Physics of Semiconductor Devices

Lecture 7

Achintya Dhar

Department of Physics

Email-id: achintya.dhar@gmail.com

Five Basic Equations

The 5 basic equations of semiconductor device physics:

We will in general be faced with finding 5 quantities:

$n(x,t)$, $p(x,t)$, $J_e(x,t)$, $J_h(x,t)$, and $E(x,t)$,

and we have five independent equations that relate them:

$$(1) \qquad J_e(x,t) = q\,\mu_e\,n(x,t)\,E(x,t) + q\,D_e\frac{\partial n(x,t)}{\partial x}$$

$$(2) \qquad J_h(x,t) = q\,\mu_h\,p(x,t)\,E(x,t) - q\,D_h\frac{\partial p(x,t)}{\partial x}$$

$$(3) \qquad \frac{\partial n(x,t)}{\partial t} - \frac{1}{q}\frac{\partial J_e(x,t)}{\partial x} = g_L(x,t) - [n(x,t)\,p(x,t) - n_i^2]\,r$$

$$(4) \qquad \frac{\partial p(x,t)}{\partial t} + \frac{1}{q}\frac{\partial J_h(x,t)}{\partial x} = g_L(x,t) - [n(x,t)\,p(x,t) - n_i^2]\,r$$

$$(5) \qquad \varepsilon\frac{\partial E(x,t)}{\partial x} = q\,[p(x,t) - n(x,t) + N_d(x) - N_a(x)]$$

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}$$

$$\nabla \cdot \mathbf{B} = 0$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

$$J_n = qn\mu_n E + qD_n\nabla n$$

$$J_p = qp\mu_p E - qD_n\nabla p$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot J_n + U_n$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot J_p + U_p$$

$$\frac{\partial n}{\partial t} = n\mu_n \frac{\partial E}{\partial x} + \mu_n E \frac{\partial n}{\partial x} + D_n \frac{\partial^2 n}{\partial x^2} + G_n - \frac{(n - n_0)}{\tau_n}$$

$$\frac{\partial p}{\partial t} = -p\mu_p \frac{\partial E}{\partial x} - \mu_p E \frac{\partial p}{\partial x} + D_p \frac{\partial^2 p}{\partial x^2} + G_p - \frac{(p - p_0)}{\tau_p}$$

$$\frac{\partial p}{\partial t} = 0 = D_p \frac{\partial^2 p}{\partial x^2} - \frac{(p - p_0)}{\tau_p}$$

$$\frac{d^2 p}{dx^2} = \frac{(p - p_0)}{D_p \tau_p} = \frac{(p - p_0)}{L_p^2}$$

where the assumptions we made getting these equations are:

$$N_d^+(x) \approx N_d(x), \quad N_a^-(x) \approx N_a(x), \quad R(x,t) \approx n(x,t)\,p(x,t)\,r$$

r, ε, μ_e, μ_h, D_e, and D_h are assumed to be independent of position.

Temperature is assumed to be constant (isothermal).
(Note: n_i, r, μ_e, μ_h, D_e, and D_h all depend on temperature).

With an understanding of these solutions to the five equations we will be able to model and understand all of the important semiconductor devices, including diodes, bipolar transistors, and MOSFETs.

Diodes involve flow problems in two regions and the depletion approximation about one junction:
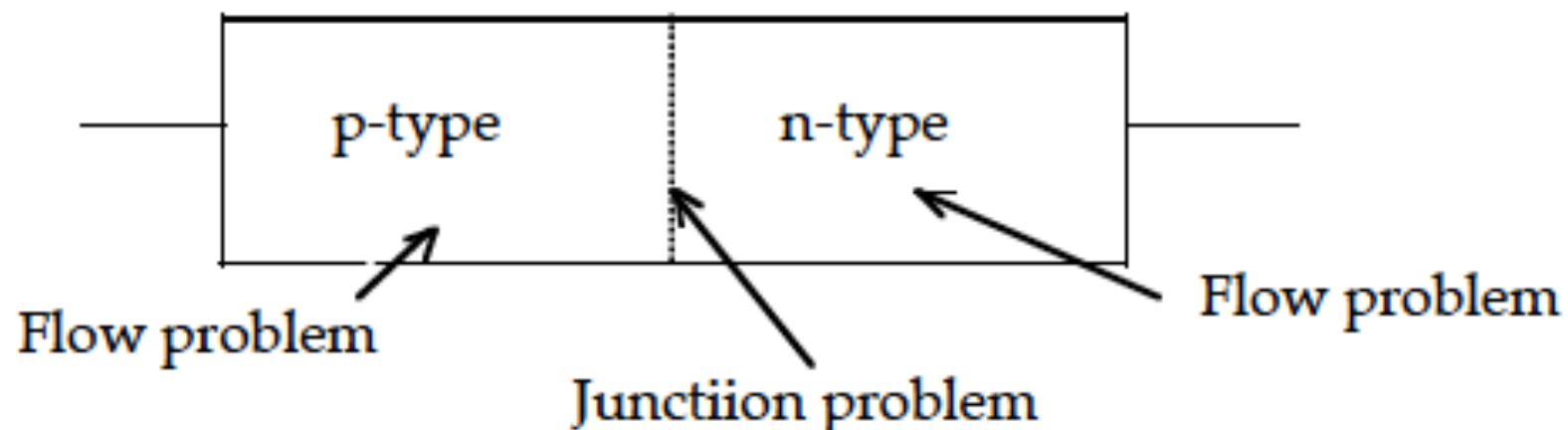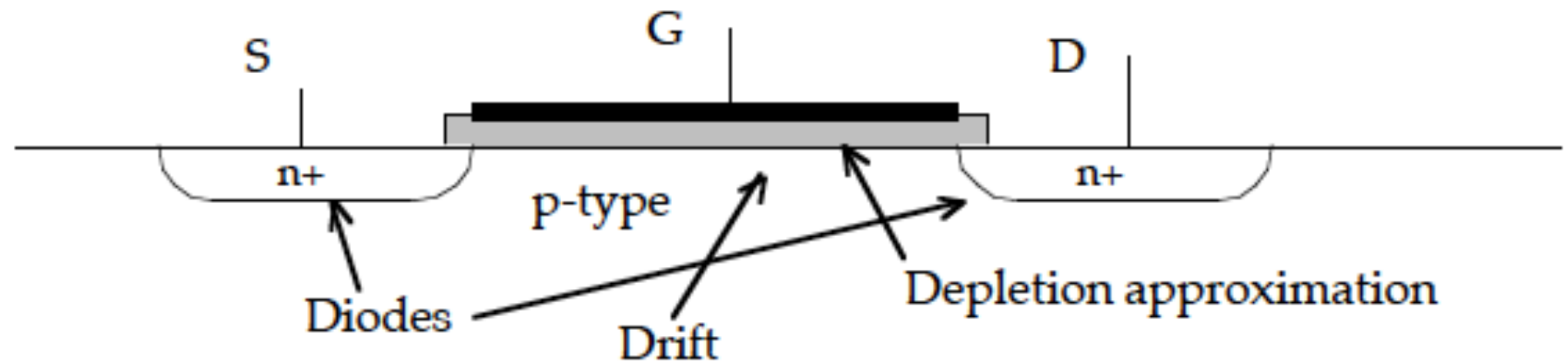

Note: This is true not only for simple electronic diodes, but also for light emitting and laser diodes, and for photodiodes and solar cells.

