CHAPTER

Electronics

10

CHAPTER OUTLINE

10.1 INTRODUCTION

The importance of semiconductors is due to their use in modern electronics. The original semiconductor device was the Schottky diode, constructed from a metal-semiconductor interface, which was used instead of a thermoionic diode for rectification. The Schottky diode was constructed by placing a metal whisker against an n-doped semiconductor crystal (the p-doped semiconductor-metal contact does not have any rectifying properties). The work function of the semiconductor is less than that of the metal. Because of the higher chemical potential, the electrons rush from the semiconductor to the metal, and the voltage of the metal is lowered until further motion of charges is prevented by electrostatic forces. The increase in voltage $(-eV(x))$ of the semiconductor compensates for the difference in the chemical potential μ, which is equal to the Fermi energy ε_F of the metal. This also adds an electrostatic potential $-eV(x)$ to the conduction and valence band levels, which are bent by the potentials formed across the junction. This scenario is shown in Figure 10.1.

Physics of Condensed Matter. DOI: 10.1016/B978-0-12-384954-0.00010-4

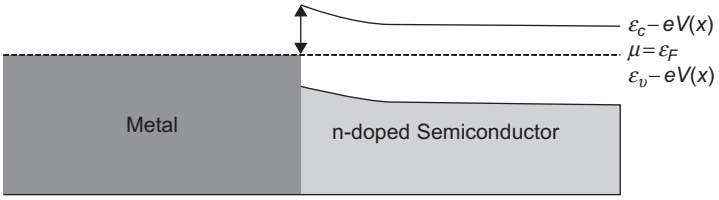

FIGURE 10.1

The charges move from the n-doped semiconductor to the metal, and both the valence and conduction bands are increased by $-eV(x)$. The electrons flow to the metal when the voltage of the metal is raised relative to the semiconductor by eV_A.

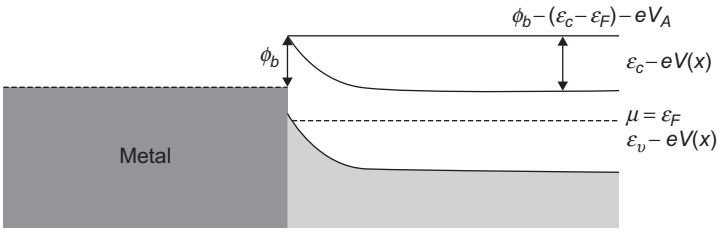

FIGURE 10.2

There is no flow of current when V_A is lowered. The height of the barrier is ϕ_b.

The effect of an external voltage V_A (which can be positive or negative) can be understood in a simple manner. When V_A is positive, the electrons flow from the semiconductor into the metal because the barrier is lowered. However, when V_A is negative, there is no change in the barrier, and hence, the current does not flow. This rectifying effect of the Schottky diode is shown in Figure 10.2.

In practice, it is difficult to construct an ideal Schottky diode because semiconductor surfaces acquire oxide layers after cleaving, and it is virtually impossible to produce atomic flat surfaces. The age of modern electronics was started with the creation of the p-n junction in a semiconductor.

10.2 p-n JUNCTION

10.2.1 Introduction

A simple p-n junction is fabricated by taking a single intrinsic semiconductor such as Si in which donor impurities are introduced in one region and acceptor impurities are introduced into another. It is easier to visualize the p-n junction as equivalent to two Si crystals doped with donor (n-type) and acceptor (p-type) impurities joined together with polished surfaces. In practice, this is not a good idea because the surface effects would interfere with the properties of the semiconductor. These semiconductors have different chemical potentials, $\mu(\mathrm{n})$ and $\mu(\mathrm{p})$, respectively. Initially, most of the conduction electrons are at the n-type, and the holes are at the p-type semiconductor. This scenario is shown in Figure 10.3a. However, when the two materials are joined together, some of the conduction electrons will diffuse to the p-type material while some of the holes will diffuse to the n-type material. When an electron

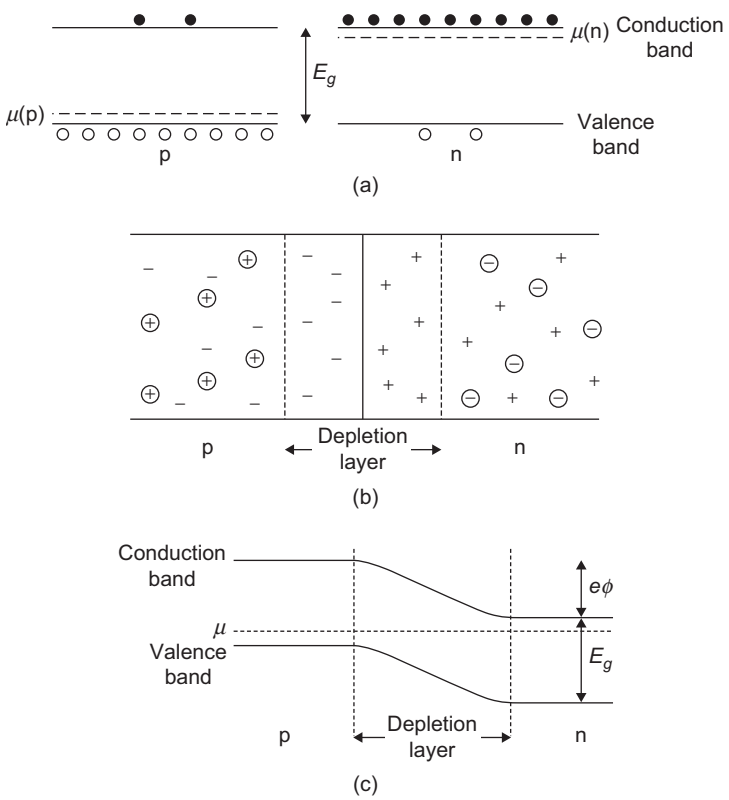

FIGURE 10.3

(a) The electrons (solids) and holes (open circles) in p- and n-type semiconductors. (b) When the p- and n-type regions are in contact, there are practically no carriers in the depletion region. The acceptor and donor ions are represented by – and + and electrons and holes are represented by circles with ⊖ and ⊕ signs. (c) The distribution of charge at a p-n junction produces a contact potential $\phi(x)$ across the junction.

diffuses to the p-type layer, it will recombine with a hole due to the large concentration of holes in that layer. A similar situation will arise when a hole diffuses to the n-type layer. As more and more electrons and holes diffuse to the p- and n-type layers, the number of carriers in the junction will be greatly reduced. Thus, the layer on either side of the junction has much lower concentration of carriers than the rest of the semiconductor. This is known as the depletion layer, which is shown in Figure 10.3b and usually has an approximate length of 10^2 Å. We note that the donor and acceptor ions are much too heavy to diffuse like the electrons and holes. Thus, the p-n junction is formed in the semiconductor with a common chemical potential μ. This is shown in Figure 10.3c.

10.2.2 p-n Junction in Equilibrium

We assume that the impurity concentration varies along the x-axis only in a small region around $x = 0$. If the "abrupt junction," which is defined as the region about $x = 0$ where the impurity

concentrations change, is narrow compared with the "depletion layer" in which the carrier densities are not uniform, the donor impurities dominate at positive x, and the acceptor impurities dominate at negative x (see Figure 10.4).

The depletion layer, which extends from x_p to x_n, is shown in Figure 10.5. The abrupt junction is at $x = 0$. The region to the left of x_p is the p-doped mobile holes, and the region to right of x_n is the n-doped mobile electrons (note that x_p is negative). In the beginning, when the p-n junction is formed, the carrier concentration would be such that there would be charge neutrality everywhere in the crystal. However, the concentration of the electrons in the n-side would be very high and that of the holes in the p-side would be very low. Thus, the electrons would diffuse to the p-side, and the holes would diffuse to the n-side. The charge transfer will eventually build up an electric field that will prevent further diffusion. In fact, the effect of the field cancels the effect of the diffusion.

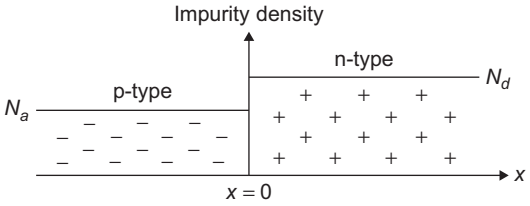

FIGURE 10.4

"Abrupt" p-n junction in which the impurity-concentration change is narrow compared with the depletion layer shown in Figure 10.5.

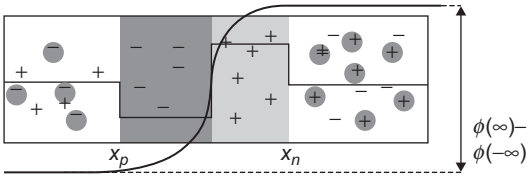

FIGURE 10.5

The depletion layer has length between 10^2 Å and 10^4 Å extending from x_p to x_n.

To derive expressions for the carrier densities at a position x at temperature T in the presence of a potential $\phi(x)$, one can use Eqs. (9.32) and (9.35) subject to the semiclassical condition that each one-electron energy level is shifted by $-e\phi(x)$. Thus, one obtains

$$\varepsilon_c \to \varepsilon_c - e\phi(x) \tag{10.1}$$

and

$$\varepsilon_v \to \varepsilon_v - e\phi(x). \tag{10.2}$$

From Eqs. (9.32) and (10.1), we obtain

$$n_c(x) = \aleph_c(T) e^{-\beta(\varepsilon_c - e\phi(x) - \mu)}, \tag{10.3}$$

and from Eqs. (9.35) and (10.2), we obtain

$$p_v(x) = \wp_v(T) e^{\beta(\varepsilon_v - e\phi(x) - \mu)}, \tag{10.4}$$

where $\beta \equiv 1/k_B T$. If we define the electrochemical potential $\mu_e(x)$ as

$$\mu_e(x) = \mu + e\phi(x), \tag{10.5}$$

Eqs. (10.3) and (10.4) can be rewritten in the alternate form,

$$n_c(x) = \aleph_c(T)e^{-\beta(\varepsilon_c - \mu_e(x))} \tag{10.6}$$

and

$$p_v(x) = \wp_v(T)e^{\beta(\varepsilon_v - \mu_e(x))}. \tag{10.7}$$

The total charge density must vanish in both the limits $x \to -\infty$ and $x \to \infty$, which requires that $N_d = n_c(\infty)$ and $N_a = p_v(-\infty)$. From Eqs. (10.3) through (10.5), one can also derive (Problem 10.1)

$$\mu_e(\infty) - \mu_e(-\infty) = e\Delta\phi = E_g + \frac{1}{\beta}\ln\left[\frac{N_d N_a}{\aleph_c(T)\wp_v(T)}\right], \tag{10.8}$$

where

$$\Delta\phi \equiv \phi(\infty) - \phi(-\infty). \tag{10.9}$$

The effect of the internal potential $\phi(x)$ on the electron and hole densities of a p-n junction is shown in Figure 10.6, in which the electrochemical potential $\mu_e(x)$ is plotted along the p-n junction. The carrier densities at a point x are obtained by using $\mu_e(x)$ as the equivalent chemical potential.

The alternate method of representing the carrier densities at any point x is shown in Figure 10.7. Here, $\varepsilon_d(x) = \varepsilon_d - e\phi(x)$ $\varepsilon_a(x) = \varepsilon_a - e\phi(x)$ and μ is the constant chemical potential.

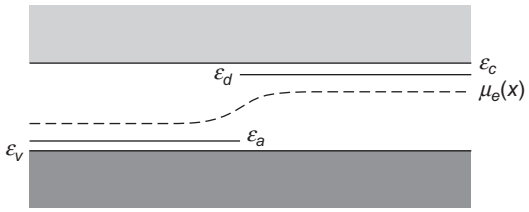

FIGURE 10.6

ε_c and ε_v are conduction and valence bands; ε_d and ε_a are donor and acceptor levels.

