particles. It may seem strange, at times, that we concern ourselves so much with minority carriers rather than with the vast number of majority carriers, but the reason for this can he found in the results derived from the ambipolar transport theory.

TEST YOUR UNDERSTANDING

E8.6 Consider the silicon pn junction diode described in E8.4. Calculate the electron and hole currents at (*a*) x = x_n, (*b*) x = $x_n + L_p$, and (*c*) $x = x_n + 10L_p$ (see Figure **8.9**).

[Ans. (*a*) $I_n = 0.154$ mA, $I_p = 1.09$ mA; (*b*) $I_n = 0.843$ mA, $I_p = 0.401$ mA; (*c*) $I_n \approx 0$, $I_p = 1.244$ mA]

The fact that we now have drift current densities in the p and n regions implies that the electric field in these regions is not zero as we had originally assumed. We can calculate the electric field in the neutral regions and determine the validity of our zero-field approximation.

Objective

EXAMPLE 8.4

To calculate the electric field required to produce a given majority carrier drift current.

Consider a silicon pn junction at T = 300 K with the parameters given in Example 8.2 and with an applied forward-hias voltage $V_a = 0.65$ V.

■ Solution

The total forward-hias current density is given by

$$J = J_s\left[\exp\left(\frac{eV}{kT}\right) - 1\right]$$

We determined the reverse saturation current density in Example 8.2, so we can write

$$J = (4.15 \times 10^{-11})\left[\exp\left(\frac{0.65}{0.0259}\right) - 1\right] = 3.29 \text{ A/cm}^2$$

The total current far from the junction in the n-region will he majority carrier electron drift current. so we can write

$$J = J_n \approx e\mu_n N_d E$$

The doping concentration is $N_d = 10^{16}$ cm^{-3}, and, if we assume $\mu_n = 1350$ cm^2/V-s, then the electric field must he

$$E = \frac{J_n}{e\mu_n N_d} = \frac{3.29}{(1.6 \times 10^{-19})(1350)(10^{16})} = 1.52 \text{ V/cm}$$

■ Comment

We assumed. in the derivation of the current–voltage equation, that the electric field in the neutral p and n regions was rero. Although the electric field is not zero, this example shows that the magnitude is very small—thus the approximation of zero electric field is very good.

8.1.7 Temperature Effects

The ideal reverse saturation current density J_s, given by Equation (8.22), is a function of the thermal-equilibrium minority carrier concentrations n_{p0} and p_{n0}. These minority carrier concentrations are proportional to n_i^2, which is a very strong function of temperature. For a silicon pn junction, the ideal reverse saturation current density will increase by approximately a factor of four for every 10°C increase in temperature.

The forward-bias current-voltage relation was given by Equation (8.23). This relation includes J_s as well as the $\exp(eV_a/kT)$ factor, making the forward-bias current–voltage relation a function of temperature also. As temperature increases, less forward-bias voltage is required to obtain the same diode current. If the voltage is held constant, the diode current will increase as temperature increases. The change in forward-bias current with temperature is less sensitive than the reverse saturation current.

EXAMPLE 8.5

Objective

To determine the change in the forward-bias voltage on a pn junction with a change in temperature.

Consider a silicon pn junction initially biased at 0.60 V at T = 300 K. Assume the temperature increases to T = 310 K. Calculate the change in the forward-bias voltage required to maintain a constant current through the junction.

■ **Solution**

The forward-bias current can be written as follows:

$$J \propto \text{enp}\left(\frac{-E_g}{kT}\right)\exp\left(\frac{eV_a}{kT}\right)$$

If the temperature changes, we may take the ratio of the diode currents at the two temperatures. This ratio is

$$\frac{J_2}{J_1} = \frac{\exp(-E_g/kT_2)\exp(eV_{a2}/kT_2)}{\exp(-E_g/kT_1)\exp(eV_{a1}/kT_1)}$$

If current is to be held constant, then $J_1 = J_2$ and we must have

$$\frac{E_g - eV_{a2}}{kT_2} = \frac{E_g - eV_{a1}}{kT_1}$$

Let $T_1 = 300$ K, $T_2 = 310$ K, $E_g = 1.12$ eV, and $V_{a1} = 0.60$ V. Then, solving for V_{a2}, we obtain $V_{a2} = 0.5827$ V.

■ **Comment**

The change in the forward-bias voltage is −17.3 mV for a 10°C temperature change.

8.1.8 The "Short" Diode

We assumed in the previous analysis that both p and n regions were long compared with the minority carrier diffusion lengths. In many pn junction structures, one region

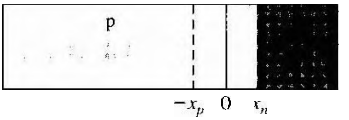


Figure 8.10 | Geometry of a "shon" diode.

may. in fact. he shon compared with the minority carrier diffusion length. Figure 8.10 shows one such example: the length W„ is assumed to be much smaller than the minority carrier hole diffusion length, L,.

The steady-state excess minority carrier hole concentration in then region is determined from Equation (8.9), which was given as

$$\frac{d^2(\delta p_n)}{dx^2} - \frac{\delta p_n}{L_p^2} = 0$$

The original houndary condition at $x = x_n$ still applies, given by Equation (8.11a) as

$$p_n(x_n) = p_{n0} \exp\left(\frac{eV_a}{kT}\right)$$

A second boundary condition needs to be determined. In many cases we will assume that an ohmic contact exists at $x = (x_n + W_n)$, implying an infinite surface-recombination velocity and therefore an excess minority carrier concentration of zero. The second boundary condition is then written as

$$p_n(x = x_n + W_n) = p_{n0} \tag{8.26}$$

The general solution to Equation (8.9) is again given by Equation (8.12). which was

$$\delta p_n(x) = p_n(x) - p_{n0} = Ae^{x/L_p} + Be^{-x/L_p} \qquad (x \geq x_n)$$

In this case, because of the finite length of the n region, both terms of the general solution must be retained. Applying the boundary conditions of Equations (8.11b) and (8.26), the excess minority carrier concentration is given by

$$\delta p_n(x) = p_{n0}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right] \frac{\sinh[(x_n + W_n - x)/L_p]}{\sinh[W_n/L_p]} \tag{8.27}$$

Equation (8.27) is the general solution for the excess minority carrier hole concentration in then region of aforward-biased pnjunction. If $W_n \gg L_p$ the assumption for the long diode, Equation (8.27) reduces to the previous result given by Equation (8.14). If $W_n \ll L_p$, we can approximate the hyperbolic sine terms by

$$\sinh\left(\frac{x_n + W_n - x}{L_p}\right) \approx \left(\frac{x_n + W_n - x}{L_p}\right) \tag{8.28a}$$

and

$$\sinh\left(\frac{W_n}{L_p}\right) \approx \left(\frac{W_n}{L_p}\right) \tag{8.28b}$$

Then Equation (8.27) becomes

$$\delta p_n(x) = p_{n0}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right]\left(\frac{x_n + W_n - x}{W_n}\right) \tag{8.29}$$

The minority carrier concentration becomes a linear function of distance

The minority carrier hole diffusion current density is given by

$$J_p = -eD_p\frac{d(\delta p_n(x))}{dx}$$

so that in the short n region, we have

$$J_p(x) = \frac{eD_p p_{n0}}{W_n}\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right] \tag{8.30}$$

The minority carrier hole diffusion current density now contains the length W_n in the denominator. rather than the diffusion length L_n. The diffusion current density is larger for a short diode than for a long diode since $W_n \ll L_p$. In addition, since the minority carrier concentration is approximately a linear function of distance through the n region, the minority carrier diffusion current density is a constant. This constant current implies that there is no recombination of minority carriers in the short region.

TEST YOUR UNDERSTANDING

E8.7 Consider the silicon pn junction diode described in ER.4. The p region is long and the n region is short with $W_n = 2\ \mu$m. **(a)** Calculate the electron and hole currents in the depletion region. (*b*) Why has the hole current increased compared to that found in E8.4?

[Ans. (*a*) $I_n = 0.154$ mA, $I_p = 5.44$ mA; (*b*) the hole density gradient has increased]

8.2 | SMALL-SIGNAL MODEL OF THE pn JUNCTION

We have been considering the dc characteristics of the pn junction diode. When semiconductor devices with pn junctions are used in linear amplifier circuits, for example, sinusoidal signals are superimposed on the dc currents and voltages, so that the small-signal characteristics of the pn junction become important.

8.2.1 Diffusion Resistance

The ideal current–voltage relationship of the pn junction diode was given by Equation (8.23), where J and J_s are current densities. If we multiply both sides of the

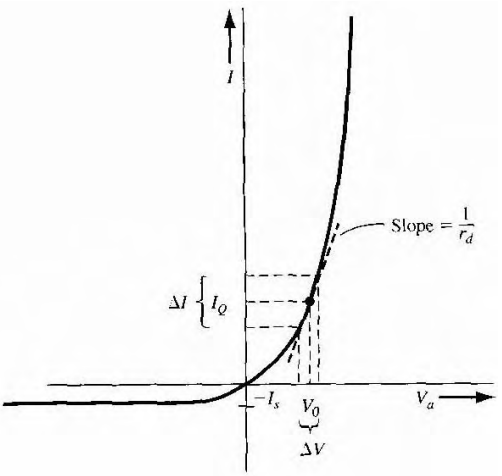


Figure 8.11 | Curve showing the concept of the small-signal diffusion resistance.

equation by the junction cross-sectional area, we have

$$I_D = I_s\left[\exp\left(\frac{eV_a}{kT}\right) - 1\right] \tag{8.31}$$

where I_D is the diode current and I_s is the diode reverse saturation current.

Assume that the diode is forward-biased with a dc voltage V_0 producing a dc diode current I_{DQ}. If we now superimposes small, low-frequency sinusoidal voltage as shown in Figure 8.11, then a small sinusoidal current will be produced, superimposed on the dc current. The ratio of sinusoidal current to sinusoidal voltage is called the incremental conductance. In the limit of a very small sinusoidal current and voltage, the small-signal incremental conductance is just the slope of the dc current-voltage curve, or

$$g_d = \left.\frac{dI_D}{dV_a}\right|_{V_a=V_0} \tag{8.32}$$

The reciprocal of the incremental conductance is the incremental resistance, defined as

$$r_d = \left.\frac{dV_a}{dI_D}\right|_{I_D=I_{DQ}} \tag{8.33}$$

where I_{DQ} is the dc quiescent diode current.