Bipolar transistors involve flow problems in three regions and the depletion approximation about two junctions:
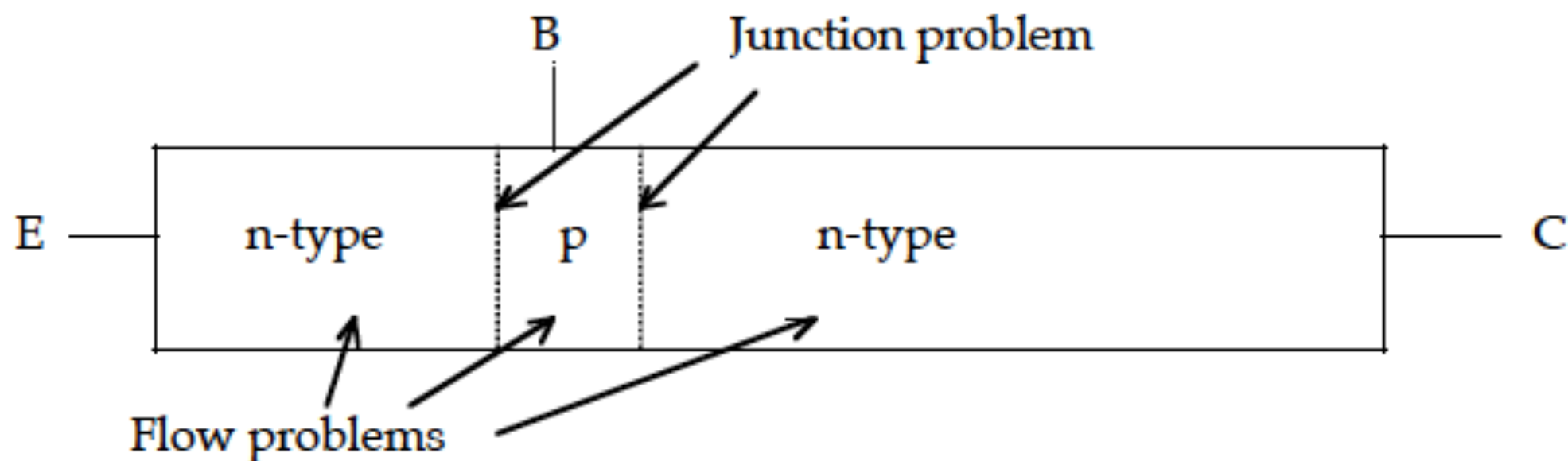
MOSFETs involve two diodes, the depletion approximation in the gate region, and drift in the channel:

pn junctions

pn junctions are found in:
    diodes
    solar cells
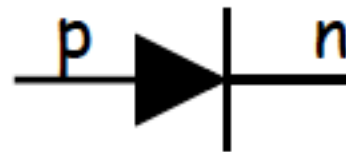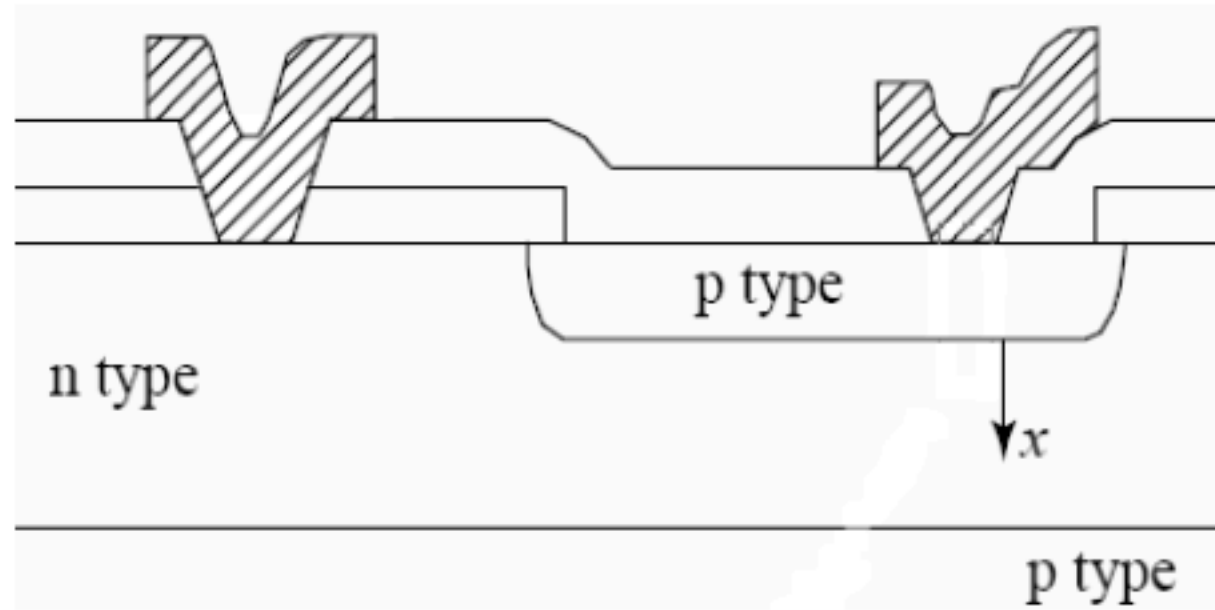   LEDs
    isolation
    JFETs
    bipolar transistors
    MOSFETs
    solid state lasers

Constant-source diffusion

One particular instance of this basic approach is the constant-source diffusion, in which a source of the dopant is applied to the surface and we assume that this has the effect of holding the surface concentration at some high, constant value. The initial and boundary conditions are:

$$N(x > 0, t = 0) = 0 \quad \text{(no dopant in the wafer before the diffusion)}$$

$$N(x = 0, t) = N_s \quad \text{(surface concentration is held constant)}$$

$$N(x \to \infty, t) = 0 \quad \text{(nothing gets to backside)}$$

The solution to the simple diffusion equation with these conditions is in the form of the *error function*, which is well known in the field of statistics, but it is a bit unwieldy.

$$N(x, t) = N_s \left[1 - \text{erf}\left(\frac{x}{2\sqrt{Dt}} \right) \right]$$

Dose

We can define a quantity called the dose, which is the total amount of a dopant diffused into a wafer. Mathematically,

$$Q = \int_0^\infty N(x)\, dx$$

Clearly, the units of dose are m^{-2} (or cm^{-2} or µm^{-2}, etc.) Physically, it means that, if you mapped out a square on the surface of the wafer and then stood within the square on the surface and looked down, the dose would represent all of the dopant that you could see below your feet and within the square, irrespective of distance from the surface.