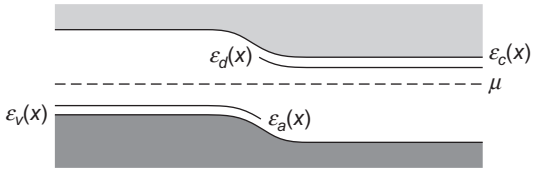

FIGURE 10.7

Carrier densities at a point x obtained by using a fixed chemical potential μ.

To calculate $\phi(x)$, one uses the Poisson's equation in one dimension,

$$\frac{\partial^2\phi}{\partial x^2} = -\frac{4\pi\rho(x)}{\in}, \tag{10.10}$$

where the charge density $\rho(x)$ is obtained from

$$\rho(x) = N_d(x) - n_c(x) - N_a(x) + p_v(x), \tag{10.11}$$

and $\in$ is the dielectric constant. It can be easily shown that (Problem 10.2)

$$n_c(x) = N_d e^{-e\beta[\phi(\infty) - \phi(x)]} \tag{10.12}$$

and

$$p_v(x) = N_a e^{e\beta[\phi(-\infty) - \phi(x)]}. \tag{10.13}$$

It can also be shown that (Problem 10.3; except at the boundaries of the depletion layer), outside the depletion layer,

$$\rho(x) = 0, \tag{10.14}$$

and inside the depletion layer,

$$\rho(x) = e[N_d(x) - N_a(x)]. \tag{10.15}$$

The solution for $\phi(x)$ is obtained in Problem 10.4,

$$\phi(x) = \left\{ \begin{array}{ll} \phi(-\infty), & x < x_p \\ \phi(-\infty) + \left(\dfrac{2\pi e N_a}{\in}\right)(x - x_p)^2, & 0 > x > x_p \\ \phi(\infty) - \left(\dfrac{2\pi e N_d}{\in}\right)(x - x_n)^2, & 0 < x < x_n \\ \phi(\infty), & x > x_n \end{array} \right\}. \tag{10.16}$$

In addition, both $\phi(x)$ and $\phi'(x)$ must be continuous at $x = 0$. From the condition $\phi(x)|_{x=-\in} = \phi(x)|_{x=\in}$, where $\in \to 0$, we obtain

$$\phi(0) = \phi(\infty) - \frac{2\pi e N_d}{\in} x_n^2 = \phi(-\infty) + \frac{2\pi e N_a}{\in} x_p^2, \tag{10.17}$$

which can be rewritten in the alternate form

$$\Delta\phi = \frac{2\pi e(N_a x_p^2 + N_d x_n^2)}{\in}. \tag{10.18}$$

From the condition $\phi'(x)|_{x=-\in} = \phi'(x)|_{x=\in}$, we obtain

$$N_d x_n = -N_a x_p. \tag{10.19}$$

From Eqs. (10.18) and (10.19), we obtain expressions for the lengths x_n and x_p,

$$x_{n,p}(0) = \pm \left[\frac{\in (N_a/N_d)^{\pm 1} \Delta\phi}{2\pi e(N_a + N_d)}\right]^{1/2}, \tag{10.20}$$

where $x_{n,p}(0)$ denotes the values of x_n and x_p when the external potential $V = 0$. Because N_a and N_d are of the order of $10^{18}\,\text{cm}^{-3}$ and $e\Delta\phi$ is of the order of 0.1 eV, the depletion layer is of the order of 10^2 to 10^4 Å. It is important to note that because the depletion layer has no mobile charges, the resistance of this region is considerably higher than that of the doped regions. It is equivalent to a series circuit in which a high resistance is sandwiched between two low resistances. When there is no external potential ($V = 0$), the p-n junction is known as the unbiased junction. The potential $\phi(x)$ for the unbiased junction is plotted versus the position x in Figure 10.8.

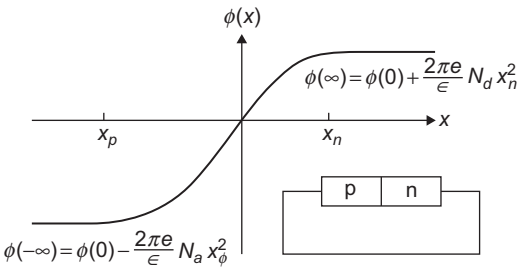

FIGURE 10.8

The potential for the unbiased junction $\phi(x)$ plotted against position x. The depletion layer is from $x = x_p$ to $x = x_n$ (note that x_p is negative).

10.3 RECTIFICATION BY A p-n Junction

10.3.1 Equilibrium Case

When the external potential $V = 0$, the change in potential from the p-side to the n-side of the depletion layer is given by $\Delta\phi$, the expression for which is obtained in Eq. (10.8). When the external potential $V \neq 0$, the potential drop (or gain) $\Delta\phi'$ will occur mostly across the high-resistance region, which is the depletion layer. Thus, $\Delta\phi'$ is given by

$$\Delta\phi' = \Delta\phi - V. \tag{10.21}$$

The size of the layer, which extends from x_p to x_n, would accordingly change because in Eq. (10.20), $\Delta\phi$ would be replaced by $\Delta\phi'$ in the expressions for x_n and x_p. From Eqs. (10.20) and (10.21), we obtain

$$x_{n,p}(V) = \pm\left[\frac{\in(N_a/N_d)^{\pm 1}\Delta\phi'}{2\pi e(N_a + N_d)}\right]^{1/2}. \tag{10.22}$$

Eq. (10.22) can be rewritten in the alternate form, by using the expressions for $x_{n,p}(0)$ and $\Delta\phi'$ from Eqs. (10.20) and (10.21),

$$x_{n,p}(V) = x_{n,p}(0)\left[1 - \frac{V}{\Delta\phi}\right]^{1/2}. \tag{10.23}$$

This scenario is shown in Figures 10.9 and 10.10.

We will first discuss the scenario when there is no external potential $(V = 0)$ in the p-n junction. Due to the thermal excitation of electrons in the valence band on the n-side of the depletion layer, holes that move at random are generated. These holes are the "minority carriers" in the n-side compared to the electrons that are the "majority carriers," but they play an important role. The holes are swept away from the n-side to the p-side as soon as they wander into the depletion layer because there is a strong electric field $\Delta\phi$ across the layer. The current due to the motion of the holes from the n-side to the p-side is known as the hole generation current, $j_{hg}(0)$. In addition, the holes

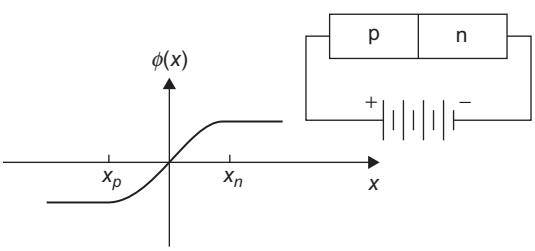

FIGURE 10.9

The potential $\phi(x)$ versus x when $V > 0$ (forward bias). $\Delta\phi' = \phi(\infty) - \phi(-\infty) - V$.

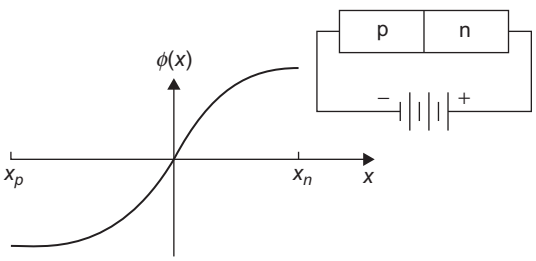

FIGURE 10.10

The potential $\phi(x)$ versus x when $V < 0$ (reverse bias). $\Delta\phi' = \phi(\infty) - \phi(-\infty) + V$.

that are thermally generated and are the majority carriers in the p-side flow to the n-side of the junction. The motion of these holes, which are positively charged, is opposed by the electric field of the depletion layer, but those holes of which the thermal energy is greater than the potential barrier would reach the n-side and would recombine with the electrons. The current due to the motion of the holes from the p-side to the n-side of the depletion layer is known as the hole recombination current, $j_{hr}(0)$. Because there is no external potential, the sum of the two currents is zero. Thus, we obtain

$$j_{hg}(0) + j_{hr}(0) = 0. \tag{10.24}$$

If we write $j_{hg} = e\aleph_{hg}$ and $j_{hr} = e\aleph_{hr}$, where $\aleph_{hg}$ and $\aleph_{hr}$ are the number density of holes generated in the n-side and the p-side of the depletion layer, respectively, it follows that at $V = 0, \aleph_{hg}(0) = \aleph_{hr}(0)$ and Eq. (10.24) is satisfied because the positively charged holes flow in opposite directions. We also note that $\aleph_{hr}$ is approximately the same as N_a. Therefore, when an external potential V is applied across the p-n junction,

$$\aleph_{hr} \propto e^{-e\beta\phi'}. \tag{10.25}$$

From Eqs. (10.21) and (10.25), we obtain

$$\aleph_{hr}(V) = \aleph_{hr}(0)\, e^{\beta eV}. \tag{10.26}$$

In contrast, the external potential V would have no effect on the number of holes crossing the depletion layer. Thus, we obtain

$$\aleph_{hg}(V) = \aleph_{hg}(0) = \aleph_{hg}. \tag{10.27}$$

The net number of holes moving across the depletion layer is given by

$$\aleph_h(V) = \aleph_{hr}(V) + \aleph_{hg}(V). \tag{10.28}$$

The expression for the hole current density, $j_h = e\aleph_h(V)$, is obtained from Eqs. (10.26) through (10.28),

$$j_h = j_{hr} - j_{hg} = e\aleph_{hg}(e^{\beta eV} - 1). \tag{10.29a}$$

We can make a similar argument for the motion of the electrons that are the majority carriers in the n-side and the minority carriers in the p-side. However, the electrons are negatively charged particles, and therefore, only those electrons that have enough thermal energy to cross the potential barrier can move from the n-side to the p-side. The current produced by the motion of the negatively charged electrons is known as the electron generation current $j_{eg}(0)$. In contrast, the electrons that reach the edge of the p-side of the p-n junction will be swept across the depletion layer and will eventually recombine with the holes in the n-side. This is known as the electron recombination current, $j_{er}(0)$. We can write

$$j_{eg}(0) + j_{er}(0) = 0. \tag{10.29b}$$

If we write $j_{eg} = -e\aleph_{eg}$ and $j_{er} = -e\aleph_{er}$, where $\aleph_{eg}$ and $\aleph_{er}$ are the number current density of electrons generated in the p-side and the n-side, respectively, it follows that at $V = 0$, $\aleph_{eg}(0) = \aleph_{er}(0)$ and

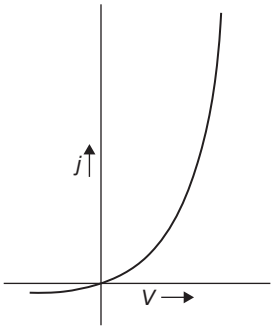

FIGURE 10.11

The $j(V)$ characteristic of the p-n junction.

Eq. (10.25) is satisfied because the electrons flow in opposite directions. We also note that $\aleph_{er}$ is approximately the same as N_d. Therefore, we can express

$$\aleph_{er}(V) = \aleph_{er}(0)e^{\beta eV}. \tag{10.30}$$

However, the external potential does not make any difference in the motion of the electrons generated in the p-side and moving across the depletion layer to the n-side. Thus, we obtain

$$\aleph_{eg}(V) = \aleph_{eg}(0) = \aleph_{eg}. \tag{10.31}$$

The net number of electrons moving across the depletion layer is given by

$$\aleph_e(V) = \aleph_{er}(V) + \aleph_{eg}(V). \tag{10.32}$$

From Eqs. (10.30) through (10.32), we obtain

$$j_e(V) = j_{er}(V) - j_{eg}(V) = e\aleph_{eg}(e^{\beta eV} - 1). \tag{10.33}$$

The total electrical current density is obtained from Eqs. (10.29) and (10.33),

$$j(V) = j_h(V) + j_e(V) = e(\aleph_{hg} + \aleph_{eg})(e^{\beta eV} - 1). \tag{10.34}$$

Eq. (10.34) is the characteristic property of rectifiers and is at the heart of modern transistors. A schematic diagram of the $j(V)$ characteristic for the p-n junction is shown in Figure 10.11.

