9

Metal–Semiconductor and Semiconductor Heterojunctions

In the preceding two chapters, we have considered the pn junction and assumed that the semiconductor material was the same throughout the structure. This type of junction is referred to as a *homojunction.* We developed the electrostatics of the junction and derived the current–voltage relationship. In this chapter, we consider the metal–semiconductor junction and the semiconductor heterojunction, in which the material on each side of the junction is not the same. These junctions can also produce diodes.

Semiconductor devices, or integrated circuits, must make contact with the outside world. This contact is made through nonrectifying metal–semiconductor junctions, or *ohmic contacts.* An ohmic contact is a low-resistance junction providing current conduction in both directions. We examine in this chapter the conditions that yield metal–semiconductor ohmic contacts. ■

9.0 | PREVIEW

In this chapter, we will:

■ Determine the energy-band diagram of a metal–semiconductor junction.

■ Investigate the electrostatics of the rectifying metal–semiconductor junction, which is known as the Schottky barrier diode.

■ Derive the ideal current–voltage relation of the Schottky barrier diode.

■ Discuss differences in the current transport mechanism between the Schottky barrier diode and pn junction diode, and discuss differences in turn-on voltage and switching times.

■ Discuss ohmic contacts, which are low-resistance, nonrectifying metal–semiconductor junctions.

■ Investigate the characteristics of a semiconductor heterojunction.

9.1 | THE SCHOTTKY BARRIER DIODE

One of the first practical semiconductor devices used in the early 1900s was the metal–semiconductor diode. This diode, also called a *point contact diode,* was made by touching a metallic whisker to an exposed semiconductor surface. These metal–semiconductor diodes were not easily reproduced or mechanically reliable and were replaced by the pn junction in the 1950s. However, semiconductor and vacuum technology is now used to fabricate reproducible and reliable metal–semiconductor contacts. In this section, we consider the metal–semiconductor rectifying contact, or Schottky barrier diode. In most cases, the rectifying contacts are made on n-type semiconductors; for this reason, we concentrate on this type of diode.

9.1.1 Qualitative Characteristics

The ideal energy-band diagram for a particular metal and n-type semiconductor before making contact is shown in Figure 9.1a. The vacuum level is used as a reference level. The parameter ϕ_m is the metal work function (measured in volts), ϕ_s is
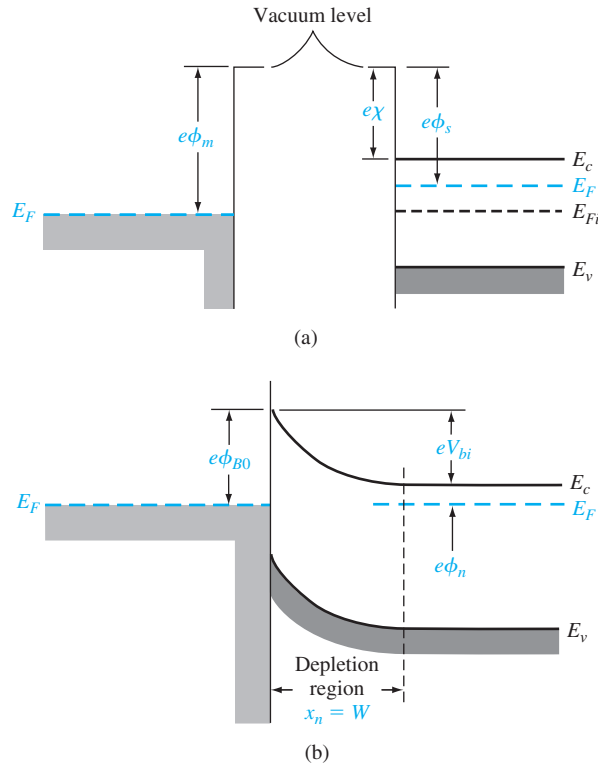

(a)

(b)

Figure 9.1 | (a) Energy-band diagram of a metal and semiconductor before contact; (b) ideal energy-band diagram of a metal–n-semiconductor junction for $\phi_m > \phi_s$.

Table 9.1 | Work functions of some elements

Element	Work function, ϕ_m
Ag, silver	4.26
Al, aluminum	4.28
Au, gold	5.1
Cr, chromium	4.5
Mo, molybdenum	4.6
Ni, nickel	5.15
Pd, palladium	5.12
Pt, platinum	5.65
Ti, titanium	4.33
W, tungsten	4.55

Table 9.2 | Electron affinity of some semiconductors

Element	Electron affinity, χ
Ge, germanium	4.13
Si, silicon	4.01
GaAs, gallium arsenide	4.07
AlAs, aluminum arsenide	3.5

the semiconductor work function, and χ is known as the *electron affinity*. The work functions of various metals are given in Table 9.1 and the electron affinities of several semiconductors are given in Table 9.2. In Figure 9.1a, we have assumed that $\phi_m > \phi_s$. The ideal thermal-equilibrium metal–semiconductor energy-band diagram, for this situation, is shown in Figure 9.1b. Before contact, the Fermi level in the semiconductor was above that in the metal. In order for the Fermi level to become a constant through the system in thermal equilibrium, electrons from the semiconductor flow into the lower energy states in the metal. Positively charged donor atoms remain in the semiconductor, creating a space charge region.

The parameter ϕ_{B0} is the ideal barrier height of the semiconductor contact, the potential barrier seen by electrons in the metal trying to move into the semiconductor. This barrier is known as the *Schottky barrier* and is given, ideally, by

$$\phi_{B0} = (\phi_m - \chi) \tag{9.1}$$

On the semiconductor side, V_{bi} is the built-in potential barrier. This barrier, similar to the case of the pn junction, is the barrier seen by electrons in the conduction band trying to move into the metal. The built-in potential barrier is given by

$$V_{bi} = \phi_{B0} - \phi_n \tag{9.2}$$

which makes V_{bi} a slight function of the semiconductor doping, as is the case in a pn junction.

If we apply a positive voltage to the semiconductor with respect to the metal, the semiconductor-to-metal barrier height increases, while ϕ_{B0} remains constant in this idealized case. This bias condition is the reverse bias. If a positive voltage is applied to the metal with respect to the semiconductor, the semiconductor-to-metal barrier V_{bi} is reduced while ϕ_{B0} again remains essentially constant. In this situation, electrons can more easily flow from the semiconductor into the metal since the barrier has been reduced. This bias condition is the forward bias. The energy-band diagrams for the reverse and forward bias are shown in Figures 9.2a,b where V_R is the magnitude of the reverse-biased voltage and V_a is the magnitude of the forward-bias voltage.
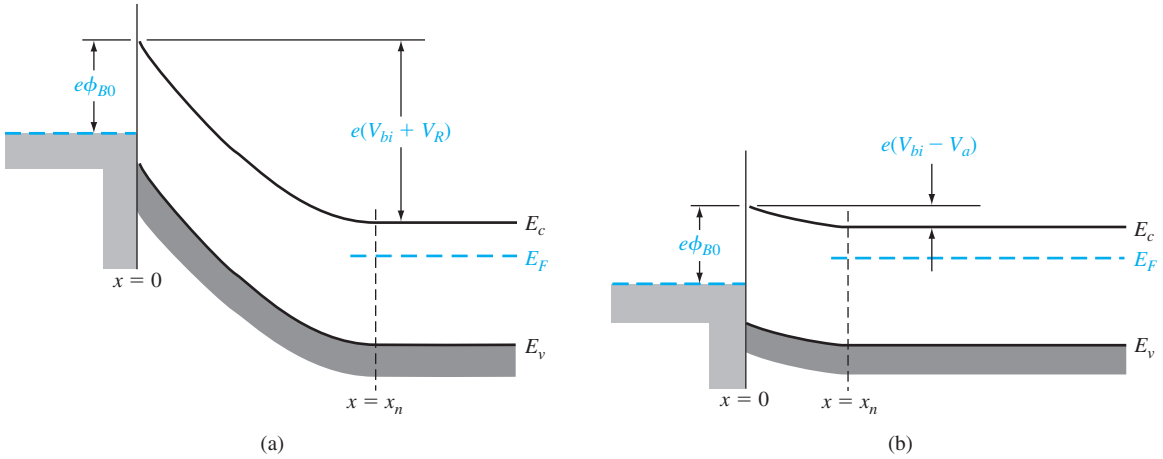
Figure 9.2 | Ideal energy-band diagram of a metal–semiconductor junction (a) under reverse bias and (b) under forward bias.

The energy-band diagrams versus voltage for the metal–semiconductor junction shown in Figure 9.2 are very similar to those of the pn junction given in the previous chapter. Because of this similarity, we expect the current–voltage characteristics of the Schottky barrier junction to be similar to the exponential behavior of the pn junction diode. The current mechanism here, however, is due to the flow of majority carrier electrons. In forward bias, the barrier seen by the electrons in the semiconductor is reduced, so majority carrier electrons flow more easily from the semiconductor into the metal. The forward-bias current is in the direction from metal to semiconductor: It is an exponential function of the forward-bias voltage V_a.

9.1.2   Ideal Junction Properties

We can determine the electrostatic properties of the junction in the same way as we do for the pn junction. The electric field in the space charge region is determined from Poisson's equation. We have that

$$\frac{d\mathrm{E}}{dx} = \frac{\rho(x)}{\epsilon_s} \tag{9.3}$$

where $\rho(x)$ is the space charge volume density and ϵ_s is the permittivity of the semiconductor. If we assume that the semiconductor doping is uniform, then by integrating Equation (9.3), we obtain

$$\mathrm{E} = \int \frac{eN_d}{\epsilon_s} dx = \frac{eN_d x}{\epsilon_s} + C_1 \tag{9.4}$$

where C_1 is a constant of integration. The electric field is zero at the space charge edge in the semiconductor, so the constant of integration can be found as

$$C_1 = -\frac{eN_d x_n}{\epsilon_s} \tag{9.5}$$

The electric field can then be written as

$$E = -\frac{eN_d}{\epsilon_s}(x_n - x)$$ (9.6)

which is a linear function of distance, for the uniformly doped semiconductor, and reaches a peak value at the metal–semiconductor interface. Since the E-field is zero inside the metal, a negative surface charge must exist in the metal at the metal–semiconductor junction.

The space charge region width, W, may be calculated as we do for the pn junction. The result is identical to that of a one-sided p^+n junction. For the uniformly doped semiconductor, we have

$$W = x_n = \left[\frac{2\epsilon_s(V_{bi} + V_R)}{eN_d}\right]^{1/2}$$ (9.7)

where V_R is the magnitude of the applied reverse-biased voltage. We are again assuming an abrupt junction approximation.

Objective: Determine the theoretical barrier height, built-in potential barrier, and maximum electric field in a metal–semiconductor diode for zero applied bias.

EXAMPLE 9.1

Consider a contact between tungsten and n-type silicon doped to $N_d = 10^{16}$ cm^{-3} at $T = 300$ K.