Surprisingly, the dose of a constant-source diffusion has a simple analytic form, even though the error function itself is unfriendly.

$$Q = \int_0^\infty N_s \left[1 - \text{erf}\left(\frac{x}{2\sqrt{Dt}} \right) \right] dx$$

$$= \frac{2N_s}{\sqrt{\pi}} \sqrt{Dt}$$

Constant-dose diffusion

A second important diffusion configuration is the constant-dose diffusion. As the name implies, a constant dose of dopant is introduced into the semiconductor from some source. The source is removed and then diffusion proceeds with a fixed amount of dopant available. Generally, we assume that initial dose is located exactly at the surface of the wafer.

The constant-dose boundary condition is

$$Q = \int_0^\infty N(x)\, dx = \text{constant}$$

The other boundary condition and the initial condition are identical to the constant-source case:

$$N(x > 0, t = 0) = 0 \qquad \text{(no dopant in the wafer prior to diffusion)}$$

$$N(x \to \infty, t) = 0 \qquad \text{(nothing gets to backside)}$$

The diffusion equation solution under these conditions is a Gaussian function:

$$N(x, t) = \frac{Q}{\sqrt{\pi D t}} \exp\left(-\frac{x^2}{4Dt}\right)$$

The surface concentration for the Gaussian profile is

$$N(0, t) = \frac{Q}{\sqrt{\pi D t}}$$

Again, we see that the Dt-product determines the shape of the profile. With a bigger Dt (hotter or longer diffusion), more dopant moves deeper into the wafer.

