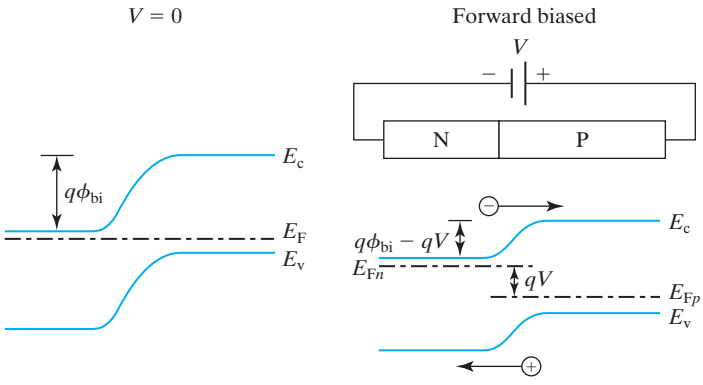


FIGURE 4–14 A forward bias reduces the junction barrier to ϕ_{bi}–V and allows electrons and holes to be injected over the reduced barrier.

[3] It can also be shown that dE_{Fn}/dx is very small in the depletion layer [1].

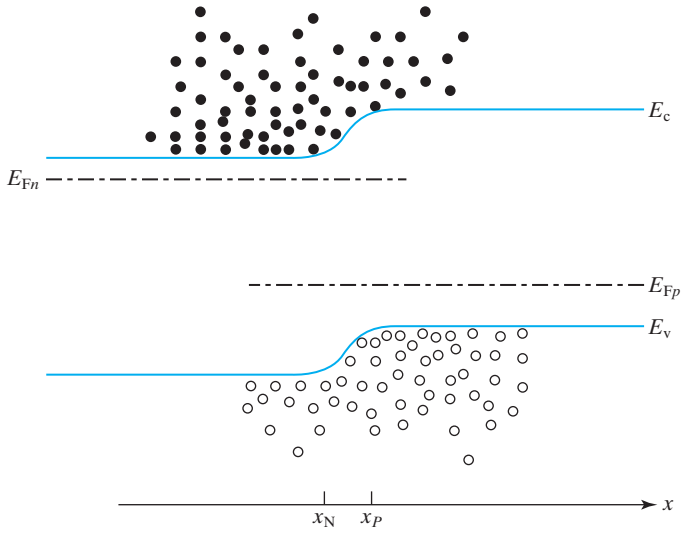


FIGURE 4–15 n at x_P (electron density at the edge of the neutral P region) is determined by $E_c - E_{Fn}$. Similarly, p at x_N is determined by $E_v - E_{Fp}$.

Equation (4.6.2) is called the **quasi-equilibrium boundary condition** or the **Shockley boundary condition**. It states that *a forward bias, V, raises the minority carrier densities at the edges of the depletion layer by the factor* $e^{qV/kT}$. This factor is 10^{10} for a moderate forward bias of 0.6 V. The derivation of Eq. (4.6.2) is also valid for a reverse bias, i.e., V can be either positive (forward bias) or negative (reverse bias). When V is a large negative number, $n(x_P)$ and $p(x_N)$ become essentially zero. This situation is sometimes called **minority carrier extraction** as opposed to carrier injection.

Another version of the quasi-equilibrium boundary condition expresses the **excess minority carrier** concentrations.

$$\begin{aligned} n'(x_P) &\equiv n(x_P) - n_{P0} = n_{P0}(e^{qV/kT} - 1) \\ p'(x_N) &\equiv p(x_N) - p_{N0} = p_{N0}(e^{qV/kT} - 1) \end{aligned} \tag{4.6.3}$$

Commit Eq. (4.6.3) to memory. The next two sections will analyze what happens in the neutral regions.

EXAMPLE 4–3 Carrier Injection

A PN junction has $N_a = 10^{19}$ cm^{-3} and $N_d = 10^{16}$ cm^{-3}. (a) With V = 0, what are the minority carrier densities at the depletion region edges? Assume $V = 0.6$ V for (b)–(d). (b) What are the minority carrier densities at the depletion region edges? (c) What are the excess minority carrier densities? (d) What are the majority carrier densities? (e) Under the reverse bias of 1.8 V, what are the minority carrier concentrations at the depletion region edges?

a. *On the P-side* $\quad n_{P0} = n_i^2/N_a = 10^{20}/10^{19} = 10$ cm^{-3}

On the N-side $\quad p_{N0} = 10^{20}/10^{16} = 10^4$cm^{-3}

b. $n(x_P) = n_{P0}e^{qV/kT} = 10 \times e^{0.6/0.026} = 10^{11}\text{cm}^{-3}$

$p(x_N) = p_{N0}e^{qV/kT} = 10^4 \times e^{0.6/0.026} = 10^{14}\text{cm}^{-3}$

We see that a moderate forward bias can increase the minority carrier densities dramatically.

c. $n'(x_P) = n(x_P) - n_{P0} = 10^{11} - 10 = 10^{11}\text{cm}^{-3}$

$p'(x_N) = p(x_N) - p_{N0} = 10^{14} - 10^4 = 10^{14}\text{cm}^{-3}$

Carrier injection into the heavily doped side is negligible when compared with injection into the lightly doped side.

d. $n' = p'$ due to charge neutrality [Eq. (2.6.2)]

On the P-side: $p(x_P) = N_a + p' = N_a + n'(x_P) = 10^{19} + 10^{11} = 10^{19}\text{cm}^{-3}$

On the N-side: $n(x_N) = N_d + n' = N_d + p'(x_N) = 10^{16} + 10^{14}$

$= 1.01 \times 10^{16}\text{cm}^{-3}$

e. $n(x_P) = n_{P0}e^{qV/kT} = 10 \times e^{-1.8/0.026} = 10^{-29}\text{cm}^{-3}$

$p(x_N) = p_{N0}e^{qV/kT} = 10^4 \times e^{-1.8/0.026} = 10^{-26}\text{cm}^{-3}$

These are meaninglessly small concentrations. We conclude that $n = p = 0$ at the junction edge under reverse bias.

4.7 • CURRENT CONTINUITY EQUATION •

In the interiors of the neutral N and P regions the minority carrier densities will be determined by the equation developed in this section. Consider the box shown in Fig. 4–16. $A \cdot \Delta x \cdot p$ is the number of holes in the box and $A \cdot J_p/q$ is the number of holes

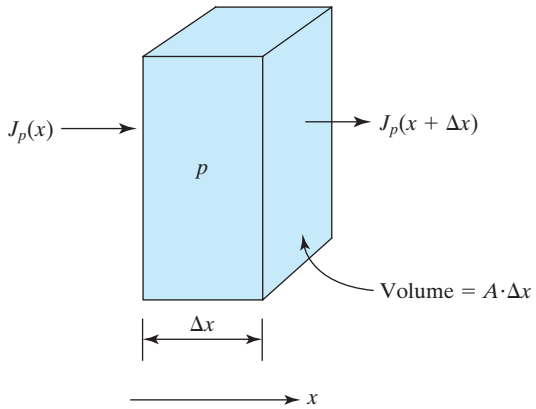


FIGURE 4–16 In steady state, the number of holes flowing into the box per second is equal to the number of holes flowing out per second plus the number of holes lost to recombination in the box per second.

flowing into the box per second. In steady state, the number of holes flowing into the box per second = number of holes flowing out of the box per second + number of holes recombining in the box per second.

$$A \cdot \frac{J_p(x)}{q} = A \cdot \frac{J_p(x+\Delta x)}{q} + A \cdot \Delta x \cdot \frac{p'}{\tau} \tag{4.7.1}$$

$$-\frac{J_p(x+\Delta x) - J_p(x)}{\Delta x} = q\frac{p'}{\tau} \tag{4.7.2}$$

Taking the limit of $\Delta x \to 0$,

$$-\frac{dJ_p}{dx} = q\frac{p'}{\tau} \tag{4.7.3}$$

τ is the recombination lifetime of the carriers. Equation (4.7.3) says that dJ_p/dx is zero only if there is no recombination. $J_p(x)$ must be larger than $J_p(x + \Delta x)$ in order to supply the holes lost to recombination in the box.

Equation (4.7.3) is valid for both the majority and minority carriers. However, it is particularly easy to apply it to the minority carriers. The minority carrier current (but not the majority carrier current) is dominated by diffusion and the drift component can be ignored. Appendix III at the end of this book provides a self-consistency proof. Let us apply Eq. (4.7.3) to the N side of the PN junction by substituting J_p, the minority carrier current, with the diffusion current [Eq. (2.3.3)].

$$qD_p\frac{d^2p}{dx^2} = q\frac{p'}{\tau_p} \tag{4.7.4}$$

A subscript p is given to τ to indicate the recombination lifetime in the N type semiconductor, in which the minority carriers are the holes.

$$\boxed{\frac{d^2p'}{dx^2} = \frac{p'}{D_p\tau_p} = \frac{p'}{L_p^2}} \tag{4.7.5}$$

$$\boxed{L_p \equiv \sqrt{D_p\tau_p}} \tag{4.7.6}$$

In Eq. (4.7.5), we replaced d^2p/dx^2 with d^2p'/dx. This assumes that the equilibrium hole concentration, p_0, is not a function of x. In other words, N_a is assumed to be uniform. (If N_a is not uniform, the mathematics becomes complicated but the result is qualitatively the same as those presented here.) Similarly, for electrons,

$$\boxed{\frac{d^2n'}{dx^2} = \frac{n'}{L_n^2}} \tag{4.7.7}$$

$$\boxed{L_n \equiv \sqrt{D_n\tau_n}} \tag{4.7.8}$$

L_n and L_p have the dimension of length. They are called the hole and electron **diffusion lengths**. They vary from a few μm to hundreds of μm depending on τ. Equations (4.7.6) and (4.7.8) are only valid for the minority carriers. Fortunately, we will not have to solve the continuity equations for the majority carriers.

4.8 • EXCESS CARRIERS IN FORWARD-BIASED PN JUNCTION •

In Fig. 4–17, we put the P side (positively biased side) on the left so that the current will flow in the positive x direction. The drawing shows x_P and x_N close to $x = 0$ because the length scale involved in this section is usually two orders of magnitude larger than W_{dep}. On the N side of the PN junction, we analyze the movement of the minority carriers (holes) by solving

$$\frac{d^2p'}{dx^2} = \frac{p'}{L_p^2} \tag{4.7.5}$$

for the boundary conditions

$$\begin{aligned} p'(\infty) &= 0 \\ p'(x_N) &= p_{N0}(e^{qV/kT} - 1) \end{aligned} \tag{4.6.3}$$

The general solution of Eq. (4.7.5) is

$$p'(x) = Ae^{x/L_p} + Be^{-x/L_p} \tag{4.8.1}$$

The first boundary condition demands $A = 0$. The second [Eq. (4.6.3)] determines B and leads to

$$p'(x) = p_{N0}(e^{qV/kT} - 1)e^{-(x-x_N)/L_p}, \quad x > x_N \tag{4.8.2}$$

Similarly, on the P side,

$$n'(x) = n_{P0}(e^{qV/kT} - 1)e^{(x-x_P)/L_n}, \quad x < x_P \tag{4.8.3}$$

Figure 4–18 presents sample minority carrier profiles. They are simple exponential functions of x. The characteristic lengths are the **diffusion lengths,** L_n *and* L_p. The carrier concentrations at the depletion-layer edges are determined by the quasi-equilibrium boundary conditions (i.e., by N_a or N_d and V). Carriers are

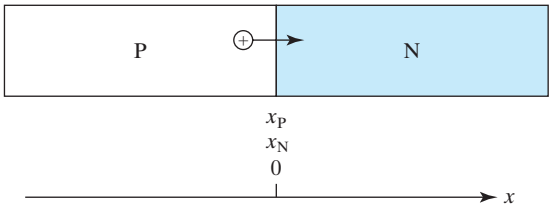


FIGURE 4–17 PN diode structure for analyzing the motion of holes after they are injected into the N side.

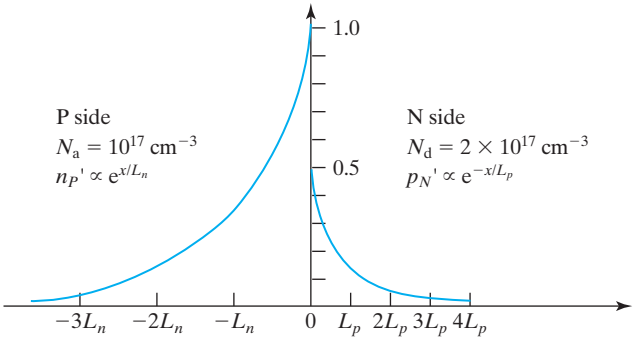


FIGURE 4–18 Normalized n' and p'. $n'(0) = 2p'(0)$ because $N_d = 2N_a$. $L_n = 2L_p$ is assumed.

mostly injected into the lighter doping side. From the depletion-layer edges, the injected minority carriers move outward by diffusion. As they diffuse, their densities are reduced due to recombination, thus creating the exponentially decaying density profiles. Beyond a few diffusion lengths from the junction, n' and p' have decayed to negligible values.

EXAMPLE 4–4 Minority and Majority Carrier Distribution and Quasi-Fermi Level

N-type	P-type
$N_d = 5 \times 10^{17}$ cm^{-3}	$N_a = 10^{17}$ cm^{-3}
$D_p = 12$ cm^2/s	$D_n = 36.4$ cm^2/s
$\tau_p = 1$ μs	$\tau_n = 2$ μs

A 0.6 V forward bias is applied to the diode. (a) What are the diffusion lengths on the N side and the P side? (b) What are the injected excess minority carrier concentrations at the junction edge? (c) What is the majority carrier profile on the N side? (d) Sketch the excess carrier densities, $p'(x)$ and $n'(x)$. (e) Sketch the energy diagram including $E_{Fp}(x)$ and $E_{Fn}(x)$, the quasi-Fermi levels. Note that $W_{dep} < 0.1$ μm may be assumed to be zero in the diagram.

SOLUTION:

a. On the N side,

$$L_p \equiv \sqrt{D_p \tau_p} = \sqrt{12\text{cm}^2/\text{s} \times 10^{-6}\text{s}} = \sqrt{12 \times 10^{-6}\text{cm}^2} = 3.5 \times 10^{-3}\text{cm} = 35\,\mu\text{m}$$

On the P side,

$$L_n \equiv \sqrt{D_n \tau_n} = \sqrt{36 \times 2 \times 10^{-6}} = 8.5 \times 10^{-3}\text{cm} = 85\,\mu\text{m}$$

Please note that these diffusion lengths are much larger than the dimensions of IC devices, which are less than 1 μm.

b. Using Eqs. (4.8.2) and (4.8.3),

$$\textit{Excess hole density on the N-side, } p'(x_N) = p_{N0}(e^{qV/kT} - 1)$$

$$= \frac{n_i^2}{N_d}(e^{qV/kT} - 1) = \frac{10^{20}}{5\times 10^{17}} e^{0.6/0.026} = 2\times 10^{12}\,\text{cm}^{-3}$$

$$\textit{Excess electron density on the P-side, } n'(x_P) = n_{P0}(e^{qV/kT} - 1)$$

$$= \frac{n_i^2}{N_a}(e^{qV/kT} - 1) = \frac{10^{20}}{10^{17}} e^{0.6/0.026} = 10^{13}\,\text{cm}^{-3}$$

c. Charge neutrality requires $p'(x) = n'(x)$. If charge neutrality is not maintained, the net charge would create an electric field (see Eq. 4.1.5) that would drive the majority carriers to redistribute themselves (by drift) until charge neutrality is achieved. On the N side, electrons are the majority carriers,

$$n_N(x) = n_{N0} + n'(x) = N_d + p'(x) = 5\times 10^{17} + p'(x_N)e^{-x/L_p}$$

$$= 5\times 10^{17}\,\text{cm}^{-3} + 2\times 10^{12} e^{-x/35\ \mu\text{m}}$$

Here $x_N \approx 0$ is assumed.
The excess carrier density is often much smaller than the doping density as is the case here. This is called **low-level injection**.

The reader is requested to write down the answer for $p_P(x)$.

d.