If we assume that the diode is biased sufficiently far in the fonvard-bias region, then the (-1) term can be neglected and the incremental conductance becomes

$$g_d = \left.\frac{dI_D}{dV_a}\right|_{V_a=V_0} = \left(\frac{e}{kT}\right)I_s\exp\left(\frac{eV_0}{kT}\right) \approx \frac{I_{DQ}}{V_t} \tag{8.34}$$

The small-signal incremental resistance is then the reciprocal function, or

$$r_d = \frac{V_t}{I_{DQ}} \tag{8.35}$$

The incremental resistance decreases as the bias current increases, and is inversely proportional to the slope of the I–V characteristic as shown in Figure 8.11. The incremental resistance is also known as the ***diffusion resistance.***

8.2.2 Small-Signal Admittance

In the last chapter, we considered the pn junction capacitance as a function of the reverse-bias voltage. When the pn junction diode is forward-biased, another capacitance becomes a factor in the diode admittance. The small-signal admittance, or impedance, of the pn junction under forward bias is derived using the minority carrier diffusion current relations we have already considered.

Qualitative Analysis Before we delve into the mathematical analysis, we can qualitatively understand the physical processes that lead to a diffusion capacitance, which is one component of the junction admittance. Figure 8.12a schematically shows a pn junction forward biased with a dc voltage. A small ac voltage is also superimposed on the dc voltage so that the total forward-biased voltage can he written as $V_a = V_{dc} + \hat{v} \sin \omega t$.

As the voltage across the junction changes, the number of holes injected across the space charge region into the n region also changes. Figure 8.12b shows the hole concentration at the space charge edge as a function of time. At $t = t_0$, the ac voltage

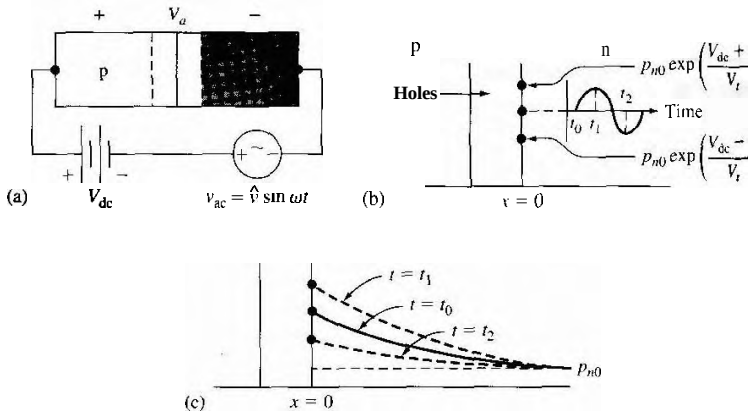


Figure 8.12 | (a) A pn junction with an ac voltage superimposed on a forward-biased dc value; (b) the hole concentration versus time at the space charge edge; (c) the hole concentration versus distance in then region at three different times.

is zero so that the concentration of holes at $x = 0$ is just given by $p_n(0) = p_{n0} \exp(V_{dc}/V_t)$, which is what we have seen previously.

Now, as the ac voltage increases during its positive half cycle, the concentration of holes at x = 0 will increase and reach a peak value at $t = t_1$, which corresponds to the peak value of the ac voltage. When the ac voltage is on its negative half cycle, the total voltage across the junction decreases so that the concentration of holes at $x = 0$ decreases. The concentration reaches a minimum value at $t = t_2$, which corresponds to the time that the ac voltage reaches its maximum negative value. The minority carrier hole concentration at $x = 0$, then, has an ac component superimposed on the dc value as indicated in Figure 8.12b.

As previously discussed, the holes at the space charge edge ($x = 0$) diffuse into the n region where they recombine with the majority carrier electrons. We will assume that the period of the ac voltage is large compared to the time it takes carriers to diffuse into the n region. The hole concentration as a function of distance into the n region can then be treated as a steady-state distribution. Figure 8.12c shows the steady-state hole concentrations at three different times. At $t = t_0$, the ac voltage is zero, so the $t = t_0$ curve corresponds to the hole distribution established by the dc voltage. The $t = t_1$ curve corresponds to the distribution established when the ac voltage has reached its peak positive value, and the $t = t_2$ curve corresponds to the distribution established when the ac voltage has reached its maximum negative value. The shaded areas represents the charge ΔQ that is alternately charged and discharged during the ac voltage cycle.

Exactly the same process is occurring in the p region with the electron concentration. The mechanism of charging and discharging of holes in the n region and electrons in the p region leads to a capacitance. This capacitance is called *diffusion capacitance*. The physical mechanism of this diffusion capacitance is different from that of the junction capacitance discussed in the last chapter. We will show that the magnitude of the diffusion capacitance in a forward-biased pn junction is usually substantially larger than the junction capacitance.

Mathematical Analysis The minority carrier distribution in the pn junction will be derived for the case when a small sinusoidal voltage is superimposed on the dc junction voltage. We can then determine small signal, or ac, diffusion currents from these minority carrier functions. Figure 8.13 shows the minority carrier distribution in a pn junction when a forward-biased dc voltage is applied. The origin, $x = 0$, is set at the edge of the space charge region on the n-side for convenience. The minority carrier hole concentration at $x = 0$ is given by Equation (8.7) as $p_n(0) = p_{n0} \exp(eV_a/kT)$, where V_a is the applied voltage across the junction.

Now let

$$V_a = V_0 + v_1(t) \tag{8.36}$$

where V_0 is the dc quiescent bias voltage and $v_1(t)$ is the ac signal voltage which is superimposed on this dc level. We may now write

$$p_n(x = 0) = p_{n0} \exp\left[\frac{e(V_0 + v_1(t))}{kT}\right] = p_n(0, t) \tag{8.37}$$

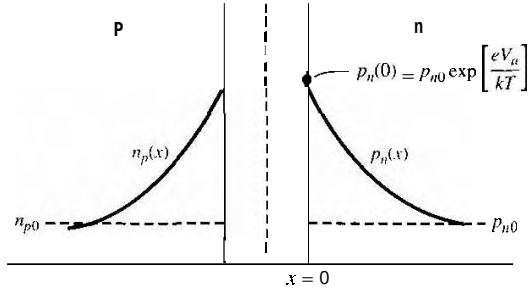


Figure 8.13 | The dc characteristics of a forward-biased pn junction used in the small-signal admittance calculations.

Equation (8.37) may he written as

$$p_n(0, t) = p_{\text{dc}} \exp\left(\frac{ev_1(t)}{kT}\right) \tag{8.38}$$

where

$$p_{\text{dc}} = p_{n0} \exp\left(\frac{eV_0}{kT}\right) \tag{8.39}$$

If we assume that $|v_1(t)| \ll (kT/e) = V_t$, then the exponential term in Equation (8.38) may be expanded into a Taylor series retaining only the linear terms, the minority carrier hole concentration at $x = 0$ can he written as

$$p_n(0, t) \approx p_{\text{dc}}\left(1 + \frac{v_1(t)}{V_t}\right) \tag{8.40}$$

If we assume that the time-varying voltage $v_1(t)$ is a sinusoidal signal, we can write Equation (8.40) as

$$p_n(0, t) = p_{\text{dc}}\left(1 + \frac{\hat{V}_1}{V_t} e^{j\omega t}\right) \tag{8.41}$$

where $\hat{V}_1$ is the phasor of the applied sinusoidal voltage. Equation (8.41) will he used as the boundary condition in the solution of the time-dependent diffusion equation for the minority canier holes in the n region.

In the neutral n region ($x > 0$), the electric field is assumed to be zero, thus the behavior of the excess minority carrier holes is determined from the equation

$$D_p \frac{\partial^2(\delta p_n)}{\partial x^2} - \frac{\delta p_n}{\tau_{p0}} = \frac{\partial(\delta p_n)}{\partial t} \tag{8.42}$$

where δp_n is the excess hole concentration in the n region. We are assuming that the ac signal voltage $v_1(t)$ is sinusoidal. We then expect the steady-state solution for δp_n to he of the form of a sinusoidal solution superimposed on the dc solution, or

$$\delta p_n(x, t) = \delta p_0(x) + p_1(x)e^{j\omega t} \tag{8.43}$$

where $\delta p_0(x)$ is the dc excess carrier concentration and $p_1(x)$ is the magnitude of the ac component of the excess carrier concentration. The expression for $\delta p_0(x)$ is the same as that given in Equation (8.14).

Substituting Equation (8.43) into the differential Equation (8.42), we obtain

$$D_p\left[\frac{\partial^2(\delta p_0(x))}{\partial x^2}+\frac{\partial^2 p_1(x)}{\partial x^2}e^{j\omega t}\right]-\frac{\delta p_0(x)+p_1(x)e^{j\omega t}}{\tau_{p0}}=j\omega p_1(x)e^{j\omega t} \qquad (8.44)$$

We may rewrite this equation, combining the time-dependent and time-independent terms, as

$$\left[D_p\frac{\partial^2(\delta p_0(x))}{\partial x^2}-\frac{\delta p_0(x)}{\tau_{p0}}\right]+\left[D_p\frac{\partial^2 p_1(x)}{\partial x^2}-\frac{p_1(x)}{\tau_{p0}}-j\omega p_1(x)\right]e^{j\omega t}=0 \qquad (8.45)$$

If the ac component, $p_1(x)$, is zero, then the first bracketed term is just the differential Equation *(8.10)*. which is identically zero. Then we have. from the second bracketed term.

$$D_p\frac{d^2p_1(x)}{dx^2}-\frac{p_1(x)}{\tau_{p0}}-j\omega p_1(x)=0 \qquad (8.46)$$