10.3.2 Nonequilibrium Case ($V \neq 0$)

In the preceding discussion, we considered the equilibrium case $(V = 0)$ and then introduced the external potential V in an ad hoc manner to explain the elementary theory of rectification by a p-n junction. We will now consider the nonequilibrium problem $(V \neq 0)$. The nonequilibrium problem is the direct generalization of the equilibrium problem $(V = 0)$. When $V \neq 0$, one has to solve the Boltzmann equation for semiconductors. In the relaxation-time approximation, the Boltzmann equation for semiconductors is different from that of metals because of the presence of electrons and holes. In the relaxation-time approximation, the Boltzmann equation for electrons can be written as (in analogy with Eqs. 6.73 and 6.74)

$$\frac{\partial g}{\partial t} = -\dot{\mathbf{r}} \cdot \frac{\partial g}{\partial \mathbf{r}} - \dot{\mathbf{k}} \cdot \frac{\partial g}{\partial \mathbf{k}} - \frac{g - f}{\tau_n}, \tag{10.35}$$

where g is the density of particles at position $\mathbf{r}$. In the presence of weak applied fields, g is replaced by $g_{\mathbf{rk}}(t)$, which is the occupation number of electrons at position $\mathbf{r}$, which have a wave vector $\mathbf{k}$ at time t. It can be easily shown from Eq. (6.83) that at constant temperature

$$g_{\mathbf{rk}} \approx f_{\mathbf{rk}} - e\tau_n \frac{\partial f}{\partial \mu} \mathbf{v}_{\mathbf{k}} \cdot \mathbf{E}. \tag{10.36}$$

The electron density n_c at position $\mathbf{r}$ in the conduction band can be written as

$$n_c = \frac{1}{4\pi^3}\int g_{r\mathbf{k}}\, d\mathbf{k}, \tag{10.37}$$

while the equilibrium density of electrons in the conduction band can be written as

$$n_c^0 = \frac{1}{4\pi^3}\int f\, d\mathbf{k}. \tag{10.38}$$

Integrating Eq. (10.35) over $d\mathbf{k}$, we obtain from Eqs. (10.35) through (10.38) (see Problem 10.6)

$$\frac{\partial n_c}{\partial t} = -\frac{\partial}{\partial \mathbf{r}}\cdot <\dot{\mathbf{r}}> + \frac{n_c^0 - n_c}{\tau_n}. \tag{10.39}$$

Here, n_c^0 is the equilibrium density of the electrons in the conduction band, and $<\dot{\mathbf{r}}>$ is $\mathbf{v_k}$ averaged over the Brillouin zone,

$$<\dot{\mathbf{r}}> = \frac{1}{4\pi^3 n_c}\int d\mathbf{k}\; g_{r\mathbf{k}}\mathbf{v_k} = \frac{1}{4\pi^3 n_c}\int d\mathbf{k}\left[f - \tau_n \mathbf{v_k}\cdot\left\{e\mathbf{E}\frac{\partial f}{\partial \mu} + \frac{\partial f}{\partial \mathbf{r}}\right\}\right]\mathbf{v_k}. \tag{10.40}$$

Because $\int f d\mathbf{k}$ vanishes by symmetry, $\partial f/\partial\mu = \beta f$ and $f \approx g$, Eq. (10.40) can be rewritten as

$$<\dot{\mathbf{r}}> \approx \frac{1}{4\pi^3 n_c}\int d\mathbf{k}\left[-\tau_n \mathbf{v_k}\cdot\left\{e\mathbf{E}\beta g + \frac{\partial g}{\partial \mathbf{r}}\right\}\right]\mathbf{v_k}. \tag{10.41}$$

We define the mobility μ_n and the diffusion constant D_n as

$$\mu_n = \frac{e\beta}{3}<\tau_n v_k^2> \tag{10.42}$$

and

$$D_n = \frac{1}{3}<\tau_n v_k^2> = \frac{\mu_n}{e\beta}. \tag{10.43}$$

From Eqs. (10.41) through (10.43), we obtain

$$<\dot{\mathbf{r}}> = -\mu_n \mathbf{E} - \frac{D_n}{n_c}\frac{\partial n_c}{\partial \mathbf{r}}. \tag{10.44}$$

The electron current is given by

$$\mathbf{j}_e = -en_c<\dot{\mathbf{r}}> = en_c\mu_n\mathbf{E} + eD_n\overrightarrow{\nabla} n_c. \tag{10.45}$$

Similarly, the hole current can be obtained as

$$\mathbf{j}_h = ep_v\mu_p\mathbf{E} - eD_p\overrightarrow{\nabla} p_v. \tag{10.46}$$

From Eq. (10.39) and its analogous equations for holes, Eqs. (10.45) and (10.46), we obtain

$$\frac{\partial n_c}{\partial t} = \frac{1}{e}\overrightarrow{\nabla}\cdot\mathbf{j}_e + \frac{n_c^0 - n_c}{\tau_n} \tag{10.47}$$

and

$$\frac{\partial p_v}{\partial t} = -\frac{1}{e}\overrightarrow{\nabla}\cdot\mathbf{j}_h + \frac{p_v^0 - p_v}{\tau_p}. \tag{10.48}$$

Because we have assumed that the carriers in the doped semiconductor primarily move in the x direction, Eqs. (10.45) and (10.46) can be rewritten as

$$j_e = en_c\mu_n E + eD_n\frac{dn_c}{dx} \tag{10.49}$$

and

$$j_h = ep_v\mu_p E - eD_p\frac{dp_v}{dx}. \tag{10.50}$$

Similarly, Eqs. (10.47) and (10.48) can be rewritten as

$$\frac{\partial n_c}{\partial t} = \frac{1}{e}\frac{\partial j_e}{\partial x} + \frac{n_c^0 - n_c}{\tau_n} \tag{10.51}$$

and

$$\frac{\partial p_v}{\partial t} = -\frac{1}{e}\frac{\partial j_h}{\partial x} + \frac{p_v^0 - p_v}{\tau_p}. \tag{10.52}$$

We note that

$$n_i^2 = n_c^0 p_v = p_v^0 n_c. \tag{10.53}$$

In the steady state, $\frac{\partial n_c}{\partial t} = \frac{\partial p_v}{\partial t} = 0$. Thus, in the steady state, we can rewrite Eqs. (10.51) and (10.52) as

$$\frac{1}{e}\frac{\partial j_e}{\partial x} + \frac{n_c^0 - n_c}{\tau_n} = 0 \tag{10.54}$$

and

$$-\frac{1}{e}\frac{\partial j_h}{\partial x} + \frac{p_v^0 - p_v}{\tau_p} = 0. \tag{10.55}$$

In the regions, where $E \to 0$, the majority carrier density is constant. In these regions, from Eqs. (10.49) and (10.50), we obtain

$$j_e(x_p) \approx eD_n\frac{dn_c}{dx}|_{x=x_p} \tag{10.56}$$

and

$$j_h(x_n) \approx -eD_p\frac{dp_v}{dx}|_{x=x_n}. \tag{10.57}$$

Eqs. (10.56) and (10.57) essentially state that in the steady state, the minority carrier drift current can be ignored compared to the minority carrier diffusion current. From Eqs. (10.54) through (10.57), we obtain

$$D_n \frac{d^2 n_c}{dx^2} + \frac{n_c^0 - n_c}{\tau_n} = 0 \tag{10.58}$$

and

$$D_p \frac{d^2 p_v}{dx^2} + \frac{p_v^0 - p_v}{\tau_p} = 0. \tag{10.59}$$

It can be easily shown (Problem 10.7) that in the n-side of the depletion layer, for $x \geq x_n$,

$$p_v(x) - p_v(\infty) = [p_v(x_n) - p_v(\infty)]e^{-(x-x_n)/d_p}, \tag{10.60}$$

and in the p-side of the depletion layer, for $x \leq x_p$,

$$n_c(x) - n_c(-\infty) = [n_c(x_p) - n_c(-\infty)]e^{(x-x_p)/d_n}, \tag{10.61}$$

where the diffusion lengths for holes in the n-doped region and electrons in the p-doped region are defined as

$$d_p = (D_p \tau_p)^{1/2} \text{ and } d_n = (D_n \tau_n)^{1/2}. \tag{10.62}$$

In fact, we can define two diffusion regions, which extend over a distance of the order of the diffusion length in either side of the depletion layer, sandwiched between the depletion layer and the homogeneous regions. The diffusion regions do not exist when $V = 0$.

Thus, we have three regions when $V \neq 0$. The difference between the three regions is that in the depletion layer, the electric field, the space charge, and the carrier density gradients are large (Eq. 10.23), but there is no current in the equilibrium case $(V = 0)$ because the drift and the diffusion currents are equal and opposite for both electrons and holes. When $V \neq 0$, there is a net current in the depletion layer due to the difference between the drift and diffusion currents of each carrier type, which are large. In the diffusion regions, the drift current of the majority carrier density is quite large, whereas that of the minority carrier is small. However, the diffusion current due to both the carriers is appreciable. In contrast, the only current in the homogeneous regions is the majority carrier drift current. These regions are shown in Figure 10.12.

We will now calculate the total current j flowing in the p-n junction for a given value of V. We make the assumption that the passage of carriers across the depletion layer is so swift that the generation and recombination of electrons and holes within the layer are negligible. In the steady state, the total current $j = j_e + j_h$ of electrons and holes will be constant. Thus, j_e and j_h can be evaluated separately at any arbitrary point in the depletion layer. The electron current j_e can easily be calculated at the boundary between the depletion layer and the diffusion region on the p-side (x_p), and the hole current j_h can easily be calculated at the boundary between the two regions on the n-side

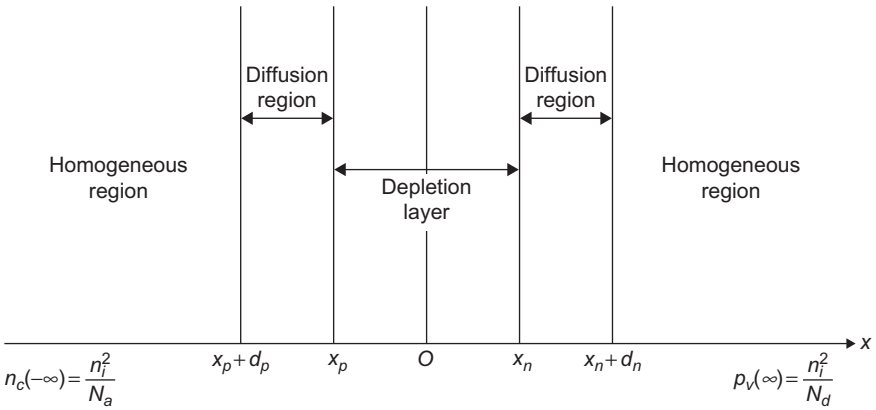

FIGURE 10.12

The depletion layer, the two diffusion and homogeneous regions when $V \neq 0$.