Since there is a fixed dose of dopant, as more atoms move deeper into the wafer, the surface concentration must decrease. The expression for $N(0)$ confirms that expectation – as Dt increases, the surface concentration decreases.
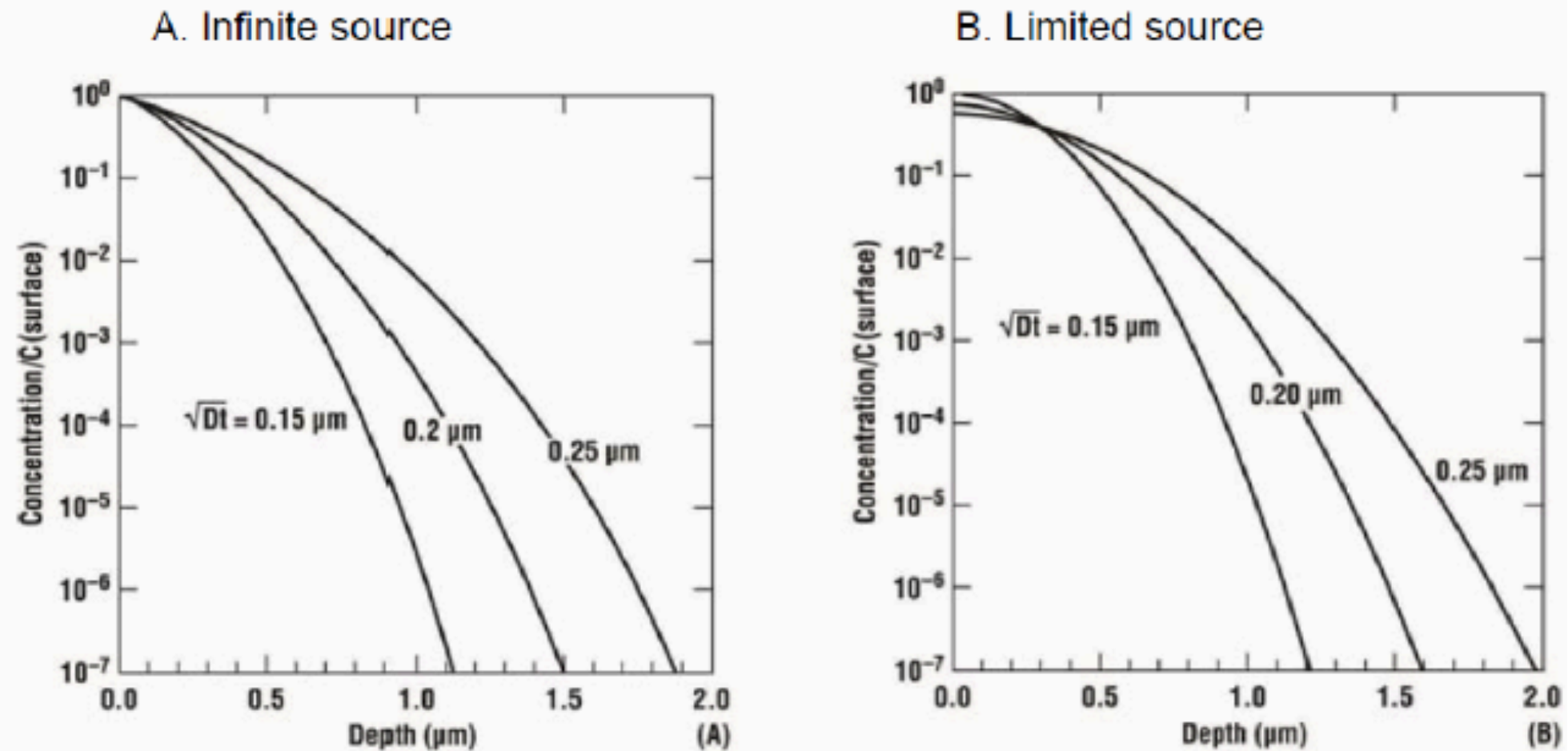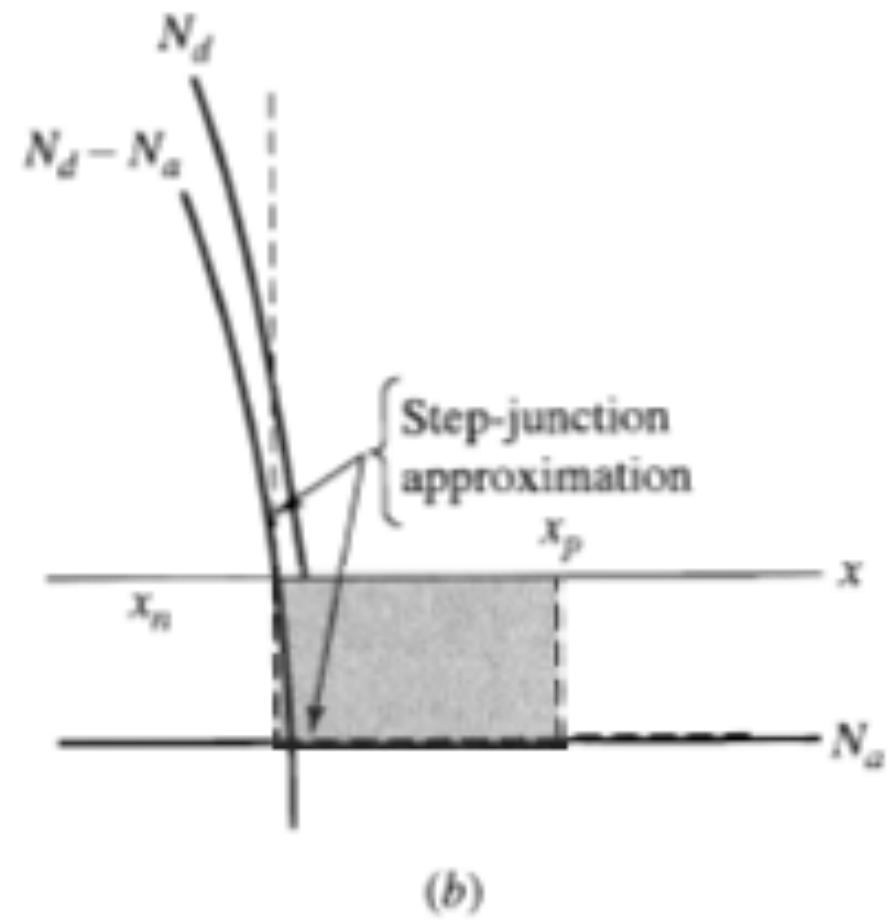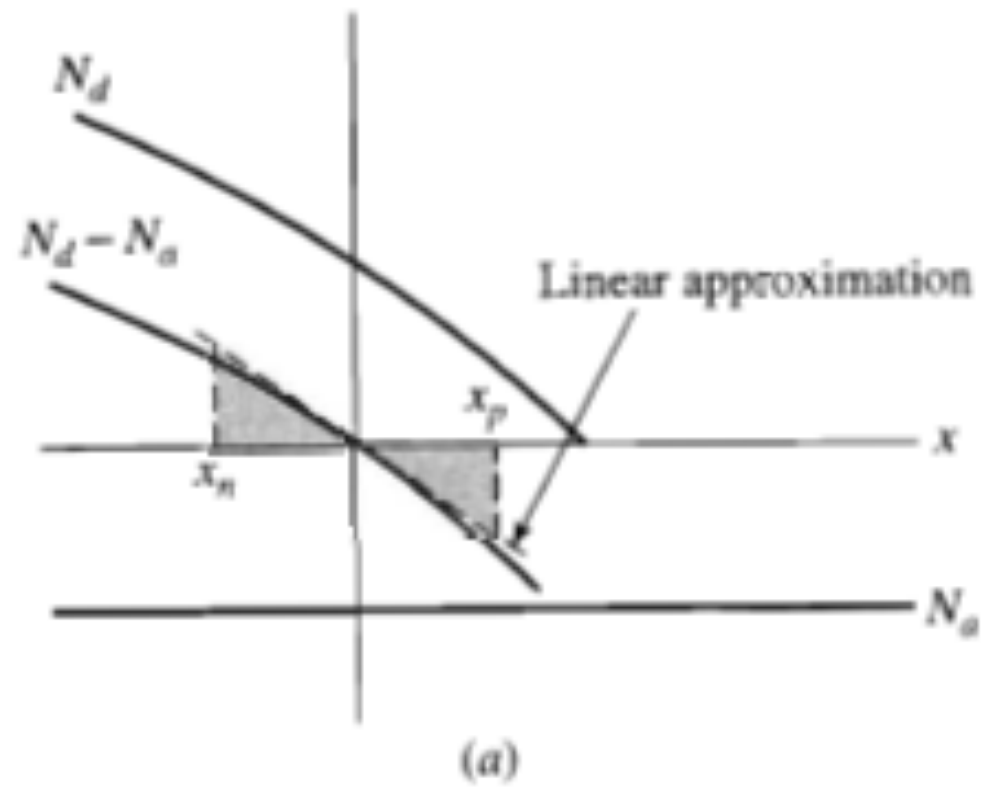
Figure 3.7 Concentration as a function of depth for (A) predeposition and (B) drive-in diffusions for several values of the characteristic diffusion length.
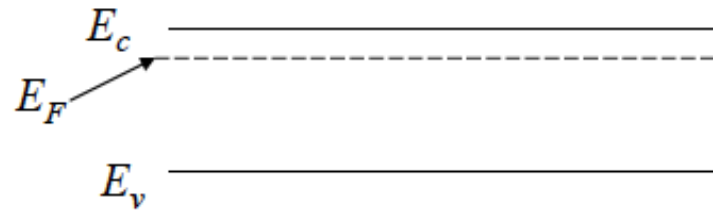
N_d

$N_d - N_a$

Linear approximation

x_p

x_n

x

N_a

(a)

N_d

$N_d - N_a$

Step-junction approximation

x_p

x_n

x

N_a

(b)