Noting that $L_p = D_p\tau_{p0}$. Equation *(8.46)* may be rewritten in the form

$$\frac{d^2p_1(x)}{dx^2}-\frac{(1+j\omega\tau_{p0})}{L_p^2}p_1(x)=0 \qquad (8.47)$$

$$\frac{d^2p_1(x)}{dx^2}-C_p^2p_1(x)=0 \qquad (8.48)$$

where

$$C_p^2=\frac{(1+j\omega\tau_{p0})}{L_p^2} \qquad (8.49)$$

The general solution to Equation *(8.48)* is

$$p_1(x)=K_1e^{-C_px}+K_2e^{+C_px} \qquad (8.50)$$

One boundary condition is that $p_1(x\to+\infty)=0$, which implies that the coefficient $K_2 = 0$. Then

$$p_1(x)=K_1e^{-C_px} \qquad (8.51)$$

Applying the boundary condition at $x = 0$ from Equation *(8.41)* we obtain

$$p_1(0)=K_1=p_{dc}\left(\frac{\hat{V}_1}{V_t}\right) \qquad (8.52)$$

The hole diffusion current density can he calculated at x = 0. This will be given by

$$J_p = -eD_p \frac{\partial p_n}{\partial x}\bigg|_{x=0} \tag{8.53}$$

If we consider a homogeneous semiconductor, the derivative of the hole concentration will be just the derivative of the excess hole concentration. Then

$$J_p = -eD_p \frac{\partial(\delta p_n)}{\partial x}\bigg|_{x=0} = -eD_p \frac{\partial(\delta p_0(x))}{\partial x}\bigg|_{x=0} - eD_p \frac{\partial p_1(x)}{\partial x}\bigg|_{x=0} e^{j\omega t} \tag{8.54}$$

We can write this equation in the form

$$J_p = J_{p0} + j_p(t) \tag{8.55}$$

where

$$J_{p0} = -eD_p \frac{\partial(\delta p_0(x))}{\partial x}\bigg|_{x=0} = \frac{eD_p p_{n0}}{L_p}\left[\exp\left(\frac{eV_0}{kT}\right) - 1\right] \tag{8.56}$$

Equation (8.56) is the dc component of the hole diffusion current density and is exactly the same as in the ideal *I–V* relation derived previously.

The sinusoidal component of the diffusion current density is then found from

$$j_p(t) = \hat{J}_p e^{j\omega t} = -eD_p \frac{\partial p_1(x)}{\partial x} e^{j\omega t}\bigg|_{x=0} \tag{8.57}$$

where $\hat{J}_p$ is the current density phasor. Combining Equations (8.57). (8.51), and (8.52), we have

$$\hat{J}_p = -eD_p(-C_p)\left[p_{dc}\left(\frac{\hat{V}_1}{V_t}\right)\right] e^{-C_p x}\bigg|_{x=0} \tag{8.58}$$

We can write the total ac hole current phasor as

$$\hat{I}_p = A\hat{J}_p = eAD_p C_p p_{dc}\left(\frac{\hat{V}_1}{V_t}\right) \tag{8.59}$$

where A is the cross-sectional area of the pn junction. Substituting the expression for $C_{„}$ we obtain

$$\hat{I}_p = \frac{eAD_p p_{dc}}{L_p}\sqrt{1 + j\omega\tau_{p0}}\left(\frac{\hat{V}_1}{V_t}\right) \tag{8.60}$$

If we define

$$I_{p0} = \frac{eAD_p p_{dc}}{L_p} = \frac{eAD_p p_{n0}}{L_p}\exp\left(\frac{eV_0}{kT}\right) \tag{8.61}$$

then Equation (8.60) becomes

$$\hat{I}_p = I_{p0}\sqrt{1 + j\omega\tau_{p0}}\left(\frac{\hat{V}_1}{V_t}\right) \tag{8.62}$$

We can go through the same type of analysis for the minority carrier electrons in the p region. We will obtain

$$\hat{I}_n = I_{n0}\sqrt{1 + j\omega\tau_{n0}}\left(\frac{\hat{V}_1}{V_t}\right) \tag{8.63}$$

where

$$I_{n0} = \frac{eAD_n n_{p0}}{L_n}\exp\left(\frac{eV_0}{kT}\right) \tag{8.64}$$

The total ac current phasor is the sum of $\hat{I}_p$ and $\hat{I}_n$. The pn junction admittance is the total ac current phasor divided by the ac voltage phasor, or

$$Y = \frac{\hat{I}}{\hat{V}_1} = \frac{\hat{I}_p + \hat{I}_n}{\hat{V}_1} = \left(\frac{1}{V_t}\right)\left[I_{p0}\sqrt{1 + j\omega\tau_{p0}} + I_{n0}\sqrt{1 + j\omega\tau_{n0}}\right] \tag{8.65}$$

There is not a linear, lumped, finite, passive, bilateral network that can be synthesized to give this admittance function. However, we may make the following approximations. Assume that

$$\omega\tau_{p0} \ll 1 \tag{8.66a}$$

and

$$\omega\tau_{n0} \ll 1 \tag{8.66b}$$

These two assumptions imply that the frequency of the ac signal is not too large. Then we may write

$$\sqrt{1 + j\omega\tau_{p0}} \approx 1 + \frac{j\omega\tau_{p0}}{2} \tag{8.67a}$$

and

$$\sqrt{1 + j\omega\tau_{n0}} \approx 1 + \frac{j\omega\tau_{n0}}{2} \tag{8.67b}$$

Substituting Equations (8.67a) and (8.67b) into the admittance Equation (8.65) yields

$$Y = \left(\frac{1}{V_t}\right)\left[I_{p0}\left(1 + \frac{j\omega\tau_{p0}}{2}\right) + I_{n0}\left(1 + \frac{j\omega\tau_{n0}}{2}\right)\right] \tag{8.68}$$

If we combine the real and imaginary portions, we get

$$Y = \left(\frac{1}{V_t}\right)(I_{p0} + I_{n0}) + j\omega\left\{\left(\frac{1}{2V_t}\right)[I_{p0}\tau_{p0} + I_{n0}\tau_{n0}]\right\} \tag{8.69}$$

Equation (8.69) may be written in the form

$$Y = g_d + j\omega C_d \tag{8.70}$$

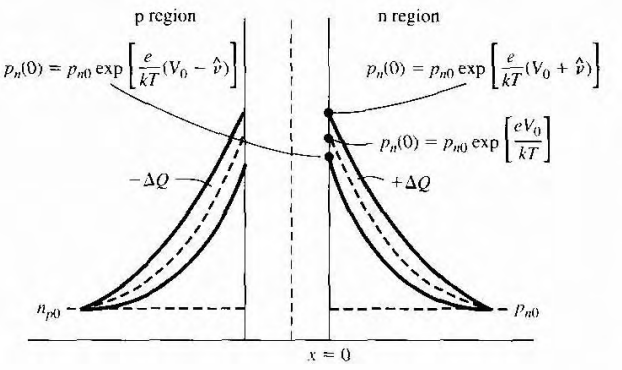


Figure 8.14 | Minority carrier concentration changes with changing forward-bias voltage.

The parameter g_d is called the *diffusion conductance* and is given by

$$g_d = \left(\frac{1}{V_t}\right)(I_{p0} + I_{n0}) = \frac{I_{DQ}}{V_t} \tag{8.71}$$

where I_{DQ} is the dc bias current. Equation (8.71) is exactly the same conductance as we obtained previously in Equation (8.34). The parameter C_d is called the *diffusion capacitance* and is given by

$$\boxed{C_d = \left(\frac{1}{2V_t}\right)(I_{p0}\tau_{p0} + I_{n0}\tau_{n0})} \tag{8.72}$$

The physics of the diffusion capacitance may be seen in Figure 8.14. The dc values of the minority carrier concentrations are shown along with the changes due to the ac component of voltage. The A Q charge is alternately being charged and discharged through the junction as the voltage across the junction changes. The change in the stored minority carrier charge as a function of the change in voltage is the diffusion capacitance. One consequence of the approximations $\omega\tau_{p0} \ll$ I and $\omega\tau_{n0} \ll 1$ is that there are no "wiggles" in the minority carrier curves. The sinusoidal frequency is low enough so that the exponential curves are maintained at all times.

EXAMPLE 8.6 | Objective

To calculate the small-signal admittance of a pn junction diode.

This example is intended to give an indication of the magnitude of the diffusion capacitance as compared with the junction capacitance considered in the last chapter. The diffusion resistance will also be calculated. Assume that $N_a \gg N_d$ so that $p_{n0} \gg n_{p0}$. This assumption implies that $I_{„„} \gg I_{n0}$. Let T = 300 K, $\tau_{p0} = 10^{-7}$ s, and $I_{„„} = I_{DQ} = 1$ mA.

Solution

The diffusion capacitance, with these assumptions, is given by

$$C_d \approx \left(\frac{1}{2V_t}\right)(I_{p0}\tau_{p0}) = \frac{1}{(2)(0.0259)}(10^{-3})(10^{-7}) = 1.93 \times 10^{-9}\ \text{F}$$

The diffusion resistance is

$$r_d = \frac{V_t}{I_{DQ}} = \frac{0.0259\ \text{V}}{\text{I mA}} = 25.9\ \Omega$$

Comment

The value of 1.93 *nF* for the diffusion capacitance of a forward-biased pn junction is 3 to 4 orders of magnitude larger than the junction capacitance of the reverse-biased pn junctiun, which we calculated in Example 7.5.

The diffusion capacitance tends to dominate the capacitance terms in a forward-biased pn junction. The small-signal diffusion resistance can be fairly small if the diode current is a fairly large value. As the diode current decreases, the diffusion resistance increases. We will consider the impedance of forward-biased pn junctions again when we discuss bipolar transistors.

TEST YOUR UNDERSTANDING

E8.8 A silicon pn junction diode at T = 300 K has the Following parameters: $N_d = 8 \times 10^{16}\ \text{cm}^{-3}$, $N_a = 2 \times 10^{15}\ \text{cm}^{-3}$, $D_n = 25\ \text{cm}^2/\text{s}$, $D_p = 10\ \text{cm}^2/\text{s}$, $\tau_{n0} = 5 \times 10^{-7}$ s, and $\tau_{p0} = 10^{-7}$ s. The cross-sectional area is $A = 10^{-3}\ \text{cm}^2$. Determine the diffusion resistance and diffusion capacitance if the diode is forward biased at (*a*) $V_a = 0.550$ V and (*b*) $V_a = 0.610$ V.