(x_n) in Figure 10.12. From Eq. (10.60) and Problem 10.4, we obtain the solution of the diffusion equations

$$p_v(x) = \frac{n_i^2}{N_d} + [p_v(x_n) - \frac{n_i^2}{N_d}]e^{-(x-x_n)/d_p}, \quad x \geq x_n \tag{10.63}$$

and

$$n_c(x) = \frac{n_i^2}{N_a} + \left[n_c(x_p) - \frac{n_i^2}{N_d}\right] e^{(x-x_p)/d_n}, \quad x \leq x_p. \tag{10.64}$$

Note that both x and x_p are negative in Eq. (10.64). From Eqs. (10.56), (10.57), (10.63), and (10.64), we obtain expressions for the minority carrier currents at the edges of the depletion layer ($x = x_n$ and $x = x_p$),

$$j_e(x_p) = \frac{eD_n}{d_n}\left[n_c(x_p) - \frac{n_i^2}{N_a}\right] \tag{10.65}$$

and

$$j_h(x_n) = \frac{eD_p}{d_p}\left[p_v(x_n) - \frac{n_i^2}{N_d}\right]. \tag{10.66}$$

From Problem 10.8,

$$n_c(x) \approx N_d e^{e\beta[V(x)-V(x_n)]} \tag{10.67}$$

and

$$p_v(x) \approx N_a e^{-e\beta[V(x)-V(x_p)]}. \tag{10.68}$$

From Eqs. (10.67), (10.68), and Problem 10.9,

$$n_c(x_p) = N_d e^{e\beta(V-\Delta\phi)} = \frac{n_i^2}{N_a} e^{e\beta V} \tag{10.69}$$

and

$$p_v(x_n) = N_a e^{e\beta(V-\Delta\phi)} = \frac{n_i^2}{N_d} e^{e\beta V}. \tag{10.70}$$

From Eqs. (10.65) and (10.69),

$$j_e(x_p) = e n_i^2 \frac{D_n}{d_n N_a}(e^{e\beta V} - 1), \tag{10.71}$$

and from Eqs. (10.66) and (10.70),

$$j_h(x_n) = e n_i^2 \frac{D_p}{d_p N_d}(e^{e\beta V} - 1). \tag{10.72}$$

The total current, $j = j_e(x_p) + j_h(x_n)$, is the sum of Eqs. (10.71) and (10.72),

$$j = e n_i^2 \left(\frac{D_n}{d_n N_a} + \frac{D_p}{d_p N_d}\right)(e^{e\beta V} - 1). \tag{10.73}$$

Eq. (10.73) is the ideal diode, or Shockley, equation. We note that because N_a and N_d appear in the denominator, the heavily doped side acts as a short circuit, whereas the current flow is in the lightly doped side.

10.4 TRANSISTORS

10.4.1 Bipolar Transistors

In the bipolar transistor, three alternately doped layers of semiconductor (n-p-n or p-n-p) are sandwiched. We will discuss the n-p-n bipolar transistor in the following section. The middle layer (p) is the base that is very narrow, and the other two layers (n) are known as the emitter and the collector. Energy band diagrams of an n-p-n bipolar transistor with no applied voltage and with an applied voltage, which makes the emitter negative with respect to the collector, are shown in Figure 10.13.

In Figure 10.13b, a voltage is applied such that the emitter is negative with respect to the collector. In such a scenario, the emitter-base junction is forward biased, whereas the

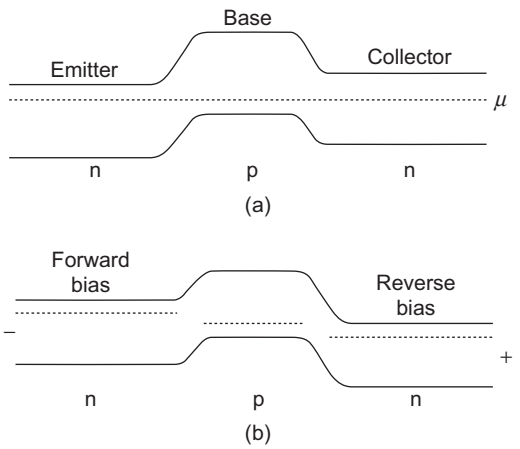

FIGURE 10.13

(a) An n-p-n bipolar transistor with no applied voltage. (b) The same transistor with the emitter negative with respect to the collector.

base-collector junction is reverse biased. In a p-n junction, each time an electron recombines with a hole in the base region, an acceptor ion is left behind that accumulates as more electrons flow through the junction and ultimately inhibits the flow of electrons. However, in a transistor, due to the contact with the base region, a current of holes flows into the base. The base remains electrically neutral because the number of holes drawn into the base compensates those that are lost due to recombination. The way in which a bipolar transistor acts as an amplifier is shown in Figure 10.14.

When a small input of hole current i_B is supplied to the base of the transistor, the base is positively charged, due to which the potential barrier between the base and the emitter is lowered. This results in an increase in the flow of electrons from the emitter to the base. Even if only a small percentage of the conduction electrons (from the emitter) recombine with the holes in the base while most of them flow into the collector, $i_E \gg i_B$, so that base remains neutral. Thus, the flow of current due to the electrons in the collector, i_C, is nearly equal to i_E, and the current gain in the amplifier, α, which is defined as the ratio of change in the output current Δi_C to the change in the input current Δi_B, is given by

$$\alpha = \frac{\Delta i_C}{\Delta i_B}. \tag{10.74}$$

Thus, a signal can be easily amplified by the bipolar transistor because a small change in the base current produces a very large change in the collector current.

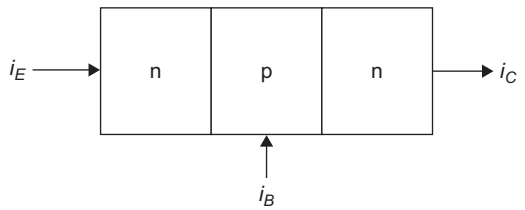

FIGURE 10.14

Amplification of current in a bipolar transistor.

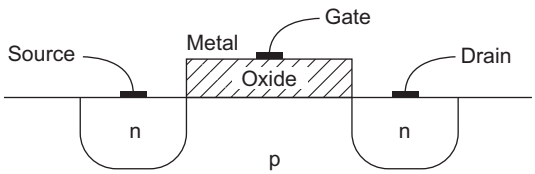

FIGURE 10.15

The structure of an n-type MOSFET.

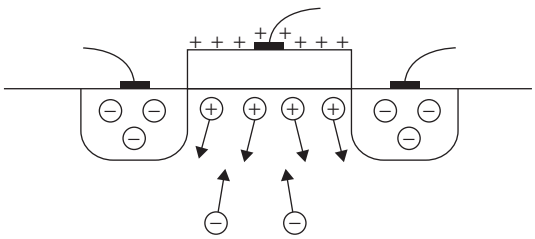

FIGURE 10.16

Electrons are attracted to the surface when a positive voltage is applied to the gate.

10.4.2 Field-Effect Transistor

The most common field-effect transistor (FET) is the metal-oxide-semiconductor FET, or the MOSFET. The structure of an n-type MOSFET is shown in Figure 10.15. In a MOSFET, a metal contact above the gate region is separated from the p-type semiconductor (Si) substrate by a thin oxide layer (usually SiO_2), which is an insulator.

When a positive voltage is applied to the metal contact above the gate region, the holes in the p-type semiconductor are repelled from the surface, whereas the electrons, which are the minority carriers, are attracted to the surface. This scenario is shown in Figure 10.16.

As the voltage is increased, more and more electrons, which are the minority carriers in the p-type semiconductor, are attracted to the surface. When the voltage increases beyond a threshold value, the number of electrons becomes greater than the number of holes near the surface even though it is a p-type semiconductor. This is known as the inversion layer, and the semiconductor behaves as if it is an n-type. This scenario is shown in Figure 10.17.

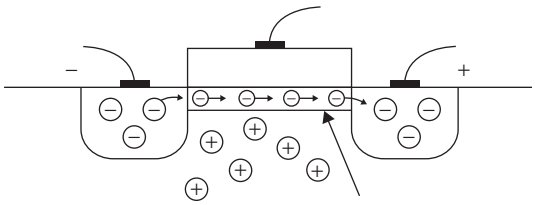

FIGURE 10.17

Inversion layer near the surface, when the gate voltage is greater than the threshold voltage.

Figure 10.17 explains how the MOSFET acts as a switch. The inversion layer allows the conduction electrons to flow from the source to the drain. The device is on when the current flows and off when the flow of current is stopped by decreasing the gate voltage below the threshold voltage. In such a situation, there is no inversion layer. An explanation of how the inversion layer is obtained is shown in Figure 10.18.

The Fermi energy is close to the valence band edge in the p-type semiconductor when it is well below the interface with the oxide layer. However, the electron energies are lowered near the surface of the semiconductor because there is a positive charge on the other side of the oxide layer, and both the conduction and valence bands are bent. If the Fermi energy is closer to the conduction band than to the valence band in this region, as shown in Figure 10.18, the n-type behavior is restricted to this region. Because the semiconductor displays n-type behavior, the probability of inversion depends on the degree of band bending.

The switching time of a transistor is defined as the minimum time period over which a transistor can be switched from off to on and again to the off position. The most important factor in designing integrated circuits is that a transistor must be in the on state until an electron (or hole) moves from the source to the drain. If it does not reach the drain before the transistor reaches the off position, no current will reach the drain, and the electronic device would not be switched on for any length of time. The MOSFET is eminently suitable to reduce the switching time in a

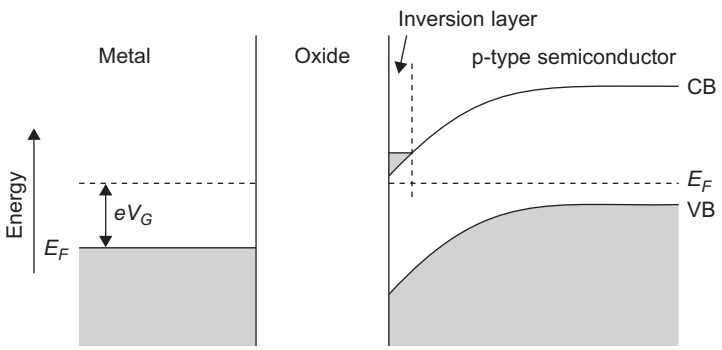

FIGURE 10.18

The band picture of the inversion layer in the MOSFET.

transistor. It may be noted that there are p-type MOSFETs, but because the mobility of the holes is much less than the mobility of the conduction electrons, the n-type MOSFETs can switch much faster than the p-type.

10.4.3 Single-Electron Transistor

As we have noted, a MOSFET is essentially a parallel-plate capacitor, one plate of which is a metal and the other of which is silicon, a semiconductor. SiO_2 is the insulator of the capacitor. Kastner et al.[7] designed a single-electron transistor (SET) that had two metal gates. A schematic diagram of their device is shown in Figure 10.19.

The electron gas, which is formed in the p-type Si by the positively biased upper gate, which is continuous as in the conventional MOSFET, is confined by the lower gate, which has a narrow gap of about 70 nm width. The narrow channel is approximately 20 nm wide and 1–10 μm long. The surface electrons are isolated from the bulk by the p-n junctions. The electron-rich region is called an inversion layer. When the voltage in the upper gate is positive with respect to the semiconductor while the bottom gate is neutral or negative, electrons are added to the semiconductor only under the gap in the bottom gate. Figure 10.19b shows how the electron gas is confined to move in one direction.

If V_g is the gate voltage, it can be easily shown that

$$e\Delta(N/L) = (C/L)\Delta V_g, \tag{10.75}$$

where C is the capacitance, L is the length, e is the electronic charge, and N is the number of electrons added. It was observed that when conductance g was plotted against the electron density (N/L), oscillations of the conductance were periodic. This is shown in Figure 10.20.

From Eq. (10.75) and Figure 10.20, it was observed that the transistor's conductance oscillated as a function of the number of electrons per unit length. In fact, there is a special length L_0, such

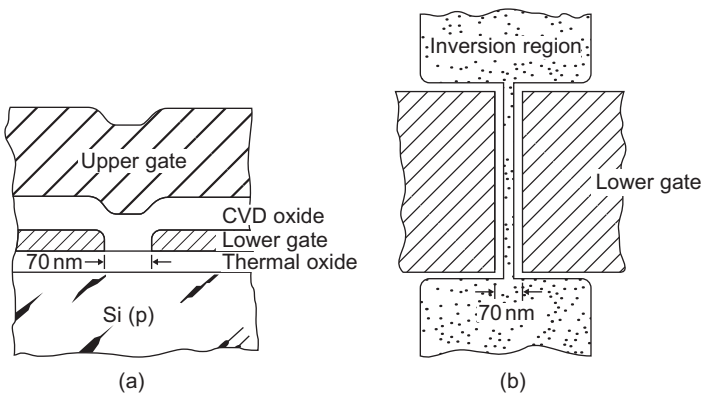

FIGURE 10.19

Schematic diagram of the (a) cross-section and (b) top view of the silicon transistor with a continuous upper gate and a gap in the lower gate.