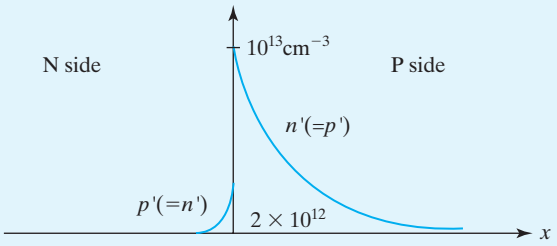


FIGURE 4–19 (a) Clearly, more carriers are injected into the lower-doping side than the higher-doping side. The ratio $p'(x_N)/n'(x_P)$ is equal to N_a/N_d. Majority carriers adjust themselves such that $n' \approx p'$ to maintain charge neutrality.

e. On the N side,

$$N_v e^{-(E_{Fp} - E_v)/kT} = p(x) = p_{N0} + p'(x)$$

$$E_{Fp}(x) = E_v + kT\ln\frac{N_v}{200 + 2\times 10^{12}\exp(x/35\ \mu\text{m})}$$

Similarly, on the P side,

$$E_{Fn}(x) = E_c - kT\ln\frac{N_c}{10^3 + 10^{13}\exp(-x/85\ \mu\text{m})}$$

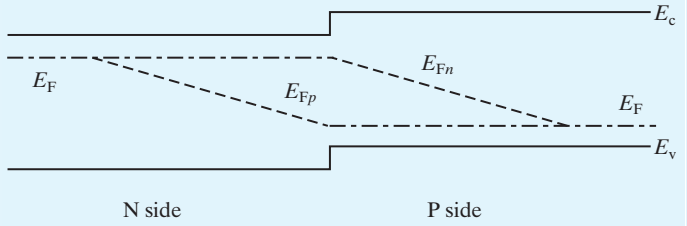


FIGURE 4–19 (b)

4.9 • PN DIODE IV CHARACTERISTICS •

Using Eqs. (4.8.2) and (4.8.3), J_p on the N side and J_n on the P side are

$$J_{p\mathrm{N}} = -qD_p\frac{\mathrm{d}p'(x)}{\mathrm{d}x} = q\frac{D_p}{L_p}p_{\mathrm{N0}}(\mathrm{e}^{qV/kT}-1)\mathrm{e}^{-(x-x_{\mathrm{N}})/L_p} \tag{4.9.1}$$

$$J_{n\mathrm{P}} = qD_n\frac{\mathrm{d}n'(x)}{\mathrm{d}x} = q\frac{D_n}{L_n}n_{\mathrm{P0}}(\mathrm{e}^{qV/kT}-1)\mathrm{e}^{(x-x_{\mathrm{P}})/L_n} \tag{4.9.2}$$

The two current components are shown in Fig. 4–20 a. Since both J_n and J_p are known at $x \approx 0$, the total current, J, can be determined at that location.

$$\text{Total current} = J_{p\mathrm{N}}(x_{\mathrm{N}}) + J_{n\mathrm{P}}(x_{\mathrm{P}}) = \left(q\frac{D_p}{L_p}p_{\mathrm{N0}} + q\frac{D_{\mathrm{n}}}{L_{\mathrm{n}}}n_{\mathrm{P0}}\right)(\mathrm{e}^{qV/kT}-1) \tag{4.9.3}$$

$$= \text{Total current at all } x$$

We know that J is not a function of x because the current that goes into one end of the diode must come out the other end and that the current is continuous in between. This fact is illustrated with the horizontal line that represents J in Fig. 4–20a. Therefore, the expression in Eq. (4.9.3) applies to all x. Once the total J is known, Fig. 4–20b shows how the remaining (majority) current components can be determined. We can rewrite the last equation in the form

$$I = I_0(\mathrm{e}^{qV/kT}-1) \tag{4.9.4}$$

$$I_0 = Aqn_{\mathrm{i}}^2\left(\frac{D_p}{L_pN_{\mathrm{d}}} + \frac{D_n}{L_nN_{\mathrm{a}}}\right) \tag{4.9.5}$$

A is the diode area and I_0 is known as the **reverse saturation current** because the diode current saturates at $-I_0$ at large reverse bias (negative V). The concepts of carrier injection, carrier recombination, and reverse leakage current are illustrated with interactive animation at http://jas.eng.buffalo.edu/education/pn/biasedPN/index.html.

Several diode IV curves are plotted in Fig. 4–21. Note that the diodes have relatively sharp turn-on characteristics. It is often said that *Si PN diodes have a turn-on voltage of about 0.6 V at room temperature*. The turn-on voltage is lower at higher temperature.

The diode IV model and data can be examined more quantitatively and over a wide range of current when plotted in a semi-log plot as shown in Fig. 4–22. The diode

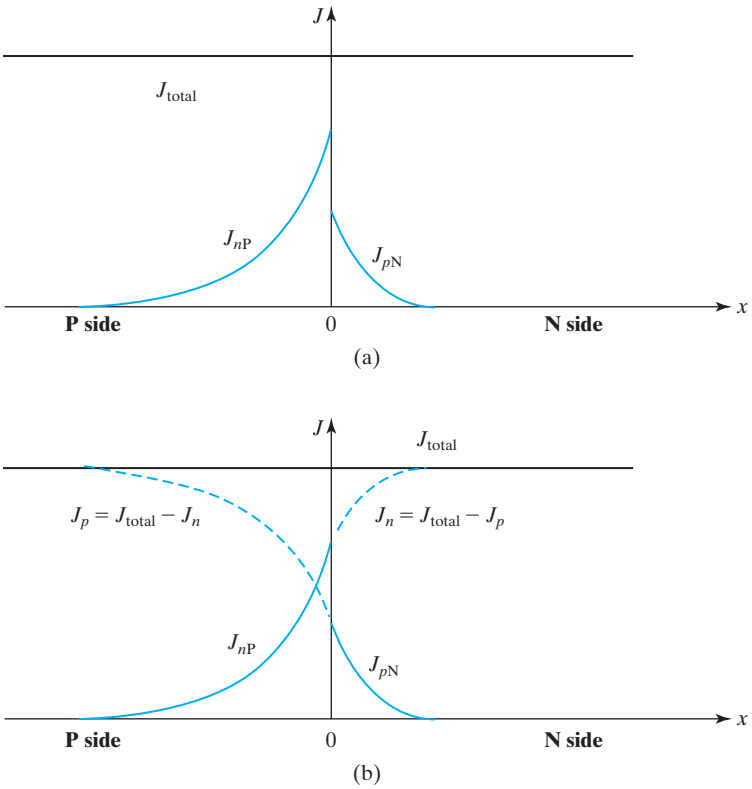


FIGURE 4–20 (a) Total J can be determined by summing $J_{n\text{P}}$ and $J_{p\text{N}}$ at the junction where both are known; (b) the other majority current components can now be determined.

model provides excellent fit to the measured diode current over a large current range under **forward bias**. The model predicts a straight line with a slope that is $(\ln 10)kT/q$ or *60 mV per decade* at room temperature. Equation (4.9.4) is used in circuit simulators to represent a diode. The I_0 for this purpose, however, is usually determined by fitting the forward IV data rather than calculated using Eq. (4.9.5) because N_{a}, N_{d}, τ_n, and τ_p are generally not accurately known and can vary with x.

● The PN Junction as a Temperature Sensor ●

Equation (4.9.4) is plotted in Fig. 4–21 for a Si diode at several temperatures. The figure shows that at a higher temperature it takes a smaller V for the diode to conduct a given I, due to a larger n_{i} in Eq. (4.9.5). This effect can be used to make a simple IC-based thermometer. A problem at the end of this chapter asks you to analyze $\text{d}V/\text{d}T$.

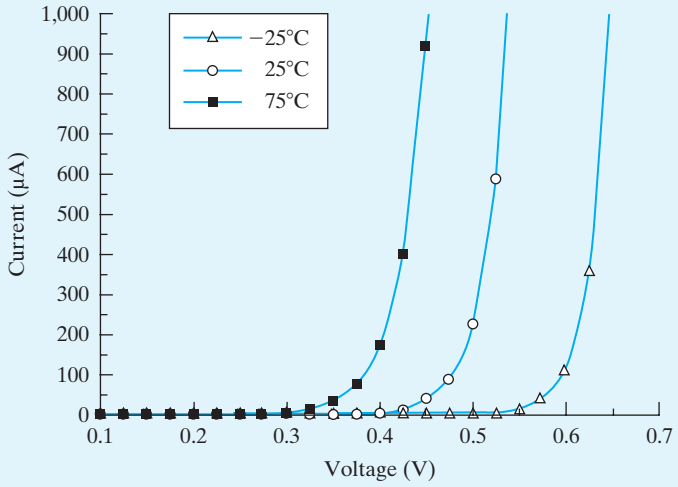


FIGURE 4–21 The *IV* curves of the silicon PN diode shift to lower voltages with increasing temperature.

4.9.1 Contributions from the Depletion Region

Equation (4.9.3) assumes that J_p and J_n do not change between x_N and x_P. As a result, Eq. (4.9.5) assumes that electrons and holes neither recombine, nor get generated, in the depletion region. In reality, there is net carrier recombination or generation in the depletion region, which contributes to the forward current and the reverse current. This contribution is called the **SCR current** for **space-charge region current**. **Space-charge region** is just another name for the depletion region. The rate of recombination (generation) in the SCR may be understood this way. Inside the depletion region, multiplying Eqs. (2.8.1) and (2.8.2) leads to

$$pn = N_c N_v e^{-(E_c - E_v)/kT} e^{(E_{Fn} - E_{Fp})/kT} = n_i^2 e^{\frac{qV}{kT}} \tag{4.9.6}$$

In the last step, Eq. (1.8.10) and $E_{Fn} - E_{Fp} = qV$ (see Fig. 4–14) were used. Equation (4.6.2) is a special case of Eq. (4.9.6) when p is known (N_a) or n is known (N_d). In the SCR, neither is known. In fact, n and p vary through the depletion-layer width. However, recombination requires the presence of both electrons and holes as shown in Fig. 2–12. It stands to reason that the recombination rate is the largest where $n \approx p$. This reasoning together with Eq. (4.9.6) suggests that the recombination rate is the largest where

$$n \approx p \approx n_i e^{\frac{qV}{2kT}} \tag{4.9.7}$$

$$\text{Net recombination (generation) rate per unit volume} = \frac{n_i}{\tau_{dep}}\left(e^{\frac{qV}{2kT}} - 1\right) \tag{4.9.8}$$

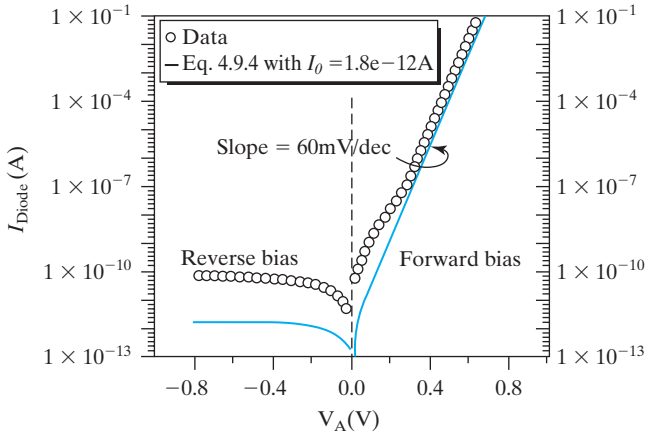


FIGURE 4–22 A semi-log plot of measured diode *IV* curve normalized to 1 cm^2. Eq. (4.9.4) represented by the color curves is accurate for the forward current except for very low current region. Reverse current is raised by thermal generation in the depletion layer.

τ_{dep} is the generation/recombination lifetime in the depletion layer. The –1 term ensures the recombination/generation rate is zero at $V = 0$ (equilibrium). When the rate is negative, there is net generation. The carriers so generated are swept by the field into the N and P regions as an additional current component to Eq. (4.9.4).

$$I = I_0\left(e^{\frac{qV}{kT}} - 1\right) + A\frac{qn_i W_{dep}}{\tau_{dep}}\left(e^{\frac{qV}{2kT}} - 1\right) \tag{4.9.9}$$

The second term is the SCR current. Under forward bias, it is an extra current with a slope corresponding to 120 mV/decade as shown in Fig. 4–22 below 10^{-7} A. This **nonideal current** is responsible for the low gain of bipolar transistor at low current (see Section 8.4). The reverse leakage current, i.e., Eq. (4.9.9) under reverse bias, is

$$I_{leakage} = I_0 + A\frac{qn_i W_{dep}}{\tau_{dep}} \tag{4.9.10}$$

Junction leakage current is a very important issue in DRAM (dynamic random access memory) technology (see Section 6.15.2) and generates noise in imager devices (see Section 5.10). Manufacturing these devices requires special care to make the generation/recombination lifetime long with super-clean and nearly crystal-defects free processing to minimize the density of recombination traps (see Fig. 2–12).

4.10 • CHARGE STORAGE[4] •

Figure 4–19 shows that excess electrons and holes are present in a PN diode when it is forward biased. This phenomenon is called **charge storage.** Clearly the **stored charge** is proportional to $p_N'(0)$ and $n_P'(0)$ (i.e., to $e^{qV/kT} - 1$). Therefore, the

[4] This section may be omitted in an accelerated course. The charge storage concept is presented more thoroughly in Section 8.7.

stored charge, Q (Coulombs), is proportional to *I*, which is also proportional to $e^{qV/kT} - 1$.

$$Q \propto I \tag{4.10.1}$$

There is a simple explanation to this proportionality. *I* is the rate of minority charge injection into the diode. In steady state, this rate must be equal to the rate of charge recombination, which is Q/τ_s (see Section 2.6).

$$\boxed{\begin{aligned} I &= Q/\tau_s \\ Q &= I\tau_s \end{aligned}} \tag{4.10.2}$$

τ_s is called the **charge-storage time**. In a one-sided junction, τ_s is the recombination lifetime on the lighter-doping side, where charge injection and recombination take place. In general, τ_s is an average of the recombination lifetimes on the N side and the P side. *In any event, I and Q are simply linked through a charge-storage time.*

4.11 • SMALL-SIGNAL MODEL OF THE DIODE[5] •

In a class of circuits called **analog circuits**, the diode is biased at a DC current, I_{DC}, and the circuit behavior is determined by how the diode reacts to small changes in the diode voltage or current.

The diode appears to the circuit as a parallel *RC* circuit as shown in Fig. 4–23. The conductance is, using Eq. (4.9.4) and assuming $qV/kT >> 1$,

$$\begin{aligned} G \equiv \frac{1}{R} &= \frac{dI}{dV} = \frac{d}{dV} I_0(e^{qV/kT} - 1) \approx \frac{d}{dV} I_0 e^{qV/kT} \\ &= \frac{q}{kT} I_0 e^{qV/kT} = I_{DC}/\frac{kT}{q} \end{aligned} \tag{4.11.1}$$

At room temperature, $G = I_{DC}/26$ mV. The small-signal conductance and resistance can be altered by adjusting the bias current, I_{DC}. The small-signal capacitance is, using Eq. (4.10.2),

$$C = \frac{dQ}{dV} = \tau_s \frac{dI}{dV} = \tau_s G = \tau_s I_{DC}/\frac{kT}{q} \tag{4.11.2}$$

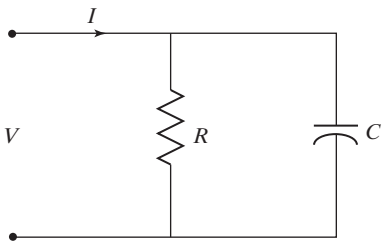


FIGURE 4–23 The small-signal equivalent circuit of a PN diode.

[5] This section may be omitted in an accelerated course.

The diode RC delay is therefore just the charge storage time, τ_s. Measuring the diode capacitance provides a convenient way to determine τ_s. This capacitance is often called the **diffusion capacitance** or the **charge-storage capacitance** because it is related to charge storage, which is related to the diffusion process. To be more accurate, one can add to Eq. (4.11.2) a term $A\varepsilon_s/W_{dep}$, representing the depletion-layer capacitance, C_{dep} [see Eq. (4.4.1)]. Under strong forward bias, the diffusion capacitance usually overwhelms C_{dep}. Both diffusion capacitance and depletion capacitance are undesirable because they slow down the devices and the circuits.

PART II: APPLICATION TO OPTOELECTRONIC DEVICES

4.12 • SOLAR CELLS •

A PN rectifier is useful, for example, to convert AC utility power to the DC power for powering the electronic equipment. Several other useful and even more interesting devices are also based on the PN junction. They are all **optoelectronic devices**. The first is the solar cell.

4.12.1 Solar Cell Basics

Commonly made of silicon, **solar cells**, also known as **photovoltaic cells**, can covert sunlight to electricity with 15 to 30% energy efficiency. Panels of solar cells are often installed on rooftops or open fields to generate electricity. A solar cell's structure is identical to a PN junction diode but with finger-shaped or transparent electrodes so that light can strike the semiconductor.

We have already seen that a voltage, without light, can produce a diode current

$$I = I_0(e^{qV/kT} - 1) \tag{4.9.4}$$

Light is also a driving force that can produce a diode current without voltage, i.e., with the two diode terminals short-circuited so that $V = 0$. Figure 4–25 shows that when light shines on the PN junction, the minority carriers that are generated by light within a diffusion length (more or less) from the junction can diffuse to the junction, be swept across the junction by the built-in field, and cause a current to flow out of the P terminal through the external short circuit and back into the N-terminal. This current is called the **short-circuit current**, I_{sc}. I_{sc} is proportional to the light intensity and to the cell area, of course.

The total diode (solar cell) current is the sum of the current generated by the voltage and that generated by light.

$$I = I_0(e^{qV/kT} - 1) - I_{sc} \tag{4.12.1}$$

The negative sign indicates that the direction of I_{sc} (see Fig. 4–25a) is opposite to that of the voltage-generated current (see Fig. 4–2). The solar cell *IV* curve is shown in Fig. 4–25b. Solar cell operates in the fourth quadrant of the *IV* plot. Since *I* and *V* have opposite signs, solar cell *generates* power. Each silicon solar cell produces about 0.6 V. Many cells are connected in series to obtain the desired voltage. Many such series strings are connected in parallel into a solar cell panel. There is a

● Earth's Energy Reserves ●

"Semiconductor Devices Save the Earth." This unusually brazen title of a scientific paper [3] called attention to the potential of solar cells as one solution to the global energy and climate warming problems. Suppose the rate of world consumption of energy does not grow with time but stays at its present level. The confirmed global oil reserve will last about 35 years. It is 60 years for natural gas, 170 years for coal, and 60 years for uranium. If our consumption rate increases by only 3% a year, the projected energy consumption (Fig. 4–24) will deplete the fossil fuel reserves even sooner. Although new reserves may be discovered from time to time, there is the additional problem of global warming (greenhouse effect) caused by the emission of carbon dioxide from burning fossil fuels. It would be foolhardy not to aggressively conserve energy and develop **renewable energy sources** such as solar cells.

Solar energy can be converted into electricity through many means besides photovoltaics. For example, it is converted into heat that drives a thermal engine that drives an electric generator in a solar thermal-electric system. Wind electricity generation harnesses the energy of the wind, which is created by solar heating of the earth. Growing plants and then burning them to generate electricity is another way. They all generate electricity without net emission of carbon dioxide.