[Ans. (*a*) $r_d = 118\ \Omega$, $C_d = 2.07\ \text{nF}$; (*b*) $r_d = 11.6\ \Omega$, $C_d = 20.9\ \text{nF}$]

E8.9 A GaAs pn junction diode at T = 300 K has the same parameters given in E8.8 except thet $D_n = 207$ cm'ls and $D_p = 9.80\ \text{cm}^2/\text{s}$. Determine the diffusion resistance and diffusion capacitance if the diode is forward biased at (*a*) $V_a = 0.970$ V and (b) $V_a = 1.045$ V.

[Ans. (*a*) $r_d = 263\ \Omega$, $C_d = 0.940\ \text{nF}$; (*b*) $r_d = 14.6\ \Omega$, $C_d = 17.0\ \text{nF}$]

8.2.3 Equivalent Circuit

The small-signal equivalent circuit of the forward-biased pn junction is derived from Equation (8.70).This circuit is shown in Figure 8.15a. We need to add the junction capacitance, which will be in parallel with the diffusion resistance and diffusion capacitance. The last element we add, to complete the equivalent circuit, is a series resistance. The neutral n and p regions have finite resistances so the actual pn junction will include a series resistance. The complete equivalent circuit is given in Figure 8.15b.

The voltage across the actual junction is V_a and the total voltage applied to the pn diode is given by V_{app}. The junction voltage V_a is the voltage in the ideal

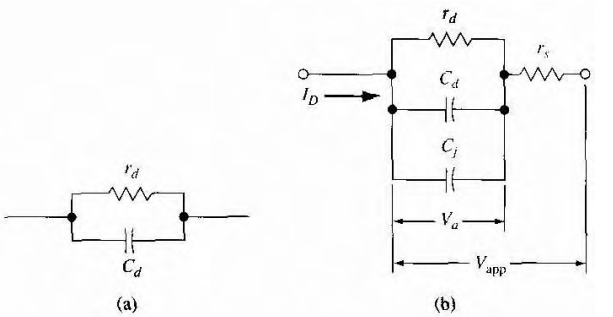


Figure 8.15 | (a) Small-signal equivalent circuit of ideal forward-biased pn junction diode; (b) Complete small-signal equivalent circuit of pn junction.

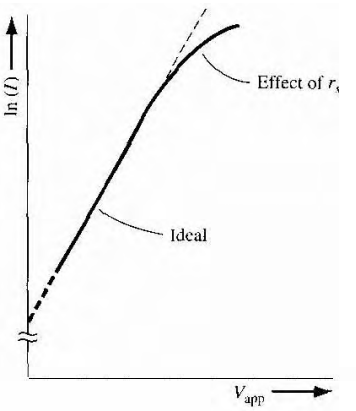


Figure 8.161 Forward-biased I–V characteristics of a pn junction diode showing the effect of series resistance.

current–voltage expression. We can write the expression

$$V_{app} = V_a + Ir_s \tag{8.73}$$

Figure 8.16 is a plot of the current–voltage characteristic from Equation (8.73) showing the effect of the series resistance. A larger applied voltage is required to achieve the same current value when a series resistance is included. In most diodes, the series resistance will be negligible. In some semiconductor devices with pn junctions, however, the series resistance will be in a feedback loop: in these cases, the resistance is multiplied by a gain factor and becomes non-negligible.

TEST YOUR UNDERSTANDING

E8.10 A silicon pn junction diode at T = 300 K has the same parameters as those described in E8.8. The neutral n-region and neutral p-region lengths are 0.01 cm. Estimate the series resistance of the diode (neglect ohmic contacts). (Ans. 99 Ω)

8.3 GENERATION–RECOMBINATION CURRENTS

In the derivation of the ideal current–voltage relationship, we neglected any effects occurring within the space charge region. Since other current components are generated within the space charge region, the actual *I–V* characteristics of a pn junction diode deviate from the ideal expression. The additional currents are generated from the recombination processes discussed in Chapter 6.

The recombination rate of excess electrons and holes, given by the Shockley–Read–Hall recombination theory, was written as

$$R = \frac{C_n C_p N_t (np - n_i^2)}{C_n(n + n') + C_p(p + p')} \tag{8.74}$$

The parameters n and p are, as usual. the concentrations of electrons and holes, respectively.

8.3.1 Reverse-Bias Generation Current

For a pn junction under reverse bias, we have argued that the mobile electrons and holes have essentially been swept out of the space charge region. Accordingly, within the space charge region, $n \approx p \approx 0$. The recombination rate from Equation (8.74) becomes

$$R = \frac{-C_n C_p N_t n_i^2}{C_n n' + C_p p'} \tag{8.75}$$

The negative sign implies a negative recombination rate; hence, we are really generating electron–hole pairs within the reverse-biased space charge region. The recombination of excess electrons and holes is the process whereby we are trying to reestablish thermal equilibrium. Since the concentration of electrons and holes is essentially zero within the reverse-biased space charge region. electrons and holes are being generated via the trap level to also try to reestablish thermal equilibrium. This generation process is schematically shown in Figure 8.17. As the electrons and holes are generated. they are swept out of the space charge region by the electric field. The flow of charge is in the direction of a reverse-bias current. This reverse-bias *generation current*, caused by the generation of electrons and holes in the space charge region. is in addition to the ideal reverse-bias saturation current.

We may calculate the density of the reverse-bias generation current by considering Equation (8.75). If we make a simplifying assumption and let the trap level be at the intrinsic Fermi level, then from Equations (6.92) and (6.97), we have that $n' = n_i$

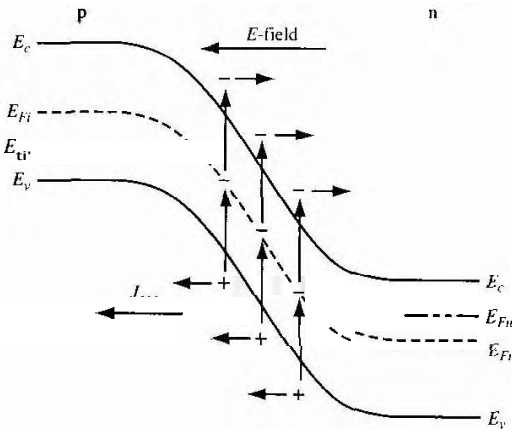


Figure 8.17 | Generation process in a reverse-biased pn junction.

and $p' = n_i$. Equation (8.75) now becomes

$$R = \frac{-n_i}{\dfrac{1}{N_t C_p} + \dfrac{1}{N_t C_n}} \tag{8.76}$$

Using the definitions of lifetimes from Equations (6.101) and (6.104), we may write Equation (8.76) as

$$R = \frac{-n_i}{\tau_{p0} + \tau_{n0}} \tag{8.77}$$

If we define a new lifetime as the average of τ_{p0} and τ_{n0}, or

$$\tau_0 = \frac{\tau_{p0} + \tau_{n0}}{2} \tag{8.78}$$

then the recombination rate can he written as

$$R = \frac{-n_i}{2\tau_0} \equiv -G \tag{8.79}$$

The negative recombination rate implies a generation rate, so G is the generation rate of electrons and holes in the space charge region.

The generation current density may he determined from

$$J_{gen} = \int_0^W eG\,dx \tag{8.80}$$

where the integral is over the space charge region. If we assume that the generation rate is constant throughout the space charge region, then we obtain

$$\boxed{J_{gen} = \frac{en_i W}{2\tau_0}} \tag{8.81}$$

The total reverse-bias current density is the sum of the ideal reverse saturation current density and the generation current density, or

$$J_R = J_s + J_{gen} \tag{8.82}$$

The ideal reverse saturation current density J_s is independent of the reverse-bias voltage. However, J_{gen} is a function of the depletion width W, which in turn is a function of the reverse-bias voltage. The actual reverse-bias current density, then, is no longer independent of the reverse-bias voltage.

EXAMPLE 8.7

Objective

To determine the relative magnitudes of the ideal reverse saturation current density and the generation current density in a silicon pn junction at T = 300 K.

Consider the silicon *pn* junction described in Example 8.2 and Jet $\tau_0 = \tau_{p0} = \tau_{n0} = 5 \times 10^{-7}$ s.

■ Solution

The ideal reverse saturation current density was calculated in Example 8.2 and was found to be $J_s = 4.15 \times 10^{-11}$ A/cm^2. The generation current density is again given by Equation (8.81) as

$$J_{gen} = \frac{en_i W}{2\tau_0}$$

and the depletion width is given *hy*

$$W = \left\{ \frac{2\epsilon_s}{e} \left(\frac{N_a + N_d}{N_a N_d} \right) (V_{bi} + V_R) \right\}^{1/2}$$

If **we** assume. for example, that $V_{bi} + V_R = 5$ V. then using the parameters given in Example 8.2 we find that $W = 1.14 \times 10^{-4}$ cm. and then calculate the generation current density to he

$$J_{gen} = 2.74 \times 10^{-7} \text{ A/cm}^2$$

■ Comment

Comparing the solutions for the two current densities, it is obvious that. for the silicon pn junction diode at room temperature, the generation current density is approximately four orders of magnitude larger than the ideal saturation current density. The generation current is the dominant reverse-bias current in a silicon pn junction diode.

TEST YOUR UNDERSTANDING

E8.11 A GaAs pn junction diode has the same parameters as described in E8.9. (a) Calculate the reverse-bias generation current if the diode is reverse biased at $V_R = 5$ V. **(b)** Determine the ratio of I_{gen} calculated in **part** *(a)* to the ideal reverse-saturation current I_S. [1.93×10^{-13} A, 1.03×10^{7}]

8.3.2 Forward-Bias Recombination Current

For the reverse-biased pn junction, electrons and holes are essentially completely swept out of the space charge region so that $n \approx p \approx 0$. Under forward bias, however, electrons and holes are injected across the space charge region, so we do, in fact, have some excess carriers in the space charge region. The possibility exists that some of these electrons and holes will recombine within the space charge region and not become part of the minority carrier distribution.