Reproduced from Kastner et al.[7] with the permission of the American Physical Society.

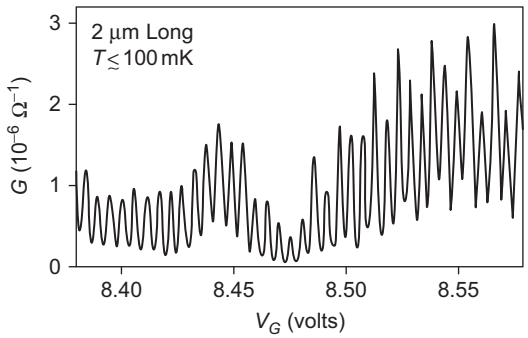

FIGURE 10.20

Conductance C versus gate voltage V_g in single-transistor Si devices.

Reproduced from Kastner et al.[7] with the permission of the American Physical Society.

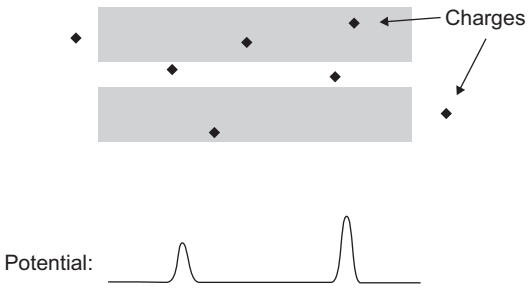

FIGURE 10.21

Narrow channel through which the electrons move. The metal gate with the gap is shown by the shaded area. The random distribution of charges near the interface is shown by black diamonds. The bottom sketch shows the electrostatic potential resulting from such impurities versus position along the channel.

Reproduced from Kastner et al.[7] with the permission of the American Physical Society.

that the voltage difference for one period is that necessary to add some fixed number of electrons to some fixed length,

$$e\Delta N = L_0(C/L)\Delta V_g. \tag{10.76}$$

In a separate experiment, it was also observed that the period was not correlated to the channel length. The sensitivity of the conductance fluctuations to thermal cycling lead to the conclusion that the period was determined by the distribution of charges at the Si-SiO_2 interfaces. These charges are always present even at densities of approximately 10^{10} per cm^2 for the best interfaces. It was calculated that for the transistor that had $2\,\mu m \times 1\,\mu m$ in size, there were approximately two charges adjacent to it. In fact, each sample had random distribution of charges that changed each time the sample was warmed to the room temperature. Kastner et al.[7] guessed that the charges create potential barriers along the length of the transistor, as shown schematically in Figure 10.21. They postulated that each period of conductance oscillations corresponded to the addition of one electron to the distance between the charges.

The proof that the period of oscillations did not correspond to the addition of two electrons due to spin degeneracy or four electrons due to the energy-band structure of Si depended on the measurement of the length L_0 defined in Eq. (10.76). The measurement of L_0 was facilitated by the discovery of small transistors in GaAs by Meirav et al.[13] Their device is shown in Figure 10.22.

Using molecular-beam epitaxy (MBE), a layer of AlGaAs, which has larger band gap than GaAs, is grown on a heavily doped GaAs crystal (n^+). Then a layer of pure GaAs is grown where the electrons accumulate. The density of the electrons is controlled by the positive voltage applied to the n^+ substrate. A metal gate, which is negatively biased so that the electrons are repelled from it, is deposited on the top by electron-beam lithography. The negative bias on the top gates creates a potential barrier for the electrons moving down the narrow channel.

The advantage of the GaAs-AlGaAs interface is that the density of charges near the semiconductor and insulator is smaller than the Si-SiO_2 interface. Figure 10.23 shows the conductance as a function of gate voltage V_g for two devices with different length.

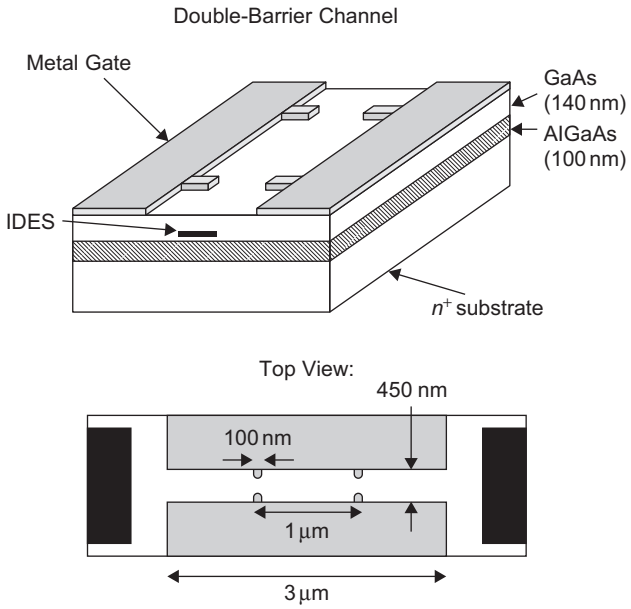

FIGURE 10.22

The double-barrier channel shows a one-dimensional electron gas (1DES) or narrow two-dimensional gas forms at the top of the GaAs-AlGaAs interface. The density is controlled by the substrate voltage V_g. The top view shows the top metal gate structure, which has a narrow channel with two constrictions.

Reproduced from Kastner et al.[7] with the permission of the American Physical Society.

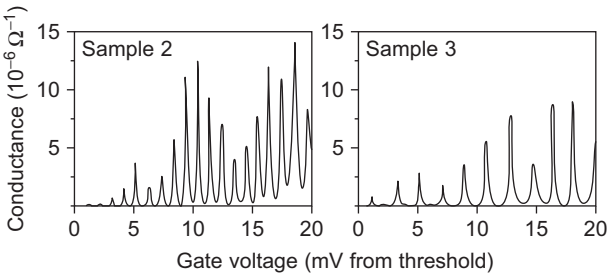

FIGURE 10.23

Conductance as a function of V_g for $L_0 = 0.8\,\mu m$ (sample 2) and $L_0 = 0.6\,\mu m$ (sample 3).

Reproduced from Kastner et al.[7] with the permission of the American Physical Society.

Each period corresponds to the addition of the same number of electrons to the regions between constrictions. The voltage to add one electron is larger for a shorter segment because the capacitance is smaller. However, it is necessary to know the absolute capacitance of a segment to be able to estimate whether the number of electrons added per period is one or two. The charge density was obtained by solving the Poisson equation and estimated by a method similar to the Thomas–Fermi approximation. The

capacitance was calculated at each gate voltage by integrating the charge contained in the region between the two constrictions. The capacitance was compared with the measured period. It was conclusively proved that the period of oscillations corresponds to the addition of one electron per oscillation. Thus, the one-electron transistor turns on and off again each time an electron is added.

We note from Figure 10.23 that the conductance of a small region of electron gas, separated by tunnel junctions from its leads, oscillates with density, which is explained by the simple Coulomb blockade model. A schematic picture of the tunnel barrier with a metal particle is shown in Figure 10.24.

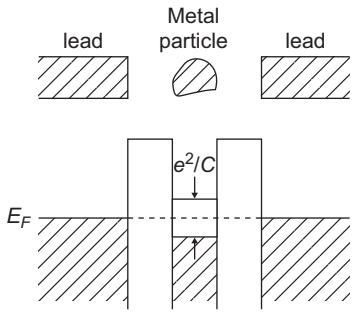

FIGURE 10.24

A Coulomb blockade system. Electrons tunnel from one lead onto a small metal particle and then to the other lead.

Reproduced from Kastner et al.[7] with the permission of the American Physical Society.

Due to the Coulomb interaction between the electrons in the metal particle, the electrons cannot tunnel from one plate of the capacitor to the other plate through the metal particle. For current to flow, an electron or hole of charge $q(\pm e)$ has to be added to the particle. This costs an energy $q^2/2C$, where C is the capacitance between the particle and the rest of the system. Thus, there is an energy gap in the single-particle density of states. If we consider the tunneling of an electron or a hole, the gap width is e^2/C. This is shown in Figure 10.24.

Because the potential difference between the gate and the electron gas is V_g, the isolated region of the transistor has an electrostatic energy

$$\varepsilon = \frac{q^2}{2C} - qV_g, \tag{10.77}$$

where $-qV_g$ is the attractive interaction between the positively charged gate and the charge in the isolated region, and $q^2/2C$ is the repulsive term between two charges in the isolated region. We note that Eq. (10.77) can be written in the alternate form

$$\varepsilon = (q - q_0)^2/2C, \tag{10.78}$$

where $q_0 = CV_g$ and a constant $q_0^2/2C$ has been added to ε. Any value of q_0 can be added to minimize the energy, but because the charge is quantized, only discrete values of energy are possible for a given q_o. Thus, either $q_0 = ne$ or $q_0 = (n + 1/2)e$, where n is an integer. These two cases are illustrated in Figure 10.25.

We note that when $q_0 = (n + 1/2)e$, the states where $q = ne$ and $q = (n + 1)e$ are degenerate. The energy gap in the tunneling density of states disappears because the charge fluctuates between the two values even at zero temperature. Because the conductance is thermally activated at all values of the gate voltage except those for $q_0 = (n + 1/2)e$, it has sharp peaks at low temperatures. The change in voltage to alter q_0 from $(n + 1/2)e$ to $(n + 3/2)e$ is $\Delta V_g = e/C$, which is the period in V_g. The activation energy at the minimum is $e^2/2C$.

To summarize, the number of electrons on the isolated segment is quantized because the time for tunneling onto and off the segment is long. The Coulomb interactions and the quantization of

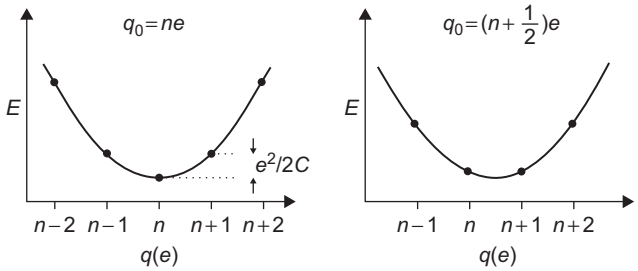

FIGURE 10.25

Energy versus charge for $q_0 = ne$ and $q_0 = (n + 1/2)e$. Because the charge is quantized, the allowed values of energy are shown by solid circles.

Reproduced from Kastner et al.[7] with the permission of the American Physical Society.

charge suppress all charge fluctuations at zero temperature for all values of q_0 not equal to $(n + 1/2)e$ and at that value the charge fluctuates only by one electron. This device is called a Coulomb island because the Coulomb interaction suppresses the charge fluctuations.

10.5 INTEGRATED CIRCUITS

An integrated circuit consists of capacitors, resistors, transistors, and other metallic connections required for a complete electrical circuit. The most popular electrical circuits are the MOSFET circuits because the switching time can be easily reduced. In addition, the transistors switch faster if the devices are made smaller. However, it is important to keep in check the power dissipated by a transistor so that the amount of heat produced in an integrated circuit can be kept under control.

Due to the improvement of the technology in building integrated circuits, primarily due to the decrease in the individual devices as well as in the increase in the area of the circuit, there has been a rapid growth in the number of transistors on an integrated circuit since the first such circuit was fabricated in 1961 with only four transistors. At present, a typical integrated circuit has about 80 million transistors. The single most important criterion is to keep in check the enormous heat produced by such circuits.

