4

PN 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*

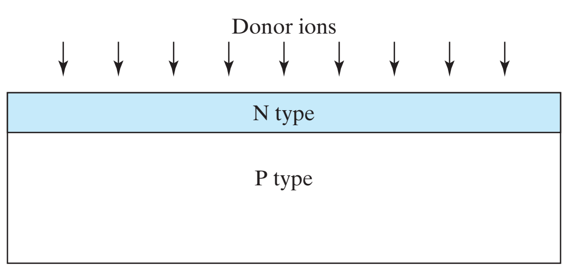

FIGURE 4–1 A PN junction can be fabricated by converting a layer of P-type semiconductor into N-type with donor implantation or diffusion.

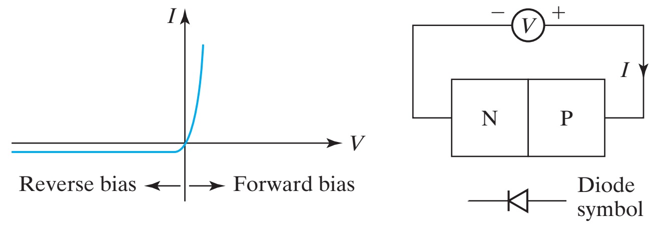

FIGURE 4–2 The rectifying *IV* characteristics of a PN junction.

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.

Hu_ch04v4.fm Page 91 Friday, February 13, 2009 5:54 PM

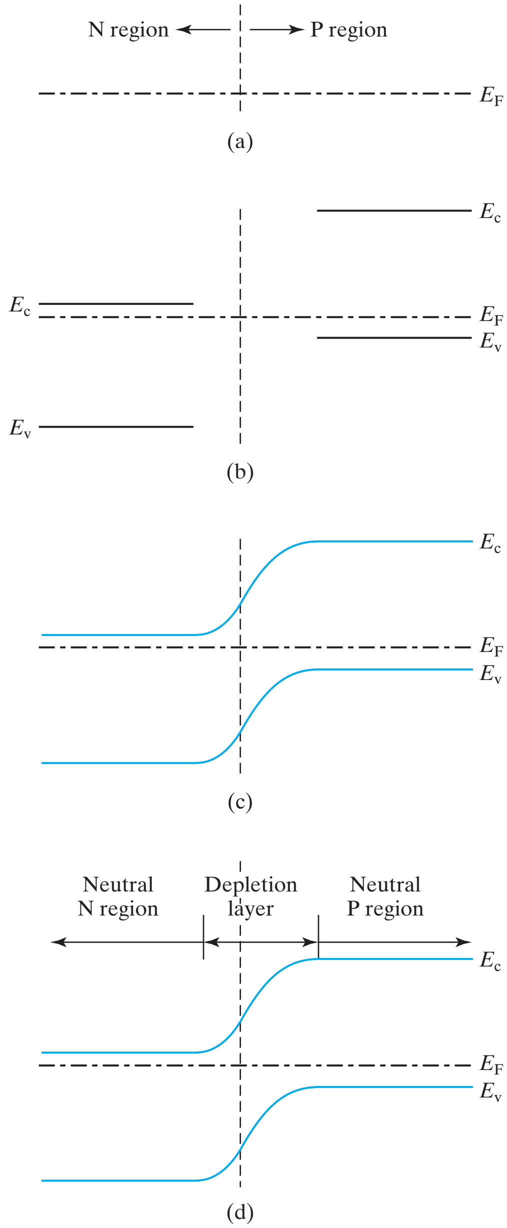

FIGURE 4–3 (a) and (b) Intermediate steps of constructing the energy band diagram of a PN junction. (c) and (d) The complete band diagram.

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

Hu_ch04v4.fm Page 92 Friday, February 13, 2009 5:54 PM

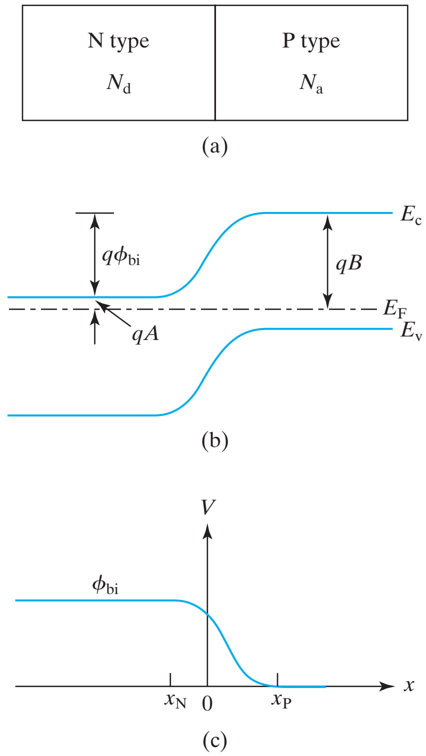

FIGURE 4–4 (a) A PN junction. The built-in potential in the energy band diagram (b) shows up as an upside down mirror image in the potential plot (c).