The recombination rate of electrons and holes is again given from Equation (8.74) as

$$R = \frac{C_n C_p N_t (np - n_i^2)}{C_n(n + n') + C_p(p + p')}$$

Dividing both numerator and denominator by $C_n C_p N_t$ and using the definitions of τ_{n0} and τ_{p0}, we may write the recombination rate as

$$R = \frac{np - n_i^2}{\tau_{p0}(n + n') + \tau_{n0}(p + p')} \tag{8.8}$$

Figure 8.18 shows the energy-band diagram of the forward-biased pn junction. Shown in the figure are the intrinsic Fermi level and the quasi-Fermi levels for

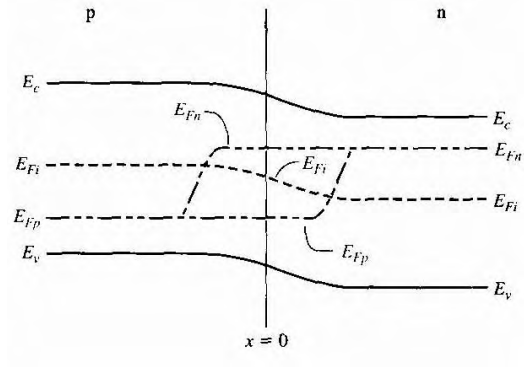


Figure 8.18 | Energy-band diagram of a forward-biased pn junction including quasi-Fermi levels.

electrons and holes. From the results of Chapter 6, we may write the electron concentration as

$$n = n_i \exp\left[\frac{E_{Fn} - E_{Fi}}{kT}\right] \tag{8.84}$$

and the hole concentration as

$$p = n_i \exp\left[\frac{E_{Fi} - E_{Fp}}{kT}\right] \tag{8.85}$$

where E_{Fn} and E_{Fp} are the quasi-Fermi levels for electrons and holes, respectively.

From Figure 8.18, we may note that

$$(E_{Fn} - E_{Fi}) + (E_{Fi} - E_{Fp}) = eV_a \tag{8.86}$$

where V_a is the applied forward-bias voltage. Again, if we assume that the trap level is at the intrinsic Fermi level, then $n' = p' = n_i$. Figure 8.19 shows a plot of the relative magnitude of the recombination rate as a function of distance through the space charge region. This plot was generated using Equations (8.83), (8.84). (8.85), and (8.86). A very sharp peak occurs at the metallurgical junction ($x = 0$).

At the center of the space charge region, we have

$$E_{Fn} - E_{Fi} = E_{Fi} - E_{Fp} = \frac{eV_a}{2} \tag{8.87}$$

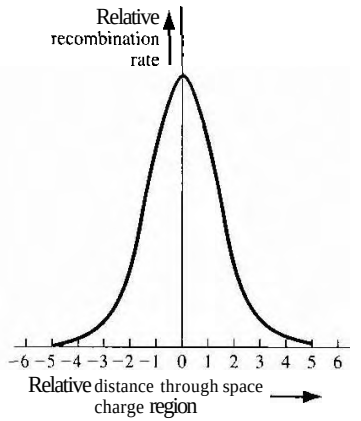


Figure 8.19 | Relative magnitude of the recombination rate through the space charge region of a forward-biased pn junction.

Equations (8.84) and (8.85) then become

$$n = n_i \exp\left(\frac{eV_a}{2kT}\right) \tag{8.88}$$

and

$$p = n_i \exp\left(\frac{eV_a}{2kT}\right) \tag{8.89}$$

If we assume that $n' = p' = 11$, and that $\tau_{n0} = \tau_{p0} = \tau_0$, then Equation (8.83) becomes

$$R_{\max} = \frac{n_i \ [\exp(eV_a/kT) - 1]}{2\tau_0 \ [\exp(eV_a/2kT) + 1]} \tag{8.90}$$

which is the maximum recombination rate for electrons and holes that occurs at the center of the forward-biased pn junction. If we assume that $V_a \gg kT/e$, we may neglect the (-1) term in the numerator and the $(+1)$ term in the denominator. Equation (8.90) then becomes

$$R_{\max} = \frac{n_i}{2\tau_0} \exp\left(\frac{eV_a}{2kT}\right) \tag{8.91}$$

The recombination current density may be calculated from

$$J_{\text{rec}} = \int_0^W eR\,dx \tag{8.92}$$

where again the integral is over the entire space charge region. In this case, however, the recombination rate is not a constant through the space charge region. We have calculated the maximum recombination rate at the center of the space charge region, so we may write

$$J_{\text{rec}} = ex' \frac{n_i}{2\tau_0} \exp\left(\frac{eV_a}{2kT}\right) \tag{8.93}$$

where x' is a length overwhich the maximum recombination rate iseffective. However, since τ_0 may not be a well-defined or known parameter, it is customary to write

$$\boxed{J_{\text{rec}} = \frac{eWn_i}{2\tau_0} \exp\left(\frac{eV_a}{2kT}\right) = J_{r0} \exp\left(\frac{eV_a}{2kT}\right)} \tag{8.94}$$

where W is the space charge width.

TEST YOUR UNDERSTANDING

E8.12 Consider a silicon pn junction diode at $T = 300$ K with the same parameters given in E8.8. The diode is forward biased at $V_a = 0.50$ V. *(a)* Calculate the forward-biased recombination current. *(b)* Determine the ratio of I_{rec} calculated in part *(a)* to the ideal diffusioncurrent. [Ans. (a) $I_{rec} = 2.30 \times 10^{-6}$ A, (b) 7.21×10^{-2}]

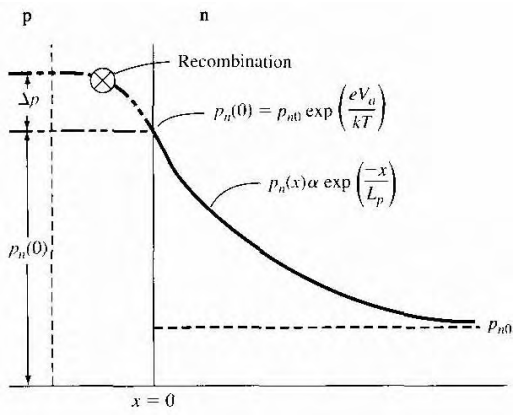


Figure **8.20** | Because of recombination, additional holes from the p region must be injected into the space charge region to establish the minority carrier hole concentration in the n region.

8.3.3 Total Forward-Bias Current

The total forward-bias current density in the pn junction is the sum of the recombination and the ideal diffusion current densities. Figure 8.20 shows a plot of the minority carrier hole concentration in the neutral n region. This distribution yields the ideal hole diffusion current density and is a function of the minority carrier hole diffusion length and the applied junction voltage. The distribution is established as a result of holes being injected across the space charge region. If, now, some of the injected holes in the space charge region are lost due to recombination, then additional holes must be injected from the p region to make up for this loss. The flow of these additional injected carriers, per unit time, results in the recombination current. This added component is schematically shown in the figure.

The total forward-bias current density is the sum of the recombination and the ideal diffusion current densities, so we can write

$$J = J_{rec} + J_D \tag{8.95}$$

where J_{rec} is given by Equation **(8.94)** and J_D is given by

$$J_D = J_s \exp\left(\frac{eV_a}{kT}\right) \tag{8.96}$$

The (-1) term in Equation **(8.23)** has been neglected. The parameter J_s is the ideal reverse saturation current density, and from previous discussion, the value of J_{r0} from the recombination current is larger than the value of J_s.

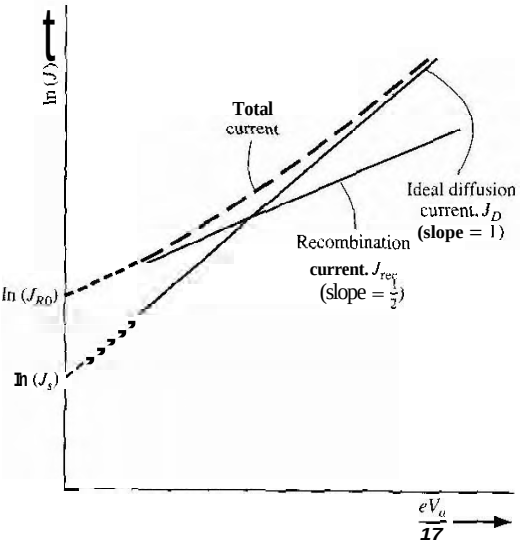


Figure 8.21 | Ideal diffusion.recornhination,and total current in a forward-biased pn junction.

If we take the natural log of Equations (8.94) and (8.96). we obtain

$$\ln J_{\text{rec}} = \ln J_{r0} + \frac{eV_a}{2kT} = \ln J_{r0} + \frac{V_a}{2V_t} \tag{8.97a}$$

and

$$\ln J_D = \ln J_s + \frac{eV_a}{kT} = \ln J_s + \frac{V_a}{V_t} \tag{8.97b}$$

Figure 8.21 shows the recombination and diffusion current components plotted on a log current scale as a function of V_a/V_t. The slopes of the two curves are not the same. Also shown in the figure is the total current density—the sum of the two current components. We may notice that, at a low current density, the recombination current dominates, and at a higher current density, the ideal diffusion current dominates.

In general, the diode current-voltage relationship may he written as

$$\boxed{I = I_s\left[\exp\left(\frac{eV_a}{nkT}\right) - 1\right]} \tag{8.98}$$

where the parameter n is called the *ideality factor.* For a large forward-bias voltage. $n \approx 1$ when diffusion dominates, and for low forward-bias voltage, $n \approx 2$ when recombination dominates. There is a transition region where $1 < n < 2$.

8.4 | JUNCTION BREAKDOWN

In the ideal pn junction, a reverse-bias voltage will result in a small reverse-bias current through the device. However, the reverse-bias voltage may not increase without limit; at some particular voltage, the reverse-bias current will increase rapidly. The applied voltage at this point is called the *breakdown* ***voltage.***

Two physical mechanisms give rise to the reverse-bias breakdown in a pn junction: the ***Zener*** *effect* and the *avalanche effect.* Zener breakdown occurs in highly doped pn junctions through a tunneling mechanism. In a highly doped junction, the conduction and valence bands on opposite sides of the junction are sufficiently close during reverse bias that electrons may tunnel directly from the valence band on the p side into the conduction band on the n side. This tunneling process is schematically shown in Figure 8.22a.