10.6 OPTOELECTRONIC DEVICES

Semiconductors can be utilized for a variety of optical properties. The simplest way to consider the role of a photon is that if it has energy $\in = h\nu$ greater than the band gap, it can excite an electron from the valence band to the conduction band, leaving behind a hole in the valence band. It may be noted that both the laws of conservation of momentum and energy have to be obeyed by the photons and electrons involved in the collision process where the energy is absorbed from the photon by the electron. This is usually possible only in direct band-gap semiconductors such as GaAs where the minimum of the conduction band is directly above the maximum of the valence band. This process, shown in Figure 10.26a, is known as photoconductivity because a beam of light with appropriate frequency can produce a large number of electrons and holes, thereby significantly

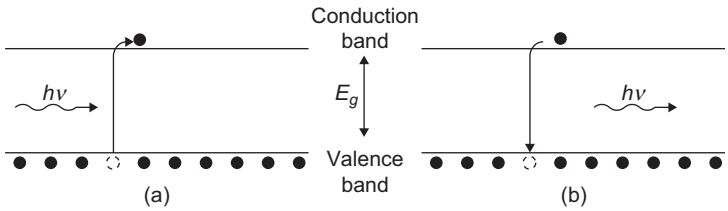

FIGURE 10.26

(a) A photon excites an electron from the valence band to the conduction band. (b) An electron in the conduction band recombines with a hole and produces a photon.

increasing the conductivity. In the reverse process, shown in Figure 10.26b, an electron in the conduction band recombines with a hole in the valence band and produces a photon of energy $\in = h\nu$. This is the basis of both the light-emitting diode (LED) and the semiconductor laser. As emphasized earlier, only semiconductors of which the minimum of the conduction band lies directly above the conduction band can be used for such a process.

A light-emitting diode constitutes a forward-biased p-n junction made from a direct band-gap material. The band gap of the semiconductor determines the wavelength of the emitted light. One can obtain light across most of the visible spectrum by using GaAsP if one varies the phosphorus content of the alloy semiconductor.

A photodiode, shown in Figure 10.27, is used to detect light. One uses a reverse-biased p-n junction. In this junction, when a valence electron in the depletion region on the p-type region absorbs a photon of the appropriate frequency, it is excited to the conduction band, where it is the minority carrier. The strong electric field sweeps it across the depletion region to the n-type region, and hence, the electron contributes to the drift current. The magnitude of this current is proportional to the intensity of light. This process can be used to construct a solar cell.

A laser (light amplification by stimulation of radiation) that produces coherent light is different from an LED in the sense that although the production of photons in an LED by the recombination of electrons and holes is a spontaneous process, in a laser the recombination of one-electron-hole pair triggers similar events in other electrons and holes, simultaneously producing a large number of photons of the same frequency. This is known as stimulated emission. The original diode laser, shown in Figure 10.28, had more electrons at the bottom of the conduction band than at the top of the valence band, which is obtained in a heavily doped p-n junction.

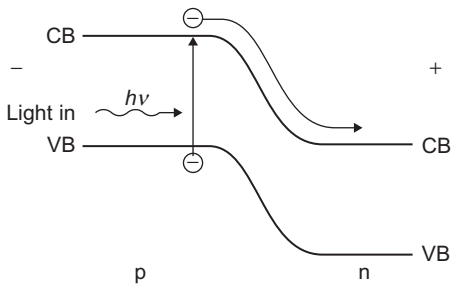

FIGURE 10.27

A reverse-biased p-n junction used as a photodetector.

In Figure 10.28, the population inversion exists only near the middle of the depletion region even though the recombination events take place in a wide region. Consequently, most of these events have spontaneous emission rather than stimulated emission as required in lasers. The efficiency of the lasing process is low. Because a major proportion of the electrons and holes are lost due to spontaneous emission, to replenish their loss, a large current must flow through the junction in a continuous basis. This large current produces so

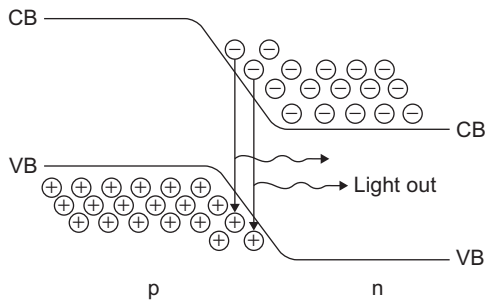

FIGURE 10.28

Schematic diagram of a diode laser.

much heat that the diode laser can be used for only short bursts or pulses.

The efficiency of the diode laser is very low. However, the efficiency of the laser is considerably improved by using two heterostructures instead of a p-n junction. A schematic diagram of the double-heterostructure laser is shown in Figure 10.29.

In a double-heterostructure laser, the conduction electrons and holes are confined in the GaAs layer. Due to the fact that both the conduction electrons and holes collect in the same region, the proportion of stimulated emission is much larger than a diode laser. In view of the above, this type of laser does not get heated very quickly, and therefore, it can be used for a longer period.

Semiconducting lasers are very widely used in communication systems using optical fibers. For best results, the wavelength is matched to the performance of the fiber. An InGaAsP laser, which is widely used in optical communication systems, is shown in Figure 10.30.

Recent studies on semiconductor lasers has shown that a semiconductor laser generates a number-phased squeezed state rather than a squeezed state, primarily due to the fact that a semiconductor laser is pumped by a shot-noise-free electric current. When the squeezed state becomes phase coherent with an independent local laser oscillator, the squeezed light can be detected by optical homodyne detectors and used for various interferometric measurements. Thus, a phase-coherent squeezed state can be generated by injection locking the squeezed slave laser with an external master laser.

A setup for generating a squeezed vacuum state by destructively interfering with an amplitude-squeezed state from an injection-locked slave laser with strong coherent light from a master laser is shown in Figure 10.31. A constant-current-driven semiconductor laser, which is denoted as a *slave laser,* is injection locked by an external *master laser*. Thus, the two signals, which are phase coherent, are combined at a high transmission mirror where the coherent excitation of the squeezed

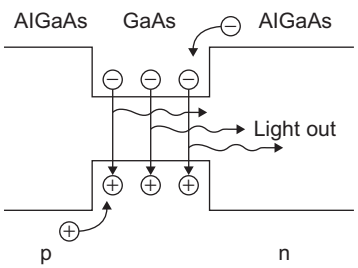

FIGURE 10.29

Schematic diagram of a double-heterostructure laser.

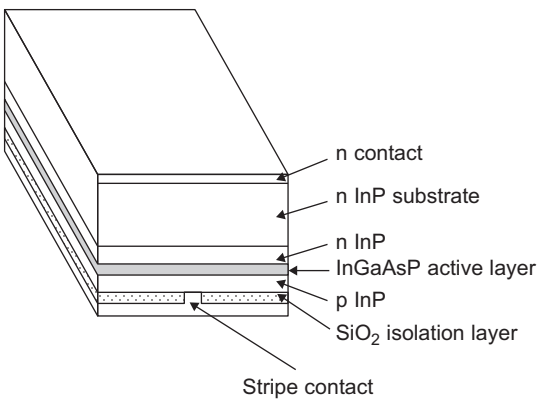

FIGURE 10.30

Schematic diagram of an InGaAsP laser.

output signal from the injection-locked slave laser is canceled by the destructive interference with the master laser signal. The squeezed output signal of the slave laser is not degraded because the noise of the master laser signal is attenuated by the mirror.

The actual experimental setup is shown in Figure 10.32. The master laser is an AlGaAs single-mode high-power semiconductor laser, and the slave laser is a single-mode low-power GaAs

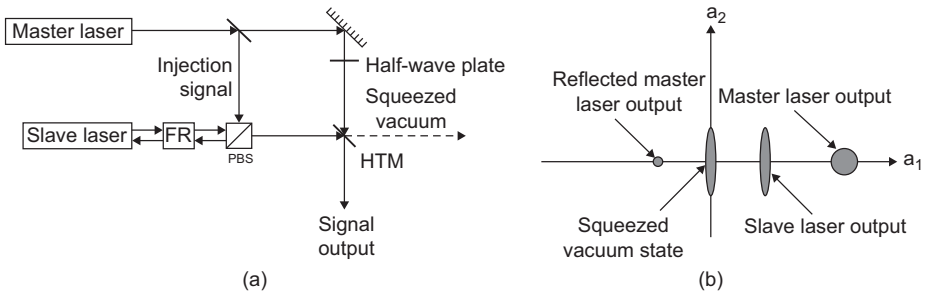

FIGURE 10.31

Squeezed-vacuum-state generation by mixing an amplitude-squeezed state with a coherent state. FR, Faraday rotator; PBS, polarization beam splitter, HTM, high-transmission mirror.

Reproduced from Yamamoto et al.[6] with the permission of Elsevier.

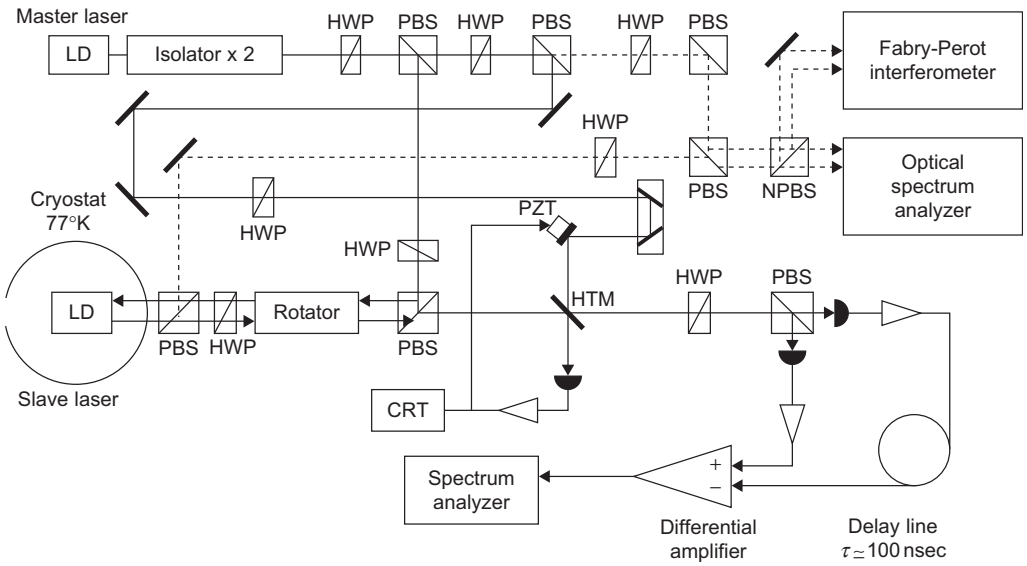

FIGURE 10.32

The experimental setup for the squeezed-vacuum-state generation by a semiconductor laser system; HWP, half-wave plate; PBS, polarization beam splitter; NPBS, nonpolarization bean splitter; HTM, high-transmission mirror; PZT, pizeo translator.

Reproduced from Yamamoto et al.[6] with the permission of Elsevier.

transverse-junction strip semiconductor laser with antireflection coating (∼10%) on the front facet and high-reflection coating (∼90%) on the rear facet.

10.7 GRAPHENE

Graphene is a flat monolayer of carbon atoms that is tightly packed into a two-dimensional honeycomb lattice. Graphene can be wrapped up into zero-dimensional fullerenes or rolled into one-dimensional nanotubes. It can also be stacked into three-dimensional graphite. Figure 10.33 shows the different ways in which graphene can be wrapped or stacked.