16

pn junction

isolated semiconductors
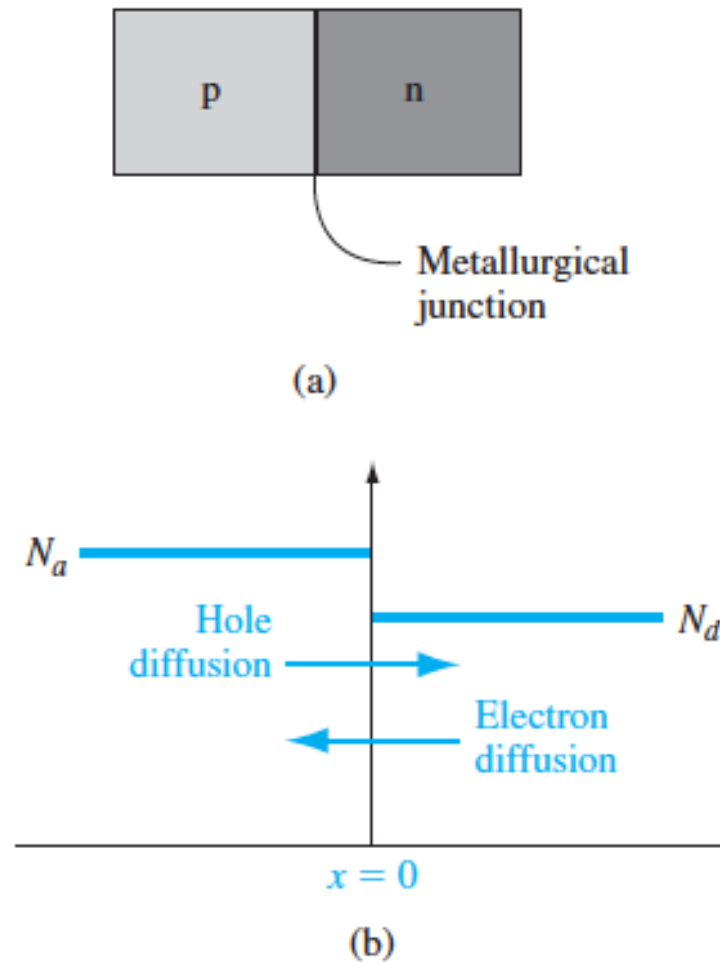
p-type   n-type

$$E_F = E_v + k_BT \ln\left(\frac{N_v}{N_A}\right)$$

$$n = N_c \exp\left(\frac{E_F - E_c}{k_BT}\right)$$

$$p = N_v \exp\left(\frac{E_v - E_F}{k_BT}\right)$$

valid for both n and p doping

$$E_F = E_c - k_BT \ln\left(\frac{N_c}{N_D}\right)$$

(a)

$x = 0$

(b)

$$\rho = \begin{cases} -qN_a & \text{for } -x_p < x < 0 \\ +qN_d & \text{for } 0 < x < x_n \\ 0 & \text{for } x < -x_p \text{ and } x > x_n \end{cases}$$

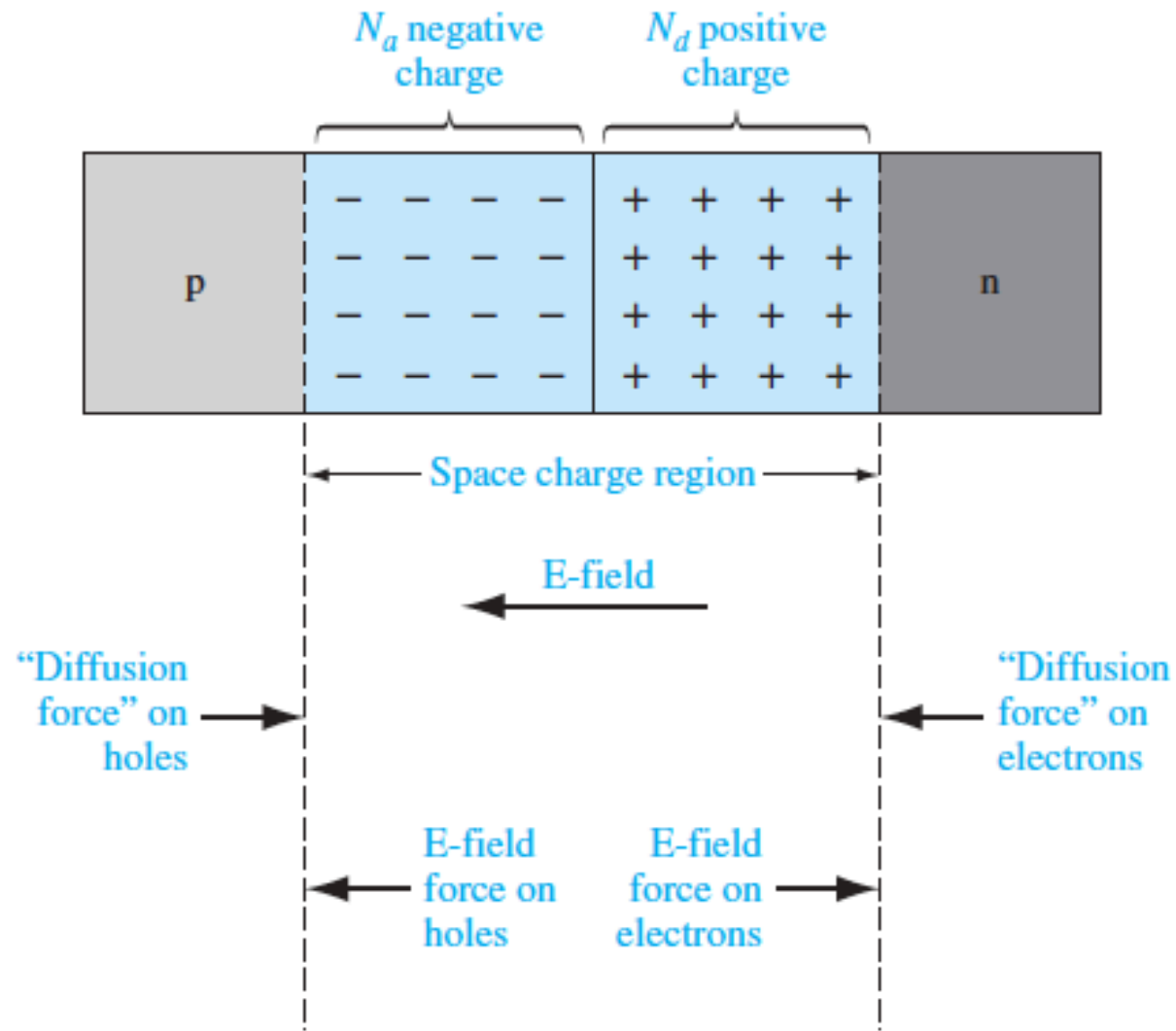**Figure 7.1** | (a) Simplified geometry of a pn junction; (b) doping profile of an ideal uniformly doped pn junction.

Figure 7.2 | The space charge region, the electric field, and the forces acting on the charged carriers.
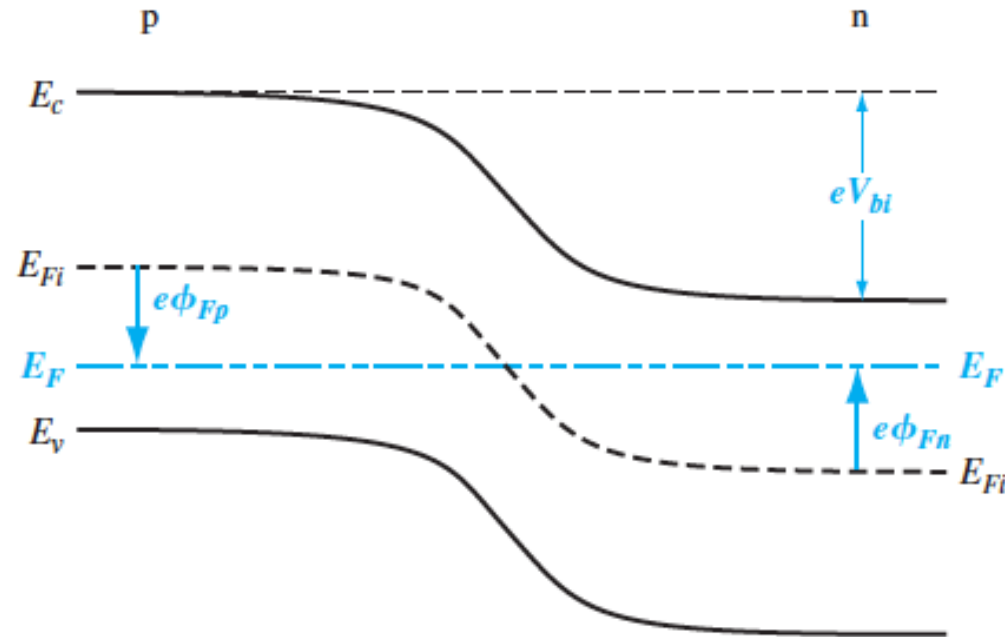
Figure 7.3 | Energy-band diagram of a pn junction in thermal equilibrium.