The avalanche breakdown process occurs when electrons and/or holes, moving across the space charge region, acquire sufficient energy from the electric field to create electron–hole pairs by colliding with atomic electrons within the depletion region. The avalanche process is schematically shown in Figure 8.22b. The newly created electrons and holes move in opposite directions due to the electric field and thereby add to the existing reverse-bias current. In addition, the newly generated electrons and/or holes may acquire sufficient energy to ionize other atoms, leading to the avalanche process. For most pn junctions, the predominant breakdown mechanism will be the avalanche effect.

If we assume that a reverse-bias electron current I_{n0} enters the depletion region at $x = 0$ as shown in Figure 8.23, the electron current I_n will increase with distance through the depletion region due to the avalanche process. At $x = W$, the electron

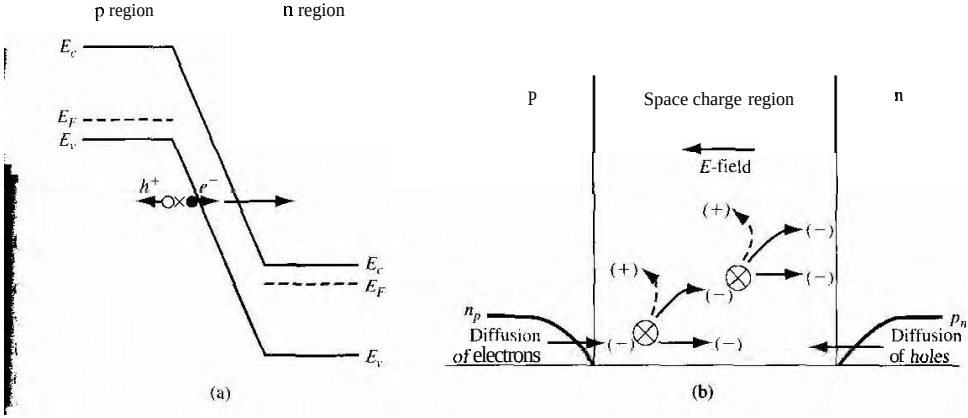


Figure 8.22 | **(a)** Zener breakdown mechanism in a reverse-biased pn junction; (b) avalanche breakdown process in a reverse-biascd pn junction.

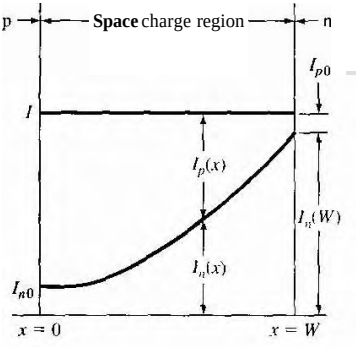


Figure **8.23** | Electron and hole current components through the space charge region during avalanche multiplication.

current may he written as

$$I_n(W) = M_n I_{n0} \tag{8.99}$$

where M_n is a multiplication factor. The hole current is increasing through the depletion region from the n to p region and reaches a maximum value at $x = 0$. The total current is constant through the pn junction in steady state.

We can write an expression for the incremental electron current at some point x as

$$dI_n(x) = I_n(x)\alpha_n dx + I_p(x)\alpha_p dx \tag{8.100}$$

where α_n and a, are the electron and hole ionization rates, respectively. The ionization rates are the number of electron-hole pairs generated per unit length by an electron $(a,,)$or by a hole (α_p). Equation (8.100) may he written as

$$\frac{dI_n(x)}{dx} = I_n(x)\alpha_n + I_p(x)\alpha_p \tag{8.101}$$

The total current I is given by

$$I = I_n(x) + I_p(x) \tag{8.102}$$

which is a constant. Solving for $I_p(x)$ from Equation (8.102) and substituting into Equation (8.101), we obtain

$$\frac{dI_n(x)}{dx} + (\alpha_p - \alpha_n)I_n(x) = \alpha_p I \tag{8.103}$$

If we make the assumption that the electron and hole ionization rates are equal so

then Equation (8.103) may be simplified and integrated through the space charge region. We will obtain

$$I_n W) - I,(0) = I \int_0^W \alpha \, dx \tag{8.105}$$

Using Equation (8.99), Equation (8.105) may bc written as

$$\frac{M_n I_{n0} - I_n(0)}{I} = \int_0^W \alpha \, dx \tag{8.106}$$

Since $M_n I_{n0} \approx I$ and since $I_n(0) = I_{n0}$, Equation (8 106) becomes

$$1 - \frac{1}{M_n} = \int_0^W \alpha \, dx \tag{8.107}$$

The avalanche breakdown voltage is defined to be the voltage at which M,, approaches intinity. The avalanche breakdown condition is then given by

$$\int_0^W \alpha \, dx = 1 \tag{8.108}$$

The ioniration rates are strong functions of electric field and, since the electric field is notconstant through the space charge region. Equation (8.108) is not easy toevaluate.

If we consider, for example, a one-sided p+n junction, the maximum electric field is given by

$$E_{max} = \frac{eN_d x_n}{\epsilon_s} \tag{8.109}$$

The depletion width r,, is given approximately as

$$x_n \approx \left\{ \frac{2\epsilon_s V_R}{e} \cdot \frac{1}{N_d} \right\}^{1/2} \tag{8.110}$$

where V_R is the magnitude of the applied reverse-hias voltage. We have neglected the built-in potential V_{bi}.

If we now dctine V_R to he the breakdown voltage V_B, the maximum clcctric field. E_{max}, will be defined as a critical electric field, E_{crit}, at breakdown. Combining Equations (8.109) and (8.110), wc may write

$$\boxed{V_B = \frac{\epsilon_s E_{crit}^2}{2eN_B}} \tag{8.111}$$

where N_B is the semiconductordoping in the low-doped region of the one-sided junction. The critical clcctric field, plotted in Figure 8.24, is a slight function of doping.

We have been considering a uniformly doped planar junction. The breakdown voltage will decrease for a linearly graded junction. Figure 8.25 shows a plot of the breakdown voltage for a one-sided abrupt junction and a linearly graded junction. If we take into account the curvature of a diffused junction as well, the breakdown voltage will be further degraded.

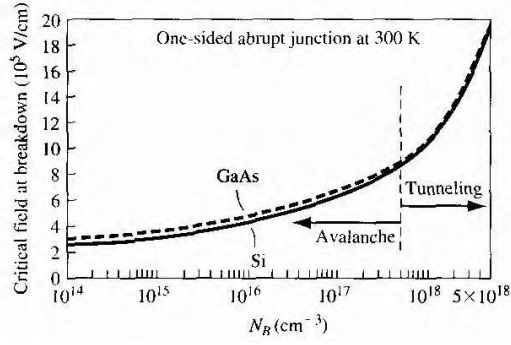


Figure 8.24 | Critical electric field at breakdown in a one sided junction as a function of impurity doping concentrations. *(From Sze [12].)*

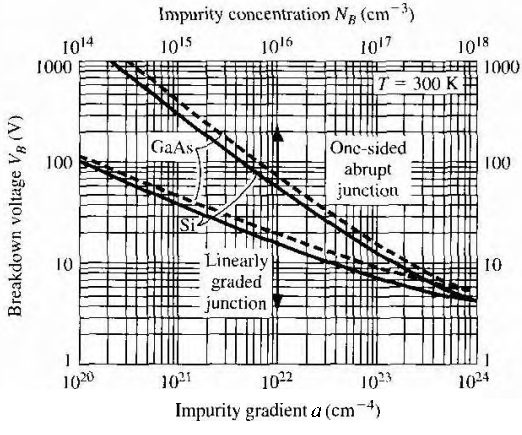


Figure 8.25 | Breakdown voltage versus impurity concentration in uniformly doped and linearly graded junctions. *(From Sze [12].)*

DESIGN EXAMPLE 8.8



Objective

To design an ideal one-sided n^+p junction diode to meet a breakdown voltage specification.

Consider a silicon pn junction diode at $T = 300$ K. Assume that $N_d = 3 \times 10^{18}$ cm^{-3}. Design the diode such that the breakdown voltage is $V_B = 100$ V.

■ Solution

From Figure 8.25, we find that the doping concentration in the low-doped side of a one-sided abrupt junction should be approximately 4×10^{15} cm^{-3} for a breakdown voltage of 100 V.

For a doping concentration of 4×10^{15} cm^{-3}, the critical electric field, from Figure 8.24, is approximately 3.7×10^5 V/cm. Then from Equation (8.111), the breakdown voltage is 110 V, which correlates quite well with the results from Figure 8.25.

Conclusion

As Figure 8.25 showc. the breakdown voltage increases as the doping concentration decreases in the low-doped region.

TEST YOUR UNDERSTANDING

E8.13 A one-sided, planar, uniformly doped silicon pn junction diode is required to have a reverse-bias breakdown voltage of $V_B = 60$ V. What is the maximum doping concentration in the low-doped region such that this specification is met? ($_{\varepsilon-}$ɯɔ $_{\varsigma}$10^{15} × 8 ≈ ^{q}N 'suA)

E8.14 Repeat E8.13 for a GaAs diode. ($_{\varepsilon-}$ɯɔ $_{9}$10^{1} × ς'1 ≈ ^{q}N 'suA)

*8.5 CHARGE STORAGE AND DIODE TRANSIENTS

The pn junction is typically used as an electrical switch. In forward bias, referred to as the *on* state, a relatively large current can be produced by a small applied voltage; in reverse bias, referred to as the *off* state, only a very small current will exist. Of primary interest in circuit applications is the speed of the pn junction diode in switching states. We will qualitatively discuss the transients that occur and the charge storage effects. We will simply state the equations that describe the switching times without any mathematical derivations.

8.5.1 The Turn-off Transient

Suppose we want to switch a diode from the forward bias on state to the reverse-bias off state. Figure 8.26 shows a simple circuit that will switch the applied bias at t = 0. For $t < 0$, the forward-bias current is

$$I = I_F = \frac{V_F - V_a}{R_F} \tag{8.112}$$

The minority carrier concentrations in the device, for the applied forward voltage V_F, are shown in Figure 8.27a. There is excess minority carrier charge stored in both the p and n regions of the diode. The excess minority carrier concentrations at the space charge edges are supported by the forward-bias junction voltage V_a. When the voltage is switched from the forward- to the reverse-bias state, the excess minority carrier concentrations at the space charge edges can no longer be supported and they start to decrease, as shown in Figure 8.27b.