■ **Solution**

The metal work function for tungsten (W) from Table 9.1 is $\phi_m = 4.55$ V and the electron affinity for silicon from Table 9.2 is $\chi = 4.01$ V. The barrier height is then

$$\phi_{B0} = \phi_m - \chi = 4.55 - 4.01 = 0.54 \text{ V}$$

where ϕ_{B0} is the ideal Schottky barrier height. We can calculate ϕ_n as

$$\phi_n = \frac{kT}{e}\ln\left(\frac{N_c}{N_d}\right) = 0.0259 \ln\left(\frac{2.8 \times 10^{19}}{10^{16}}\right) = 0.206 \text{ V}$$

Then

$$V_{bi} = \phi_{B0} - \phi_n = 0.54 - 0.206 = 0.334 \text{ V}$$

The space charge width at zero bias is

$$x_n = \left[\frac{2\epsilon_s V_{bi}}{eN_d}\right]^{1/2} = \left[\frac{2(11.7)(8.85 \times 10^{-14})(0.334)}{(1.6 \times 10^{-19})(10^{16})}\right]^{1/2}$$

or

$$x_n = 0.208 \times 10^{-4} \text{ cm}$$

Then the maximum electric field is

$$|E_{max}| = \frac{eN_d x_n}{\epsilon_s} = \frac{(1.6 \times 10^{-19})(10^{16})(0.208 \times 10^{-4})}{(11.7)(8.85 \times 10^{-14})}$$

or finally

$$|E_{max}| = 3.21 \times 10^4 \text{ V/cm}$$

■ Comment

The values of space charge width and electric field are very similar to those obtained for a pn junction.

■ EXERCISE PROBLEM

Ex 9.1 Consider an ideal tungsten-to-n-type GaAs junction. Assume the GaAs is doped to a concentration of $N_d = 5 \times 10^{15}$ cm^{-3}. Determine the theoretical barrier height, the built-in potential barrier, and maximum electric field for the case of zero applied bias.

(Ans. $\phi_{B0} = 0.48$ V, $V_{bi} = 0.3623$ V, $|E_{max}| = 2.24 \times 10^4$ V/cm)

A junction capacitance can also be determined in the same way as we do for the pn junction. We have that

$$C' = eN_d \frac{dx_n}{dV_R} = \left[\frac{e\epsilon_s N_d}{2(V_{bi} + V_R)} \right]^{1/2} \tag{9.8}$$

where C' is the capacitance per unit area. If we square the reciprocal of Equation (9.8), we obtain

$$\left(\frac{1}{C'} \right)^2 = \frac{2(V_{bi} + V_R)}{e\epsilon_s N_d} \tag{9.9}$$

We can use Equation (9.9) to obtain, to a first approximation, the built-in potential barrier V_{bi}, and the slope of the curve from Equation (9.9) to yield the semiconductor doping N_d. We can calculate the potential ϕ_n and then determine the Schottky barrier ϕ_{B0} from Equation (9.2).

EXAMPLE 9.2

Objective: To calculate the semiconductor doping concentration and Schottky barrier height from the silicon diode experimental capacitance data shown in Figure 9.3. Assume $T = 300$ K.

■ Solution

The intercept of the tungsten–silicon curve is approximately at $V_{bi} = 0.40$ V. From Equation (9.9), we can write

$$\frac{d(1/C')^2}{dV_R} \approx \frac{\Delta(1/C')^2}{\Delta V_R} = \frac{2}{e\epsilon_s N_d}$$

Then, from the figure, we have

$$\frac{\Delta(1/C')^2}{\Delta V_R} \approx 4.4 \times 10^{13}$$

so that

$$N_d = \frac{2}{(1.6 \times 10^{-19})(11.7)(8.85 \times 10^{-14})(4.4 \times 10^{13})} = 2.7 \times 10^{17} \text{ cm}^{-3}$$

We can calculate

$$\phi_n = \frac{kT}{e} \ln \left(\frac{N_c}{N_d} \right) = (0.0259) \ln \left(\frac{2.8 \times 10^{19}}{2.7 \times 10^{17}} \right) = 0.12 \text{ V}$$

so that

$$\phi_{Bn} = V_{bi} + \phi_n = 0.40 + 0.12 = 0.52 \text{ V}$$

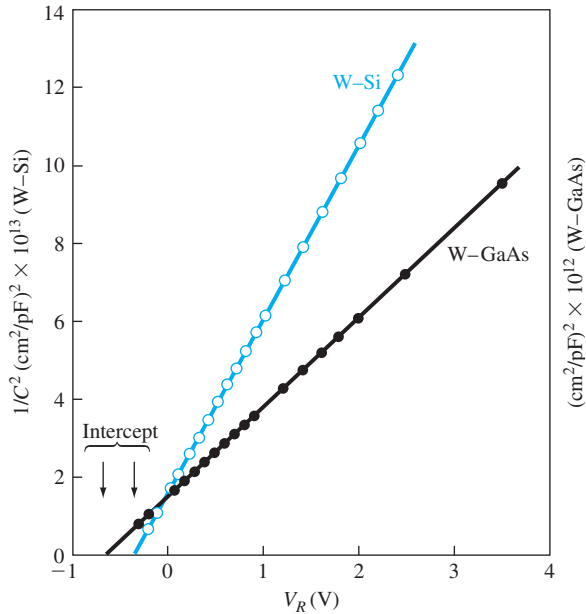where ϕ_{Bn} is the actual Schottky barrier height.

Figure 9.3 | $1/C^2$ versus V_R for W–Si and W–GaAs Schottky barrier diodes.
(From Sze and Ng [15].)

■ Comment

The experimental value of 0.52 V can be compared with the ideal barrier height of $\phi_{B0} = 0.54$ V found in Example 9.1. These results agree fairly well. For other metals, the discrepancy between experiment and theory is larger.

■ EXERCISE PROBLEM

Ex 9.2 Repeat Example 9.2 for the GaAs diode capacitance data shown in Figure 9.3.

(Ans. $V_{bi} \cong 0.64$ V, $N_d = 4.62 \times 10^{18}$ cm^{-3}.)

We can see that the built-in potential barrier of the gallium arsenide Schottky diode is larger than that of the silicon diode. This experimental result is normally observed for all types of metal contacts.

TEST YOUR UNDERSTANDING

TYU 9.1 Consider an ideal chromium-to-n-type silicon Schottky diode at $T = 300$ K. Assume the semiconductor is doped at a concentration of $N_d = 3 \times 10^{15}$ cm^{-3}. Determine the (*a*) ideal Schottky barrier height, (*b*) built-in potential barrier, (*c*) peak electric field with an applied reverse-biased voltage of $V_R = 5$ V, and (*d*) junction capacitance per unit area for $V_R = 5$ V. [Ans. (*a*) $\phi_{B0} = 0.49$ V; (*b*) $V_{bi} = 0.253$ V; (*c*) $|E_{max}| = 6.98 \times 10^4$ V/cm; (*d*) $C' = 6.88 \times 10^{-9}$ F/cm^2.]

TYU 9.2   Repeat TYU 9.1 for an ideal palladium-to-n-type GaAs Schottky diode with the
same impurity concentration. [$_z$ɯɔ/Ⅎ $_{6-}$0I × 98·9 = ,Ɔ (p)
:ɯɔ/Λ $_4$0I × ᘔ = | ^{xɐɯ}Ǝ| (ɔ) :Λ 6I6·0 = $^{!q}$Λ (q) :Λ ϛ0·I = $^{0ℵ}ϕ$ (ɒ) ·su∀]

9.1.3   Nonideal Effects on the Barrier Height

Schottky Barrier Lowering   Several effects alter the actual Schottky barrier height
from the theoretical value given by Equation (9.1). The first such effect that we con-
sider is the Schottky effect, or image-force-induced lowering of the potential barrier.

An electron in a dielectric material at a distance x from the metal will create an
electric field. The electric field can be determined by adding an image charge, $+e$, inside
the metal located at the same distance, $|x|$, from the interface. This image effect is shown
in Figure 9.4a. Note that the E–field lines are perpendicular to the metal surface as we
expect. The force on the electron, due to the coulomb attraction with the image force, is

$$F = \frac{-e^2}{4\pi\epsilon_s(2x)^2} = -e\mathrm{E} \tag{9.10}$$

Figure 9.4 | (a) Image charge and electric field lines at a metal–dielectric interface.
(b) Distortion of the potential barrier due to image forces with zero electric field and (c) with
a constant electric field.

The potential can then be found as

$$-\phi(x) = + \int_x^\infty E\,dx' = + \int_x^\infty \frac{e}{4\pi\epsilon_s \cdot 4(x')^2}\,dx' = \frac{-e}{16\pi\epsilon_s x} \qquad (9.11)$$

where x' is the integration variable and where we have assumed that the potential is zero at $x = \infty$.

The potential energy of the electron is $-e\phi(x)$: Figure 9.4b is a plot of the potential energy assuming that no other electric fields exist. With an electric field present in the dielectric, the potential is modified and can be written as

$$-\phi(x) = \frac{-e}{16\pi\epsilon_s x} - Ex \qquad (9.12)$$

The potential energy of the electron, including the effect of a constant electric field, is plotted in Figure 9.4c. The peak potential barrier is now lowered. This lowering of the potential barrier is the Schottky effect, or image force–induced lowering.

We can find the Schottky barrier lowering, $\Delta\phi$, and the position of the maximum barrier, x_m, from the condition that

$$\frac{d[e\phi(x)]}{dx} = 0 \qquad (9.13)$$

We find that

$$x_m = \sqrt{\frac{e}{16\pi\epsilon_s E}} \qquad (9.14)$$

and

$$\boxed{\Delta\phi = \sqrt{\frac{eE}{4\pi\epsilon_s}}} \qquad (9.15)$$

Objective: Calculate the Schottky barrier lowering and the position of the maximum barrier height.

Consider a gallium arsenide metal–semiconductor contact in which the electric field in the semiconductor is assumed to be $E = 6.8 \times 10^4$ V/cm.

■ **Solution**
The Schottky barrier lowering is given by Equation (9.15), which in this case yields

$$\Delta\phi = \sqrt{\frac{eE}{4\pi\epsilon_s}} = \sqrt{\frac{(1.6 \times 10^{-19})(6.8 \times 10^4)}{4\pi\,(13.1)(8.85 \times 10^{-14})}} = 0.0273 \text{ V}$$

The position of the maximum barrier height is

$$x_m = \sqrt{\frac{e}{16\pi\epsilon_s E}} = \sqrt{\frac{(1.6 \times 10^{-19})}{16\pi(13.1)(8.85 \times 10^{-14})(6.8 \times 10^4)}}$$

or

$$x_m = 2 \times 10^{-7} \text{ cm} = 20 \text{ Å}$$

■ Comment

Although the Schottky barrier lowering may seem like a small value, the barrier height and the barrier lowering will appear in exponential terms in the current–voltage relationship. A small change in the barrier height can thus have a significant effect on the current in a Schottky barrier diode.

■ EXERCISE PROBLEM

Ex 9.3   Consider the Schottky diode described in Example 9.1. Calculate the Schottky barrier lowering for a reverse-biased voltage of (*a*) $V_R = 1$ V and (*b*) $V_R = 5$ V.

[Ans. (*a*) $\Delta\phi = 0.0281$ V; (*b*) $\Delta\phi = 0.0397$ V]

Interface States   Figure 9.5 shows the measured barrier heights in gallium arsenide and silicon Schottky diodes as a function of metal work functions. There is a monotonic relation between the measured barrier height and the metal work function, but the curves do not fit the simple relation given in Equation (9.1). The barrier height of the metal–semiconductor junction is determined by both the metal work function and the semiconductor surface or interface states.