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

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

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

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

Hu_ch04v4.fm Page 93 Friday, February 13, 2009 5:54 PM

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

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.

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

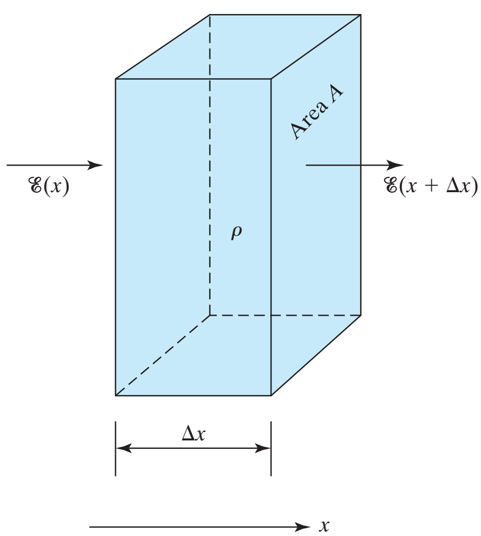

FIGURE 4–5 A small volume in a semiconductor, used to derive the Poisson's equation.

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

Hu_ch04v4.fm Page 95 Friday, February 13, 2009 5:54 PM

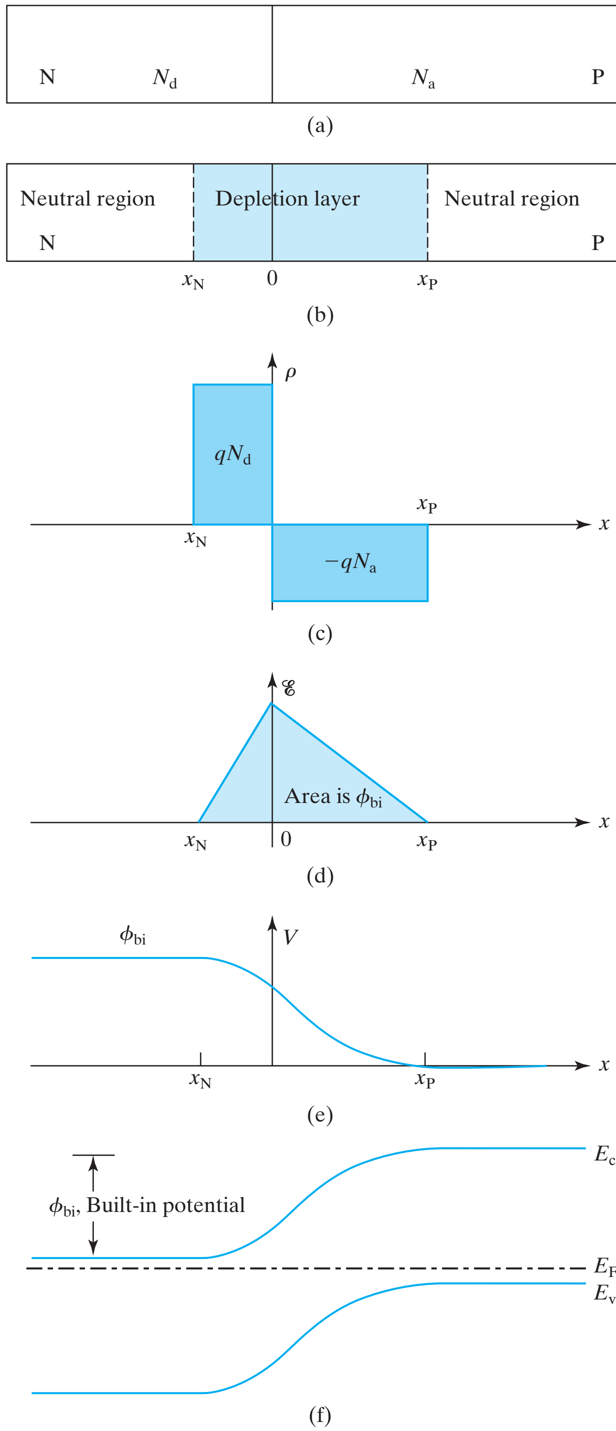

FIGURE 4–6 (a) Step PN junction; (b) depletion approximation; (c) space charge profile; (d) electric field from integration of ρ/ε_s (Poisson's equation); (e) electric potential from integrating $-\mathscr{E}$; and (f) energy band diagram.