The intrinsic Fermi level is equidistant from the conduction band edge through the junction; thus, the built-in potential barrier can be determined as the difference between the intrinsic Fermi levels in the p and n regions. We can define the potentials ϕ_{Fn} and ϕ_{Fp} as shown in Figure 7.3, so we have

$$V_{bi} = |\phi_{Fn}| + |\phi_{Fp}| \tag{7.1}$$

In the n region, the electron concentration in the conduction band is given by

$$n_0 = N_c \exp\left[\frac{-(E_c - E_F)}{kT}\right] \tag{7.2}$$

which can also be written in the form

$$n_0 = n_i \exp\left[\frac{E_F - E_{Fi}}{kT}\right] \tag{7.3}$$

where n_i and E_{Fi} are the intrinsic carrier concentration and the intrinsic Fermi energy, respectively. We may define the potential ϕ_{Fn} in the n region as

$$e\phi_{Fn} = E_{Fi} - E_F \qquad (7.4)$$

Equation (7.3) may then be written as

$$n_0 = n_i \exp\left[\frac{-(e\phi_{Fn})}{kT}\right] \qquad (7.5)$$

Taking the natural log of both sides of Equation (7.5), setting $n_0 = N_d$, and solving for the potential, we obtain

$$\phi_{Fn} = \frac{-kT}{e} \ln\left(\frac{N_d}{n_i}\right) \qquad (7.6)$$

$$\phi_i = \frac{kT}{q} \ln\left(\frac{N_d N_a}{n_i^2}\right)$$

Similarly, in the p region, the hole concentration is given by

$$p_0 = N_a = n_i \exp\left[\frac{E_{Fi} - E_F}{kT}\right] \tag{7.7}$$

where N_a is the acceptor concentration. We can define the potential ϕ_{Fp} in the p region as

$$e\phi_{Fp} = E_{Fi} - E_F \tag{7.8}$$

Combining Equations (7.7) and (7.8), we find that

$$\phi_{Fp} = +\frac{kT}{e} \ln\left(\frac{N_a}{n_i}\right) \tag{7.9}$$

Finally, the built-in potential barrier for the step junction is found by substituting Equations (7.6) and (7.9) into Equation (7.1), which yields

$$V_{bi} = \frac{kT}{e} \ln\left(\frac{N_a N_d}{n_i^2}\right) = V_t \ln\left(\frac{N_a N_d}{n_i^2}\right) \tag{7.10}$$

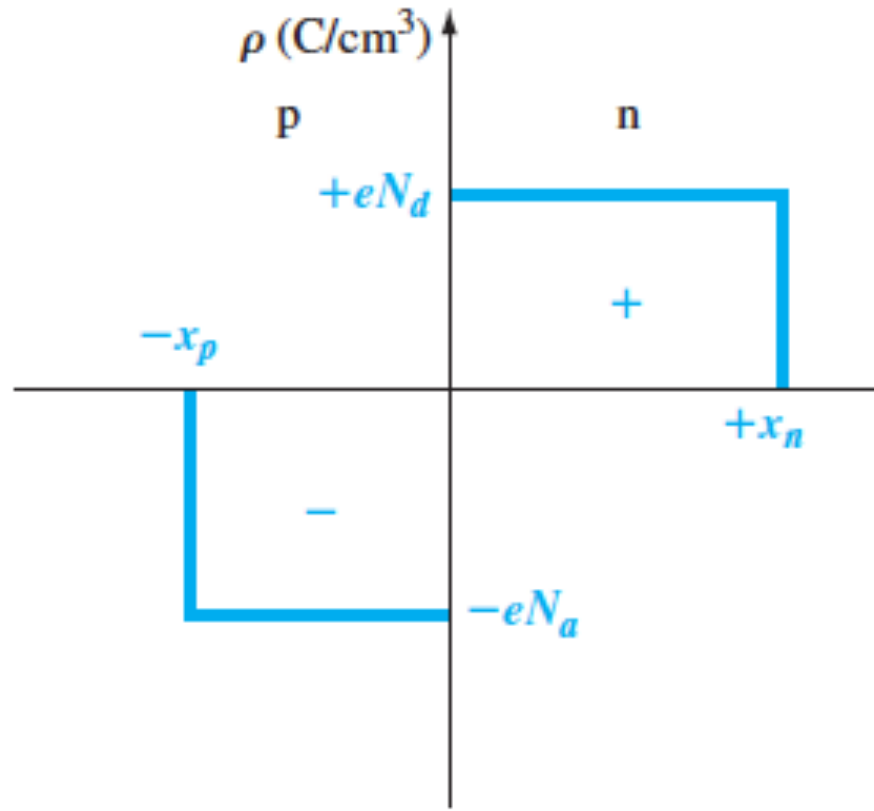where $V_t = kT/e$ and is defined as the thermal voltage.

Figure 7.4 | The space charge density in a uniformly doped pn junction assuming the abrupt junction approximation.
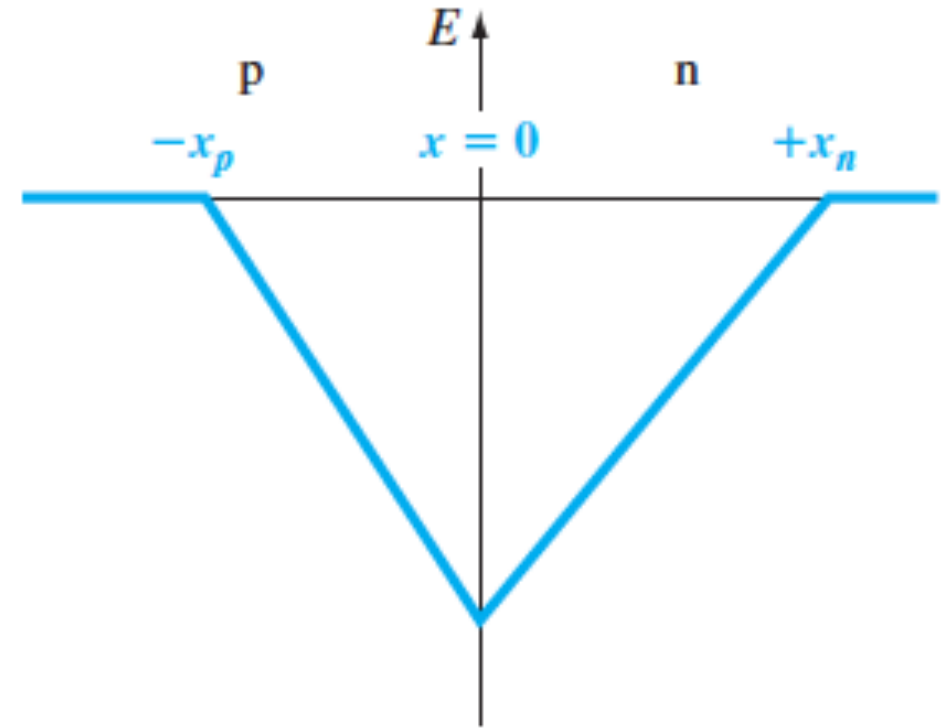