A more detailed energy-band diagram of a metal to n-type semiconductor contact in thermal equilibrium is shown in Figure 9.6. We assume that a narrow interfacial layer of insulator exists between the metal and semiconductor. The interfacial layer can support a potential difference, but will be transparent to the flow of electrons between the metal and semiconductor. The semiconductor also shows a distribution of surface states at the metal–semiconductor interface. We assume that all states below
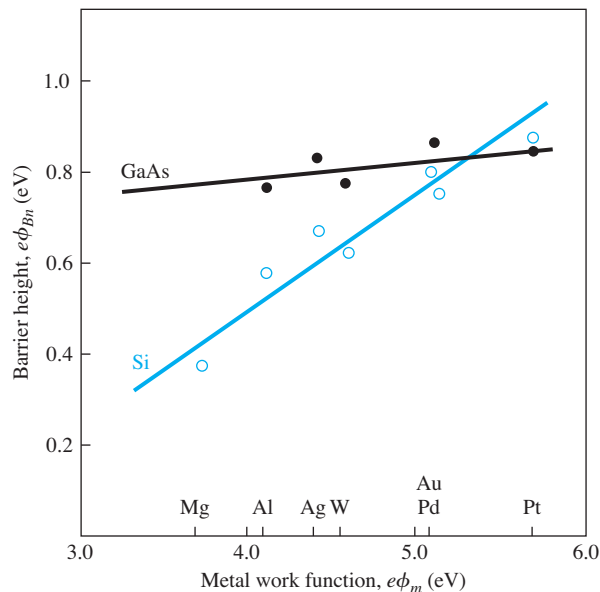
Figure 9.5 | Experimental barrier heights as a function of metal work functions for GaAs and Si.
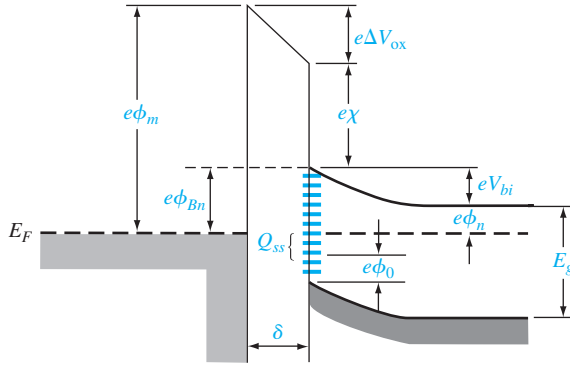*(From Crowley and Sze [2].)*

Figure 9.6 | Energy-band diagram of a metal–semiconductor junction with an interfacial layer and interface states.

the surface potential ϕ_0 are donor states, which will be neutral if the state contains an electron and positively charged if the state does not contain an electron. We also assume that all states above ϕ_0 are acceptor states, which will be neutral if the state does not contain an electron and negatively charged if the state contains an electron.

The diagram in Figure 9.6 shows some acceptor states above ϕ_0 and below E_F. These states tend to contain electrons and are negatively charged. We may assume that the surface state density is constant and equal to D_{it} states/cm²-eV. The relation between the surface potential, surface state density, and other semiconductor parameters is found to be

$$(E_g - e\phi_0 - e\phi_{Bn}) = \frac{1}{eD_{it}} \sqrt{2e\epsilon_s N_d(\phi_{Bn} - \phi_n)} - \frac{\epsilon_i}{eD_{it}\delta} [\phi_m - (\chi + \phi_{Bn})]$$

(9.16)

We consider two limiting cases.

Case 1 Let $D_{it} \to \infty$. In this case, the right side of Equation (9.16) goes to zero. We then have

$$\phi_{Bn} = \frac{1}{e} (E_g - e\phi_0)$$

(9.17)

The barrier height is now fixed by the bandgap energy and the potential ϕ_0. The barrier height is totally independent of the metal work function and the semiconductor electron affinity. The Fermi level becomes "pinned" at the surface, at the surface potential ϕ_0.

Case 2 Let $D_{it}\delta \to 0$. Equation (9.16) reduces to

$$\phi_{Bn} = (\phi_m - \chi)$$

which is the original ideal expression.

The Schottky barrier height is a function of the electric field in the semiconductor through the barrier lowering effect. The barrier height is also a function of the surface states in the semiconductor. The barrier height, then, is modified from the ideal theoretical value. Since the surface state density is not predictable with any degree of certainty, the barrier height must be an experimentally determined parameter.

TEST YOUR UNDERSTANDING

TYU 9.3   Determine the Schottky barrier lowering and the position of the maximum barrier height for the junction described in TYU 9.1. Use the value of the electric field found in TYU 9.1. (Ans. $\Delta\phi = 0.0293$ eV, $x_m = 21$ Å)

9.1.4   Current–Voltage Relationship

The current transport in a metal–semiconductor junction is due mainly to majority carriers as opposed to minority carriers in a pn junction. The basic process in the rectifying contact with an n-type semiconductor is by transport of electrons over the potential barrier, which can be described by the thermionic emission theory.

The thermionic emission characteristics are derived by assuming that the barrier height is much larger than kT, so that the Maxwell–Boltzmann approximation applies and that thermal equilibrium is not affected by this process. Figure 9.7 shows the one-dimensional barrier with an applied forward-bias voltage V_a and two electron current density components. The current $J_{s \to m}$ is the electron current density due to the flow of electrons from the semiconductor into the metal, and the current $J_{m \to s}$ is the electron current density due to the flow of electrons from the metal into the semiconductor. The subscripts of the currents indicate the direction of electron flow. The conventional current direction is opposite to electron flow.
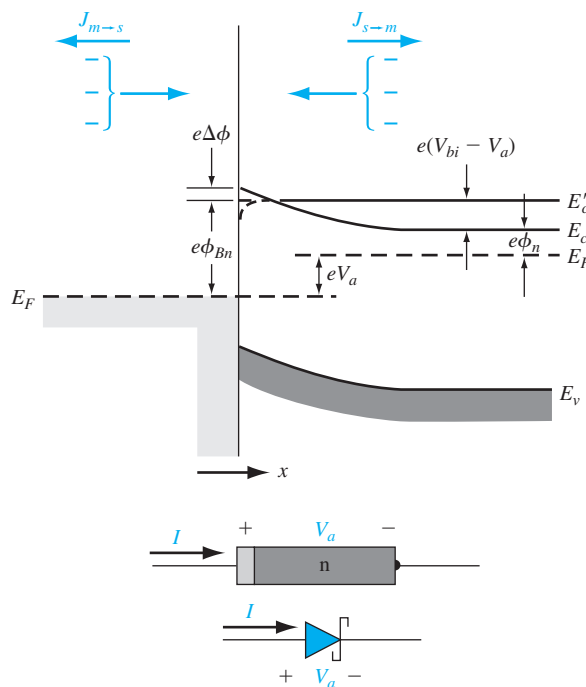

Figure 9.7 | Energy-band diagram of a forward-biased metal–semiconductor junction including the image lowering effect.

The current density $J_{s \to m}$ is a function of the concentration of electrons which have x-directed velocities sufficient to overcome the barrier. We may write

$$J_{s \to m} = e \int_{E_c'}^{\infty} v_x \, dn \tag{9.18}$$

where E_c' is the minimum energy required for thermionic emission into the metal, v_x is the carrier velocity in the direction of transport, and e is the magnitude of the electronic charge. The incremental electron concentration is given by

$$dn = g_c(E) f_F(E) \, dE \tag{9.19}$$

where $g_c(E)$ is the density of states in the conduction band and $f_F(E)$ is the Fermi–Dirac probability function. Assuming that the Maxwell–Boltzmann approximation applies, we may write

$$dn = \frac{4\pi (2m_n^*)^{3/2}}{h^3} \sqrt{E - E_c} \, \exp\left[\frac{-(E - E_F)}{kT}\right] dE \tag{9.20}$$

If all of the electron energy above E_c is assumed to be kinetic energy, then we have

$$\frac{1}{2} m_n^* v^2 = E - E_c \tag{9.21}$$

The net current density in the metal-to-semiconductor junction can be written as

$$J = J_{s \to m} - J_{m \to s} \tag{9.22}$$

which is defined to be positive in the direction from the metal to the semiconductor. We find that

$$J = \left[A^* T^2 \exp\left(\frac{-e\phi_{Bn}}{kT}\right)\right]\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right] \tag{9.23}$$

where

$$A^* \equiv \frac{4\pi e m_n^* k^2}{h^3} \tag{9.24}$$

The parameter A^* is called the effective Richardson constant for thermionic emission. Equation (9.23) can be written in the usual diode form as

$$J = J_{sT}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right] \tag{9.25}$$

where J_{sT} is the reverse-saturation current density and is given by

$$J_{sT} = A^* T^2 \exp\left(\frac{-e\phi_{Bn}}{kT}\right) \tag{9.26}$$

We may recall that the Schottky barrier height ϕ_{Bn} changes because of the image-force lowering. We have that $\phi_{Bn} = \phi_{B0} - \Delta\phi$. Then we can write Equation (9.26) as
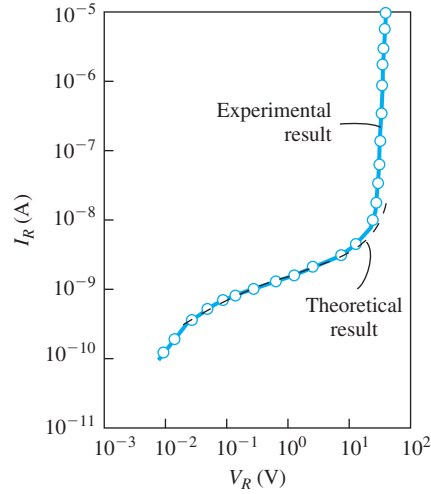
Figure 9.8 | Experimental and theoretical reverse-biased currents in a PtSi–Si diode. (From Sze and Ng [15].)

$$J_{sT} = A^* T^2 \exp\left(\frac{-e\phi_{B0}}{kT}\right) \exp\left(\frac{e\Delta\phi}{kT}\right) \tag{9.27}$$

The change in barrier height, $\Delta\phi$, will increase with an increase in the electric field, or with an increase in the applied reverse-biased voltage. Figure 9.8 shows a typical reverse-biased current–voltage characteristic of a Schottky barrier diode. The reverse-biased current increases with reverse-biased voltage because of the barrier lowering effect. This figure also shows the Schottky barrier diode going into breakdown.

EXAMPLE 9.4

Objective: Determine the effective Richardson constant from the current–voltage characteristics.

Consider the tungsten–silicon diode curve in Figure 9.9 and assume a barrier height of $\phi_{Bn} = 0.67$ V. From the figure, $J_{sT} \approx 6 \times 10^{-5}$ A/cm².