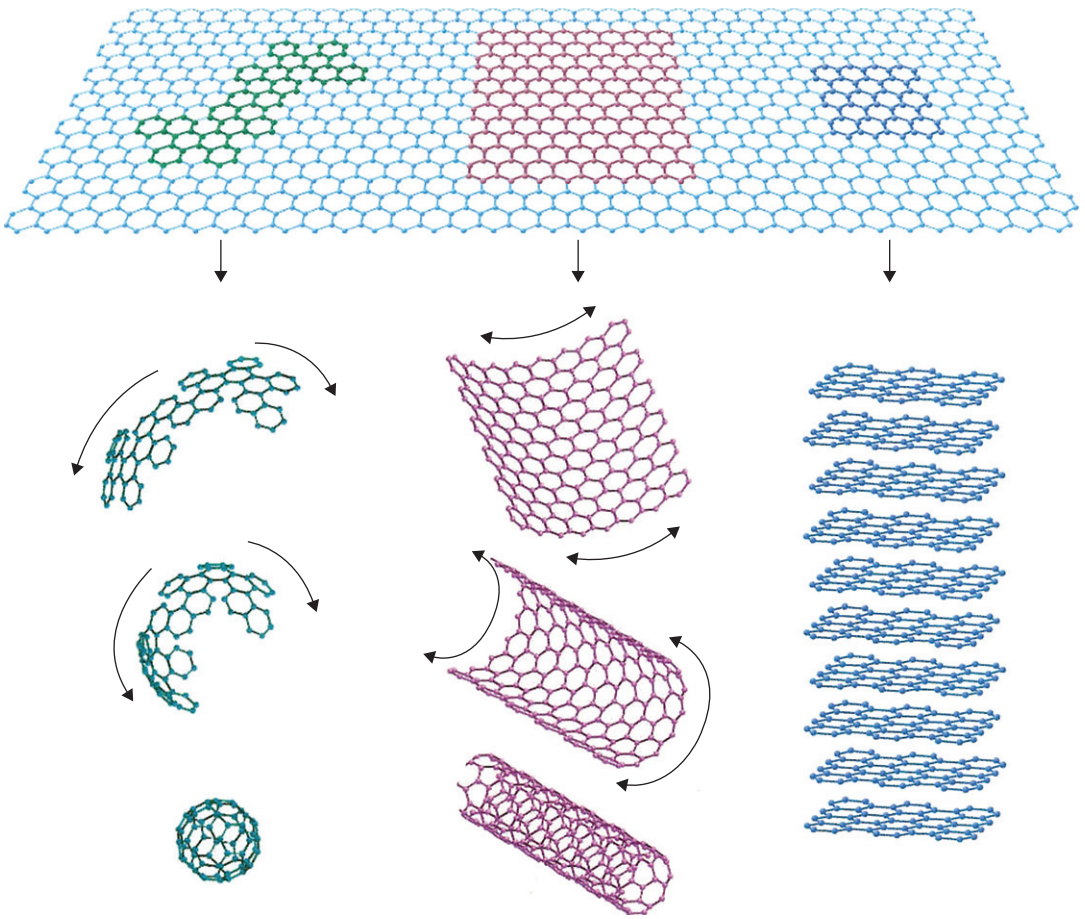

FIGURE 10.33

Graphene is a 2D carbon material that can be wrapped into 0D buckyballs, rolled into 1D nanotubes, or stacked into 3D graphites.

Reproduced from Geim[5] with the permission of MacMillan Publishers Ltd. Copyright 2011.

Graphene and its bilayer are both zero-gap semiconductors (which can also be considered as zero-overlap semimetals) with one type of electron and one type of hole. The conduction and valence bands start overlapping for three or more layers. Thus, single-, double-, and few- (3- to 9-) layer graphene can be considered as three types of 2D crystals. In these crystals, the charge carriers can travel thousands of interatomic distances.

Single- and few-layer graphenes have been grown epitaxially by chemical vapor deposition of hydrocarbons on metal substrates and thermal decomposition of SiC. Few-layer graphene obtained on SiC shows high-mobility charge carriers. Although epitaxial growth of graphene is the viable route for experimental applications, current experiments mostly use samples obtained by micromechanical cleavage of bulk graphite, which provides high-quality graphene crystallites up to 100 μm in size. Graphene becomes visible in an optical microscope if placed on an Si wafer with an appropriate thickness of SiO_2. This is due to feeble interference-like contrast with an empty wafer. These high-quality graphene crystallites are shown in Figure 10.34.

There is a pronounced ambipolar electric field effect in graphene that is shown in Figure 10.35.

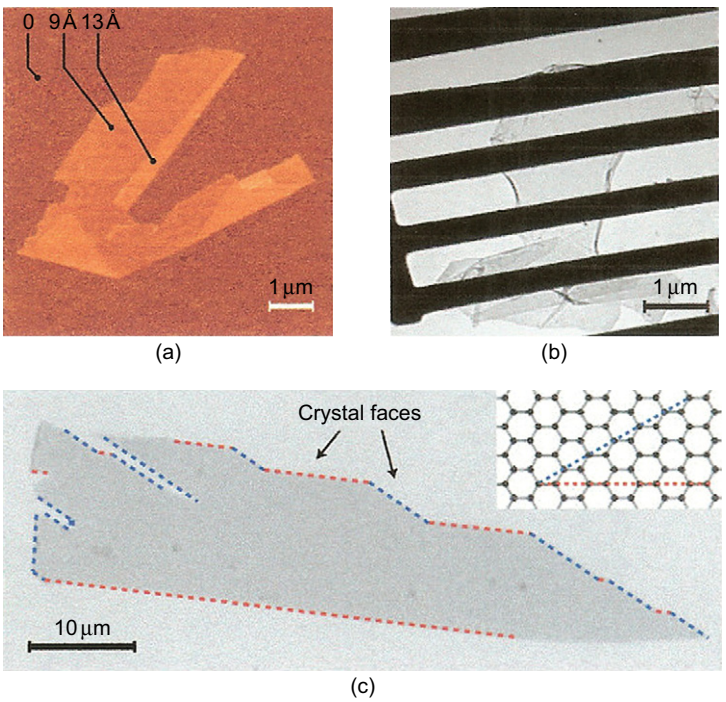

FIGURE 10.34

(a) Graphene visualized by atomic force microscopy. The folded region exhibiting a relative height of ≈ 4 Å indicates that it is a single crystal. (b) Transmission electron microscopy image of a graphene sheet freely suspended on a micrometer-sized metallic scaffold. (c) Scanning electron microscopy of relatively large graphene crystal, which shows that most of the crystal's faces are zigzag and armchair edges.

Reproduced from Geim[5] with the permission of MacMillan Publishers Ltd. Copyright 2011.

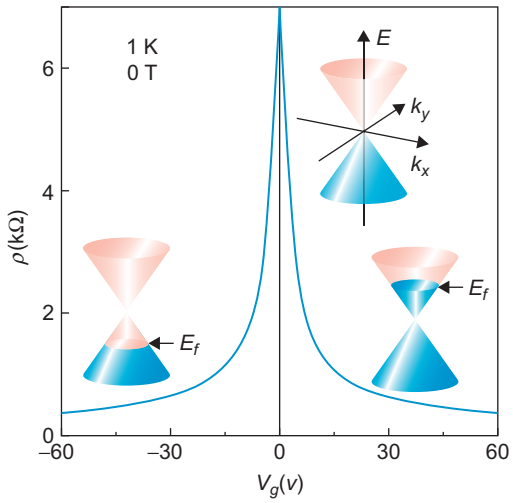

FIGURE 10.35

Ambipolar electric field effect in single-layer graphene.

Reproduced from Geim[5] with the permission of MacMillan Publishers. Copyright 2011.

The insets in Figure 10.35 show the conical low-energy spectrum $E(\mathbf{k})$, which shows the change in the Fermi energy E_F with gate voltage V_g. Positive V_g induces electrons and negative V_g induces holes. The concentration $n = \alpha V_g$, where $\alpha = 7.2 \times 10^{10}\ \text{cm}^{-2}\text{V}^{-1}$ for field-effect devices with a 300 nm SiO_2 layer used as a dielectric. The rapid decrease in resistivity ρ on adding charge carriers indicates their high mobility and does not change significantly when the temperature is increased to 300° K. The charge carriers in graphene can be tuned continuously between electrons and holes in concentrations $n \sim 10^{13}\ \text{cm}^{-2}$ and mobilities $\mu \geq 15{,}000\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$.

A unique property of graphene is that the interaction of the electrons with the periodic potential of the honeycomb lattice gives rise to quasiparticles that are described at low energies E by a $(2 + 1)$–dimensional Dirac equation with an effective speed of light $v_F = 10^6\ \text{m}^{-1}\text{s}^{-1}$. These quasiparticles are essentially electrons that have lost their rest mass m_e and are called massless Dirac fermions.

The conical sections of the energy spectrum for $|E| < 1$ eV are a consequence of the fact that graphene is a zero-gap semiconductor. The low-E quasiparticles within each valley can be described by the Dirac-like Hamiltonian, which is a direct consequence of graphene's crystal symmetry. A detailed analysis of the physical properties of graphene, including the origin of Dirac fermions, is given in Chapter 18 (on novel materials). The Hamiltonian H can be written as (for details, see Chapter 18):

$$H = \hbar v_F \begin{pmatrix} 0 & k_x - ik_y \\ k_x + ik_y & 0 \end{pmatrix} = \hbar v_F\ \vec{\sigma} \cdot \mathbf{k}. \tag{10.79}$$

Here, $\mathbf{k}$ is the quasiparticle momentum, $\vec{\sigma}$ is the 2D Pauli matrix (but pseudospin rather than real spin), and the Fermi velocity v_F, which is independent of k, plays the role of the speed of light. The honeycomb lattice of graphene is made up of two equivalent carbon sublattices, A and B. The cosine-like energy bands associated with the sublattices give rise to the conical sections of the energy spectrum for $|E| < 1$ eV (Figure 10.33) because they intersect at zero E near the edges of the Brillouin zone.

In addition to the unique feature of the band structure $E = \hbar v_F k$, the electronic states at zero E, where bands intersect, are a mixture of states of different sublattices. One has to use two-component wave functions (spinors) by requiring an index to identify sublattices A and B, to account for their relative contributions in the makeup of the quasiparticles. Thus, in Eq. (10.79), $\vec{\sigma}$ is referred as the pseudospin. The real spin of the electrons is described by additional terms in the Hamiltonian. However, the pseudospin effects, which are inversely proportional to the speed of light c (note that v_F plays the role of c) in quantum electrodynamics (QED), dominate the effects due to real spin because $c/v_F \approx 300$.

In QED, chirality (which is positive or negative for electrons or holes, respectively) is defined as the projection of $\overrightarrow{\sigma}$ on the direction of motion **k**. In graphene, the intricate connection of chirality is a consequence of the fact that k electrons and $-k$ hole states originate from the same carbon sublattices. Both chirality and pseudospin are conserved in graphene.

10.8 GRAPHENE-BASED ELECTRONICS

The high mobility (μ) of the charge carriers of graphene does not decrease even in the highest-field-induced concentrations and remains unchanged by chemical doping. Thus, there can be ballistic transport on a submicrometer scale at room temperature. The switching time is reduced due to the large value of v_F and low-resistance contacts without a Schottky barrier. The on–off ratios, which are comparatively low for graphene because of poor conductivity (≈ 100), do not create any problem for high-frequency applications. In fact, it has been shown that transistors can be operational at THz frequencies.

The fact that graphene remains metallic even at the neutrality point creates a problem for mainstream logic applications. However, it has been shown that significant semiconductor gaps, $\Delta E(\approx 0.3\ eV)$, can be induced in bilayer graphene, which can be used in tunable infrared lasers and detectors. ΔE can also be induced in single-layer graphene by spatial confinement or lateral-superlattice potential. If graphene is epitaxially grown on top of crystals with matching lattices such as BN or SiC, such superlattice effects are likely to occur.

It can be shown that the confinement gap for graphene is

$$\Delta E(eV) \approx \alpha \hbar v_F / d \approx 1/d\ (\text{nm}), \tag{10.80}$$

where the coefficient $\alpha \approx 0.5$ for Dirac fermions. For room-temperature operations, $d \approx 10$ nm, which is achievable with the rapidly advancing Si-based technology. However, hitherto, no technique has been found to make anisotropic etching of graphene to make devices with crystallographic-defined faces in order to avoid irregular edges. Electronic states associated with short irregular edges in short channels usually induce a significant sample-dependent conductance, whereas those associated in long channels lead to additional scattering.

Graphene can also be used as a conductive sheet where the different nanometer-size structures can be carved to make a single-electron transistor circuit. Graphene nanostructures are stable down to nanometer sizes that allow the exploration of a region between SET and molecular electronics. Figure 10.36 shows an SET made from graphene by using electron-beam lithography and dry etching.