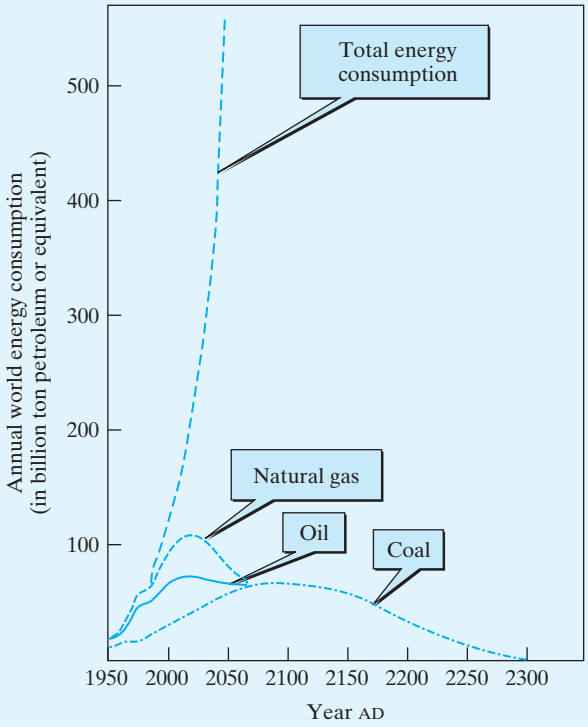


FIGURE 4–24 Depletion of fossil-fuel deposits and recent history and projection of world energy consumption assuming 3% annual growth. (From [3]. © 1992 IEEE.)

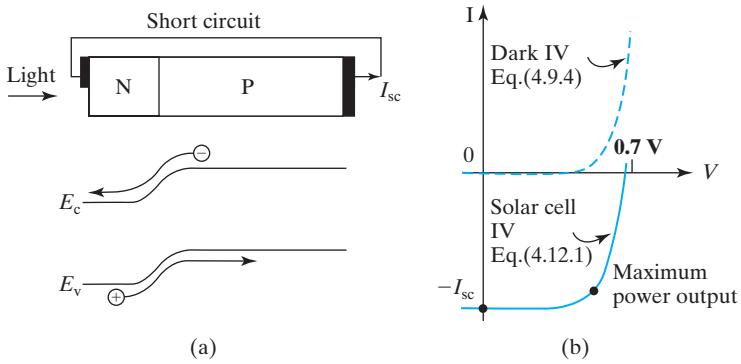


FIGURE 4–25 (a) Light can produce a current in PN junction at $V = 0$. (b) Solar cell IV product is negative, indicating power generation. (After [4].)

particular operating point on the IV curve (see Fig. 4–25b) that maximizes the output power, $|I \times V|$. A load-matching circuit ensures that the cell operates at that point.

Unfortunately, solar energy is diffuse. Each square meter of solar cells can produce about 25 W of continuous power when averaged over day, night, and sunny/cloudy days. The average electricity consumption of the world is over 10^{12} W. If all the electricity is to be provided by solar cells, the cells need to cover a huge 200 km by 200 km area. Therefore, low cost and high **energy conversion efficiency**, defined as the ratio of the electric energy output to the solar energy input, are important. Solar cells can be made of amorphous or polycrystalline (see the sidebar in Section 3.7) as well as single-crystalline semiconductors. The first two types are less expensive to manufacture but also less efficient in electricity generation.

4.12.2 Light Penetration Depth—Direct-Gap and Indirect-Gap Semiconductors

The sunlight spans a range of wavelength or photon energy mostly from infrared to violet. The photon energy and the wavelength, λ, are related by

$$\text{Photon energy(eV)} = \frac{hc}{\lambda} = \frac{1.24}{\lambda}(\mu\text{m}) \tag{4.12.2}$$

Photons with energy less than E_g are not absorbed by the semiconductor as shown in Fig. 4–26. Photons with energy larger than E_g are absorbed but some photons may travel a considerable distance in the semiconductor before being absorbed. The light intensity decreases exponentially with the distance of travel x

$$\text{Light intensity}(x) \propto e^{-\alpha x} \tag{4.12.3}$$

α is called the **absorption coefficient**. $1\backslash\alpha$ may be called the light **penetration depth**. A solar cell must have a thickness significantly larger than the light penetration depth in order to capture nearly all the photons. Figure 4–26 shows that a Si solar cell should be no thinner than 50 μm in order to absorb most of the photons with energy above E_g. On the other hand, a GaAs solar cell needs to be

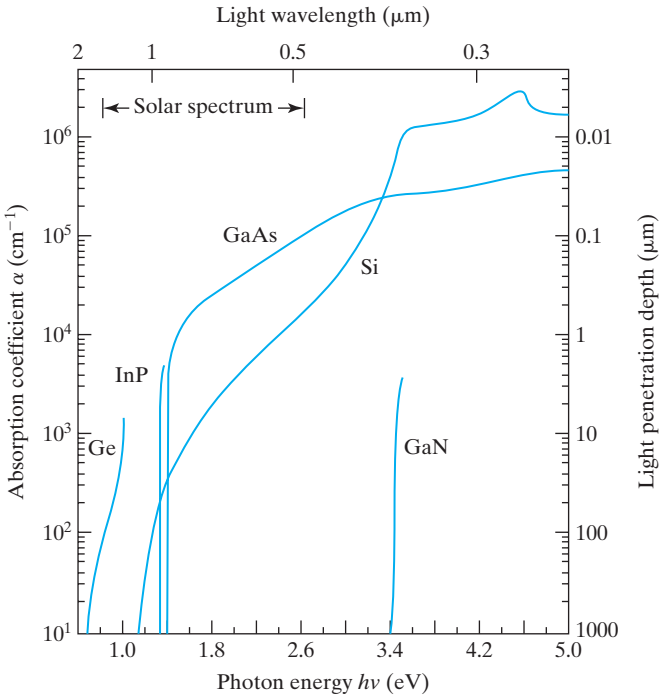


FIGURE 4–26 Light absorption coefficient as a function of photon energy. Si and Ge are indirect-gap semiconductors. InAs, GaAs, and GaN are direct-gap semiconductors, which exhibit steeply rising absorption coefficients.

only 1 μm thick because its absorption coefficient rises steeply when the photon energy exceeds E_g. Si and Ge have lower absorption coefficients because of a characteristic of their energy band structures.

Electrons have both particle and wave properties. An electron has energy E and wave vector k. k represents the direction and the wavelength of the electron wave (k = 2π/electron wavelength). The upper branch of Fig. 4–27a shows the schematic energy versus k plot of the electrons near the bottom of the GaAs conduction band. The lower branch shows the E versus k plot of the electrons near the top of its valence band. This E–k relationship is the solution of the Schrödinger equation [see Equation (1.5.3)] of quantum mechanics for the GaAs crystal. Because the bottom of the conduction band and the top of the valence band occur at the same k, a characteristic shared also by InP and GaN, they are called **direct-gap semiconductors**. "Gap" refers to the energy band gap. A photon can move an electron from the valence band to the conduction band efficiently because k conservation (the equivalent of momentum conservation) is satisfied. Hence, the absorption coefficient is large.

Figure 4–27b shows a schematic E–k plot of Si and Ge. They are called **indirect-gap semiconductors.** Light absorption is inefficient because assistance by phonons is required to satisfy k conservation.

Everything else being equal, direct-gap semiconductors are preferred for solar cell applications because only a very thin film and therefore a small amount of

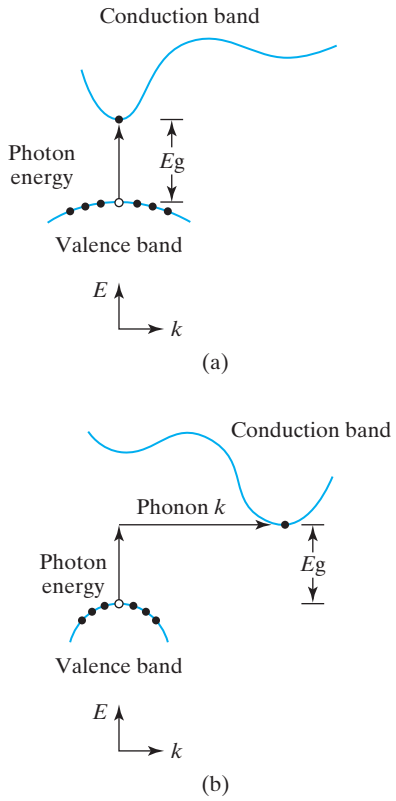


FIGURE 4–27 The *E–k* diagrams of (a) direct-gap semiconductor and (b) indirect-gap semiconductor.

semiconductor material is needed. This has positive cost implications. In reality, silicon is the most prevalent solar cell material by far because silicon is inexpensive. Nonetheless, intense research is going on to search for inexpensive inorganic or **organic semiconductor** materials to drive thin-film solar cell system cost, including packaging and installation, lower than the silicon solar cell system cost.

A related property of the direct-gap semiconductors makes them the only suitable materials for LEDs and diode lasers (see Sections 4.13 and 4.14).

4.12.3 Short-Circuit Current and Open-Circuit Voltage

If light shines on the semiconductor in Fig. 4–16 and generates holes (and electrons) at the rate of $G\ \mathrm{s}^{-1}\mathrm{cm}^{-3}$, an additional term $G \cdot A \cdot \Delta x$ should be added to the left-hand side of Eq. (4.7.1). As a result, Eq. (4.7.5) is modified to

$$\frac{\mathrm{d}^2 p'}{\mathrm{d}x^2} = \frac{p'}{L_p^2} - \frac{G}{D_p} \tag{4.12.4}$$

Assume that the P^+N solar cell in Fig. 4–28a has a very thin P^+ layer for simplicity and that all the electron–hole pairs are generated in the N region at a uniform rate

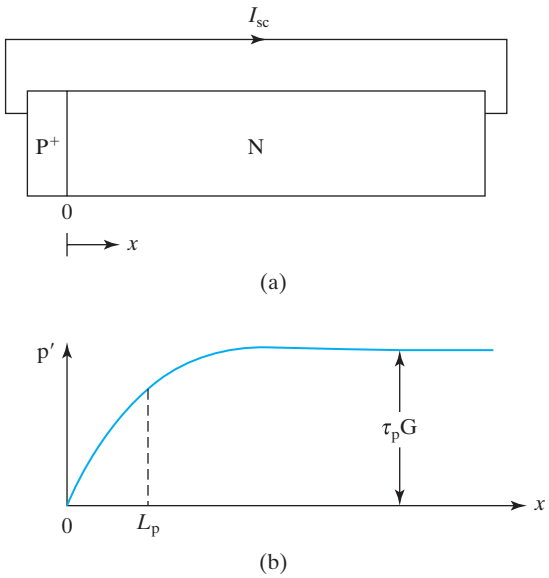


FIGURE 4–28 (a) A P^+N solar cell under the short-circuit condition and (b) the excess carrier concentration profile. Effectively only the carriers generated within $x < L_p$ can diffuse to the junction and contribute to the short-circuit current.

of G pairs $s^{-1}cm^{-3}$. Under the short-circuit condition ($V = 0$), the boundary condition at $x = 0$ is (see Eq.4.6.3)

$$p'(0) = 0 \tag{4.12.5}$$

At $x = \infty$, p' should reach a constant value and therefore the left-hand side of Eq. (4.12.4) should be zero.

$$p'(\infty) = L_p^2 \frac{G}{D_p} = \tau_p G \tag{4.12.6}$$

The solution of Eq. (4.12.4) is

$$p'(x) = \tau_p G(1 - e^{-x/L_p}) \tag{4.12.7}$$

One can easily verify that Eq. (4.12.7) satisfies Eqs. (4.12.4), (4.12.5), and (4.12.6) by substitution. $p'(x)$ is plotted in Fig. 4–28b.

$$J_p = -qD_p \frac{dp'(x)}{dx} = q\frac{D_p}{L_p}\tau_p G e^{-x/L_p} \tag{4.12.8}$$

$$I_{sc} = AJ_p(0) = AqL_pG \tag{4.12.9}$$

There is an insightful interpretation of Eq. (4.12.9). Only the holes generated within a distance L_p from the junction, i.e., in the volume $A \cdot L_p$, are collected by the PN junction and contribute to the short-circuit current. The carriers generated farther from the junction are lost to recombination. The conclusion is that a large minority carrier diffusion length is good for the solar cell current. This is always

true, although the relationship between I_{sc} and the diffusion length is more complex in a realistic solar cell structure than the simple proportionality in Eq. (4.12.9) derived under the assumption of uniform light absorption.

The solar cell material should have a large carrier diffusion length, i.e., a long carrier recombination time. In other words, the material should be quite free of defects and impurities that serve as recombination centers (see Fig. 2–12). This is particularly important for indirect-gap materials. For direct-gap semiconductors, light does not penetrate deep. In that case, all carriers are generated in a narrow region and the carrier diffusion length does not have to be large if the junction is properly positioned in this narrow region to collect the carriers.

Substituting Eqs. (4.12.9) and Eq. (4.9.5) (assuming a P^+N solar cell) into Eq. (4.12.1) leads to

$$I = Aq\frac{n_i^2}{N_d}\frac{D_p}{L_p}(e^{qV/kT} - 1) - AqL_pG \tag{4.12.10}$$

By setting $I = 0$, we can solve for the open-circuit voltage V_{oc} (assuming $e^{qV_{oc}/kT} >> 1$ for simplicity).

$$0 = \frac{n_i^2}{N_d}\frac{D_p}{L_p}e^{qV_{oc}/kT} - L_pG \tag{4.12.11}$$

$$V_{oc} = \frac{KT}{q}\ln(\tau_pGN_d/n_i^2) \tag{4.12.12}$$

4.12.4 Output Power

Equation (4.12.10) is sketched in Fig. 4–25b. There is a particular operating point on the solar cell *IV* curve that maximizes the output power, $|I \times V|$. A load-matching circuit is usually employed to ensure that the cell operates at that point.

$$\text{Output Power} = I_{sc} \times V_{oc} \times FF \tag{4.12.13}$$

where *FF* (called the **fill factor**) is simply the ratio of the maximum $|I \times V|$ to $I_{sc} \times V_{oc}$. *FF* is typically around 0.75. The short-circuit current, I_{sc}, is proportional to the light intensity as shown in Eq. (4.12.9).

Increasing N_d can raise V_{oc} according to Equation (4.12.12). Solar cells should therefore be doped fairly heavily. Large carrier generation rate, *G*, is good for V_{oc}. Using optical concentrators to focus sunlight on a solar cell can raise *G* and improve V_{oc}. Besides reducing the solar cell area and cell cost, light concentration can thus increase the cell efficiency provided that the cell can be effectively cooled. If the cell becomes hot, n_i^2 increases and V_{oc} drops. A larger band-gap energy, E_g, reduces n_i^2 exponentially (see Eq. 1.8.12). V_{oc} therefore increases linearly with E_g. On the other hand, if E_g is too large, the material would not absorb the photons in a large long-wavelength (red and infrared) portion of the solar spectrum (see Fig. 4–26) and I_{sc} drops. The best solar cell efficiency (~24%) is obtained with E_g values between 1.2 and 1.9 eV. Commercial rooftop silicon solar-cell panels have conversion efficiencies between 15 and 20%. **Tandem solar cells** can achieve very high (over

30%) energy conversion efficiency by using two or more semiconductor materials. One material with a larger E_g absorbs and converts the short-wavelength portion of the solar radiation to electricity and another smaller E_g material, positioned behind the first, does the same to the solar radiation that is not absorbed by the first material.

4.13 • LIGHT-EMITTING DIODES AND SOLID-STATE LIGHTING •

LEDs or **light-emitting diodes** of various colors are used for such applications as traffic lights, indicator lights, and video billboards. LEDs can also provide space lighting at much higher energy efficiency than the incandescent lamps. The electrons and holes recombine by emitting photons (light) with $h\nu \approx E_g$. By adjusting the composition of the semiconductor, E_g can be altered to make blue, green, yellow, red, infrared, and UV LEDs possible. They are made of compound semiconductors involving In, Ga, Al, As, P, and N.

Figure 4–29 shows a basic LED. A PN junction made of an appropriate semiconductor is forward biased to inject minority carriers. When the injected minority carriers recombine with the majority carriers, photons are emitted. The light is emitted to all directions. To reduce the reflection of light at the semiconductor and air interface (back into the semiconductor) and therefore project more light into the forward direction, the semiconductor surface may be textured or a dome shaped lens may be provided.

4.13.1 LED Materials and Structures

Direct-gap semiconductors such as GaN (gallium nitride) are much better for LED applications than indirect-gap semiconductors such as Si. The electrons and holes in direct-gap materials have matching wave vectors (see Section 4.12.2) and can recombine easily. Figure 4–27 illustrates how a photon generates an electron–hole pair in the two types of semiconductors. If the arrows were reversed, this figure also explains how a photon is generated when an electron

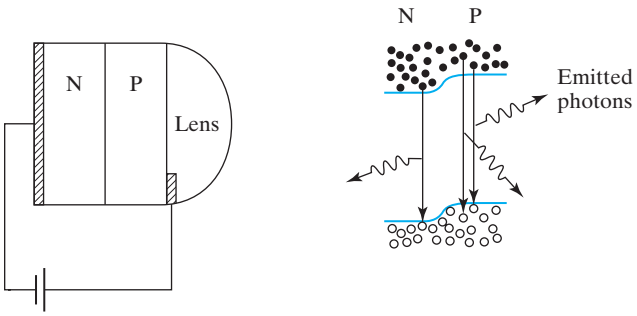


FIGURE 4–29 Schematic drawing of an LED. Photons are generated when the electrons and holes injected by the PN junction recombine.

and a hole recombine. The photon generation process, called radiative recombination, is straightforward and fast in direct-gap semiconductors with nanosecond lifetime. Therefore, the radiative recombination process is the dominant recombination process, i.e., a high percentage of the injected carriers generate photons. This percentage is known as the **quantum efficiency**. The quantum efficiency of photon generation is much lower in indirect-gap semiconductors because the radiative recombination is slow with millisecond lifetime. As a result, the recombination-through-traps process (see Fig. 2–13), which generates phonons rather than photons, is the faster and dominant process of recombination.

The next consideration is the band-gap energy. From Eq. (4.12.2)

$$\text{LED wavelength } (\mu\text{m}) = \frac{1.24}{\text{photon energy}} \approx \frac{1.24}{E_g(\text{e}V)} \tag{4.13.1}$$

Table 4–1 shows the semiconductors commonly used in LEDs. Materials with band gaps in the infrared range such as InP and GaAs are popular for LEDs used in optical communication applications (See Section 4.14). There are a few suitable semiconductors for visible-light applications. For example, GaP has a band-gap energy corresponding to yellow light. Mixing GaP and GaAs at varying ratios, i.e., using $GaAs_{1-x}P_x$, we can make LEDs that emit yellow ($x \approx 1$), orange, red, and infrared ($x \approx 0$) light.