■ **Solution**

We have that

$$J_{sT} = A^* T^2 \exp\left(\frac{-e\phi_{Bn}}{kT}\right)$$

so that

$$A^* = \frac{J_{sT}}{T^2} \exp\left(\frac{+e\phi_{Bn}}{kT}\right)$$

Then

$$A^* = \frac{6 \times 10^{-5}}{(300)^2} \exp\left(\frac{0.67}{0.0259}\right) = 114 \text{ A/K}^2\text{-cm}^2$$

■ **Comment**

The experimentally determined value of A^* is a very strong function of ϕ_{Bn}, since ϕ_{Bn} is in the exponential term. A small change in ϕ_{Bn} will change the value of the Richardson constant substantially.
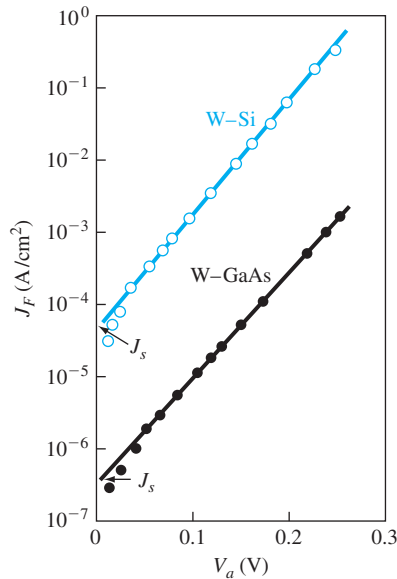
Figure 9.9 | Forward-bias current density J_F versus V_a for Wi–Si and W–GaAs diodes. *(From Sze and Ng [15].)*

■ **EXERCISE PROBLEM**

Ex 9.4   Calculate the ideal Richardson constant for a free electron.

(Ans. $A* = 120$ A/K²-cm²)

We may note that the reverse-saturation current densities of the tungsten–silicon and tungsten–gallium arsenide diodes in Figure 9.9 differ by approximately two orders of magnitude. This two order of magnitude difference will be reflected in the effective Richardson constant, assuming the barrier heights in the two diodes are essentially the same. The definition of the effective Richardson constant, given by Equation (9.24), contains the electron effective mass, which differs substantially between silicon and gallium arsenide. The fact that the effective mass is in the expression for the Richardson constant is a direct result of using the effective density of states function in the thermionic emission theory. The net result is that $A*$ and J_{sT} will vary widely between silicon and gallium arsenide.

9.1.5   Comparison of the Schottky Barrier Diode and the pn Junction Diode

Although the ideal current–voltage relationship of the Schottky barrier diode given by Equation (9.25) is of the same form as that of the pn junction diode, there are two important differences between a Schottky diode and a pn junction diode: The first is in the magnitudes of the reverse-saturation current densities and the second is in the switching characteristics.

The reverse-saturation current density of the Schottky barrier diode was given by Equation (9.26) and is

$$J_{sT} = A^* T^2 \exp\left(\frac{-e\phi_{Bn}}{kT}\right)$$

The ideal reverse-saturation current density of the pn junction diode can be written as

$$J_s = \frac{eD_n n_{po}}{L_n} + \frac{eD_p p_{no}}{L_p} \tag{9.28}$$

The form of the two equations is vastly different, and the current mechanism in the two devices is different. The current in a pn junction is determined by the diffusion of minority carriers while the current in a Schottky barrier diode is determined by thermionic emission of majority carriers over a potential barrier.

EXAMPLE 9.5

Objective: Calculate the ideal reverse-saturation current densities of a Schottky barrier diode and a pn junction diode.

Consider a tungsten barrier on silicon with a measured barrier height of $e\phi_{Bn} = 0.67$ eV. The effective Richardson constant is $A^* = 114$ A/K^2-cm^2. Let $T = 300$ K.

■ **Solution**

If we neglect the barrier lowering effect, we have for the Schottky barrier diode

$$J_{sT} = A^* T^2 \exp\left(\frac{-e\phi_{Bn}}{kT}\right) = (114)(300)^2 \exp\left(\frac{-0.67}{0.0259}\right) = 5.98 \times 10^{-5} \text{ A/cm}^2$$

Consider a silicon pn junction with the following parameters at $T = 300$ K.

$$N_a = 10^{18} \text{ cm}^{-3} \quad N_d = 10^{16} \text{ cm}^{-3}$$
$$D_p = 10 \text{ cm}^2/\text{s} \quad D_n = 25 \text{ cm}^2/\text{s}$$
$$\tau_{po} = 10^{-7} \text{ s} \quad \tau_{no} = 10^{-7} \text{ s}$$

We can then calculate the following parameters:

$$L_p = 1.0 \times 10^{-3} \text{ cm} \quad L_n = 1.58 \times 10^{-3} \text{ cm}$$
$$p_{no} = 2.25 \times 10^4 \text{ cm}^{-3} \quad n_{po} = 2.25 \times 10^2 \text{ cm}^{-3}$$

The ideal reverse-saturation current density of the pn junction diode can be determined from Equation (9.28) as

$$J_s = \frac{(1.6 \times 10^{-19})(25)(2.25 \times 10^2)}{(1.58 \times 10^{-3})} + \frac{(1.6 \times 10^{-19})(10)(2.25 \times 10^4)}{(1.0 \times 10^{-3})}$$
$$= 5.7 \times 10^{-13} + 3.6 \times 10^{-11} = 3.66 \times 10^{-11} \text{ A/cm}^2$$

■ **Comment**

The ideal reverse-saturation current density of the Schottky barrier junction is orders of magnitude larger than that of the ideal pn junction diode.

■ **EXERCISE PROBLEM**

Ex 9.5   Using the results of Example 9.5, determine the forward-bias voltages required to produce a current of 10 μA in each diode. Assume each cross-sectional area is 10^{-4} cm^2.

(Ans. pn junction, $V_a = 0.5628$ V; Schottky junction, $V_a = 0.1922$ V)
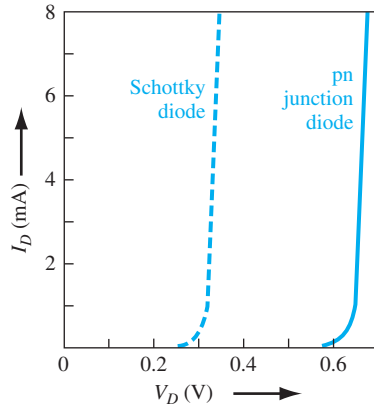
Figure 9.10 | Comparison of forward-bias *I–V* characteristics between a Schottky diode and a pn junction diode.

Recall that the reverse-biased current in a silicon pn junction diode is dominated by the generation current. A typical generation current density is approximately 10^{-7} A/cm², which is still two to three orders of magnitude less than the reverse-saturation current density of the Schottky barrier diode. A generation current also exists in the reverse-biased Schottky barrier diode; however, the generation current is negligible compared with the J_{sT} value.

Since $J_{sT} \gg J_s$, the forward-bias characteristics of the two types of diodes will also be different. Figure 9.10 shows typical *I–V* characteristics of a Schottky barrier diode and a pn junction diode. The effective turn-on voltage of the Schottky diode is less than that of the pn junction diode.

Objective: Calculate the forward-bias voltage required to induce a forward-bias current density of 10 A/cm² in a Schottky barrier diode and a pn junction diode.

EXAMPLE 9.6

Consider diodes with the parameters given in Example 9.5. We can assume that the pn junction diode will be sufficiently forward biased so that the ideal diffusion current will dominate. Let $T = 300$ K.

■ **Solution**

For the Schottky barrier diode, we have

$$J = J_{sT}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]$$

Neglecting the (-1) term, we can solve for the forward-bias voltage. We find

$$V_a = \left(\frac{kT}{e}\right)\ln\left(\frac{J}{J_{sT}}\right) = V_t \ln\left(\frac{J}{J_{sT}}\right) = (0.0259)\ln\left(\frac{10}{5.98 \times 10^{-5}}\right) = 0.312 \text{ V}$$

For the pn junction diode, we have

$$V_a = V_t \ln\left(\frac{J}{J_s}\right) = (0.0259)\ln\left(\frac{10}{3.66 \times 10^{-11}}\right) = 0.682 \text{ V}$$

■ **Comment**

A comparison of the two forward-bias voltages shows that the Schottky barrier diode has a turn-on voltage that, in this case, is approximately 0.37 V smaller than the turn-on voltage of the pn junction diode.

■ **EXERCISE PROBLEM**

Ex 9.6   A pn junction diode and a Schottky diode have equal cross-sectional areas and have forward-bias currents of 0.5 mA. The reverse-saturation current of the Schottky diode is 5×10^{-7} A. The difference in forward-bias voltage between the two diodes is 0.30 V. Determine the reverse-saturation current of the pn junction diode.

(Ans. 4.66×10^{-12} A)

The actual difference between the turn-on voltages will be a function of the barrier height of the metal–semiconductor contact and the doping concentrations in the pn junction, but the relatively large difference will always be realized. We will consider one application that utilizes the difference in turn-on voltage in Chapter 12, in what is referred to as a *Schottky clamped transistor.*

The second major difference between a Schottky barrier diode and a pn junction diode is in the frequency response, or switching characteristics. In our discussion, we have considered the current in a Schottky diode as being due to the injection of majority carriers over a potential barrier. The energy-band diagram of Figure 9.1, for example, shows that there can be electrons in the metal directly adjacent to empty states in the semiconductor. If an electron from the valence band of the semiconductor were to flow into the metal, this effect would be equivalent to holes being injected into the semiconductor. This injection of holes would create excess minority carrier holes in the n region. However, calculations as well as measurements have shown that the ratio of the minority carrier hole current to the total current is extremely low in most cases.

The Schottky barrier diode, then, is a majority carrier device. This fact means that there is no diffusion capacitance associated with a forward-biased Schottky diode. The elimination of the diffusion capacitance makes the Schottky diode a higher-frequency device than the pn junction diode. Also, when switching a Schottky diode from forward to reverse bias, there is no minority carrier stored charge to remove, as is the case in the pn junction diode. Since there is no minority carrier storage time, the Schottky diodes can be used in fast-switching applications. A typical switching time for a Schottky diode is in the picosecond range, while for a pn junction it is normally in the nanosecond range.

TEST YOUR UNDERSTANDING

TYU 9.4   (*a*) The reverse-saturation currents of a pn junction and a Schottky diode are 10^{-14} A and 10^{-9} A, respectively. Determine the required forward-bias voltages in the pn junction diode and Schottky diode to produce a current of 100 μA in each diode. (*b*) Repeat part (*a*) for forward bias currents of 1 mA.

[Ans. (*a*) 0.596 V, 0.298 V; (*b*) 0.656 V, 0.358 V]

9.2 | METAL–SEMICONDUCTOR OHMIC CONTACTS

Contacts must be made between any semiconductor device, or integrated circuit, and the outside world. These contacts are made via *ohmic contacts*. Ohmic contacts are metal-to-semiconductor contacts, but in this case they are not rectifying contacts. An ohmic contact is a low-resistance junction providing conduction in both directions between the metal and the semiconductor. Ideally, the current through the ohmic contact is a linear function of applied voltage, and the applied voltage should be very small. Two general types of ohmic contacts are possible: The first type is the ideal nonrectifying barrier, and the second is the tunneling barrier. We define in this section a specific contact resistance that is used to characterize ohmic contacts.

9.2.1 Ideal Nonrectifying Barrier

We have considered an ideal metal-n-type semiconductor contact in Figure 9.1 for the case when $\phi_m > \phi_s$. Figure 9.11 shows the same ideal contact for the opposite case of $\phi_m < \phi_s$. In Figure 9.11a we see the energy levels before contact, and in Figure 9.11b, the barrier after contact for thermal equilibrium. To achieve thermal equilibrium in this junction, electrons flow from the metal into the lower energy states in the semiconductor, which makes the surface of the semiconductor more n type. The excess electron charge in the n-type semiconductor exists essentially as a surface charge density. If a positive voltage is applied to the metal, there is no barrier to electrons flowing from the semiconductor into the metal. If a positive voltage is applied to the semiconductor, the effective barrier height for electrons flowing from the metal into the semiconductor will be approximately $\phi_{Bn} = \phi_n$, which is fairly small for a moderately to heavily doped semiconductor. For this bias condition, electrons can easily flow from the metal into the semiconductor.