Hu_ch04v4.fm Page 96 Friday, February 13, 2009 5:54 PM

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

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

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

$$V(x) = D - \frac{qN_a}{2\varepsilon_s}(x - x_N)^2$$

$$= \phi_{bi} - \frac{qN_d}{2\varepsilon_s}(x - x_N)^2 \qquad x_N \le x \le 0 \qquad (4.2.7)$$

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

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

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

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

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

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

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

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

Hu_ch04v4.fm Page 97 Friday, February 13, 2009 5:54 PM

EXAMPLE 4–1 A P^+N junction has $N_a = 10^{20} \text{cm}^3$ and $N_d = 10^{17} \text{cm}^{-3}$. 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.

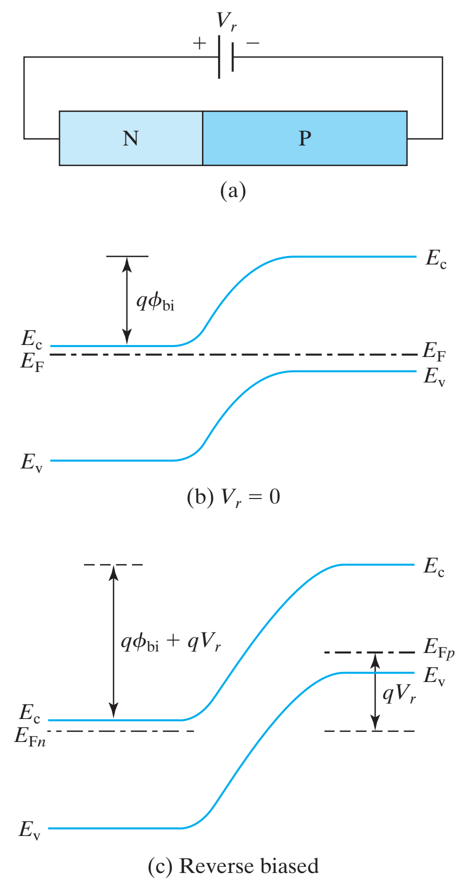

FIGURE 4–7 Reverse-biased PN junction (a) polarity of reverse bias; (b) energy band diagram without bias; and (c) energy band diagram under reverse bias.

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

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

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

$$C_{dep} = A\frac{\varepsilon_s}{W_{dep}} \qquad (4.4.1)$$

Hu_ch04v4.fm Page 99 Friday, February 13, 2009 5:54 PM

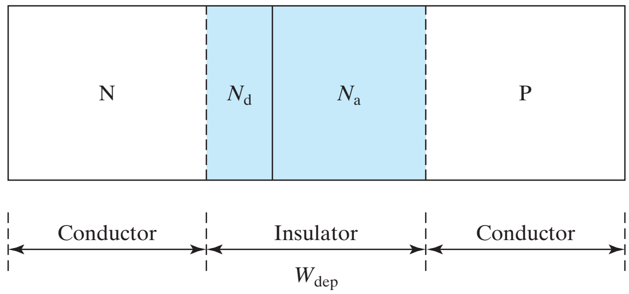

FIGURE 4–8 The PN junction as a parallel-plate capacitor.

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}.

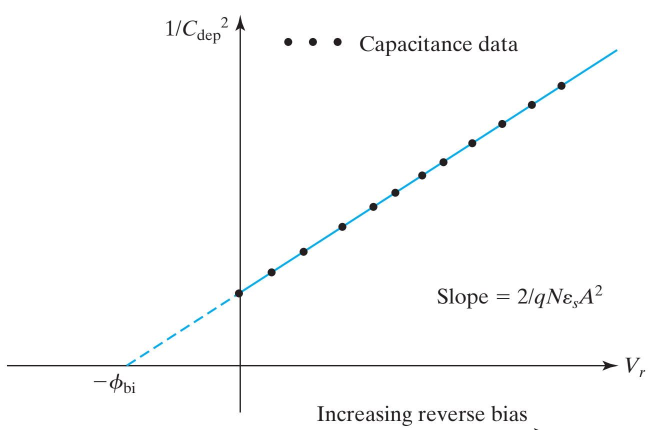

FIGURE 4–9 The common way of plotting the C–V data of a PN junction.