GaAs and GaP are called **binary semiconductors**, which are made of two chemical elements. $GaAs_{1-x}P_x$, containing three elements, is a **ternary semiconductor**. They are all known as **compound semiconductors**.

There is a third consideration—the substrate material. High-efficiency LEDs are made of thin semiconductor films grown epitaxially (see Section 3.7.3) over a substrate wafer having a crystal lattice constant closely matched to that of the films. The lattice constants are given in the last column of Table 4–1. High-quality wafers of InP, GaAs, GaP, and Al_2O_3 are available at reasonable costs.

TABLE 4–1 Optoelectronic-device materials.

	$E_g(\text{e}V)$	Wavelength (μm)	Color	Lattice constant (Å)
InAs	0.36	3.44		6.05
InN	0.65	1.91	infrared	3.45
InP	1.36	0.92		5.87
GaAs	1.42	0.87	↑	5.66
GaP	2.26	0.55	red yellow	5.46
AlP	3.39	0.51	blue violet	5.45
GaN	2.45	0.37	↓	3.19
AlN	6.20	0.20	UV	3.11

TABLE 4–2 Some common LEDs.

Spectral Range	Material System	Substrate	Example Applications
Infrared	InGaAsP	InP	Optical communication
Infrared-Red	GaAsP	GaAs	Indicator lamps. Remote control
Red-Yellow	AlInGaP	GaP	Optical communication. High-brightness traffic signal lights
Green-Blue	InGaN	Sapphire	High-brightness signal lights. Video billboards
Blue-UV	AlInGaN	GaN or sapphire	Solid-state lighting
Red-Blue	Organic semiconductors	Glass	Displays

For example, GaP film on GaAs substrate is used to produce yellow LEDs. Unfortunately, GaP does not have the same lattice constant as the GaAs substrate. The mismatch in lattice constants creates crystal defects in the expitaxial GaP film (see Section 3.7.3) and degrades LED efficiency and reliability. GaP can be mixed with InP to obtain a lattice constant matching that of the GaAs substrate. However, that mixture has an E_g that is too low for the desired visible-light emission. Fortunately, both E_g and the lattice constant can be independently tuned if one mixes GaP, InP, and AlP in varying proportions. In other words, **quaternary semiconductors**, which contain four chemical elements such as AlInGaP, can provide a range of E_g (wavelengths) while meeting the requirement of matching the lattice constant of the substrate material.

Table 4–2 summarizes some important LED material systems. GaP substrate, replacing GaAs, improves the efficiency of red and yellow LEDs because GaP is transparent while GaAs absorb light at these wavelengths. Sapphire (Al_20_3) with an epitaxial GaN overlayer is an excellent substrate for blue and UV LEDs because it is transparent to UV light.

With well-chosen materials, LED can have internal quantum efficiency close to unity, i.e., most of the injected carriers generate photons. Care must be taken to extract the light. Figure 4–30a shows a red LED with its substrate shaped into a truncated pyramid to collected light with reflector on the back and total internal reflection from the sides. The GaP substrate is transparent to red light.

Figure 4–30b illustrates the concept of energy well or **quantum well**. Because the AlInGaP film has a smaller band gap than GaP, it forms a well, called a quantum well, between the GaP on both sides. The concentrations of both electrons and holes are high in the well, a condition favorable for recombination and light emission. Often **multiple quantum wells** are used with several repeated alternating layers of different semiconductors.

Another class of LEDs is the **organic light-emitting diodes** (**OLEDs**). Certain organic compounds have semiconductor properties and can be made into PN junction diodes, LEDs, and transistors [5]. Different colors can be obtained by modifying the molecules or adding phosphorescent materials that emit longer wavelength light when excited with a shorter wavelength light. OLED single and multicolor displays on glass substrates compete with LCD displays in some applications such as car dashboard displays.

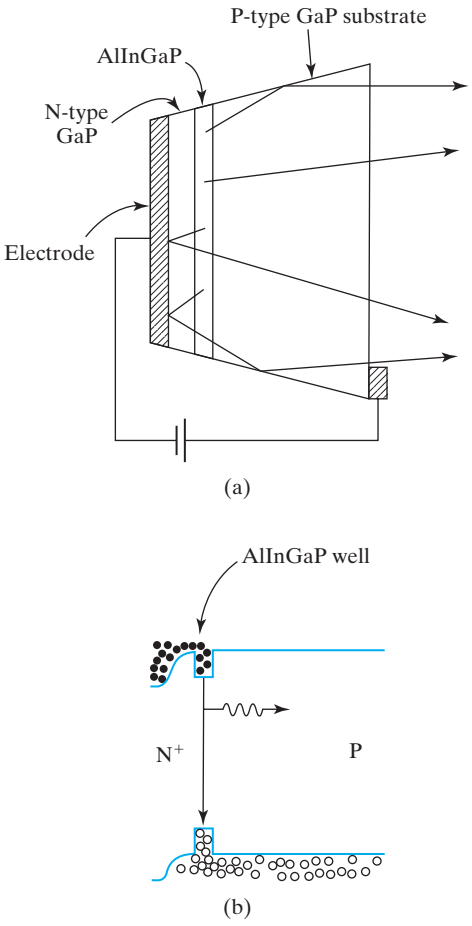


FIGURE 4–30 A red LED with sloped sides for better light extraction; and (b) energy band diagram showing the quantum well.

4.13.2 Solid-State Lighting

Solid-state lighting refers to space lighting using LEDs in lieu of traditional light sources such as incandescent light bulbs and fluorescent lamps, which involve gas or vacuum. About 25% of the world electricity usage is attributable to lighting (more if the air-conditioning load due to heat generated by lamps is included). Improving the energy efficiency of light can significantly reduce energy consumption and greenhouse gas emission.

Lumen (lm) is a measure of the visible light energy (normalized to the sensitivity of the human eye at different wavelengths). A 100 W incandescent light bulb produces about 1,700 lm, therefore the **luminous efficacy** of the light bulb is 17 lm/W. Table 4–3 compares the efficacy of white LED with other light sources and the theoretical limit. The efficacy of LED continues to improve through material and device innovations.

TABLE 4–3 Luminous efficacy of lamps in lumen/watt.

Incandescent lamp	Compact fluorescent lamp	Tube fluorescent lamp	White LED	Theoretical limit at peak of eye sensitivity (λ = 555 nm)	Theoretical limit (white light)
17	60	50-100	90-?	683	~340

White light may be obtained by packaging red, green, and blue LEDs in one lamp housing. An economically attractive alternative is to use a UV LED and phosphors that convert the UV light into red, green, and blue, i.e., white light. Efficient UV LEDs can be fabricated with AlInGaN/InGaN/AlInGaN quantum wells grown on sapphire substrate (see Table 4–2). Light is extracted from the UV-transparent sapphire substrate.

White LED efficacy up to ten times that of the incandescent lamps are achievable. The technical challenge is to reduce the cost including the cost of the substrate, epitaxial film growth, and packaging. One approach of cost reduction is to use OLED. Organic semiconductors are polymers and can be printed on large sheets of glass or even flexible substrates at low cost. Different polymers can generate light of different colors. Using several layers of different materials or adding fluorescent dopants can produce white light. OLEDs, however, have lower efficacy than nitride and aluminide based LEDs. For solid-state lighting, both high efficacy and low cost are required. White LEDs are also used as the back lighting source for LCD displays.

4.14 • DIODE LASERS •

Lasers can be made with many materials from gas to glass, and energy may be supplied to the lasers by many means such as electric discharge in gas or intense light from flash lamps. **Diode lasers, powered by the diode currents,** are by far the most compact and lowest-cost lasers. Their basic structure has a PN junction under forward-bias. Diode lasers shine in many applications from fiber-optic communications to DVD and CD-ROM readers. The applications make use of one or both of these characteristics of the laser light: single frequency and ability to be focused to a small spot size. Diode lasers, as LEDs, employ direct-gap semiconductors for efficient light emission.

4.14.1 Light Amplification

Laser operation requires light amplification. Figure 4–31 illustrates how to achieve light amplification. Part (a) shows a photon generating an electron–hole pair as in solar-cell operation. Part (b) shows that when an electron falls from the conduction band to the valence band, a photon is emitted in a random direction as in the operation of an LED. This is called **spontaneous emission**. If a photon of a suitable energy comes along as shown in Fig. 4–31c, the incident photon can stimulate the electron, causing it to fall and emit a second photon. This is called **stimulated emission**. In light-wave terms, the amplitude of the incident light wave is amplified

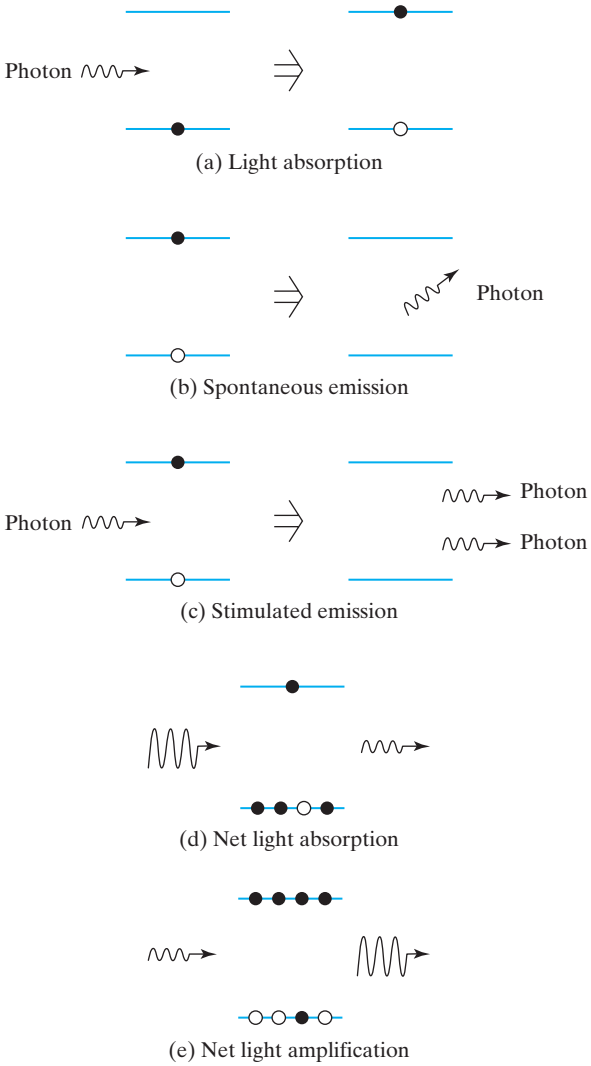


FIGURE 4–31 (a–c) Three types of light–electron interactions; (d) normally, light is absorbed in the semiconductor; and (e) under population inversion, light (wave amplitude) is amplified in the semiconductor.

just as an electrical signal waveform can be amplified. The amplified light wavefrom is identical to the incident light in frequency and direction of travel but has larger amplitude. This is the reason for the wavelength purity and directional tightness of the laser light. Stimulated emission is the foundation of laser operation.

In general, there are electrons in both the upper and the lower energy levels. If there is a higher probability of electron occupation in the lower states as shown in Fig. 4–31d (the normal case according to the Fermi function of Section 1.7), there is a higher rate of absorption than stimulated emission and the light is absorbed by the semiconductor. If there is a higher probability of electron occupation in the

upper states as shown in Fig. 4–31e, a condition called **population inversion**, the light is amplified. Population inversion is a necessary condition for laser operation.

Consider a light beam traveling along the P^+N^+ junction as shown in Fig. 4–32a. At zero bias voltage, shown in Fig. 4–32b, there is a higher probability of electron occupation near E_v than E_c and the light beam is attenuated by absorption. Population inversion can be achieved by applying a large forward bias voltage to the P^+N^+ junction as shown in Fig. 4–32c. There is now a higher probability of electron occupation of states near E_c than E_v and the light beam is amplified. Population inversion is achieved when

$$E_{Fn} - E_{Fp} > E_g \tag{4.14.1}$$

Equation (4.14.1) can be satisfied, i.e., population inversion achieved, more easily when a quantum well is inserted in or near the PN junction as shown in Fig. 4–32d and e. A quantum well or energy well is created when a thin layer of a narrower-gap semiconductor is sandwiched between two wider-gap semiconductors. The quantum well in Fig. 4–32e confines population inversion to a narrow region.

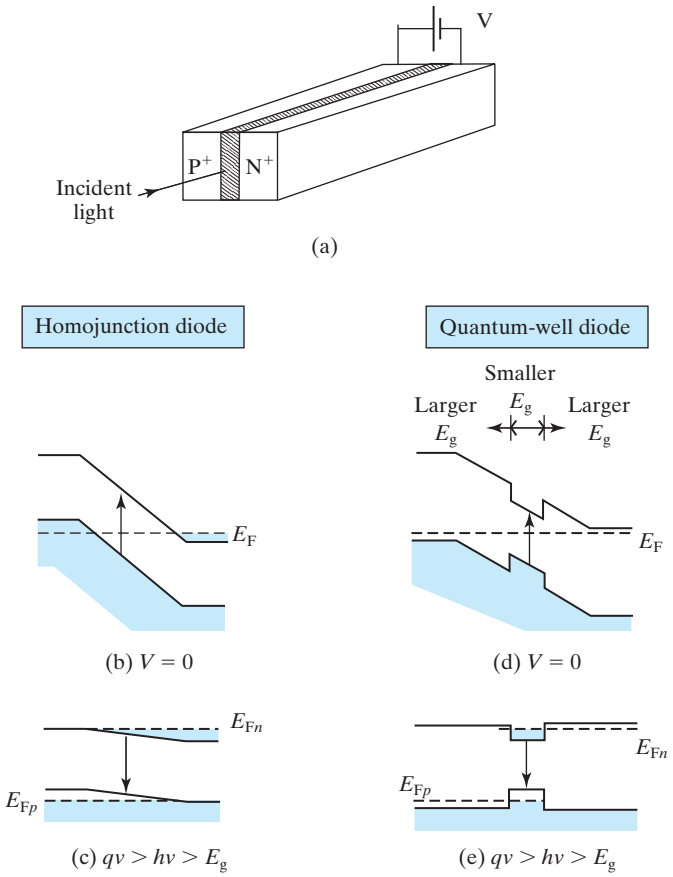


FIGURE 4–32 (a) Schematic drawing of light passing through a diode structure; (b) and (d) light is absorbed if the diode is at equilibrium. Energy states below E_F are basically filled with electrons, and those above E_F are empty. (c) and (e) under population inversion, light is amplified by stimulated emission. The arrows indicate the electron transitions caused by the light.

Fewer excess carriers are needed to achieve population inversion in the smaller volume of a narrow quantum well and the external circuitry does not need to inject carriers at a high rate, i.e., the "**threshold current**" of lasing is low.

Light amplification by stimulated emission is performed many times in a long-distance optical fiber communication line to compensate for light absorption in the fiber. To make a laser, light amplification must be paired with optical feedback as shown in the next section.

4.14.2 Optical Feedback

The operation of an electronic oscillator circuit depends on signal amplification (gain) and feedback. A laser is an optical oscillator. Besides optical amplification, it needs optical feedback. A simple way to provide optical feedback is to cleave or polish the end faces of the laser diode (see Fig. 4–33a) such that they are

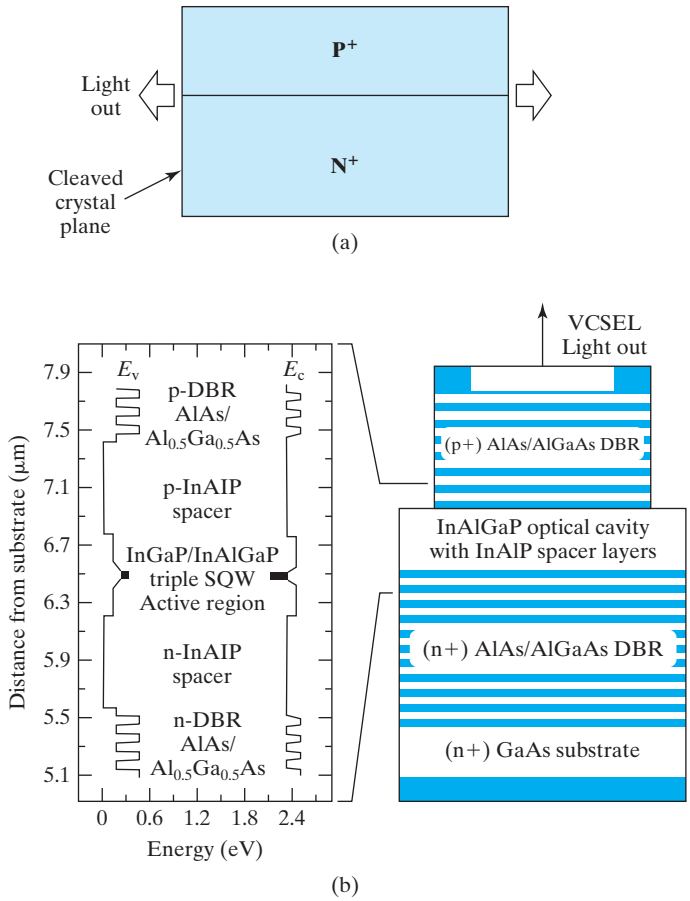


FIGURE 4–33 (a) A simple side-emitting diode laser with cleaved mirror surface. (b) The complex structure of a red-light VCSEL. Left half shows the energy band diagram of a few of the many layers of semiconductors. The energy axis is the x axis, not the usual y axis. (From [6].)

perpendicular to the PN junction. A part of the light incident on the semiconductor–air interface is reflected and amplified while traveling back through the laser diode before arriving at the other face of the diode, where a part is again reflected. The condition for oscillation is that the net round-trip gain be equal to or greater than unity.

$$R_1 \times R_2 \times G \geq 1 \tag{4.14.2}$$

Where R_1 and R_2 are the reflectivities of the two ends and G is the light amplification factor (gain) for a round-trip travel of the light from one end to the other and back. When Eq. (4.14.2) is satisfied, the internal light intensity grows until it can stimulate emissions at a sufficient rate to consume the carriers injected by the external circuitry. That determines the steady-state internal laser light intensity, a fraction of which is emitted through the end reflectors.