Figure 9.12a shows the energy-band diagram when a positive voltage is applied to the metal with respect to the semiconductor. Electrons can easily flow "downhill"
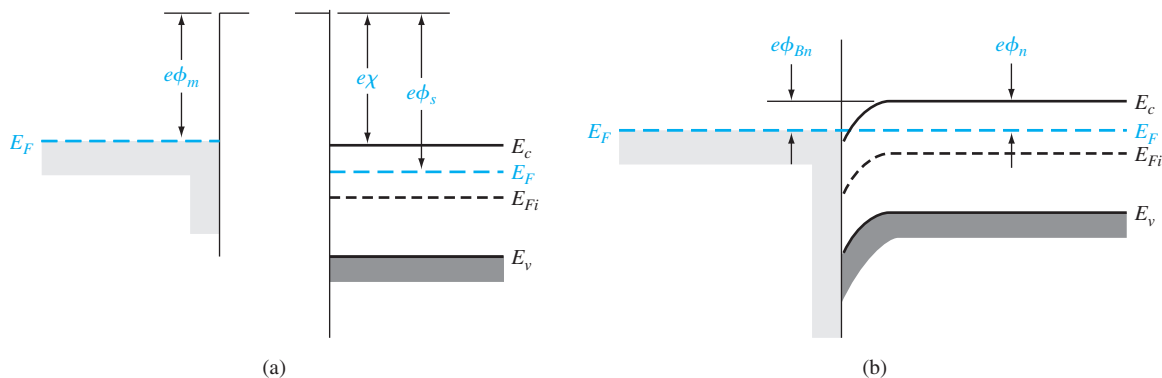

(a)   (b)

Figure 9.11 | Ideal energy-band diagram (a) before contact and (b) after contact for a metal-n-type semiconductor junction for $\phi_m < \phi_s$.
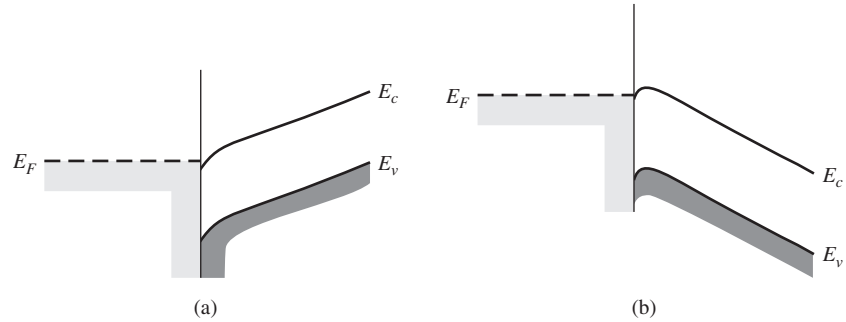
Figure 9.12 | Ideal energy-band diagram of a metal-n-type semiconductor ohmic contact (a) with a positive voltage applied to the metal and (b) with a positive voltage applied to the semiconductor.
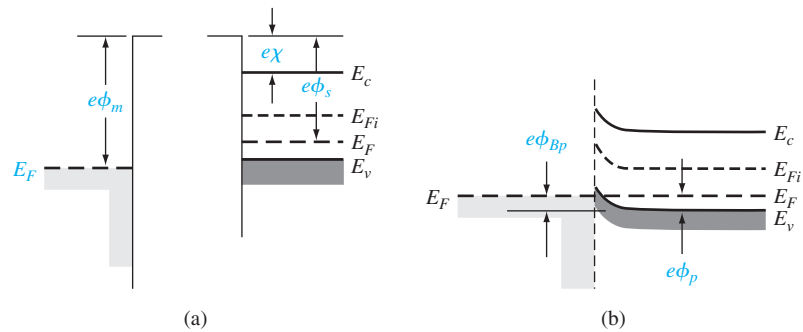

Figure 9.13 | Ideal energy-band diagram (a) before contact and (b) after contact for a metal–p-type semiconductor junction for $\phi_m < \phi_s$.

from the semiconductor into the metal. Figure 9.12b shows the case when a positive voltage is applied to the semiconductor with respect to the metal. Electrons can easily flow over the barrier from the metal into the semiconductor. This junction, then, is an ohmic contact.

Figure 9.13 shows an ideal nonrectifying contact between a metal and a p-type semiconductor. Figure 9.13a shows the energy levels before contact for the case when $\phi_m > \phi_s$. When contact is made, electrons from the semiconductor flow into the metal to achieve thermal equilibrium, leaving behind more empty states, or holes. The excess concentration of holes at the surface makes the surface of the semiconductor more p type. Electrons from the metal can readily move into the empty states in the semiconductor. This charge movement corresponds to holes flowing from the semiconductor into the metal. We can also visualize holes in the metal flowing into the semiconductor. This junction is also an ohmic contact.

The ideal energy bands shown in Figures 9.11 and 9.13 do not take into account the effect of surface states. If we assume that acceptor surface states exist in
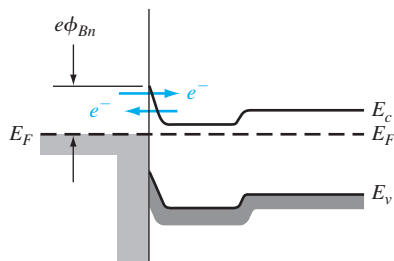
Figure 9.14 | Energy-band diagram of a heavily doped n-semiconductor-to-metal junction.

the upper half of the semiconductor bandgap, then, since all the acceptor states are below E_F for the case shown in Figure 9.11b, these surface states will be negatively charged and will alter the energy-band diagram. Similarly, if we assume that donor surface states exist in the lower half of the bandgap, then all of the donor states will be positively charged for the case shown in Figure 9.13b; the positively charged surface states will also alter this energy-band diagram. Therefore, if $\phi_m < \phi_s$ for the metal-n-type semiconductor contact, and if $\phi_m > \phi_s$ for the metal–p-type semiconductor contact, we may not necessarily form a good ohmic contact.

9.2.2 Tunneling Barrier

The space charge width in a rectifying metal–semiconductor contact is inversely proportional to the square root of the semiconductor doping. The width of the depletion region decreases as the doping concentration in the semiconductor increases; thus, as the doping concentration increases, the probability of tunneling through the barrier increases. Figure 9.14 shows a junction in which the metal is in contact with a heavily doped n-type epitaxial layer.

Objective: Calculate the space charge width for a Schottky barrier on a heavily doped semiconductor.

| **EXAMPLE 9.7**

Consider silicon at $T = 300$ K doped at $N_d = 7 \times 10^{18}$ cm^{-3}. Assume a Shottky barrier with $\phi_{Bn} = 0.67$ V. For this case, we can assume that $V_{bi} \approx \phi_{B0}$. Neglect the barrier lowering effect.

■ **Solution**
From Equation (9.7), we have for zero applied bias

$$x_n = \left[\frac{2\epsilon_s V_{bi}}{eN_d} \right]^{1/2} = \left[\frac{2(11.7)(8.85 \times 10^{-14})(0.67)}{(1.6 \times 10^{-19})(7 \times 10^{18})} \right]^{1/2}$$

or

$$x_n = 1.\,1 \times 10^{-6} \text{ cm} = 110 \text{ Å}$$

■ **Comment**

In a heavily doped semiconductor, the depletion width is on the order of angstroms, so that tunneling is now a distinct possibility. For these types of barrier widths, tunneling may become the dominant current mechanism.

■ **EXERCISE PROBLEM**

Ex 9.7   Calculate the space charge width of a rectifying metal–GaAs–semiconductor junction. Assume the n-type doping concentration is $N_d = 7 \times 10^{18}$ cm^{-3} and the built-in potential barrier is $V_{bi} = 0.80$ V.

(Ans. $x_n = 128.7$ Å)

The tunneling current has the form

$$J_t \propto \exp\left(\frac{-e\phi_{Bn}}{E_{oo}}\right) \tag{9.29}$$

where

$$E_{oo} = \frac{e\hbar}{2}\sqrt{\frac{N_d}{\epsilon_s m_n^*}} \tag{9.30}$$

The tunneling current increases exponentially with doping concentration.

9.2.3   Specific Contact Resistance

A figure of merit of ohmic contacts is the specific contact resistance, R_c. This parameter is defined as the reciprocal of the derivative of current density with respect to voltage evaluated at zero bias. We may write

$$R_c = \left(\frac{\partial J}{\partial V}\right)^{-1}\Big|_{V=0} \qquad \Omega\text{-cm}^2 \tag{9.31}$$

We want R_c to be as small as possible for an ohmic contact.

For a rectifying contact with a low to moderate semiconductor doping concentration, the current–voltage relation is given by Equation (9.23) as

$$J_n = A^* T^2 \exp\left(\frac{-e\phi_{Bn}}{kT}\right)\left[\exp\left(\frac{eV}{kT}\right) - 1\right]$$

The thermionic emission current is dominant in this junction. The specific contact resistance for this case is then

$$R_c = \frac{\left(\frac{kT}{e}\right)\exp\left(\frac{+e\phi_{Bn}}{kT}\right)}{A^* T^2} \tag{9.32}$$

The specific contact resistance decreases rapidly as the barrier height decreases.

For a metal–semiconductor junction with a high impurity doping concentration, the tunneling process will dominate. From Equations (9.29) and (9.30), the specific contact resistance is found to be

$$R_c \propto \exp\left(\frac{+2\sqrt{\epsilon_s m_n^*}}{\hbar} \cdot \frac{\phi_{Bn}}{\sqrt{N_d}}\right) \tag{9.33}$$
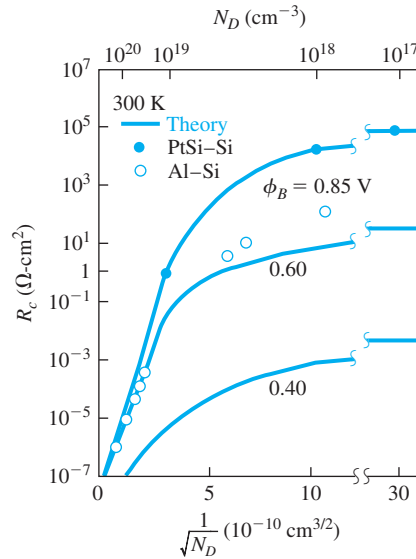
Figure 9.15 | Theoretical and experimental specific contact resistance as a function of doping.
(From Sze and Ng [15].)

which shows that the specific contact resistance is a very strong function of semiconductor doping.

Figure 9.15 shows a plot of the theoretical values of R_c as a function of semiconductor doping. For doping concentrations greater than approximately 10^{19} cm^{-3}, the tunneling process dominates and R_c shows the exponential dependence on N_d. For lower doping concentrations, the R_c values are dependent on the barrier heights and become almost independent of the doping. Also shown in the figure are experimental data for platinum silicide–silicon and aluminum–silicon junctions.