The collapse of the minority carrier concentrations at the edges of the space charge region leads to large concentration gradients and diffusion currents in the

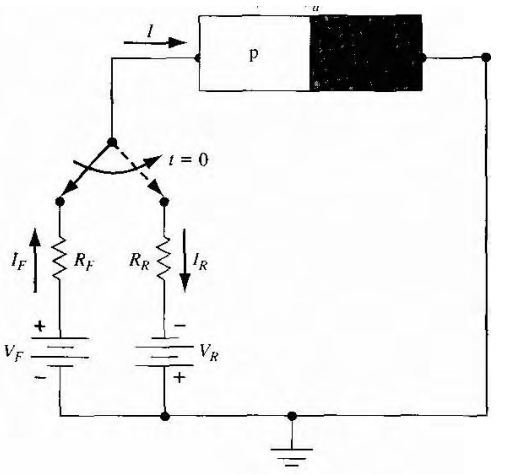


Figure **8.26** | Simple circuit for switching a diode from forward to reverse bias.

reverse-bias direction. If we assume, for the moment, that the voltage across the diode junction is small compared with V_R, then the reverse-bias current is limited to approximately

$$I = -I_R \approx \frac{-V_R}{R_R} \tag{8.113}$$

The junction capacitances do not allow the junction voltage to change instantaneously. If the current I_R were larger than this value, there would be a forward-bias voltage across the junction, which would violate our assumption of a reverse-bias current. If the current I_R were smaller than this value, there would be a reverse-bias voltage across the junction, which means that the junction voltage would have changed instantaneously. Since the reverse current is limited to the value given by Equation (8.113), the reverse-bias density gradient is constant; thus, the minority carrier concentrations at the space charge edge decrease with time as shown in Figure 8.27b.

This reverse current I_R will be approximately constant for $0^+ \leq t \leq t_s$, where t_s is called the ***storage time.*** The storage time is the length of time required for the minority carrier concentrations at the space charge edge to reach the thermal-equilibrium values. After this time, the voltage across the junction will begin to change. The current characteristic is shown in Figure 8.28. The reverse current is the Row of the stored minority carrier charge, which is the difference between the minority carrier concentrations at $t = 0^-$ and $t = \infty$, as was shown in Figure 8.27b.

The storage time t_s can be determined by solving the time-dependent continuity equation. If we consider a one-sided p^+n junction, the storage time is determined

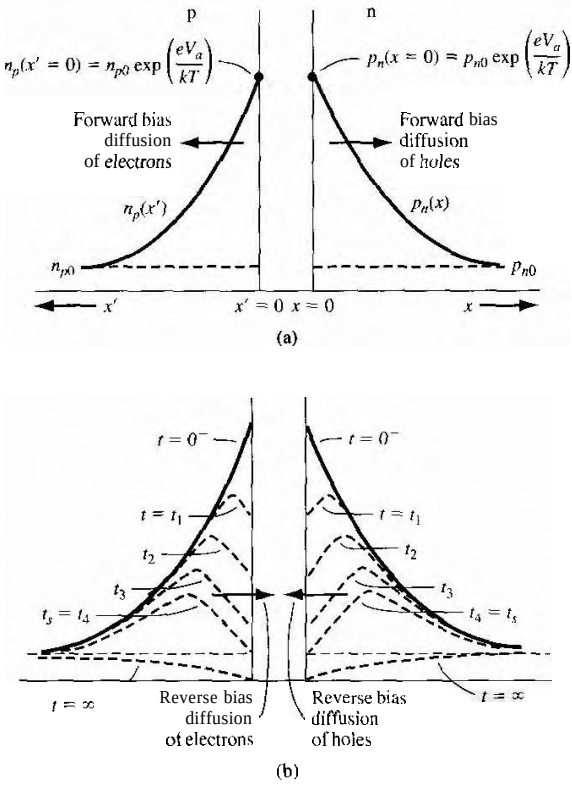


Figure **8.27** | (a) Steady-state forward-bias minority carrier concentrations: (b) minority carrier concentrations at **various** times during switching.

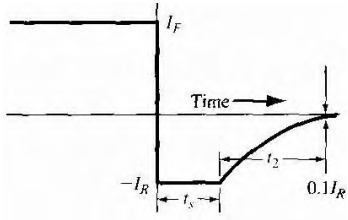


Figure 8.28 | Current characteristic versus time during diode switching.

from the equation

$$\text{erf}\sqrt{\frac{t_s}{\tau_{p0}}} = \frac{I_F}{I_F + I_R} \tag{8.11}$$

where erf (x) is known as the error function. An approximate solution for the storage time can he obtained as

$$t_s \approx \tau_{p0} \ln\left[1 + \frac{I_F}{I_R}\right] \tag{8.115}$$

The recovery phase for $t > t_s$ is the time required for the junction to reach its steady-state reverse-bias condition. The remainder of the excess charge is being removed and the space charge width is increasing to the reverse-bias value. The decay time t_2 is determined from

$$\text{erf}\sqrt{\frac{t_2}{\tau_{p0}}} + \frac{\exp(-t_2/\tau_{p0})}{\sqrt{\pi t_2/\tau_{p0}}} = 1 + 0.1\left(\frac{I_R}{I_F}\right) \tag{8.116}$$

The total turn-off time is the sum of t_s and t_2.

To switch the diode quickly, we need to he able to produce a large reverse current as well as have a small minority carrier lifetime. In the design of diode circuits, then, the designer must provide a path for the transient reverse-bias current pulse in order to he able to switch the diode quickly. These same effects will be considered when we discuss the switching of bipolar transistors.

TEST YOUR UNDERSTANDING

E8.15 A one-sided p^+n silicon diode, that has a forward-bias current of $I_F = 1.75$ mA, is switched to reverse bias with an effective reverse-bias voltage of $V_R = 2$ V and an effective series resistance of $R_R = 4\ \text{k}\Omega$. The minority carrier hole lifetime is 10^{-7} s. (a) Determine the storage time t_s. (b) Calculate the decay time t_2. (c) What is the turn-off time of the diode?

[(a) 0.746 × 10⁻⁷ s, (b) 1.25 × 10⁻⁷ s, (c) 2 × 10⁻⁷ s]

8.5.2 The Turn-on Transient

The turn-on transient occurs when the diode is switched from its "off" state into the forward-bias "on" state. The turn-on can be accomplished by applying a forward-bias current pulse. The first stage of turn-on occurs very quickly and is the length of time required to narrow the space charge width from the reverse-bias value to its thermal-equilibrium value when $V_a = 0$. During this time, ionized donors and acceptors are neutralized as the space charge width narrows.

The second stage of the turn-on process is the time required to establish the minority-carrier distributions. During this time the voltage across the junction is increasing toward its steady-state value. A small turn-on time is achieved if the minority carrier lifetime is small and if the forward-bias current is small.

*8.6 | THE TUNNEL DIODE

The *tunnel* diode is a pn junction in which both the n and p regions are degenerately doped. As we discuss the operation of this device, we will find a region that exhibits a negative differential resistance. The tunnel diode was used in oscillator circuits in the past, but other types of solid-state devices are now used as high-frequency oscillators: thus, the tunnel diode is really only of academic interest. Nevertheless, this device does demonstrate the phenomenon of tunneling we discussed in Chapter 2.

Recall the degenerately doped semiconductors we discussed in Chapter 4: the Fermi level is in the conduction band of a degenerately doped n-type material and in the valence band of a degenerately doped p-type material. Then, even at T = 0 K, electrons will exist in the conduction hand of the n-type material, and holes (empty states) will exist in the p-type material.

Figure 8.29 shows the energy-band diagram of a pn junction in thermal equilibrium for the case when both the n and p regions are degenerately doped. The depletion region width decreases as the doping increases and may be on the order of approximately 100 Å for the case shown in Figure 8.29. The potential barrier at the junction can be approximated by a triangular potential barrier, as is shown in Figure 8.30. This potential harrier is similar to the potential barrier used in Chapter 2 to illustrate the tunneling phenomenon. The barrier width is small and the electric field in the space charge region is quite large; thus, a finite probability exists that an electron may tunnel through the forbidden band from one side of the junction to the other.

We may qualitatively determine the current–voltage characteristics of the tunnel diode by considering the simplified energy-band diagrams in Figure 8.31.

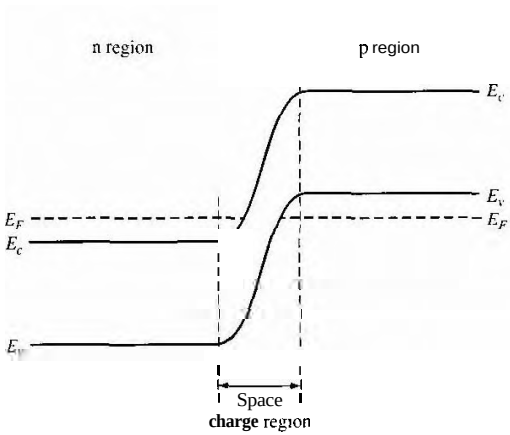


Figure 8.29 | Energy-hand diagram of a pn junction in thermal equilibrium in which both the **n and** p regions are degenerately doped.

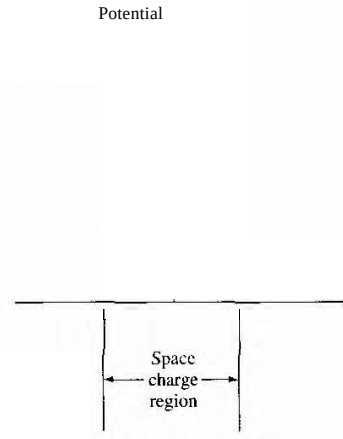


Figure 8.30 | Triangular potential harrier approximation of the potential harrier in the tunnel diode.