Often a diode laser uses a series of alternating layers of two different semiconductors of the proper thicknesses that create constructive interference and function as a reflector for a particular wavelength. The series of layers form a **distributed Bragg reflector** (**DBR**) and provide **distributed feedback**. An example is shown in Fig. 4–33b. The advantage of DBR is improved wavelength purity because it only reflects light whose wavelength is twice the period of the layer series. Another advantage is its compatibility with planar processing techniques (mirror polish and cleaving are not) that produce thousands of lasers on one wafer substrate. The laser light exits through the top in Fig. 4–33b and this structure is called a **vertical-cavity surface-emitting laser** (**VCSEL**).

The left-hand side of Fig. 4–33b is the energy-band diagram of the core region of the laser. The optical gain is provided only by the thin quantum wells in the middle of the sketch.

4.14.3 Diode Laser Applications

Red diode lasers are used in CD and DVD readers. The laser beam is focused to a tiny spot to read the indentations embossed into the plastic disks. Blue diode laser beams can be focused into even smaller spots because of the shorter wavelength. They are used in high-density or **Blu-ray** DVD readers. For **writable optical storage**, the focused laser melts a thin film and thus changes its optical reflectance for later reading.

Diode lasers are used in the highly important **fiber-optic communication** systems. The **optical fiber** is a thin (~20 μm) flexible fused quartz fiber that is extraordinarily transparent to light. Its transparency is the greatest at the 1.55 μm wavelength. Pulses of laser light of that wavelength are generated by modulating the laser diode current, the light pulse can travel more than 10 km in the fiber before losing half the intensity through absorption and scattering. With optical amplification every certain distance, optical fibers carry data between cities and nations. There are several hundred thousand kilometers of fiber-optic cables that crisscross the ocean floor to link the continents.

1.55 μm wavelength infrared diode lasers are constructed with InGaAsP materials on InP substrate. For short distance links, InGaAsP LEDs may be used (see Table 4–2). Light of different wavelengths travels at different speeds in the glass fiber. LEDs emit light of many different wavelengths, which arrives at the

destination after different times of travel. Consequently, a short LED pulse at the originating point would arrive at the destination as a longer broadened pulse. For this reason, lasers, with their extraordinary purity of wavelength, are the light source of choice for long-distance high data rate links.

4.15 • PHOTODIODES •

Figure 4–25 shows that a reverse current flows through a diode when illuminated with light and the current is proportional to the light intensity. A reverse-biased PN diode can thus be used to detect light, and the device is called a **photodiode**. If the photodiode is biased near the avalanche breakdown voltage, photo-generated carriers are multiplied by impact ionization as they travel through the depletion layer (see Fig. 4–13) and thereby the sensitivity of the detector is increased. This device is called an **avalanche photodiode**. Photodiodes are used for optical communication, DVD reader, and other light-sensing applications.

PART III: METAL–SEMICONDUCTOR JUNCTION

There are two kinds of **metal–semiconductor junction**. The junctions between metal and lightly doped semiconductors exhibit rectifying *IV* characteristics similar to those of PN junctions. They are called **Schottky diodes** and have some interesting applications. The junction between metal and heavily doped semiconductors behaves as low-resistance **ohmic contacts** (basically electrical shorts). Ohmic contacts are an important part of semiconductor devices and have a significant influence on the performance of high-speed transistors.

4.16 • SCHOTTKY BARRIERS •

The energy diagram of a metal–semiconductor junction is shown in Fig. 4–34. The Fermi level, E_F, is flat because no voltage is applied across the junction. Far to the right of the junction, the energy band diagram is simply that of an N-type silicon sample. To the left of the junction is the energy band diagram of a metal—with the energy states below E_F almost totally filled and the states above E_F almost empty. The most striking and important feature of this energy diagram is the energy barrier at the metal–semiconductor interface. It is characterized by the **Schottky barrier height**, ϕ_B. ϕ_B is a function of the metal and the semiconductor. Actually, there are two energy barriers. In Fig. 4–34a, $q\phi_{Bn}$ is the barrier against electron flow between the metal and the N-type semiconductor.[6] In Fig. 4–34b, $q\phi_{Bp}$ is the barrier against hole flow between the metal and the P-type semiconductor. In both figures, there is clearly a depletion layer adjacent to the semiconductor–metal interface, where E_F is close to neither E_c nor E_v (such that $n \approx 0$ and $p \approx 0$).

[6] The hole flow in Fig. 4–34a is usually insignificant because there are few holes in the N-type semiconductor.

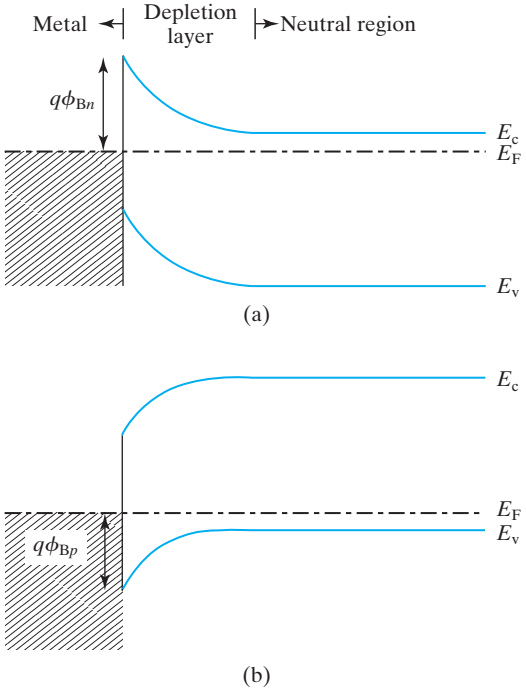


FIGURE 4–34 Energy band diagram of a metal–semiconductor contact. The Schottky barrier heights depend on the metal and semiconductor materials. (a) ϕ_{Bn} is the barrier against electron flow between the metal and the N-type semiconductor; (b) ϕ_{Bp} is the barrier against hole flow between the metal and the P-type semiconductor.

TABLE 4–4 Measured Schottky barrier heights for electrons on N-type silicon (ϕ_{Bp}) and for holes on P-type silicon (ϕ_{Bp}). (From [7].)

Metal	Mg	Ti	Cr	W	Mo	Pd	Au	Pt
ϕ_{Bn} (V)	0.4	0.5	0.61	0.67	0.68	0.77	0.8	0.9
ϕ_{Bp} (V)		0.61	0.50		0.42		0.3	
Work Function ψ_M (V)	3.7	4.3	4.5	4.6	4.6	5.1	5.1	5.7

It will become clear later that ϕ_B is the single most important parameter of a metal–semiconductor contact. Table 4–4 presents the approximate ϕ_{Bn} and ϕ_{Bp} for several metal–silicon contacts. Please note that the sum of $q\phi_{Bn}$ and $q\phi_{Bp}$ is approximately equal to E_g (1.12 eV), as suggested by Fig. 4–35.

$$\phi_{Bn} + \phi_{Bp} \approx E_g \tag{4.16.1}$$

Why does ϕ_{Bn} (and ϕ_{Bp}) vary with the choice of the metal? Notice that Table 4–4 is arranged in ascending order of ϕ_{Bn}. There is a clear trend that ϕ_{Bn} increases with increasing metal work function (last row in Table 4–4). This trend may be partially explained with Fig. 4–2a.

$$\phi_{Bn} = \psi_M - \chi_{Si} \tag{4.16.2}$$

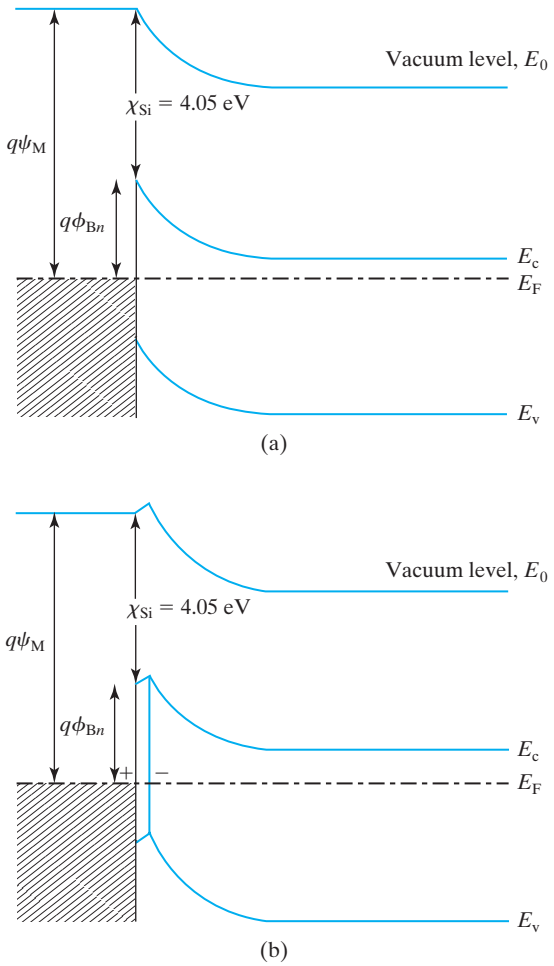


FIGURE 4–35 (a) An "ideal" metal–semiconductor contact and (b) in a real metal–semiconductor contact, there is a dipole at the interface.

ψ_M is the metal work function and χ_{Si} is the silicon electron affinity. See Sec. 5.1 for more discussion of these two material parameters. Equation (4.16.2) suggests that ϕ_{Bn} should increase with increasing ψ_M (in *qualitative* agreement with Table 4–4) by 1 eV for each 1 eV change in ψ_M (not in *quantitative* agreement with Table 4–4). The explanation for the quantitative discrepancy is that there are high densities of energy states in the band gap at the metal–semiconductor interface.[7] Some of these energy states are acceptor like and may be neutral or negative. Other energy states are donor like and may be neutral or positive. The net charge is zero when the Fermi level at the interface is around the middle of the silicon band gap. In other words, Eq. (4.16.2) is only correct for ψ_M around 4.6V, under which condition there is little interface charge. At any other ψ_M, there is a dipole at the interface as shown in

[7] In a three-dimensional crystal, there are no energy states in the band gap. Not so at the metal–semiconductor interface.

TABLE 4–5 Measured Schottky barrier heights of metal silicide on Si.

Silicide	$ErSi_{1.7}$	HfSi	$MoSi_2$	$ZrSi_2$	$TiSi_2$	$CoSi_2$	WSi_2	$NiSi_2$	Pd_2Si	PtSi
ϕ_{Bn} (V)	0.28	0.45	0.55	0.55	0.61	0.65	0.67	0.67	0.75	0.87
ϕ_{Bp} (V)			0.55	0.55	0.49	0.45	0.43	0.43	0.35	0.23

Fig. 4–35b and it prevents ϕ_{Bn} from moving very far from around 0.7 V. This phenomenon is known as **Fermi-level pinning**. Table 4–4 can be approximated with

$$\phi_{Bn} = 0.7\ \text{V} + 0.2(\psi_M - 4.75) \tag{4.16.3}$$

The factor of 0.2 in Eq. (4.16.3) is determined by the polarizability of Si and the energy state density at the metal–silicon interface [8].

● Using *C–V* Data to Determine ϕ_B ●

In Fig. 4–36a, ϕ_{bi} is the built-in potential across the depletion layer.

$$q\phi_{bi} = q\phi_{Bn} - (E_c - E_F) = q\phi_{Bn} - kT\ln\frac{N_c}{N_d} \tag{4.16.4}$$

The depletion-layer thickness is [see Eq. (4.3.1)]

$$W_{dep} = \sqrt{\frac{2\varepsilon_s(\phi_{bi} + V)}{qN_d}} \tag{4.16.5}$$

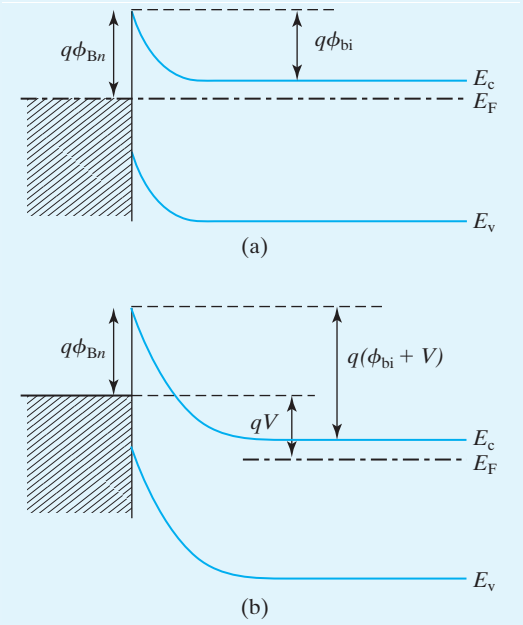


FIGURE 4–36 The potential across the depletion layer at the Schottky junction. (a) No voltage applied; (b) a negative voltage (reverse bias) is applied to the metal.

$$C = A\frac{\varepsilon_s}{W_{\text{dep}}} \tag{4.16.6}$$

$$\frac{1}{C^2} = \frac{2(\phi_{bi} + V)}{qN_{\text{d}}\varepsilon_s A^2} \tag{4.16.7}$$

Figure 4–37 shows how Eq. (4.16.7) allows us to determine ϕ_{bi} using measured C–V data. Once ϕ_{bi} is known, $\phi_{\text{B}n}$ can be determined using Eq. (4.16.4).

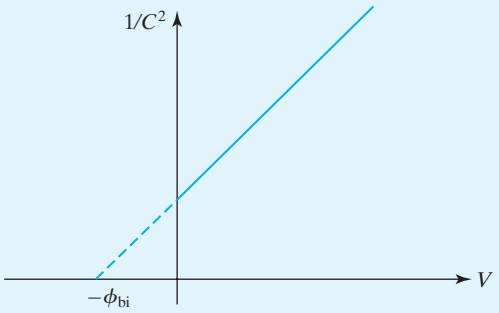


FIGURE 4–37 ϕ_{bi} (and hence ϕ_{B}) can be extracted from the C–V data as shown.

Much more prevalent in IC technology than metal–Si contacts are the silicide–Si contacts. Metals react with silicon to form metal like silicides at a moderate temperature. Silicide–Si interfaces are more stable than the metal–Si interfaces and free of native silicon dioxide. After the metal is deposited on Si by sputtering or CVD (Chemical Vapor Deposition) (see Chapter 3), an annealing step is applied to form a silicide–Si contact. *The term metal–silicon contact is understood to include silicide–silicon contacts.* Table 4–5 shows some available data of $\phi_{\text{B}n}$ and $\phi_{\text{B}p}$ of silicide–silicon contacts.

4.17 • THERMIONIC EMISSION THEORY •

Figure 4–38 presents the energy band diagram of a Schottky contact with a bias V applied to the metal. Let us analyze the current carried by the electrons flowing from Si over the energy barrier into metal, $J_{\text{S}\to\text{M}}$. This current can be predicted quite accurately by the thermionic emission theory.

In the **thermionic emission** theory, we assume that $E_{\text{F}n}$ is flat all the way to the peak of the barrier, the electron concentration at the interface (using Eqs. (1.8.5) and (1.8.6)) is

$$n = N_{\text{c}}\text{e}^{-q(\phi_{\text{B}}-V)/kT} = 2\left[\frac{2\pi m_n kT}{h^2}\right]^{3/2}\text{e}^{-q(\phi_{\text{B}}-V)/kT} \tag{4.17.1}$$

The x-component of the average electron velocity is of course smaller than the *total* thermal velocity, $\sqrt{3kT/m_n}$ [Eq. (2.1.3)], and only half of the electrons travel

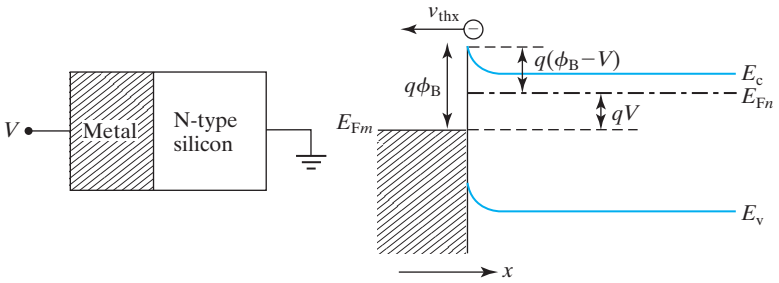


FIGURE 4–38 Energy band diagram of a Schottky contact with a forward bias V applied between the metal and the semiconductor.

toward the left (the metal). It can be shown that the average velocity of the left-traveling electrons is

$$v_{thx} = -\sqrt{2kT/\pi m_n} \tag{4.17.2}$$

Therefore,

$$J_{S\to M} = -\frac{1}{2}qnv_{thx} = \frac{4\pi q m_n k^2}{h^3}T^2 e^{-q\phi_B/kT}e^{qV/kT} \tag{4.17.3}$$

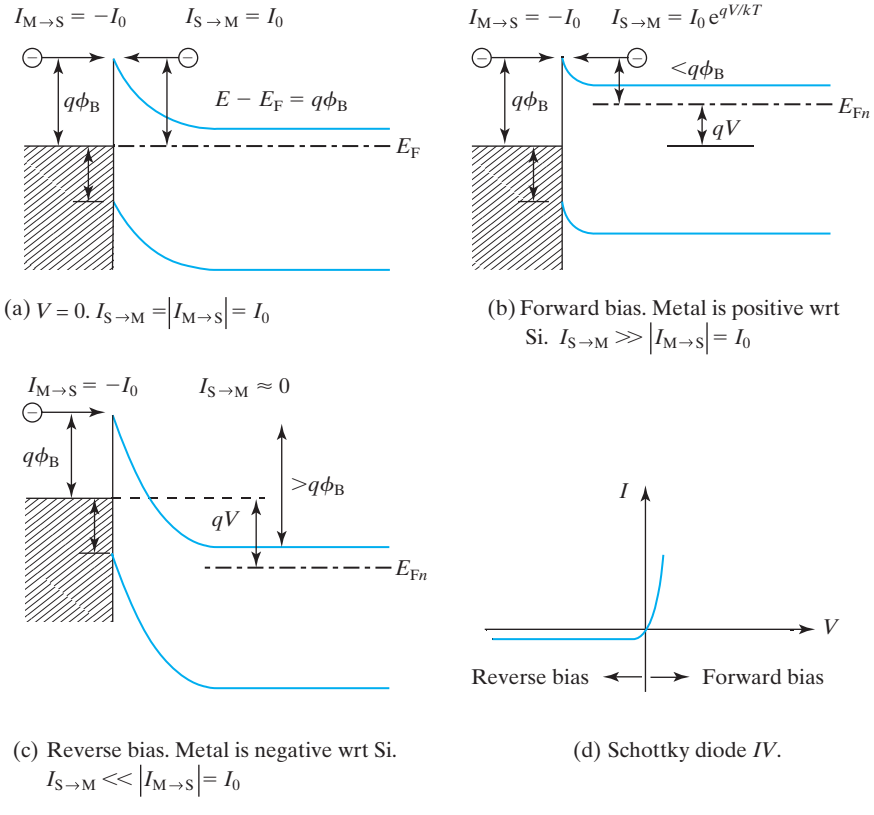
$$\equiv J_0 e^{qV/kT} \tag{4.17.4}$$

Equation (4.17.4) carries two notable messages. First $J_0 \approx 100e^{-q\phi_B/kT}$ (A/cm^2) is larger if ϕ_B is smaller. Second, $J_{S\to M}$ is only a function of $\phi_B - V$ (see Fig. 4–38). The shape of the barrier is immaterial as long as it is narrow compared to the carrier mean free path. $\phi_B - V$ determines how many electrons possess sufficient energy to surpass the peak of the energy barrier and enter the metal.