Figure 10.36b shows that for a minimum feature size of ≈ 10 nm, the combined Coulomb and confinement gap reaches $> 3\ k_B T$. This would allow a SET-like circuitry operational at room temperature, and resistive barriers can be used to induce a Coulomb blockade.

There are two impediments for growth of graphene electronics. High-quality wafers suitable for industrial applications are yet to be developed despite the recent progress in epitaxial growth of graphene. In addition, individual features in graphene devices have to be controlled to provide reasonably accurate reproducibility in their properties.

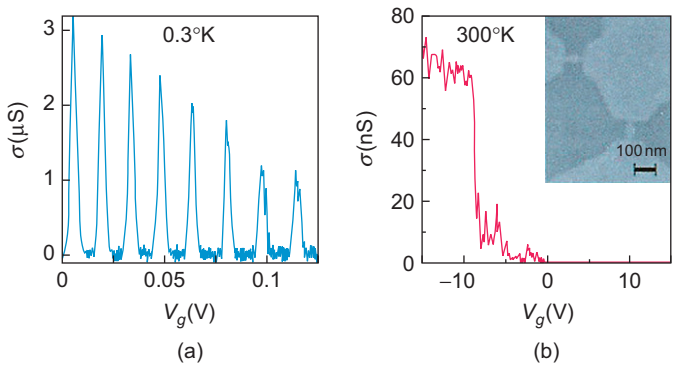

FIGURE 10.36

(a) Coulomb blockade in large quantum dots (diameter ≈0.25 μm) at low temperatures. Narrow constrictions in graphene serve as quantum barriers. (b) Here, 10 nm–scale graphene structures are stable under ambient conditions and survive thermal cycling to liquid-helium temperature. The inset shows a scanning electron micrograph of two graphene dots of ≈40 nm in diameter with narrower (<10 nm) constrictions.

Reproduced from Geim[5] with the permission of MacMillan Publishers Ltd. Copyright 2011.

PROBLEMS

10.1. In Eqs. (10.3) and (10.4), we derived

$$n_c(x) = \aleph_c(T)e^{-\beta(\varepsilon_c - e\phi(x) - \mu)} \tag{1}$$

and

$$p_v(x) = \wp_v(T)e^{\beta(\varepsilon_v - e\phi(x) - \mu)}. \tag{2}$$

Because $N_d = n_c(\infty)$ and $N_a = p_v(-\infty)$, show that

$$e\phi(\infty) - e\phi(-\infty) = E_g + \frac{1}{\beta}\ln\left[\frac{N_d N_a}{\aleph_c(T)\wp_v(T)}\right]. \tag{3}$$

10.2. From the preceding expressions, show that

$$n_c(x) = N_d e^{-e\beta[\phi(\infty) - \phi(x)]} \tag{1}$$

and

$$p_v(x) = N_a e^{e\beta[\phi(-\infty) - \phi(x)]}. \tag{2}$$

10.3. Within the boundaries of the depletion layer ($x = x_p$ and $x = x_n$), $n_c \ll N_d$ and $p_v \ll N_a$. Outside the depletion layer, $n_c = N_d$ *on the n-side* and $p_v = N_a$ *on the p-side*. Neglecting the boundary effects, show that, outside the depletion layer,

$$\rho(x) = 0, \tag{1}$$

and, inside the depletion layer,

$$\rho(x) = e[N_d(x) - N_a(x)]. \tag{2}$$

10.4. In the "abrupt junctions," one can approximate

$$N_d(x) = \left\{ \begin{matrix} N_d, & x>0 \\ 0, & x<0 \end{matrix} \right\} \tag{1}$$

and

$$N_a(x) = \left\{ \begin{matrix} N_a, & x<0 \\ 0, & x>0 \end{matrix} \right\}. \tag{2}$$

Show that the one-dimensional Poisson's equation (Eq. 10.10) can be written as

$$\frac{\partial^2 \phi}{\partial x^2} = \left\{ \begin{matrix} 0, & x<x_p \\ -\dfrac{4\pi e N_a}{\in}, & 0>x>x_p \\ \dfrac{4\pi e N_d}{\in}, & 0<x<x_n \\ 0, & x>x_n \end{matrix} \right\}. \tag{3}$$

By integrating Eq. (3), show that

$$\phi(x) = \left\{ \begin{matrix} \phi(-\infty), & x<x_p \\ \phi(-\infty) + \left(\dfrac{2\pi e N_a}{\in}\right)(x-x_p)^2, & 0>x>x_p \\ \phi(\infty) - \left(\dfrac{2\pi e N_d}{\in}\right)(x-x_n)^2, & 0<x<x_n \\ \phi(\infty), & x>x_n \end{matrix} \right\}. \tag{4}$$

10.5. From Eq. (10.6) and (10.7), prove the law of mass action,

$$n_c(x,T)p_v(x,T) = \aleph_c(T)\wp_v(T)e^{-\beta E_g}. \tag{1}$$

Since, for an intrinsic semiconductor, $n_c(x,T) = p_v(x,T) = n_i(T)$, and from Eq. (10.12), $n_c(\infty) = N_d$ (the n-side of the depletion layer), using Eq. (1), show that

$$p_v(\infty) \approx \frac{n_i^2(T)}{N_d}, \tag{2}$$

where

$$n_i(T) = n_i(0)e^{-\beta E_g/2}. \tag{3}$$

Similarly, because $p_v(-\infty) = N_a$ in the p-side of the depletion layer, show that

$$n_c(-\infty) \approx \frac{n_i^2(T)}{N_a}. \tag{4}$$

10.6. It has been shown that

$$\frac{\partial g}{\partial t} = -\dot{\mathbf{r}}.\frac{\partial g}{\partial \mathbf{r}} - \dot{\mathbf{k}} \cdot \frac{\partial g}{\partial \mathbf{k}} - \frac{g-f}{\tau_n}, \tag{1}$$

where g is the density of particles at position $\mathbf{r}$. In the presence of weak-applied fields, g is replaced by $g_{\mathbf{rk}}(t)$, which is the occupation number of electrons at position $\mathbf{r}$, which have a wave vector $\mathbf{k}$ at time t. It can be easily shown from Eq. (6.83) that at constant temperature

$$g_{\mathbf{rk}} \approx f_{\mathbf{rk}} - e\tau_n \frac{\partial f}{\partial \mu} \mathbf{v}_{\mathbf{k}} \cdot \mathbf{E}. \tag{2}$$

The electron density n_c at position $\mathbf{r}$ in the conduction band can be written as

$$n_c = \frac{1}{4\pi^3} \int g_{r\mathbf{k}} d\mathbf{k}, \tag{3}$$

while the equilibrium density of electrons in the conduction band can be written as

$$n_c^0 = \frac{1}{4\pi^3} \int f d\mathbf{k}. \tag{4}$$

Integrating Eq. (1) over $d\mathbf{k}$, show from Eqs. (1) through (4),

$$\frac{\partial n_c}{\partial t} = -\frac{\partial}{\partial \mathbf{r}} \cdot <\dot{\mathbf{r}}> + \frac{n_c^0 - n_c}{\tau_n}. \tag{5}$$

Here, n_c^0 is the equilibrium density of the electrons in the conduction band and $<\dot{\mathbf{r}}>$ is $\mathbf{v}_{\mathbf{k}}$ averaged over the Brillouin zone,

$$<\dot{\mathbf{r}}> = \frac{1}{4\pi^3 n_c} \int d\mathbf{k}\, g_{\mathbf{rk}} \mathbf{v}_{\mathbf{k}} = \frac{1}{4\pi^3 n_c} \int d\mathbf{k} \left[f - \tau_n \mathbf{v}_{\mathbf{k}} \cdot \left\{ e\mathbf{E} \frac{\partial f}{\partial \mu} + \frac{\partial f}{\partial \mathbf{r}} \right\} \right] \mathbf{v}_{\mathbf{k}}. \tag{6}$$

10.7. It has been shown that in the steady state, for the minority carriers,

$$D_n \frac{d^2 n_c}{dx^2} + \frac{n_c^0 - n_c}{\tau_n} = 0 \tag{1}$$

and

$$D_p \frac{d^2 p_v}{dx^2} + \frac{p_v^0 - p_v}{\tau_p} = 0. \tag{2}$$

Show that in the n-side of the depletion layer, for $x \geq x_n$,

$$p_v(x) - p_v(\infty) = [p_v(x_n) - p_v(\infty)] e^{-(x-x_n)/d_p}, \tag{3}$$

and in the p-side of the depletion layer, for $x \leq x_p$,

$$n_c(x) - n_c(-\infty) = [n_c(x_p) - n_c(-\infty)] e^{(x-x_p)/d_n}, \tag{4}$$

where the diffusion lengths for holes and electrons are defined as

$$d_p = (D_p \tau_p)^{1/2} \text{ and } d_n = (D_n \tau_n)^{1/2}. \tag{5}$$

10.8. From Eqs. (10.49) and (10.50), we have

$$j_e = e n_c \mu_n E + e D_n \frac{dn_c}{dx} \tag{1}$$

and

$$j_h = e p_v \mu_p E - e D_p \frac{dp_v}{dx}. \tag{2}$$

Show that

$$n_c(x) = N_d e^{\beta[V(x)-V(x_n)]} \left[1 + \frac{j_e}{e N_d D_n} \int_{x_n}^{x} dx' e^{-\beta[V(x')-V(x_n)]} \right] \tag{3}$$

and

$$p_v(x) = N_a e^{-\beta[V(x)-V(x_p)]} \left[1 - \frac{j_h}{e N_a D_p} \int_{x_p}^{x} dx' e^{\beta[V(x')-V(x_p)]} \right]. \tag{4}$$

Hence, show that the second term in the square bracket can be neglected compared to the first term in both Eqs. (3) and (4).

10.9. We derived in Chapter 9 that

$$n_i^2 = \aleph_c(T) \wp_v(T) e^{-\beta E_g}. \tag{1}$$

We also derived in Eq. (10.8),

$$e\Delta\phi = E_g + \frac{1}{\beta} \ln \left[\frac{N_d N_a}{\aleph_c(T) \wp_v(T)} \right]. \tag{2}$$

Show from Eqs. (1) and (2) that

$$n_i^2 = N_d N_a e^{-e\beta\Delta\phi}. \tag{3}$$

References

1. Ashcroft NW, Mermin ND. *Solid state physics*. New York: Brooks/Cole; 1976.
2. Bardeen J. Surface states and rectification at a metal semi-conductor contact. *Phys Rev* 1947;**71**:717.
3. Bardeen J, BrattinWH. The transistor, a semi-conductor triode. *Phys Rev* 1948;**74**:230.

4. Cohen ML, Chelikosky JR. *Electronic structure and optical properties of semiconductors*. Berlin: Springer-Verlag; 1989.
5. Geim AK, Novoselov KS. The rise of graphene. *Nat Mater* 2007;**6**:183.
6. Yamamoto Y, Inoue S, Bjork G, Heitmann H, Matinaga F. In: Kapon E, editor. *Semiconductor lasers,* vol. 1. Amsterdam: Elsevier; 1999:361.
7. Kastner MA. The single-electron transistor. *Rev Mod Phys* 1992;**64**:849.
8. Kimmerling LC, Kolenbrander KD, Michel J, Palm J. Light emission from silicon. *Solid State Phys Adv Res Appl* 1996;**50**:333.
9. Kittel C. *Introduction to solid state physics*. New York: John Wiley & Sons; 1976.
10. Klingston CF. *Semiconductor optics*. Berlin: Springer-Verlag; 1995.
11. Maiman TH. Stimulated optical radiation in ruby. *Nature* 1960;**147**:493.
12. Marder MP. *Condensed matter physics*. New York: John Wiley & Sons; 2000.
13. Meirav U, Kastner MA, Heiblum M, Wind SJ. *Phys Rev B* 1989;**40**:5871.