Figure 7.5 | Electric field in the space charge region of a uniformly doped pn junction.

$$\frac{\partial E}{\partial x} = \frac{-qN_a}{\epsilon_s} \text{ for } -x_p < x < 0$$

$$E = -\frac{qN_a x}{\epsilon_s} + C_1$$

$$0 = \frac{qN_a x_p}{\epsilon_s} + C_1$$

$$E = -\frac{qN_a}{\epsilon_s}(x + x_p)$$

$$\frac{\partial E}{\partial x} = +\frac{qN_d}{\epsilon_s} \text{ for } 0 < x < x_n$$

$$E = \frac{qN_d x}{\epsilon_s} + C_2$$

$$0 = \frac{qN_d x_n}{\epsilon_s} + C_2$$

$$E = +\frac{qN_d}{\epsilon_s}(x - x_n)$$

$$-\frac{\partial \phi}{\partial x} = -\frac{qN_a}{\epsilon_s}(x + x_p) \text{ for } -x_p < x < 0$$

$$\phi = \frac{qN_a}{\epsilon_s}(x^2/2 + x_p x)\Big|_{-x_p}^{0}$$

$$\phi_p = \frac{qN_a}{\epsilon_s}(x_p^2/2 - x_p^2)$$

$$= -\frac{qN_a}{2\epsilon_s}x_p^2$$

$$-\frac{\partial \phi}{\partial x} = +\frac{qN_d}{\epsilon_s}(x - x_n) \text{ for } 0 < x < x_n$$

$$\phi = -\frac{qN_d}{\epsilon_s}(x^2/2 - x_n x)\Big|_{0}^{x_n}$$

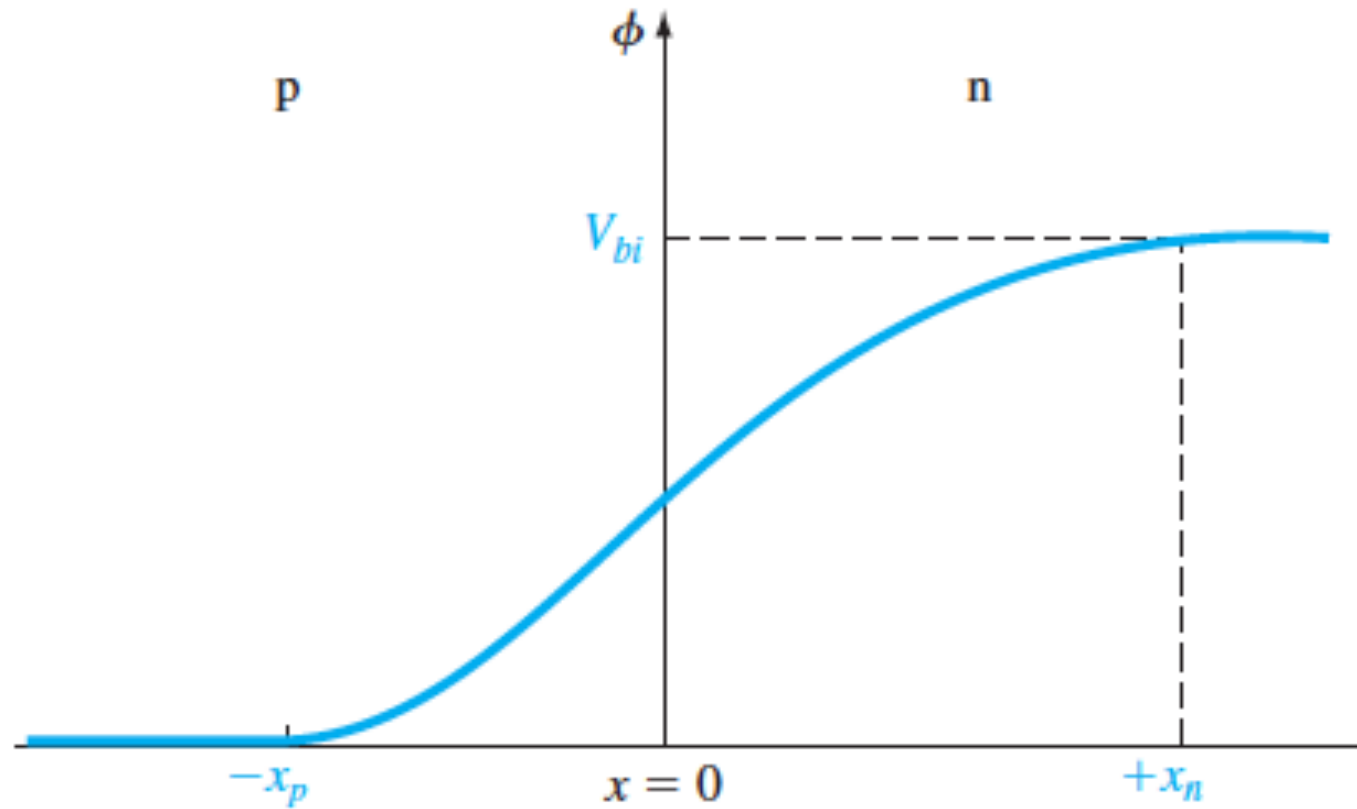$$\phi_n = -\frac{qN_d}{\epsilon_s}(x_n^2/2 - x_n^2)$$

$$= +\frac{qN_d}{2\epsilon_s}x_n^2$$

Figure 7.6 | Electric potential through the space charge region of a uniformly doped pn junction.