4.18 • SCHOTTKY DIODES •

At zero bias (Fig. 4–39a), the net current is zero because equal (and small) numbers of electrons on the metal side and on the semiconductor side have sufficient energy to cross the energy barrier and move to the other side. The probability of finding an electron at these high-energy states is $e^{-(E-E_c)/kT} = e^{-q\phi_B/kT}$ on both sides of the junction, as shown in Fig. 4–39a. Therefore, the net current is zero.[8] In other words, $I_{S\to M} = I_0$ and $I_{M\to S} = -I_0$, where $I_{S\to M}$ and $I_{M\to S}$ (see Fig. 4–39a) represent

[8] What if the densities of states are different on the two sides of the junction? Assume that the density of states at E on the metal side is twice that on the silicon side. There would be twice as many electrons on the metal side attempting to cross the barrier as on the Si side. On the other hand, there would be twice as many empty states on the metal side to *receive* the electrons coming from the Si side. Therefore, in a more detailed analysis, the net current is still zero.



(a) $V = 0$. $I_{S\to M} = |I_{M\to S}| = I_0$

(b) Forward bias. Metal is positive wrt Si. $I_{S\to M} \gg |I_{M\to S}| = I_0$

(c) Reverse bias. Metal is negative wrt Si. $I_{S\to M} \ll |I_{M\to S}| = I_0$

(d) Schottky diode *IV*.

FIGURE 4–39 Explanation of the rectifying *IV* characteristics of Schottky diodes. The arrows in the subscripts indicate the direction of electron flows.

the electron current flowing from Si to metal and from metal to Si, respectively. According to the thermionic emission theory,

$$I_0 = AKT^2 e^{-q\phi_B/kT} \tag{4.18.1}$$

A is the diode area and

$$K = \frac{4\pi q m_n k^2}{h^3} \tag{4.18.2}$$

$K \approx 100$ A/(cm^2/K^2) is known as the **Richardson constant**. In Fig. 4–39b, a positive bias is applied to the metal. $I_{M\to S}$ remains unchanged at $-I_0$ because the barrier against $I_{M\to S}$ remains unchanged at ϕ_B. $I_{S\to M}$, on the other hand, is enhanced by $e^{qV/kT}$ because the barrier is now smaller by qV. Therefore,

$$I_{S\to M} = AKT^2 e^{-(q\phi_B - qV)/kT} = AKT^2 e^{-q\phi_B/kT} e^{qV/kT} = I_0 e^{qV/kT} \tag{4.18.3}$$

$$I = I_{S\to M} + I_{M\to S} = I_0 e^{qV/kT} - I_0 = I_0(e^{qV/kT} - 1) \tag{4.18.4}$$

In summary,

$$\boxed{I = I_0(e^{qV/kT} - 1)} \tag{4.18.5}$$

$$\boxed{I_0 = AKT^2 e^{-q\phi_B/kT}} \tag{4.18.6}$$

Equation (4.18.5) is applicable to the $V < 0$ case (reverse bias, Fig. 4–39c) as well. For a large negative V, Eq. (4.18.5) predicts $I = -I_0$. Figure 4–39c explains why: $I_{S \to M}$ is suppressed by a large barrier, while $I_{M \to S}$ remains unchanged at $-I_0$. Equation (4.18.5) is qualitatively sketched in Fig. 4–39d. I_0 may be extracted using Eq. (4.18.5) and the IV data. From I_0, ϕ_B can be determined using Eq. (4.18.6).

The similarity between the Schottky diode IV and the PN junction diode IV is obvious. The difference will be discussed in Section 4.19.

4.19 • APPLICATIONS OF SCHOTTKY DIODES •

Although Schottky and PN diodes follow the same IV expression

$$I = I_0(e^{qV/kT} - 1)\,, \tag{4.19.1}$$

I_0 of a silicon Schottky diode can be 10^3–10^8 times larger than a typical PN junction diode, depending on ϕ_B (i.e., the metal employed). A smaller ϕ_B leads to a larger I_0. A larger I_0 means that a smaller forward bias, V, is required to produce a given diode current as shown in Fig. 4–40.

This property makes the Schottky diode the preferred rectifier in low-voltage and high-current applications where even a ~0.8 V forward-voltage drop across a PN junction diode would produce an undesirably large power loss. Figure 4–41 illustrates the switching power supply as an example. After the utility power is rectified, a 100 kHz pulse-width modulated (square-wave) AC waveform is produced so that a small (lightweight and cheap) high-frequency transformer can down-transform the voltage. This low-voltage AC power is rectified with Schottky diode (~0.3 V forward voltage drop) and filtered to produce the 50 A, 1 V, 50 W DC

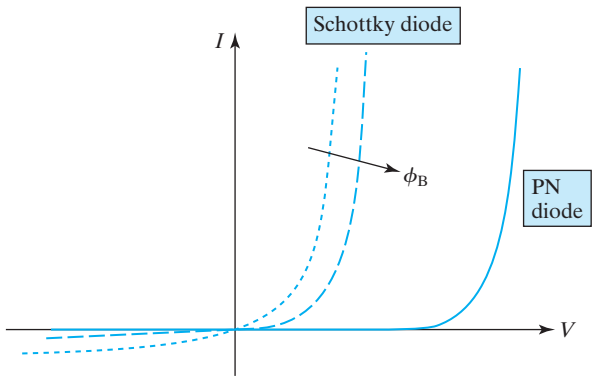


FIGURE 4–40 Schematic IV characteristics of PN and Schottky diodes having the same area.

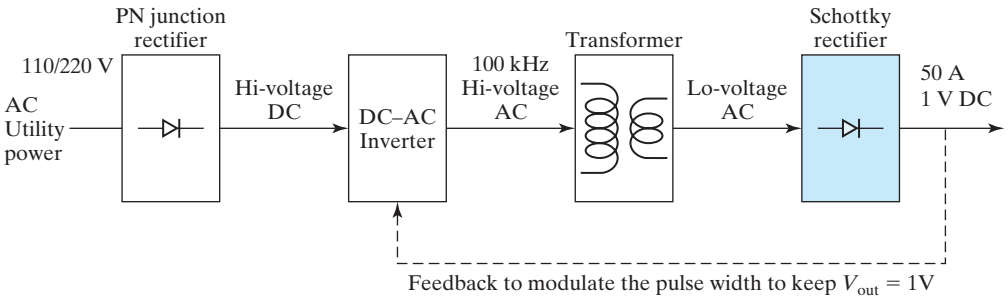


FIGURE 4–41 Block diagram of a switching power supply for electronic equipment such as PCs.

output. If a PN diode with 0.8 V forward voltage drop is used, it would consume 40 W (50 A × 0.8 V) of power and require a larger fan to cool the equipment.

For this application, a Schottky contact with a relatively small ϕ_B would be used to obtain a large I_0 and a small forward voltage drop. However, ϕ_B cannot be too small, or else the large I_0 will increase the power loss when the diode is reverse biased and can cause excessive heat generation. The resultant rise in temperature will further raise I_0 [Eq. (4.18.1)] and can lead to **thermal runaway**.

• The Transistor as a Low Voltage-Drop Rectifier •

Even a Schottky diode's forward voltage may be too large when the power-supply output voltage is, say, 1V. One solution is to replace the diode with a MOSFET transistor [9]. A MOSFET is essentially an on–off switch as shown in Fig. 6–2. A low-power circuit monitors the voltage polarity across the transistor and generates a signal to turn the switch (transistor) on or off. In this way, the transistor, with the control circuit, functions as a rectifier and is called a **synchronous rectifier**. The MOSFET in this application would have a very large channel width in order to conduct large currents. The important point to note is that a MOSFET is not subjected to the same trade-off between the reverse leakage current and forward voltage drop as a diode [Eq. (4.18.5)].

The second difference between a Schottky diode and a PN junction diode is that the basic Schottky diode operation involves only the majority carriers (only electrons in Fig. 4–39, for example). There can be negligible minority carrier injection at the Schottky junction (depending on the barrier height). Negligible injection of minority carriers also means negligible storage of excess minority carriers (see Section 4.10). Therefore, Schottky diodes can operate at higher frequencies than PN junctiondiodes.

Schottky junction is also used as a part of a type of GaAs transistor as described in Section 6.3.2.

4.20 • QUANTUM MECHANICAL TUNNELING •

Figure 4–42 illustrates the phenomenon of **quantum mechanical tunneling**. Electrons, in quantum mechanics, are represented by traveling waves. When the electrons arrive at a potential barrier with potential energy (V_H) that is higher than the electron

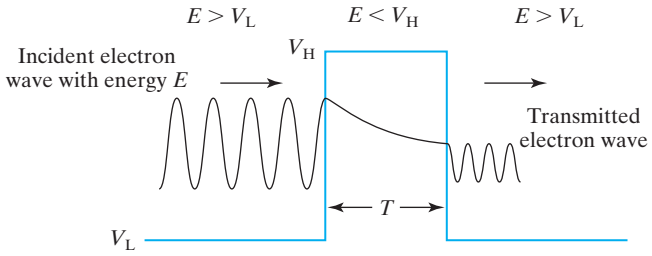


FIGURE 4–42 Illustration of quantum mechanical tunneling.

energy (E), the electron wave becomes a decaying function. Electron waves will emerge from the barrier as a traveling wave again but with reduced amplitude. In other words, there is a finite probability for electrons to tunnel through a potential barrier. The **tunneling probability** increases exponentially with decreasing barrier thickness [10] as

$$P \approx \exp\left(-2T\sqrt{\frac{8\pi^2 m}{h^2}(V_H - E)}\right) \tag{4.20.1}$$

where m is the effective mass and h is the Planck's constant. This theory of tunneling will be used to explain the ohmic contact in the next section.

4.21 • OHMIC CONTACTS •

Semiconductor devices are connected to each other in an integrated circuit through metal. The semiconductor to metal contacts should have sufficiently low resistance so that they do not overly degrade the device performance. Careful engineering is required to reach that goal. These low-resistance contacts are called **ohmic contacts**. Figure 4–43 shows the cross-section of an ohmic contact. A surface layer of a heavily doped semiconductor diffusion region is converted into a silicide such as $TiSi_2$ or $NiSi_2$ and a dielectric (usually SiO_2) film is deposited.

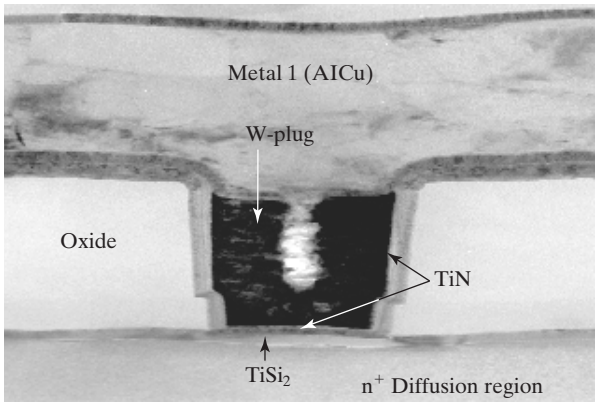


FIGURE 4–43 A contact structure. A film of metal silicide is formed before the dielectric-layer deposition and contact-hole etching. (From [11]. © 1999 IEEE.)

Lithography and plasma etching are employed to produce a contact hole through the dielectric reaching the silicide. A thin conducting layer of titanium nitride (TiN) is deposited to prevent reaction and interdiffusion between the silicide and tungsten. Tungsten is deposited by CVD to fill the contact hole. Figure 4–43 also shows what goes on top of the W plug: another layer of TiN and a layer of AlCu as the interconnect metal material.

An important feature of all good ohmic contacts is that the semiconductor is very heavily doped. The depletion layer of the heavily doped Si is only tens of Å thin because of the high dopant concentration.

When the potential barrier is very thin, the electrons can pass through the barrier by tunneling with a larger tunneling probability as shown in Fig. 4–44. The tunneling barrier height, $V_H - E$ in Eq. (4.20.1) is simply ϕ_{Bn}. The barrier thickness T may be taken as

$$T \approx W_{dep}/2 = \sqrt{\varepsilon_s \phi_{Bn}/(2qN_d)} \tag{4.21.1}$$

$$P \approx e^{-H\phi_{Bn}/\sqrt{N_d}} \tag{4.21.2}$$

$$H \equiv \frac{4\pi}{h}\sqrt{(\varepsilon_s m_n)/q} \tag{4.21.3}$$

At $V = 0$, $J_{S\to M}$ and $J_{M\to S}$ in Fig. 4–44a are equal but of opposite signs so that the net current is zero.

$$J_{S\to M}(=-J_{M\to S}) \approx \frac{1}{2}qN_d v_{thx} P \tag{4.21.4}$$

Only half of the electrons in the semiconductor, with density $N_d/2$, are in thermal motion toward the junction. The other half are moving away from the junction. v_{thx} may be found in Eq. (4.17.2). Assuming that $N_d = 10^{20}$ cm^3, P would be about 0.1and $J_{S\to M} \approx 10^8$ A/cm^2. (This is a very large current density.) If a small voltage is applied across the contact as shown in Fig. 4–44b, the balance between $J_{S\to M}$ and $J_{M\to S}$ is broken. The barrier for $J_{M\to S}$ is reduced from ϕ_{Bn} to $(\phi_{Bn} - V)$.

$$J_{S\to M} = \frac{1}{2}qN_d v_{thx}\, e^{-H(\phi_{Bn}-V)/\sqrt{N_d}} \tag{4.21.5}$$

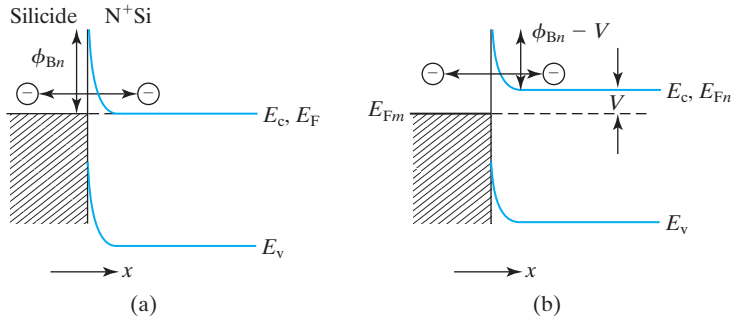


FIGURE 4–44 (a) Energy band diagram of metal–N$^+$ Si contact with no voltage applied and (b) the same contact with a voltage, V, applied to the contact.

At small V, the net current density is

$$J \approx \left.\frac{\mathrm{d}J_{S\to M}}{\mathrm{d}V}\right|_{V=0} \cdot V = V \cdot \frac{1}{2} q v_{\mathrm{thx}} H \sqrt{N_d}\, e^{-H\phi_{\mathrm{B}n}/\sqrt{N_{\mathrm{d}}}} \tag{4.21.6}$$

$$R_{\mathrm{c}} \equiv \frac{V}{J} = \frac{2 \cdot \mathrm{e}^{H\phi_{\mathrm{B}n}/\sqrt{N_{\mathrm{d}}}}}{q v_{\mathrm{thx}} H \sqrt{N_{\mathrm{d}}}} \tag{4.21.7}$$

$$\propto \mathrm{e}^{H\phi_{\mathrm{B}n}/\sqrt{N_{\mathrm{d}}}} \tag{4.21.8}$$

R_{c} is the **specific contact resistance** ($\Omega\ \mathrm{cm}^2$), the resistance of a 1 cm^2 contact. Of course, Eq. (4.21.8) is applicable to P^+ semiconductor contacts if $\phi_{\mathrm{B}n}$, m_n, and N_{d} are replaced by $\phi_{\mathrm{B}p}$, m_p, and N_{a}. Figure 4–45 shows the IV characteristics of a silicide–Si contact. The IV relationship is approximately linear, or ohmic in agreement with Eq. (4.21.6). The resistance decreases with increasing temperature in qualitative agreement with Eq. (4.21.7), due to increasing thermal velocity, v_{thx}. The contact resistance is 140 Ω and $R_{\mathrm{c}} \approx 10^7\ \Omega\ \mathrm{cm}^2$. The R_{c} model embodied in Eq. (4.2.7) is qualitatively accurate, but B and H are usually determined experimentally[9] [11]. R_{c} calculated from a more complex model is plotted in Fig. 4–46. If we want to keep the resistance of a 30 nm diameter contact below 1 kΩ, R_{c} should be less than $7 \times 10^{-10}\ \Omega\ \mathrm{cm}^2$. This will require a very high doping concentration and a low ϕ_{B}. Perhaps two different silicides will be used for N^+ and P^+ contacts, since a single metal cannot provide a low $\phi_{\mathrm{B}n}$ *and* a low $\phi_{\mathrm{B}p}$.