Equation (9.33) is the specific contact resistance of the tunneling junction, which corresponds to the metal–n^+ contact shown in Figure 9.14. However, the n^+n junction also has a specific contact resistance, since there is a barrier associated with this junction. For a fairly low doped n region, this contact resistance may actually dominate the total resistance of the junction.

The theory of forming ohmic contacts is straightforward. To form a good ohmic contact, we need to create a low barrier and use a highly doped semiconductor at the surface. However, the actual technology of fabricating good, reliable ohmic contacts is not as easy in practice as in theory. It is also more difficult to fabricate good ohmic contacts on wide-bandgap materials. In general, low barriers are not possible on these materials, so a heavily doped semiconductor at the surface must be used to form a tunneling contact. The formation of a tunneling junction requires diffusion, ion implantation, or perhaps epitaxial growth. The surface doping concentration in

the semiconductor may be limited to the impurity solubility, which is approximately 5×10^{19} cm^{-3} for n-type GaAs. Nonuniformities in the surface doping concentration may also prevent the theoretical limit of the specific contact resistance from being reached. In practice, a good deal of empirical processing is usually required before a good ohmic contact is obtained.

9.3 | HETEROJUNCTIONS

In the discussion of pn junctions in the previous chapters, we assumed that the semiconductor material is homogeneous throughout the structure. This type of junction is called a *homojunction.* When two different semiconductor materials are used to form a junction, the junction is called a *semiconductor heterojunction.*

As with many topics in this text, our goal is to provide the basic concepts concerning the heterojunction. The complete analysis of heterojunction structures involves quantum mechanics and detailed calculations that are beyond the scope of this text. The discussion of heterojunctions will, then, be limited to the introduction of some basic concepts.

9.3.1 Heterojunction Materials

Since the two materials used to form a heterojunction will have different energy bandgaps, the energy band will have a discontinuity at the junction interface. We may have an abrupt junction in which the semiconductor changes abruptly from a narrow-bandgap material to a wide-bandgap material. On the other hand, if we have a GaAs–Al$_x$Ga$_{1-x}$As system, for example, the value of x may continuously vary over a distance of several nanometers to form a graded heterojunction. Changing the value of x in the Al$_x$Ga$_{1-x}$As system allows us to engineer, or design, the bandgap energy.

In order to have a useful heterojunction, the lattice constants of the two materials must be well matched. The lattice match is important because any lattice mismatch can introduce dislocations resulting in interface states. For example, germanium and gallium arsenide have lattice constants matched to within approximately 0.13 percent. Germanium–gallium arsenide heterojunctions have been studied quite extensively. More recently, gallium arsenide–aluminum gallium arsenide (GaAs–AlGaAs) junctions have been investigated quite thoroughly, since the lattice constants of GaAs and the AlGaAs system vary by no more than 0.14 percent.

9.3.2 Energy-Band Diagrams

In the formation of a heterojunction with a narrow-bandgap material and a wide-bandgap material, the alignment of the bandgap energies is important in determining the characteristics of the junction. Figure 9.16 shows three possible situations. In Figure 9.16a, we see the case when the forbidden bandgap of the wide-gap material completely overlaps the bandgap of the narrow-gap material. This case, called *straddling,* applies to most heterojunctions. We consider only this case here. The other possibilities are called *staggered* and *broken gap* and are shown in Figure 9.16b,c.
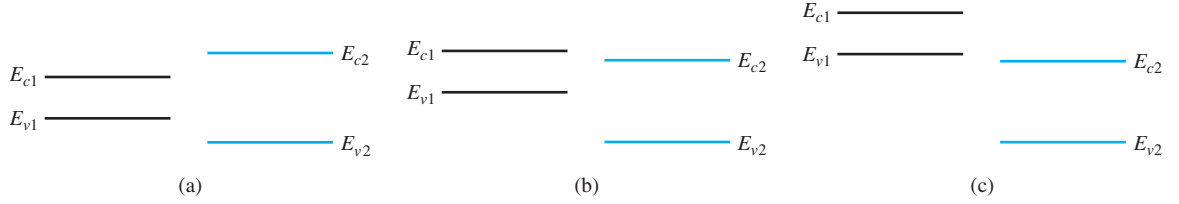
Figure 9.16 | Relation between narrow-bandgap and wide-bandgap energies: (a) straddling, (b) staggered, and (c) broken gap.

There are four basic types of heterojunction. Those in which the dopant type changes at the junction are called *anisotype*. We can form nP or Np junctions, where the capital letter indicates the larger-bandgap material. Heterojunctions with the same dopant type on either side of the junction are called *isotype*. We can form nN and pP isotype heterojunctions.

Figure 9.17 shows the energy-band diagrams of isolated n-type and P-type materials, with the vacuum level used as a reference. The electron affinity of the wide-bandgap material is less than that of the narrow-bandgap material. The difference between the two conduction band energies is denoted by ΔE_c, and the difference between the two valence band energies is denoted by ΔE_v. From Figure 9.17, we can see that

$$\Delta E_c = e(\chi_n - \chi_P) \tag{9.34a}$$

and

$$\Delta E_c + \Delta E_v = E_{gP} - E_{gn} = \Delta E_g \tag{9.34b}$$

In the ideal abrupt heterojunction using nondegenerately doped semiconductors, the vacuum level is parallel to both conduction bands and valence bands. If the vacuum level is continuous, then the same ΔE_c and ΔE_v discontinuities will exist at the
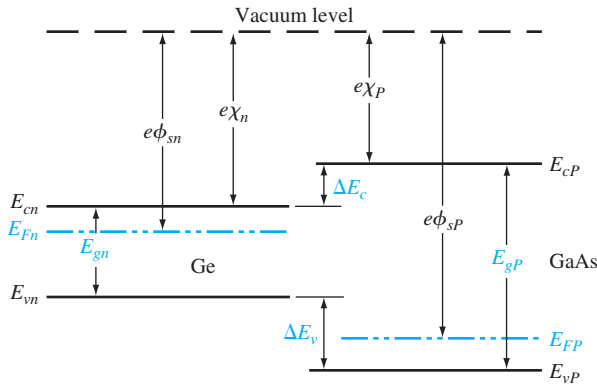

Figure 9.17 | Energy-band diagrams of a narrow-bandgap and a wide-bandgap material before contact.
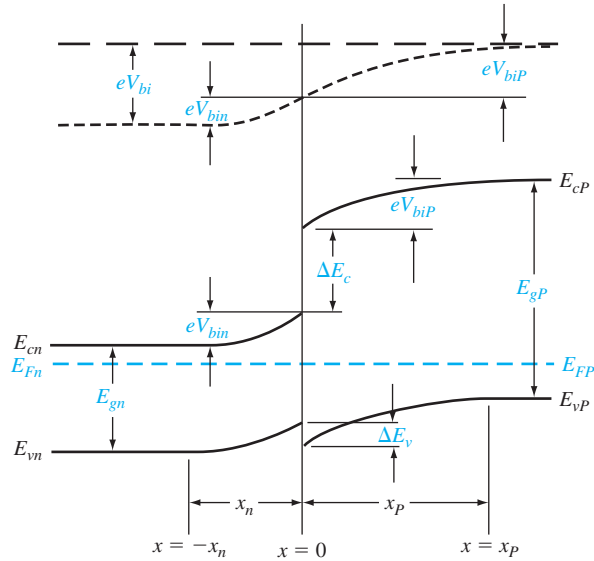
Figure 9.18 | Ideal energy-band diagram of an nP
heterojunction in thermal equilibrium.

heterojunction interface. This ideal situation is known as the *electron affinity rule.*
There is still some uncertainty about the applicability of this rule, but it provides a
good starting point for the discussion of heterojunctions.

Figure 9.18 shows a general ideal nP heterojunction in thermal equilibrium. In
order for the Fermi levels in the two materials to become aligned, electrons from
the narrow-gap n region and holes from the wide-gap P region must flow across the
junction. As in the case of a homojunction, this flow of charge creates a space charge
region in the vicinity of the metallurgical junction. The space charge width into the
n-type region is denoted by x_n and the space charge width into the P-type region is de-
noted by x_P. The discontinuities in the conduction and valence bands and the change
in the vacuum level are shown in the figure.

9.3.3   Two-Dimensional Electron Gas

Before we consider the electrostatics of the heterojunction, we will discuss a unique
characteristic of an isotype junction. Figure 9.19 shows the energy-band diagram
of an nN GaAs–AlGaAs heterojunction in thermal equilibrium. The AlGaAs can
be moderately to heavily doped n type, while the GaAs can be more lightly doped
or even intrinsic. As mentioned previously, to achieve thermal equilibrium, elec-
trons from the wide-bandgap AlGaAs flow into the GaAs, forming an accumulation
layer of electrons in the potential well adjacent to the interface. One basic quantum-
mechanical result that we have found previously is that the energy of an electron
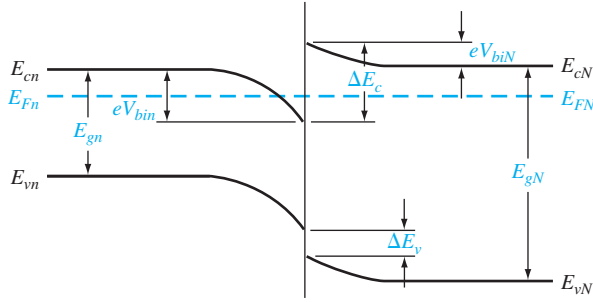contained in a potential well is quantized. The phrase *two-dimensional electron gas*

Figure 9.19 | Ideal energy-band diagram of an nN heterojunction in thermal equilibrium.

refers to the condition in which the electrons have quantized energy levels in one spatial direction (perpendicular to the interface), but are free to move in the other two spatial directions.

The potential function near the interface can be approximated by a triangular potential well. Figure 9.20a shows the conduction band edges near the abrupt junction interface and Figure 9.20b shows the approximation of the triangular potential well. We can write

$$V(x) = eEz \qquad z > 0 \qquad\qquad (9.35a)$$

$$V(z) = \infty \qquad z < 0 \qquad\qquad (9.35b)$$

Schrodinger's wave equation can be solved using this potential function. The quantized energy levels are shown in Figure 9.20b. Higher energy levels are usually not considered.
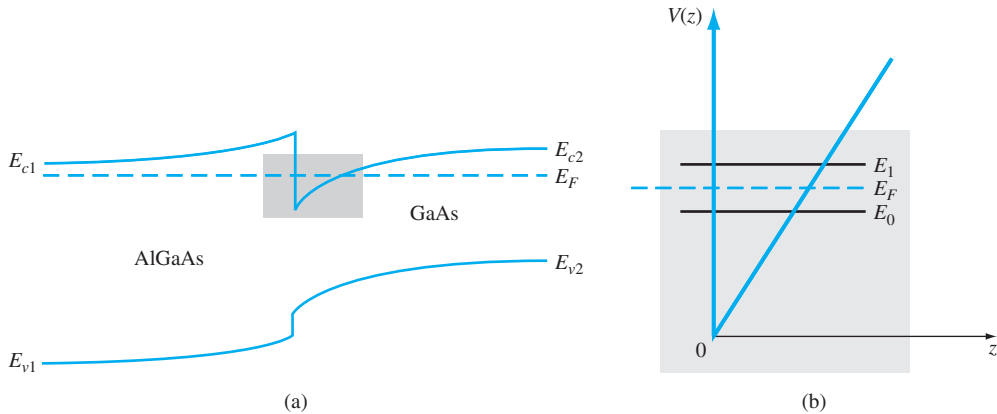

(a)   (b)

Figure 9.20 | (a) Conduction-band edge at N-AlGaAs, n-GaAs heterojunction; (b) triangular well approximation with discrete electron energies.