$$\phi_i = -\int_{-x_p}^{x_n} E dx$$

$$= \frac{q}{\epsilon_s} \left[N_a \int_{-x_p}^{0} (x + x_p) dx - N_d \int_{0}^{x_n} (x - x_n) dx \right],$$

$$= \frac{q}{\epsilon_s} \left[N_a \left(\frac{x^2}{2} + x_p x \right) \Big|_{-x_p}^{0} \right) - N_d \left(\frac{x^2}{2} - x_n x \right) \Big|_{0}^{x_n} \right) \right]$$

$$= \frac{q}{\epsilon_s} \left[N_a \left(\frac{x_p^2}{2} \right) + N_d \left(\frac{x_n^2}{2} \right) \right].$$

$$= \frac{q}{2\epsilon_s} \left[N_a x_p^2 + N_d x_n^2 \right].$$

$$\phi_i = \frac{1}{2} E_{max} (x_n + x_p) = \frac{1}{2} E_{max} x_d$$

$$\phi_i = \frac{q}{2\epsilon_s} \left[N_a x_p^2 + N_d x_n^2 \right]$$

7.2.3 Space Charge Width

We can determine the distance that the space charge region extends into the p and n regions from the metallurgical junction. This distance is known as the space charge width. From Equation (7.17), we may write, for example,

$$x_p = \frac{N_d x_n}{N_a} \tag{7.27}$$

Then, substituting Equation (7.27) into Equation (7.26) and solving for x_n, we obtain

$$x_n = \left\{ \frac{2\epsilon_s V_{bi}}{e} \left[\frac{N_a}{N_d} \right] \left[\frac{1}{N_a + N_d} \right] \right\}^{1/2} \tag{7.28}$$

Equation (7.28) gives the space charge width, or the width of the depletion region, x_n extending into the n-type region for the case of zero applied voltage.

Similarly, if we solve for x_n from Equation (7.17) and substitute into Equation (7.26), we find

$$x_p = \left\{ \frac{2\epsilon_s V_{bi}}{e} \left[\frac{N_d}{N_a} \right] \left[\frac{1}{N_a + N_d} \right] \right\}^{1/2} \tag{7.29}$$

where x_p is the width of the depletion region extending into the p region for the case of zero applied voltage.

The total depletion or space charge width W is the sum of the two components, or

$$W = x_n + x_p \tag{7.30}$$

Using Equations (7.28) and (7.29), we obtain

$$W = \left\{ \frac{2\epsilon_s V_{bi}}{e} \left[\frac{N_a + N_d}{N_a N_d} \right] \right\}^{1/2} \tag{7.31}$$

$$x_n = \left[\frac{2\epsilon_s}{q} \phi_i \left(\frac{N_a}{N_d(N_a + N_d)} \right) \right]^{1/2}$$

$$x_p = \left[\frac{2\epsilon_s}{q} \phi_i \left(\frac{N_d}{N_a(N_a + N_d)} \right) \right]^{1/2}$$

$$x_d = x_n + x_p$$

$$= \left[\frac{2\epsilon_s}{q} \phi_i \left(\frac{1}{N_a} + \frac{1}{N_d} \right) \right]^{1/2}$$
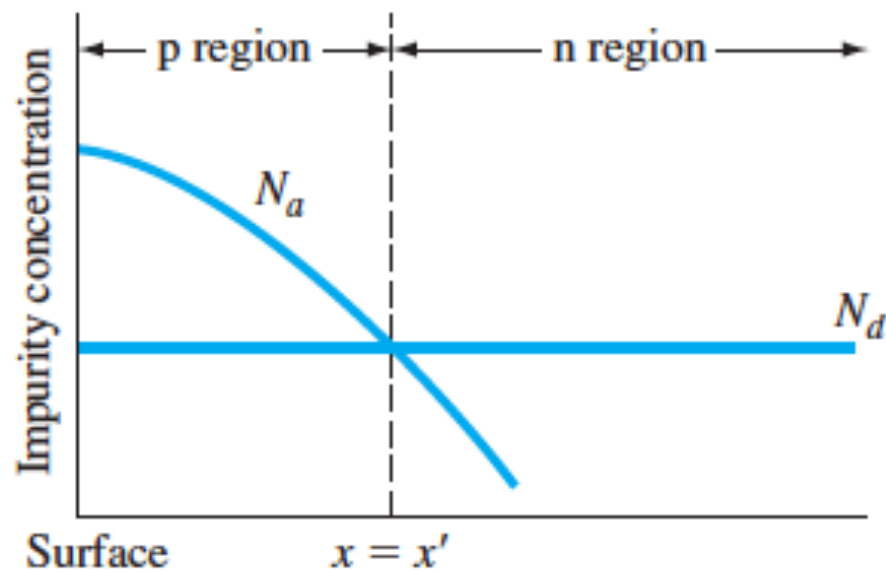
Figure 7.16 | Impurity concentrations of a pn junction with a nonuniformly doped p region.
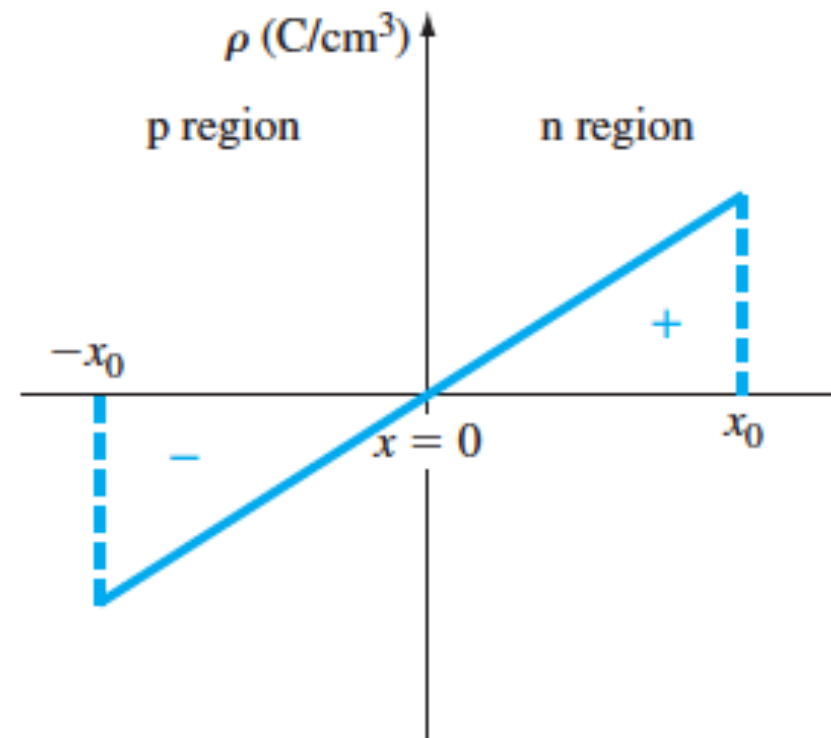

Figure 7.17 | Space charge density in a linearly graded pn junction.