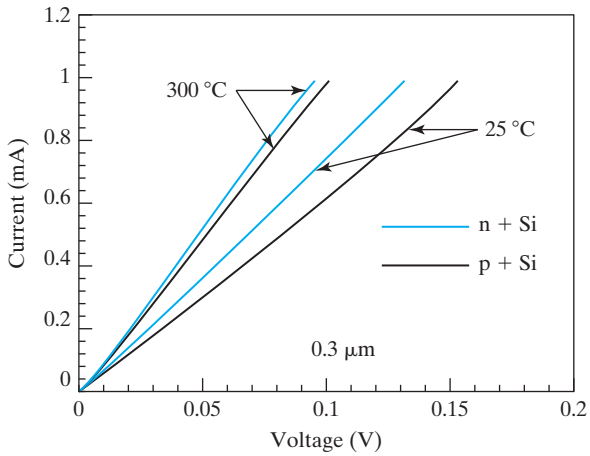


FIGURE 4–45 The IV characteristics of a 0.3 μm (diameter) $TiSi_2$ contact on N^+-Si and P^+-Si. (From [11] ©1999 IEEE.)

● Boundary Condition at an Ohmic Contact ●

The voltage across an ideal ohmic contact is zero. This means that the Fermi level cannot deviate from its equilibrium position, and therefore $n' = p' = 0$ at an ideal ohmic contact.

[9] The electron effective mass in Eq. (4.21.2) is not equal to m_n (effective mass of electron in the conduction band) while it is tunneling under the barrier (in the band gap). Also Eq. (4.17.2) overestimates v_{thx} for a heavily doped semiconductor, for which the Boltzmann approximation is not valid.

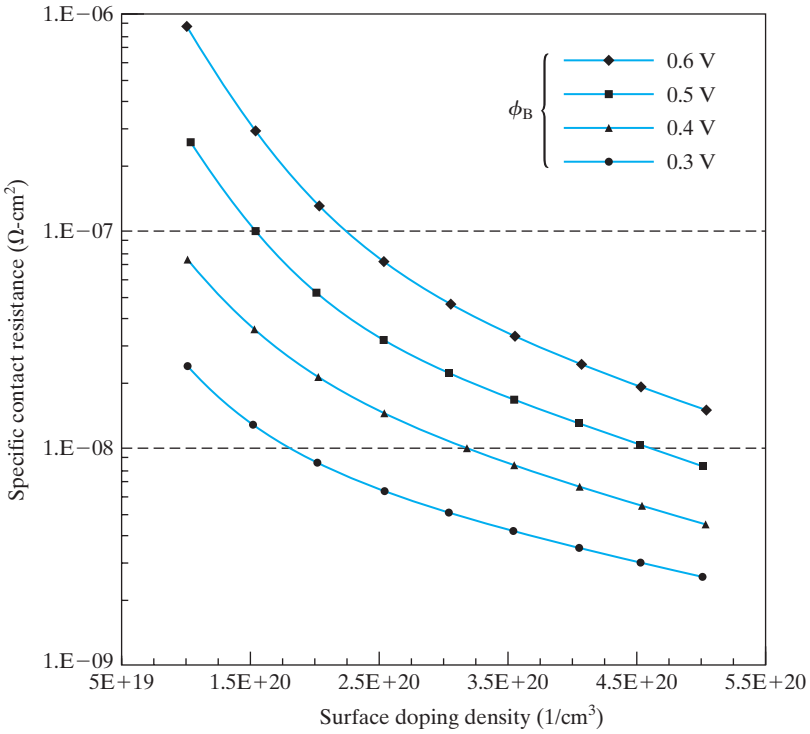


FIGURE 4–46 Theoretical specific contact resistance. (After [12].)

4.22 • CHAPTER SUMMARY •

PART I: PN JUNCTION

It is important to know how to draw the energy band diagram of a PN junction. At zero bias, the potential barrier at the junction is the *built-in potential*,

$$\phi_{bi} = \frac{kT}{q}\ln\frac{N_d N_a}{n_i^2} \tag{4.1.2}$$

The potential barrier increases beyond ϕ_{bi} by 1V if a 1V reverse bias is applied and decreases by 0.1V if a 0.1V forward bias is applied.

The width of the depletion layer is

$$W_{dep} = \sqrt{\frac{2\varepsilon_s \times \text{potential barrier}}{qN}} \tag{4.3.1}$$

N is basically the smaller of the two doping concentrations. The main significance of W_{dep} is that it determines the *junction capacitance*.

$$C_{dep} = A\frac{\varepsilon_s}{W_{dep}} \tag{4.4.1}$$

In general, C_{dep} should be minimized because it contributes to the capacitive loading that slows down the circuit. The way to reduce C_{dep} is to reduce the capacitor area or doping concentrations. Applying a reverse bias will also reduce C_{dep} because W_{dep} increases.

Under forward bias, electrons are injected from the N side to the P side and holes are injected from the P side to the N side. This is called *minority carrier injection*. $E_{\text{F}n}$ is flat from the N region through the depletion layer up to the beginning of the neutral P region. This and similar consideration for $E_{\text{F}p}$ lead to the quasi-equilibrium boundary condition of minority carrier densities:

$$\begin{aligned} n(x_{\text{P}}) &= n_{\text{P0}} e^{qV/kT} \\ p(x_{\text{N}}) &= p_{\text{N0}} e^{qV/kT} \end{aligned} \tag{4.6.2}$$

"Quasi-equilibrium" refers to the fact that $E_{\text{F}n}$ and $E_{\text{F}p}$ are flat across the depletion layer so that the electrons and the holes are separately at equilibrium within each species. Equation (4.6.2) states that *more minority carriers are injected into the lighter-doping side.*

The steady-state *continuity equations* for minority carriers are

$$\frac{d^2 p'}{dx^2} = \frac{p'}{L_p^2}, \qquad L_p \equiv \sqrt{D_p \tau_p} \tag{4.7.5}$$

$$\frac{d^2 n'}{dx^2} = \frac{n'}{L_n^2}, \qquad L_n \equiv \sqrt{D_n \tau_n} \tag{4.7.7}$$

The injected minority carriers diffuse outward from the edges of the depletion layer and decay exponentially with distance due to recombination in the manner of $e^{-|x|/L_p}$ and $e^{-|x|/L_n}$. L_p and L_n are the diffusion lengths.

$$I = I_0(e^{qV/kT} - 1) \tag{4.9.4}$$

$$I_0 = Aqn_{\text{i}}^2\left(\frac{D_p}{L_p N_{\text{d}}} + \frac{D_n}{L_n N_{\text{a}}}\right) \tag{4.9.5}$$

The charge storage concept can be expressed as

$$Q = I\tau_s \tag{4.10.2}$$

The storage charge gives rise to a diffusion capacitance under a forward current, I_{DC},

$$C = \tau_s G \tag{4.11.2}$$

where the small-signal conductance, G, is

$$G = I_{\text{DC}} / \frac{kT}{q} \tag{4.11.1}$$

PART II: APPLICATION TO OPTOELECTRONIC DEVICES

Solar cells convert light into electricity through a simple PN junction. To make the photovoltaic technology more competitive against the fossil-fuel based and other renewable energy technologies, its energy conversion efficiency and cost should be improved.

Low E_g semiconductors can collect larger portions of the solar spectrum and produce larger currents, while large E_g semiconductors can produce larger voltages. The highest theoretical energy conversion efficiency of around 24% is obtained with E_g in the range of 1.2 eV to 1.9 eV. Tandem solar cells stack multiple cells made of different E_gs can achieve even higher efficiency.

The low cost of silicon makes it a favorite solar cell material. Silicon is an indirect-gap semiconductor. Direct-gap semiconductors can collect light in a thin layer of materials and offer two potential cost advantages. First, a smaller quantity (thinner layer) of the semiconductor is needed. Second, the material purity requirement may be lower since a long diffusion length is not needed to collect the carriers generated by light at distances far from the PN junction. Low-cost organic or inorganic solar cells with high conversion efficiency and low installation cost would be an ideal renewable and carbon-emission-free electricity source.

LED generates light with photon energies about equal to the band gap energy when the injected carriers recombine in a forward biased PN junction diode. LEDs are used in signal lights, optical data links, and back lighting for LCD displays. Their potentially most important application may be space lighting replacing the incandescent lamps that are up to 10 times less efficient and fluorescent lamps that contain mercury.

In the most advanced LED, nearly every electron–hole pair recombination produces a photon, i.e., the internal quantum efficiency is 100%. This is achieved with the use of direct-gap semiconductors in which the radiative recombination lifetime is much shorter than the nonradiative recombination lifetime. The external quantum efficiency is raised by employing transparent substrates and reflectors in the back and sides of LED. The PN junction is produced in a thin film of a semiconductor having the desired band gap, which determines the emission wavelength or color. The thin film is epitaxially grown over a low cost and preferably transparent substrate. The suitable substrate materials are few. Low defect epitaxial growth requires the matching of crystal lattice constants of the substrate and the thin film. The thin film is often a quaternary compound semiconductor. Varying the composition of the compound can achieve the goals of tuning its band gap and tuning its lattice constant.

Lasers are optical oscillators. They are based on optical amplification and optical feedback. Both optical amplification and feedback can be achieved in a compact PN diode structure. A large forward bias voltage that exceeds E_g/q produces *population inversion* in a PN junction. A light wave passing through the diode under population inversion is amplified through *stimulated emission*. The amplified light retains the exact wavelength and direction of the original light wave. Population inversion can be achieved with a small forward current using the *quantum well* structure with a lower band gap semiconductor sandwiched between two wider band gap materials. The optical feedback can be provided

with multi-layer *Bragg reflectors*. Diode lasers are widely used in CD and DVD readers and writers and fiber-optic communication systems.

PART III: METAL–SEMICONDUCTOR JUNCTION

The Fermi level at the metal–silicon interface is located at 0.3–0.9 eV below E_c, depending on the metal material. A low work function metal or silicide provides a low Schottky barrier height for electrons, ϕ_{Bn}, and a large barrier height for holes, ϕ_{Bp}.

A junction between metal and lightly doped silicon usually has rectifying *IV* characteristics, and is called a **Schottky diode**. The sense of the forward/reverse bias of a metal/N-type Si diode is the same as that of a P^+N junction diode. A metal on a P-silicon contact has the same forward/reverse sense as an N^+P junction diode. Compared with PN junction diodes, typical Schottky diodes have much larger reverse saturation currents, which are determined by the Schottky barrier height.

$$I_0 = AKT^2 e^{-q\phi_B/kT} \tag{4.18.6}$$

where $K = 100$ A/cm^2/K^2. Due to the larger I_0, Schottky diodes require lower forward voltages to conduct a given current than PN junction diodes. They are often used as rectifiers in low-voltage power supplies.

Low-resistance ohmic contacts are critical to the performance of high-current devices. The key ingredient of an ohmic contact is heavy doping of the semiconductor. A low Schottky barrier height is also desirable. The carrier transport mechanism is quantum-mechanical tunneling. The contact resistance of a metal/N-Si ohmic contact is

$$R_c \propto e^{-\left(\frac{4\pi}{h}\phi_B\sqrt{(\varepsilon_s m_n)/(qN_d)}\right)} \tag{4.21.8}$$

PROBLEMS

Part I: PN Junction

Electrostatics of PN Junctions

4.1 Applying the depletion approximation to a linearly graded junction with $N_d - N_a = ax$, derive expressions for

(a) the electric field distribution,

(b) the potential distribution,

(c) the built-in potential, and

(d) the depletion-layer width.

4.2 Consider a silicon PN step junction diode with $N_d = 10^{16}\text{cm}^{-3}$ and $N_a = 5 \times 10^{15}\text{cm}^{-3}$. Assume $T = 300$ K.

(a) Calculate the built-in potential ϕ_{bi}.

(b) Calculate the depletion-layer width (W_{dep}) and its length on the N side (x_n) and P side (x_p).

(c) Calculate the maximum electric field.

(d) Sketch the energy band diagram, electric potential, electric field distribution, and the space-charge profile.

(e) Now let $N_a = 10^{18}\text{cm}^{-3}$. Repeat (a), (b), and (c). Compare these to the previous results. How have the depletion widths changed?

4.3 Consider the silicon PN junction in Fig. 4–47.

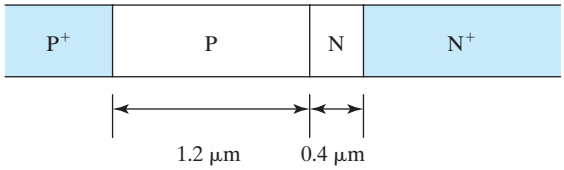


FIGURE 4–47

(a) If $N_a = 5 \times 10^{16}\text{cm}^{-3}$ in the P region and $N_d = 1 \times 10^{17}\text{cm}^{-3}$ in the N region, under increasing reverse bias, which region (N or P) will become completely depleted first? What is the reverse bias at this condition? (Hint: use $N_a x_p = N_d x_n$. The doping densities of P^+ and N^+ are immaterial).

(b) Repeat part (a) with $N_a = 1 \times 10^{16}\text{cm}^{-3}$ and $N_d = 1 \times 10^{17}\text{cm}^{-3}$.

(c) What are the small-signal capacitances (F/cm²) at the bias conditions in (a) and (b)?

4.4 A silicon sample maintained at 300 K is characterized by the energy band diagram in Fig. 4–48:

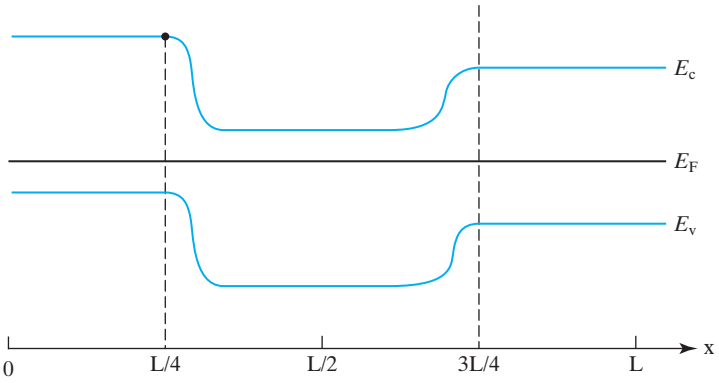


FIGURE 4–48

(a) Does the equilibrium condition prevail? How do you know?

(b) Roughly sketch n and p versus x.

(c) Sketch the electrostatic potential (Φ) as a function of x.

(d) Assume that the carrier pictured on Fig. 4–48 by the dot may move without changing its total energy. Sketch the kinetic and potential energies of the carrier as a function of its position x.

4.5

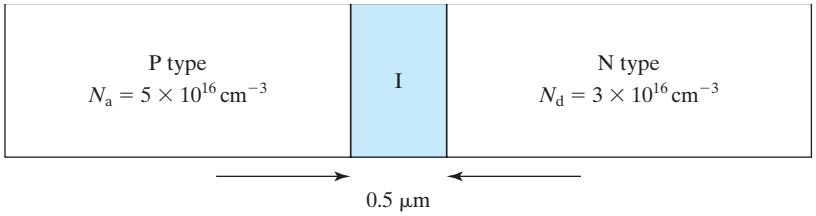


FIGURE 4–49

Consider the P-I-N structure shown in (Fig. 4–49). The I region is intrinsic. Determine the quantities in (a) and (c). Assume that no bias is applied. (Hint: It may be helpful to think of the I region as a P or N and then let the doping concentration approach zero. That is, $N_d \cong N_a \cong 0$.)

(a) Find the depletion-layer width (W_{dep}) and its widths on the N side (x_n) and the P side.

(b) Calculate the maximum electric field.

(c) Find the built-in potential.

(d) Now assume that a reverse bias is applied. If the critical field for breakdown in silicon is 2×10^5 V/cm, compare the breakdown voltages between the P-I-N structure and a P-N structure (without the I region) with the doping levels shown above.

If interested, you can find more P-I-N diode examples at http://jas.eng.buffalo.edu/education/pin/pin2/index.html.

• Diffusion Equation •

4.6 Consider a piece of infinitely long semiconductor sample shown in Fig. 4–50.

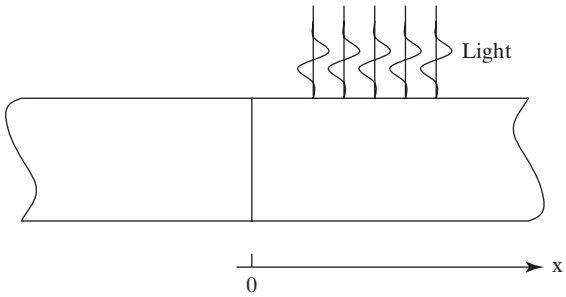


FIGURE 4–50

The $x > 0$ portion is illuminated with light. The light generates $G_L = 10^{15}$ electron–hole pairs per cm^2 per s uniformly throughout the bar in the region $x > 0$. G_L is 0 for $x < 0$. Assume that the steady-state conditions prevail, the semiconductor is made of silicon, $N_d = 10^{18}\,cm^{-3}$, $\tau = 10^{-6}$ s, and $T = 300$ K.

(a) What is the hole concentration at $x = \infty$? Explain your answer.

(b) What is the hole concentration at $x = +\infty$? Explain your answer.

(c) Do low-level injection conditions prevail? Explain your answer.