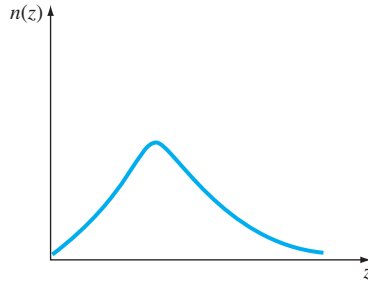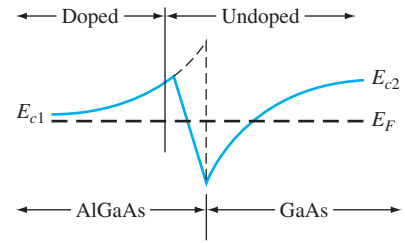**Figure 9.21** | Electron density in triangular potential well.

Figure 9.22 | Conduction-band edge at a graded heterojunction.

The qualitative distribution of electrons in the potential well is shown in Figure 9.21. A current parallel to the interface will be a function of this electron concentration and of the electron mobility. Since the GaAs can be lightly doped or intrinsic, the two-dimensional electron gas is in a region of low impurity doping so that impurity scattering effects are minimized. The electron mobility will be much larger than if the electrons were in the same region as the ionized donors.

The movement of the electrons parallel to the interface will still be influenced by the coulomb attraction of the ionized impurities in the AlGaAs. The effect of these forces can be further reduced by using a graded AlGaAs–GaAs heterojunction. The graded layer is $Al_xGa_{1-x}As$ in which the mole fraction x varies with distance. In this case, an intrinsic layer of graded AlGaAs can be sandwiched between the N-type AlGaAs and the intrinsic GaAs. Figure 9.22 shows the conduction-band edges across a graded AlGaAs–GaAs heterojunction in thermal equilibrium. The electrons in the potential well are further separated from the ionized impurities so that the electron mobility is increased above that in an abrupt heterojunction.

*9.3.4   Equilibrium Electrostatics

We now consider the electrostatics of the nP heterojunction that is shown in Figure 9.18. As in the case of the homojunction, potential differences exist across the space charge regions in both the n region and the P region. These potential differences correspond to the built-in potential barriers on either side of the junction. The built-in potential barrier for this ideal case is defined as shown in Figure 9.18 to be the potential difference across the vacuum level. The built-in potential barrier is the sum of the potential differences across each of the space charge regions. The heterojunction built-in potential barrier, however, is not equal to the difference between the conduction bands across the junction or the difference between the valence bands across the junction, as we defined for the homojunction.

Ideally, the total built-in potential barrier V_{bi} can be found as the difference between the work functions, or

$$V_{bi} = \phi_{sP} - \phi_{sn} \tag{9.36}$$

Equation (9.36), from Figure 9.17, can be written as

$$eV_{bi} = [e\chi_P + E_{gP} - (E_{FP} - E_{vP})] - [e\chi_n + E_{gn} - (E_{Fn} - E_{vn})] \quad (9.37a)$$

or

$$eV_{bi} = e(\chi_P - \chi_n) + (E_{gP} - E_{gn}) + (E_{Fn} - E_{vn}) - (E_{FP} - E_{vP}) \quad (9.37b)$$

which can be expressed as

$$eV_{bi} = -\Delta E_c + \Delta E_g + kT \ln\left(\frac{N_{vn}}{p_{no}}\right) - kT \ln\left(\frac{N_{vP}}{p_{po}}\right) \quad (9.38)$$

Finally, we can write Equation (9.38) as

$$eV_{bi} = \Delta E_v + kT \ln\left(\frac{p_{po}}{p_{no}} \cdot \frac{N_{vn}}{N_{vP}}\right) \quad (9.39)$$

where p_{po} and p_{no} are the hole concentrations in the P and n materials, respectively, and N_{vn} and N_{vP} are the effective density of states functions in the n and P materials, respectively. We can also obtain an expression for the built-in potential barrier in terms of the conduction band shift as

$$eV_{bi} = -\Delta E_c + kT \ln\left(\frac{n_{no}}{n_{po}} \cdot \frac{N_{cP}}{N_{cn}}\right) \quad (9.40)$$

Objective: Determine ΔE_c, ΔE_v, and V_{bi} for an n–Ge to P–GaAs heterojunction using the electron affinity rule. **EXAMPLE 9.8**

Consider n-type Ge doped with $N_d = 10^{16}$ cm^{-3} and P-type GaAs doped with $N_a = 10^{16}$ cm^{-3}. Let $T = 300$ K so that $n_i = 2.4 \times 10^{13}$ cm^{-3} for Ge.

■ **Solution**
From Equation (9.34a), we have

$$\Delta E_c = e(\chi_n - \chi_P) = e(4.13 - 4.07) = 0.06 \text{ eV}$$

and from Equation (9.34b), we have

$$\Delta E_v = \Delta E_g - \Delta E_c = (1.43 - 0.67) - 0.06 = 0.70 \text{ eV}$$

To determine V_{bi} using Equation (9.39), we need to determine p_{no} in Ge, or

$$p_{no} = \frac{n_i^2}{N_d} = \frac{(2.4 \times 10^{13})^2}{10^{16}} = 5.76 \times 10^{10} \text{ cm}^{-3}$$

Then

$$eV_{bi} = 0.70 + (0.0259) \ln\left[\frac{(10^{16})(6 \times 10^{18})}{(5.76 \times 10^{10})(7 \times 10^{18})}\right]$$

or, finally,

$$V_{bi} \approx 1.0 \text{ V}$$

■ **Comment**
There is a nonsymmetry in the ΔE_c and ΔE_v values that will tend to make the potential barriers seen by electrons and holes different. This nonsymmetry does not occur in homojunctions.

■ **EXERCISE PROBLEM**

Ex 9.8   Repeat Example 9.8 for an n-Ge-to-P-GaAs heterojunction. The Ge is doped with $N_d = 10^{15}$ cm^{-3} donors and the GaAs doped with $N_a = 10^{15}$ cm^{-3} acceptors. Let $T = 300$ K. (Ans. $V_{bi} = 0.889$ V)

We can determine the electric field and potential in the junction from Poisson's equation in exactly the same way as we do for the homojunction. For homogeneous doping on each side of the junction, we have in the n region

$$E_n = \frac{eN_{dn}}{\epsilon_n}(x_n + x) \qquad (-x_n \leq x < 0) \tag{9.41a}$$

and in the P region

$$E_P = \frac{eN_{aP}}{\epsilon_P}(x_P - x) \qquad (0 < x \leq x_P) \tag{9.41b}$$

where ϵ_n and ϵ_P are the permittivities of the n and P materials, respectively. We may note that $E_n = 0$ at $x = -x_n$ and $E_P = 0$ at $x = x_P$. The electric flux density D is continuous across the junction, so

$$\epsilon_n E_n(x = 0) = \epsilon_P E_P(x = 0) \tag{9.42a}$$

which gives

$$N_{dn}x_n = N_{aP}x_P \tag{9.42b}$$

Equation (9.42b) simply states that the net negative charge in the P region is equal to the net positive charge in the n region—the same condition we had in a pn homojunction. We are neglecting any interface states that may exist at the heterojunction.

The electric potential can be found by integrating the electric field through the space charge region so that the potential difference across each region can be determined. We find that

$$V_{bin} = \frac{eN_{dn}x_n^2}{2\epsilon_n} \tag{9.43a}$$

and

$$V_{biP} = \frac{eN_{aP}x_P^2}{2\epsilon_P} \tag{9.43b}$$

Equation (9.42b) can be rewritten as

$$\frac{x_n}{x_P} = \frac{N_{aP}}{N_{dn}} \tag{9.44}$$

The ratio of the built-in potential barriers can then be determined as

$$\frac{V_{bin}}{V_{biP}} = \frac{\epsilon_P}{\epsilon_n} \cdot \frac{N_{dn}}{N_{aP}} \cdot \frac{x_n^2}{x_P^2} = \frac{\epsilon_P N_{aP}}{\epsilon_n N_{dn}} \tag{9.45}$$

Assuming that ϵ_n and ϵ_P are of the same order of magnitude, the larger potential difference is across the lower-doped region.

The total built-in potential barrier is

$$V_{bi} = V_{bin} + V_{biP} = \frac{eN_{dn}x_n^2}{2\epsilon_n} + \frac{eN_{aP}x_P^2}{2\epsilon_P} \tag{9.46}$$

If we solve for x_P, for example, from Equation (9.42b) and substitute into Equation (9.46), we can solve for x_n as

$$x_n = \left[\frac{2\epsilon_n \epsilon_P N_{aP} V_{bi}}{eN_{dn}(\epsilon_n N_{dn} + \epsilon_P N_{aP})} \right]^{1/2} \tag{9.47a}$$

We can also find

$$x_P = \left[\frac{2\epsilon_n \epsilon_P N_{dn} V_{bi}}{eN_{aP}(\epsilon_n N_{dn} + \epsilon_P N_{aP})} \right]^{1/2} \tag{9.47b}$$

The total depletion width is found to be

$$W = x_n + x_P = \left[\frac{2\epsilon_n \epsilon_P (N_{dn} + N_{aP})^2 V_{bi}}{eN_{dn}N_{aP}(\epsilon_n N_{dn} + \epsilon_P N_{aP})} \right]^{1/2} \tag{9.48}$$

If a reverse-biased voltage is applied across the heterojunction, the same equations apply if V_{bi} is replaced by $V_{bi} + V_R$. Similarly, if a forward bias is applied, the same equations also apply if V_{bi} is replaced by $V_{bi} - V_a$. As explained earlier, V_R is the magnitude of the reverse-biased voltage and V_a is the magnitude of the forward-bias voltage.

As in the case of a homojunction, a change in depletion width with a change in junction voltage yields a junction capacitance. We can find for the nP junction

$$C_j' = \left[\frac{eN_{dn}N_{aP}\epsilon_n\epsilon_P}{2(\epsilon_n N_{dn} + \epsilon_P N_{aP})(V_{bi} + V_R)} \right]^{1/2} \quad (\text{F/cm}^2) \tag{9.49}$$

A plot of $(1/C_j')^2$ versus V_R again yields a straight line. The extrapolation of this plot of $(1/C_j')^2 = 0$ is used to find the built-in potential barrier, V_{bi}.