(d) Determine $p'(x)$ for all x, where $p'(x)$ is the excess minority carrier concentration. (Hint: Solve the continuity equation.)

4.7

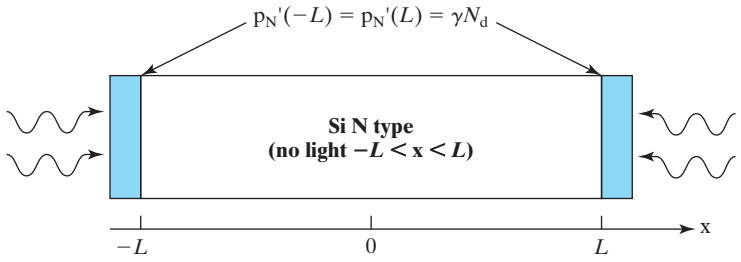


FIGURE 4–51

The two ends of a uniformly doped N-type silicon bar of length 2L are simultaneously illuminated (Fig. 4–51) so as to maintain $p' = \gamma N_d$ excess hole concentration at both $x = -L$ and $x = L$. L is the hole diffusion length. The wavelength and intensity of the illumination are such that no light penetrates into the interior ($-L < x < L$) of the bar and $\gamma = 10^{-3}$. Assume the steady-state conditions, $T = 300$ K, $N_d >> n_i$, and minority carrier lifetime of τ.

(a) Is the silicon bar at thermal equilibrium near $x = 0$? Why or why not?

(b) What are the excess concentrations of holes (p_N') and electron (n_N') at $x = -L$? What are the total electron and hole concentrations at $x = -L$?

(c) Do low-level injection conditions prevail inside the bar? Explain your answer.

(d) Write down the differential equation that you need to solve to determine $p_N'(x)$ inside the bar.

(e) Write down the general form of the $p_N'(x)$ solution and the boundary condition(s) appropriate for this particular problem.

4.8 Consider a P^+N junction diode with $N_d = 10^{16}\ cm^{-3}$ in the N region.

(a) Determine the diffusion length L on the N-type side.

(b) What are the excess hole density and excess electron density at the depletion-layer edge on the N-type side under (a) equilibrium and (b) forward bias $V = 0.4$ V?

4.9 Consider an ideal, silicon PN junction diode with uniform cross section and constant doping on both sides of the junction. The diode is made from 1 Ωcm P-type and 0.2 Ωcm N-type materials in which the recombination lifetimes are $\tau_n = 10^{-6}$ s and $\tau_p = 10^{-8}$ s, respectively.

(a) What is the value of the built-in voltage?

(b) Calculate the density of the minority carriers at the edge of the depletion layer when the applied voltage is 0.589 V (which is $23 \times kT/q$).

(c) Sketch the majority and minority carrier current as functions of distance from the junction on both sides of the junction, under the bias voltage of part (b).

(d) Calculate the location(s) of the plane (or planes) at which the minority carrier and majority carrier currents are equal in magnitude.

4.10

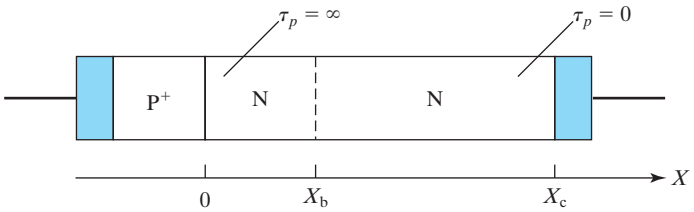


FIGURE 4–52

Consider the silicon P^+N junction diode pictured in Fig. 4–52. $\tau_p = \infty$ for $0 \le x \le x_b$ and $\tau_p = 0$ for $x_b \le x \le x_c$. Excluding biases that would cause high-level injection or breakdown, develop an expression for the IV characteristic of the diode. Assume the depletion-layer width (W_{dep}) never exceeds x_b for all biases of interest. The tinted regions are simply the metal contacts.

● Proof of Minority Drift Current Being Negligible ●

4.11 Consider an ideal, long-base, silicon abrupt P^+N junction diode with uniform cross section and constant doping on either side of the junction. The diode is made from a heavily doped P-type material and 0.5Ωcm N-type materials in which the minority carrier lifetime is $\tau_p = 10^{-8}$ s.

Answer the following questions on the n side of the junction only.

(a) Calculate the density of the minority carriers as a function of x (distance from the junction) when the applied voltage is 0.589 V (which is $23 \times kT/q$).

(b) Find and sketch the majority and minority carrier currents as functions of x (distance from the junction), under the applied bias voltage of part (a).

(c) What is the majority carrier diffusion current as a function of x?

The purpose of the following questions is to show that the minority drift current is negligible.

(d) Use the results of parts (b) and (c) to find the majority carrier drift current, $J_{n\text{drift}}$. Then find electric field $\mathscr{E}(x)$, and finally the minority drift current $J_{p\text{drift}}$. Is $J_{p\text{drift}} << J_{p\text{diff}}$? Sketch $J_{p\text{drift}}$ and $J_{p\text{diff}}$ in the same graph.

(e) Justify the assumption of $n' = p'$.

● Temperature Effect on IV ●

4.12 The forward-bias voltage (V) required to maintain a PN diode current (I) is a function of the temperature (T).

(a) Derive an expression for $\delta V/\delta T$.

(b) What is a typical value for a silicon diode?

(c) Compare the result of (b) with a numerical value extracted from Fig. 4–21.

4.13 Equation (4.9.5) can be interpreted this way: The minority carriers that are thermally generated within the diffusion length from the reverse-biased junction are collected by the junction and are responsible for the reverse-leakage current. Rewrite Eq. (4.9.5) in a way that justifies this interpretation.

4.14 Assume that the neutral regions of a PN diode present a series resistance R such that the voltage across the PN junction is not V but V-RI.

(a) How should Eq. (4.9.4) be modified?

(b) Find an expression of V as a function of I.

(c) Sketch a typical *I–V* curve without *R* for *I* from 0 to 100 mA. Sketch a second *I–V* curve in this figure for $R = 200\ \Omega$ without using a calculator.

● Charge Storage ●

4.15 A PN diode with lengths much larger than the carrier diffusion length such as shown in Fig. 4–18 is called a long-base diode. A short-base diode has lengths much shorter than the diffusion lengths, and its excess carrier concentration is similar to that shown in Fig. 8–6. A uniformly doped short-base Si diode has $N_d = 10^{17}\ cm^{-3}$ and $N_a = 10^{16}\ cm^{-3}$, $\tau_p = \tau_n = 1\ \mu s$, $D_p = 10\ cm^2/s$, $D_n = 30\ cm^2/s$, and cross-sectional area = $10^{-5}\ cm^{-2}$. The length of the quasi-neutral N-type and P-type regions $W_E' = W_B' = 1\ \mu m$. The diode is at room temperature under applied forward bias of 0.5 V. Answer the following questions:

(a) Show that the total current and the sum of the charge stored on both N and P sides of the junction are proportional to each other: $Q_t = I_t \tau_s$.

Find the expression for τ_s. Use the short-base approximation, i.e., assume that the excess minority carrier concentration decreases linearly from its maximum value at the edge of the depletion region to zero at the ohmic contacts at either end of the diode.

(b) τ_s is called the charge-storage time. Show that it is significantly smaller than τ_p and τ_n.

(c) Which diode can operate at a higher frequency, short-based or long-based?

Part II: Application to Optoelectronic Devices

● IV of Photodiode/Solar Cell ●

4.16 Photodiodes and solar cells are both specially designed PN junction diodes packaged to permit light to reach the vicinity of the junction. Consider a P^+N step junction diode where incident light is uniformly absorbed throughout the N region of the device producing photogeneration rate of G_L electron–hole pairs/cm^3s. Assume that low-level injection prevails so that the minority drift current is negligible.

(a) What is the excess minority carrier concentration on the N side at a large distance ($x \to \infty$) from the junction? [Note: $p'(x \to \infty) \neq 0$. Far away from the junction, the recombination rate is equal to the photocarrier generation rate].

(b) The usual quasi-equilibrium boundary conditions still hold at the edges of the depletion layer. Using those and the boundary condition established in part (a), device an expression for the IV characteristic of the P^+N diode under the stated conditions of illumination. Ignore all recombination/generation, including photogeneration, occurring in the depletion layer.

(c) Sketch the general form of the IV characteristics for $G_L = 0$ and $G_L = G_{L0}$. Indicate the voltage developed across the diode when the diode is an open circuit, i.e., $I = 0$. What is the current that will flow when the diode is short circuited, i.e., $V = 0$?

Part III: Metal–Semiconductor Junction

● Ohmic Contacts and Schottky Diodes ●

4.17 Sketch the energy band diagram and comment on whether a very heavy doping is important, unimportant, or unacceptable for

(a) an ohmic contact between P^+-type silicon and $TiSi_2$ at equilibrium,

(b) an ohmic contact between N^+-type silicon and $TiSi_2$, and

(c) a rectifying contact between P-type silicon and $TiSi_2$ under 2 V reverse bias.

4.18 **(a)** Draw the energy band diagram for a metal–semiconductor contact (including the vacuum level) under 0.4 V applied forward bias. The metal has a work function φ_M of 4.8 eV, and the semiconductor is N-type Si with uniform doping concentration of 10^{16} cm^{-3}. Label clearly $q\varphi_M$, $q\phi_{Bn}$, $q(\phi_{bi} + V)$, and χ_{Si} on your sketch. Assume no surface states are present. Find the numerical values for $q\varphi_M$, $q\phi_{Bn}$, $q(\phi_{bi} + V)$.

(b) Sketch the charge density ρ, electric field ε, and potential ϕ, for the device in (a). For each diagram, draw two curves: one for equilibrium case and one for $V = 0.4$ V. No numbers or calculations are required.

4.19 Consider a Schottky diode with the doping profile shown in Fig. 4–53. Assume that the built-in potential ϕ_{bi} is 0.8 V.

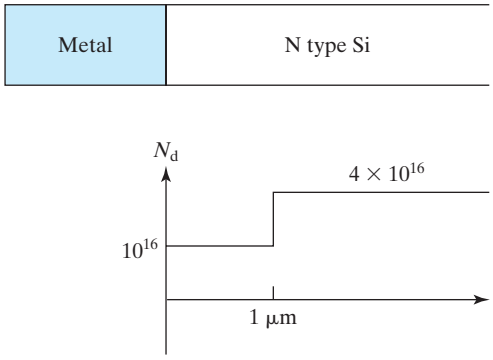


FIGURE 4–53

(a) Sketch $1/C^2$ vs. V (the reverse bias voltage) qualitatively. Do not find numerical values for C.

(b) Sketch the electric field profile for the bias condition when $W_{dep} = 2$ μm. Again, do not find numerical values for the electric field.

(c) What is the potential drop across the junction in part (b)?

(d) Derive an expression of C as a function of V for $W_{dep} > 1$ μm.

• Depletion-Layer Analysis for Schottky Diodes •

4.20 **(a)** Calculate the small signal capacitance at zero bias and 300 K for an ideal Schottky barrier [see Eq. (4.16.2)] between platinum (work function 5.3 eV) and silicon doped with $N_d = 10^{16}$cm^{-3}. The area of the Schottky diode is 10^{-5} cm^2.

(b) Calculate the reverse bias at which the capacitance is reduced by 25% from its zero-bias value.

4.21 The doping profile inside the semiconductor of a Schottky diode is linearly graded, i.e., $N_d(x) = ax$.

Derive expressions for ρ, ε, V, and W_{dep} inside the semiconductor.
Indicate how ϕ_{bi} is to be determined and computed.
Establish an expression for the junction (depletion layer) capacitance.

4.22 A metal/N-type semiconductor Schottky diode has the CV characteristic given in Fig. 4–54.

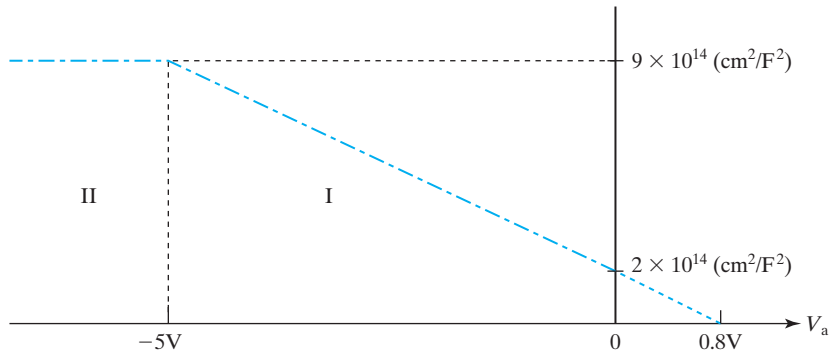


FIGURE 4–54

(a) What is the built-in voltage of the diode (from Region I data)?

(b) Find the doping profile of the N-type semiconductor.

● Comparison Between Schottky Diodes and PN Junction Diodes ●

4.23 **(a)** Qualitatively hand-sketch $\log(I)$ vs. V for a Schottky diode and a PN diode in the same figure. Comment on the similarity and difference.

(b) Calculate the I_0 of a 1 mm^2 $MoSi_2$ on N-type Si Schottky diode. Compare it with the I_0 of a 1 mm^2 P^+N diode with $N_d = 10^{18}\,cm^{-3}$ and τ_p=1 μs.

(c) Compare the forward voltage of the two diodes in (b) at a forward current of 50 A.

(d) Besides increasing the diode area (cost), what can one do to reduce the forward voltage drop of the Schottky diode?

(e) What prevents one from using a Schottky diode having a much smaller ϕ_{Bn}?

● Ohmic Contacts ●

4.24 Consider an aluminum Schottky barrier on silicon having a constant donor density N_d. The barrier height $q\Phi_B$ is 0.65 eV. The junction will be a low-resistance ohmic contact and can pass high currents by tunneling if the barrier presented to the electrons is thin enough. We assume that the onset of efficient tunneling occurs when the Fermi level extrapolated from the metal meets the edge of the conduction band (E_c) at a distance no larger than 10 nm from the interface.

(a) What is the minimum N_d such that this condition would be met at equilibrium?

(b) Draw a sketch of the energy band diagram under the condition of (a).

(c) Assume that N_d is increased four times from (a). By what factor is the tunneling distance (W_{dep}) reduced? And by what factor is R_c reduced?

4.25 A PN diode conducting 1 mA of current has an ohmic contact of area 0.08 μm^2 and surface density of $1 \times 10^{20} cm^{-3}$.

(a) What specific contact resistance can be allowed if the voltage drop at the ohmic contact is to be limited to 50 mV?

(b) Using Eq. (4.21.7), estimate the Φ_{Bn} that is allowed. Is that the maximum or minimum allowable Φ_{Bn}?

(c) Repeat (b), but this time use Fig. 4–46 to estimate Φ_{Bn}. (Fig. 4–46 is based on a more detailed model than Eq. 4.21.7)

4.26 Use Fig. 4-46, which is applicable to ohmic contacts to both N^+ and P^+ silicon, for this problem. Assume the doping concentration is 1.5×10^{20} cm^{-3}.

(a) Estimate the R_c of NiSi contact on N^+ and P^+ silicon.

(b) Estimate the R_c of PtSi contact on P^+ silicon.

(c) Estimate the R_c of $ErSi_{1.7}$ contact on N^+ silicon.

(d) If a contact resistance of $R_c \leq 4 \times 10^{-9} \Omega$ cm^2 is required for contact on both N^+ and P^+ silicon, what silicide(s) and doping concentration(s) would you have to use?

REFERENCES

1. Sze, S. M. *Physics of Semiconductor Devices*, 2nd ed. New York: John Wiley & Sons, 1981, Ch. 2.

2. Muller, R. S., and T. I. Kamins. *Device Electronics for Integrated Circuits*, 2nd ed. New York: John Wiley & Sons, 1986, 194.

3. Kuwano, Y., S. Okamoto, and S. Tsuda. "Semiconductor Devices Save the Earth," *Technical Digest of International Electron Devices Meeting*, (1992), 3–10.

4. Hu, C., and R. M. White. *Solar Cells.* New York: McGraw Hill, 1983.

5. Kalinowski, J. *Organic Light-Emitting Diodes: Principles, Characteristics & Processes.* New York: Marcel Dekker, 2005.

6. Schneider, R. P., and J. A. Lott. "Cavity Design for Improved Electrical Injection in AlGaInP/AlGaAs Visible (639–661 nm) VCSEL Diodes,"*Applied Physics Letter* 63 (1993), 917–919.

7. Beadle, W. E., J. C. Tsai, and R. D. Plummer. *Quick Reference Manual for Silicon Integrated Circuit Technology.* New York: Wiley-Interscience, 1985.

8. Monch, W. "Role of Virtual Gap States and Defects in Metal–Semiconductor Contacts," *Physics Review Letter*, 58 (12), (1987), 1260.

9. Kagen, R., M. Chi, and C. Hu. Improving Switching Power Supply Efficiency by Using MOSFET Synchronous Rectifiers. *Proceedings of Powercon*, 9, (July 1982), 5.

10. Choi, Y.-K., et al. "Ultrathin-body SOI MOSFET for deep-sub-tenth micron era," *IEEE Electron Device Letters*, 21(5), (2000), 254–255.

11. Banerjee, K., A. Amerasekera, G. Dixit, and C. Hu. "Temperature and Current Effect on Small-Geometry-Contact Resistance," *Technical Digest of International Electron Devices Meeting*, (1999), 115.

12. Ozturk, M. C. "Advanced Contact Formation," *Review of SRC Center for Front End Processes* 1999.

GENERAL REFERENCES

1. Muller, R. S., T. I. Kamins, and M. Chen. *Device Electronics for Integrated Circuits*, 3rd ed., New York: John Wiley & Sons, 2003.

2. Streetman, B., and S. K. Banerjee. *Solid State Electronic Devices*, 6th ed. Upper Saddle River, NJ: Prentice Hall, 2006.

3. Sze, S. M. *Semiconductor Devices: Physics and Technology*, 2nd ed. New York: John Wiley & Sons, 2002.