Figure 9.18 shows the ideal energy-band diagram for the nP abrupt heterojunction. The experimentally determined values of ΔE_c and ΔE_v may differ from the ideal values determined using the electron affinity rule. One possible explanation for this difference is that most heterojunctions have interface states. If we assume that the electrostatic potential is continuous through the junction, then the electric flux density will be discontinuous at the heterojunction due to the surface charge trapped in the interface states. The interface states will then change the energy-band diagram of the semiconductor heterojunction just as they changed the energy-band diagram of the metal–semiconductor junction. Another possible explanation for the deviation from the ideal is that as the two materials are brought together to form the heterojunction, the electron orbitals of each material begin to interact with each other, resulting in a transition region of a few angstroms at the interface. The energy bandgap is then continuous through this transition region and not a characteristic of either material. However, we still have the relation

$$\Delta E_c + \Delta E_v = \Delta E_g \tag{9.50}$$

for the straddling type of heterojunction, although the ΔE_c and ΔE_v values may differ from those determined from the electron affinity rule.
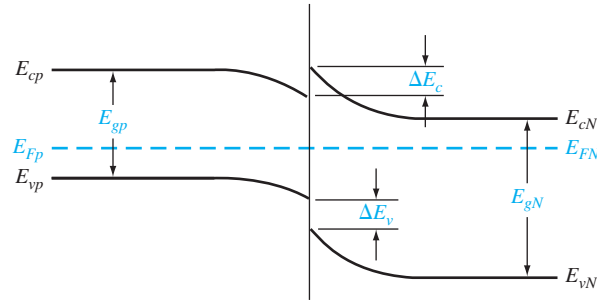
Figure 9.23 | Ideal energy-band diagram of an Np heterojunction in thermal equilibrium.

We may consider the general characteristics of the energy-band diagrams of the other types of heterojunction. Figure 9.23 shows the energy-band diagram of an Np heterojunction. The same ΔE_c and ΔE_v discontinuities exist, although the general shape of the conduction band, for example, is different in the nP and the Np junctions. This difference in energy bands will influence the I–V characteristics of the two junctions.

The other two types of heterojunctions are the nN and the pP isotype junctions. The energy-band diagram of the nN junction is shown in Figure 9.19. To achieve thermal equilibrium, electrons from the wide-bandgap material will flow into the narrow-bandgap material. A positive space charge region exists in the wide-bandgap material and an accumulation layer of electrons now exists at the interface in the narrow-bandgap material. Since there are a large number of allowed energy states in the conduction band, we expect the space charge width x_n and the built-in potential barrier V_{bin} to be small in the narrow-gap material. The energy-band diagram of the pP heterojunction in thermal equilibrium is shown in Figure 9.24. To achieve thermal equilibrium, holes from the wide-bandgap material will flow into the narrow-bandgap material, creating an accumulation layer of holes in the narrow-bandgap material at the interface. These types of isotype heterojunctions are obviously not possible in a homojunction.
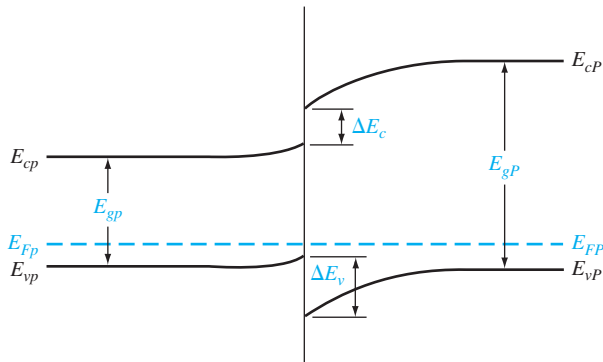

Figure 9.24 | Ideal energy-band diagram of a pP heterojunction in thermal equilibrium.

*9.3.5 Current–Voltage Characteristics

The ideal current–voltage characteristics of a pn homojunction have been developed in Chapter 8. Since the energy-band diagram of a heterojunction is more complicated than that of a homojunction, we would expect the I–V characteristics of the two junctions to differ.

One immediate difference between a homojunction and a heterojunction is in the barrier heights seen by the electrons and holes. Since the built-in potential barrier for electrons and holes in a homojunction is the same, the relative magnitude of the electron and hole currents is determined by the relative doping levels. In a heterojunction, the barrier heights seen by electrons and holes are not the same. The energy-band diagrams in Figures 9.18 and 9.23 demonstrate that the barrier heights for electrons and holes in a heterojunction can be significantly different. The barrier height for electrons in Figure 9.18 is larger than that for holes, so we would expect the current due to electrons to be insignificant compared to the hole current. If the barrier height for electrons is 0.2 eV larger than that for holes, the electron current will be approximately a factor of 10^4 smaller than the hole current, assuming all other parameters are equal. The opposite situation exists for the band diagram shown in Figure 9.23.

The conduction-band edge in Figure 9.23 and the valence-band edge in Figure 9.18 are somewhat similar to that of a rectifying metal–semiconductor contact. We derive the current–voltage characteristics of a heterojunction, in general, on the basis of thermionic emission of carriers over the barrier, as we do in the case of metal–semiconductor junction. We can then write

$$J = A^* T^2 \exp\left(\frac{-E_w}{kT}\right) \tag{9.51}$$

where E_w is an effective barrier height. The barrier height can be increased or reduced by an applied potential across the junction as in the case of a pn homojunction or a Schottky barrier junction. The heterojunction I–V characteristics, however, may need to be modified to include diffusion effects and tunneling effects. Another complicating factor is that the effective mass of a carrier changes from one side of the junction to the other. Although the actual derivation of the I–V relationship of the heterojunction is complex, the general form of the I–V equation is still similar to that of a Schottky barrier diode and is generally dominated by one type of carrier.

9.4 | SUMMARY

■ A metal on a lightly doped semiconductor can produce a rectifying contact that is known as a Schottky barrier diode. The ideal barrier height between the metal and semiconductor is the difference between the metal work function and the semiconductor electron affinity.

■ When a positive voltage is applied to an n-type semiconductor with respect to the metal (reverse bias), the barrier between the semiconductor and metal increases so that there is essentially no flow of charged carriers. When a positive voltage is applied to the metal with respect to an n-type semiconductor (forward bias), the barrier between the semiconductor and metal is lowered so that electrons can easily flow from the semiconductor into the metal by a process called thermionic emission.

■ The ideal current–voltage relationship of the Schottky barrier diode is the same as that of the pn junction diode. However, since the current mechanism is different from that of the pn junction diode, the switching speed of the Schottky diode is faster. In addition, the reverse saturation current of the Schottky diode is larger than that of the pn junction diode, so a Schottky diode requires less forward bias voltage to achieve a given current compared to a pn junction diode.

■ Metal–semiconductor junctions can also form ohmic contacts, which are low-resistance junctions providing conduction in both directions with very little voltage drop across the junction.

■ Semiconductor heterojunctions are formed between two semiconductor materials with different bandgap energies. One useful property of a heterojunction is the creation of a potential well at the interface. Electrons are confined to the potential well in the direction perpendicular to the interface, but are free to move in the other two directions.

GLOSSARY OF IMPORTANT TERMS

anisotype junction   A heterojunction in which the type of dopant changes at the metallurgical junction.

electron affinity rule   The rule stating that, in an ideal heterojunction, the discontinuity at the conduction band is the difference between the electron affinities in the two semiconductors.

heterojunction   The junction formed by the contact between two different semiconductor materials.

image force–induced lowering   The lowering of the peak potential barrier at the metal–semiconductor junction due to an electric field.

isotype junction   A heterojunction in which the type of dopant is the same on both sides of the junction.

ohmic contact   A low-resistance metal–semiconductor contact providing conduction in both directions between the metal and semiconductor.

Richardson constant   The parameter A^* in the current–voltage relation of a Schottky diode.

Schottky barrier height   The potential barrier ϕ_{Bn} from the metal to semiconductor in a metal–semiconductor junction.

Schottky effect   Another term for image force–induced lowering.

specific contact resistance   The inverse of the slope of the J versus V curve of a metal–semiconductor contact evaluated at $V = 0$.

thermionic emission   The process by which charge flows over a potential barrier as a result of carriers with sufficient thermal energy.

tunneling barrier   A thin potential barrier in which the current is dominated by the tunneling of carriers through the barrier.

two-dimensional electron gas (2-DEG)   The accumulation layer of electrons contained in a potential well at a heterojunction interface. The electrons are free to move in the "other" two spatial directions.

CHECKPOINT

After studying this chapter, the reader should have the ability to:

■ Sketch the energy-band diagram of zero-biased, reverse-biased, and forward-biased Schottky barrier diodes.

■ Describe the charge flow in a forward-biased Schottky barrier diode.

- ■ Explain the Schottky barrier lowering and its effect on the reverse saturation current in a Schottky barrier diode.
- ■ Explain the effect of interface states on the characteristics of a Schottky barrier diode.
- ■ Describe one effect of a larger reverse saturation current in a Schottky barrier diode compared to that of a pn junction diode.
- ■ Describe what is meant by an ohmic contact.
- ■ Draw the energy-band diagram of an nN heterojunction.
- ■ Explain what is meant by a two-dimensional electron gas.

REVIEW QUESTIONS

1. What is the ideal Schottky barrier height? Indicate the Schottky barrier height on an energy-band diagram.

2. Using an energy-band diagram, indicate the effect of the Schottky barrier lowering.

3. What is the mechanism of charge flow in a forward-biased Schottky barrier diode?

4. Compare the forward-biased current–voltage characteristic of a Schottky barrier diode to that of pn junction diode.

5. Explain the difference in switching characteristics between a Schottky diode and a pn junction diode. Discuss charge storage effects.

6. Sketch the ideal energy-band diagram of a metal–semiconductor junction in which $\phi_m < \phi_s$. Explain why this is an ohmic contact.

7. Sketch the energy-band diagram of a tunneling junction. Why is this an ohmic contact?

8. What is a heterojunction?

9. What is a 2-D electron gas?

PROBLEMS

(In the following problems, assume $A^* = 120$ A/K^2-cm^2 for silicon and $A^* = 1.12$ A/K^2-cm^2 for gallium arsenide Schottky diodes unless otherwise stated.)

Section 9. 1   The Schottky Barrier Diode

9.1 Consider a contact between Al and n Si doped at $N_d = 10^{16}$ cm^{-3}. $T = 300$ K. (a) Draw the energy-band diagrams of the two materials before the junction is formed. (b) Draw the ideal energy band at zero bias after the junction is formed. (c) Calculate ϕ_{B0}, x_d, and E$_{max}$ for part (b). (d) Repeat parts (b) and (c) using the data in Figure 9.5.

9.2 (a) A Schottky barrier diode formed on n-type silicon has a doping concentration of $N_d = 5 \times 10^{15}$ cm^{-3} and a barrier height of $\phi_{B0} = 0.65$ V. Determine the built-in potential barrier V_{bi}. (b) If the doping concentration changes to $N_d = 10^{16}$ cm^{-3}, determine the values of ϕ_{B0} and V_{bi}. Do these values increase, decrease, or remain the same? (c) Repeat part (b) if the doping concentration is $N_d = 10^{15}$ cm^{-3}.

9.3 Gold is deposited on n-type silicon forming an ideal rectifying junction. The doping concentration is $N_d = 10^{16}$ cm^{-3}. Assume $T = 300$ K. Determine the theoretical values of (a) ϕ_{B0}, (b) V_{bi}, and (c) x_n and $|$E$_{max}|$ at (i) $V_R = 1$ V and (ii) $V_R = 5$ V.

9.4 A Schottky diode is formed by depositing gold on n-type GaAs that is doped at a concentration of $N_d = 5 \times 10^{15}$ cm^{-3}. For $T = 300$ K, determine the theoretical values of (a) ϕ_{B0}, (b) ϕ_n, (c) V_{bi}, and (d) x_n and $|$E$_{max}|$ at (i) $V_R = 1$ V and (ii) $V_R = 5$ V.