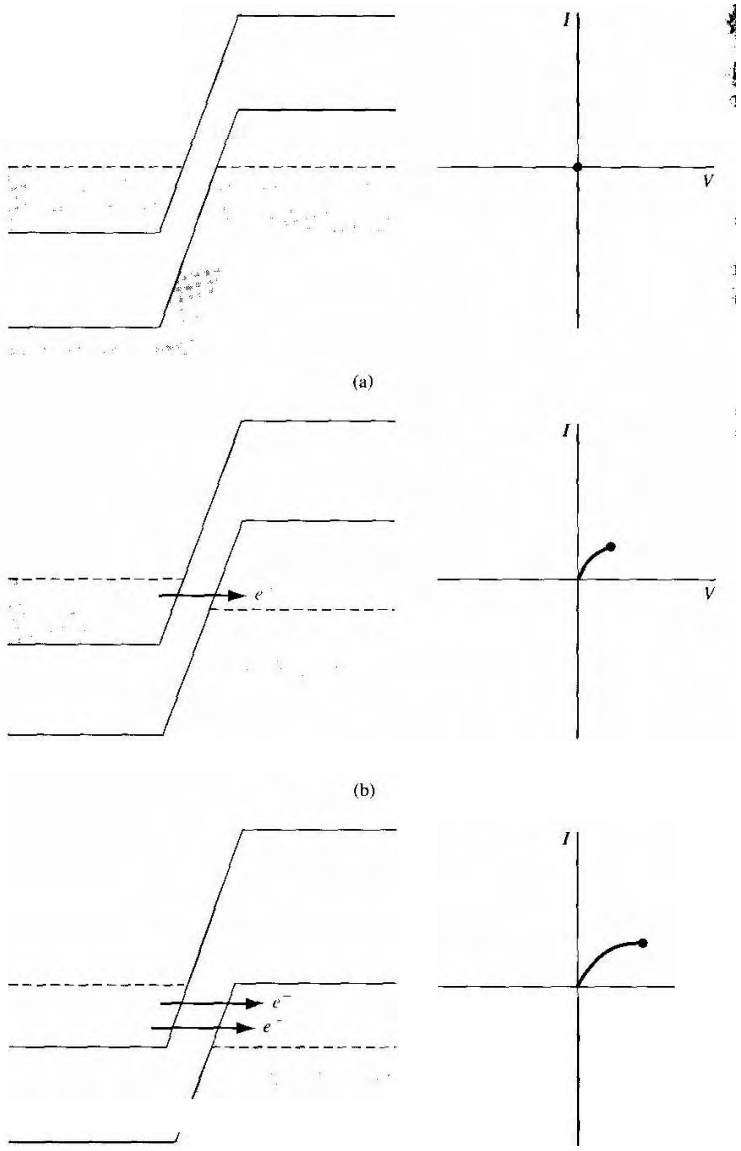


Figure 8.31 | Simplified enetgy-band diagrams and *I–V* characteristics of the tunnel diode at (a) zero bias; (b) a slight forward bias; (c) a forward bias producing maximum tunneling current.

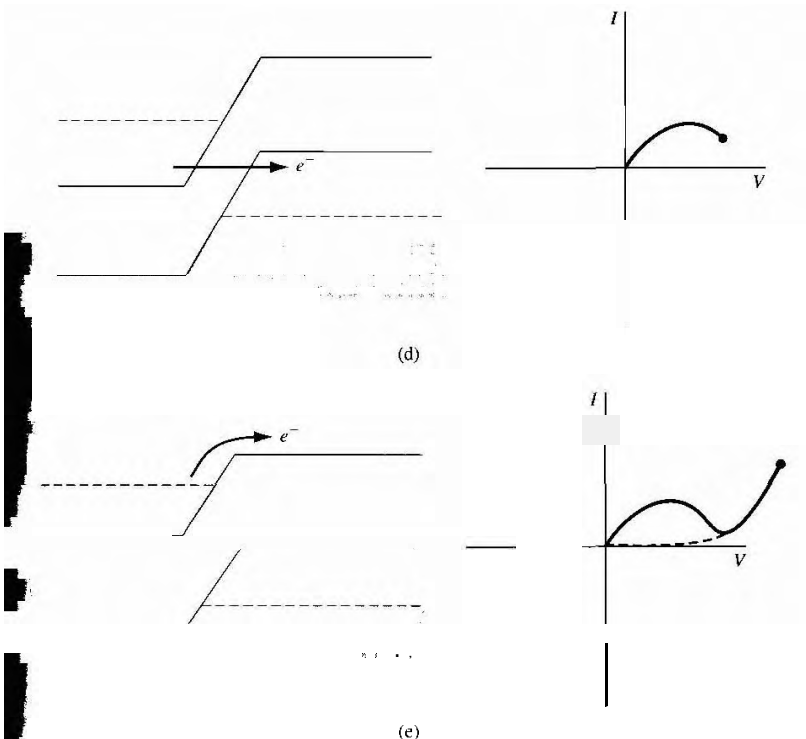


Figure 8.31 *(concluded)* (d) A higher forward bias showing less tunneling current; (e) a forward bias for which the diffusion current dominates.

Figure 8.31a shows the energy-band diagram at zero bias, which produces zero current on the *I–V* diagram. If we assume, for simplicity, that we are near 0 K, then all energy states are filled below E_F on both sides of the junction.

Figure 8.31b shows the situation when a small forward-bias voltage is applied to the junction. Electrons in the conduction band of the n region are directly opposite to empty states in the valence band of the p region. There is a finite probability that some of these electrons will tunnel directly into the empty states, producing a forward-bias tunneling current as shown. With a slightly larger forward-bias voltage, as in Figure 8.31c, the maximum number of electrons in the n region will be opposite the maximum number of empty states in the p region; this will produce a maximum tunneling current.

As the forward-bias voltage continues to increase, the number of electrons on the n side directly opposite empty states on the p side decreases, as in Figure 8.31d, and the tunneling current will decrease. In Figure 8.31e, there are no electrons on the

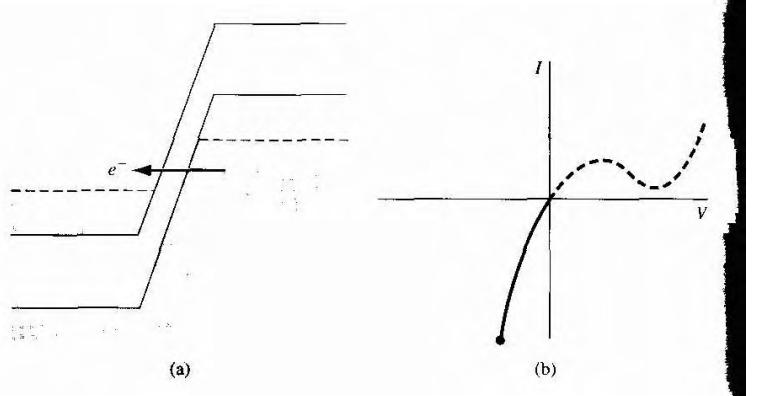


Figure 8.32 | (a) Simplified energy-band diagram of a tunnel diode with a reverse-bias voltage: (b) *I–V* characteristic of a tunnel diode with a reverse-bias voltage.

n side directly opposite available empty states on the p side. For this forward-bias voltage, the tunneling current will be zero and the normal ideal diffusion current will exist in the device as shown in the *I–V* characteristics.

The portion of the curve showing a decrease in current with an increase in voltage is the region of differential negative resistance. The range of voltage and current for this region is quite small; thus, any power generated from an oscillator using this negative resistance property would also be fairly small.

A simplified energy-band diagram of the tunnel diode with an applied reverse-bias voltage is shown in Figure 8.32a. Electrons in the valence band on the p side are directly opposite empty states in the conduction band on the n side, so electrons can now tunnel directly from the p region into the n region, resulting in a large reverse-bias tunneling current. This tunneling current will exist for any reverse-bias voltage. The reverse-bias current will increase monotonically and rapidly with reverse-bias voltage as shown in Figure 8.32b.

8.7 | SUMMARY

- When a forward-bias voltage is applied across a pn junction (p region positive with respect to the n region), the potential barrier is lowered so that holes from the p region and electrons from the n region can Row across the junction.
- The boundary conditions relating the minority carrier hole concentration in then region at the space charge edge and the minority carrier electron concentration in the p region at the space charge edge were derived.
- The holes that are injected into the n region and the electrons that are injected into the p region now become excess minority carriers. The behavior of the excess minority carrier is described bv the ambipolar transoon equation develooed and described in Chapter 6. Solving the ambipolar transpon equation and using the boundary conditions. the steady-state minority carrier hole and electron concentrations in the n region and p region, respectively, were derived.

- Gradients exist in the minority carrier hole and electron concentrations so that minority carrier diffusion currents exist in the pn junction. These diffusion currents yield the ideal current–voltage relationship of the pn junction diode.
- The small-signal equivalent circuit of the pn junction diode was developed. The twu parameters of interest are the diffusion resistance and the diffusion capacitance.
- Excess carriers are generated in the space charge region of a reverse-biased pn junction. These carriers are swept out by the electric field and create the reverse-bias generation current that is another component of the reverse-bins diode current. Excess carriers recombine in the space charge region of a forward-biased pn junction. This recombination pmcess creates the forward-bias recombination current that is another component of the forward-bias diode current.
- Avalanche hreakdown occurs when a sufficiently large reverse-bias voltage is applied to the pn junction. A large reverse-hias current may then be induced in the pn junction. The breakdown voltage as a function of the doping levels in the pn junction was derived. In a one-sided pn junction, the breakdown voltage is a function of the doping concentration in the low-doped region.
- When a pn junction is switched from forward bias to reverse bias, the stored excess minority carrier charge must be removed from the junction. The time required to remove this charge is called the storage time and is a limiting factor in the switching speed of a diode.

GLOSSARY OF IMPORTANT TERMS

avalanche breakdown The process whereby a large reverse-bias pn junction current is created due to the generation of electron–hole pairs by the collision of electrons and/or holes with atomic electrons within the space charge region.

carrier injection The flow of carriers across the space charge region of a pn junction when a voltage is applied.

critical electric field The peak electric field in the space charge region at breakdown.

diffusion capacitance The capacitance of a forward-biased pn junction due to minority carrier storage effects.

diffusion conductance The ratio of a low-frequency, small-signal sinusoidal current to voltage in a forward-biased pn junction.

diffusion resistance The inverse of diffusion conductance.

forward bias The condition in which a positive voltage is applied to the p region with respect to then region of a pn junction so that the potential barrier between the two regions is lowered below the thermal-equilibrium value.

generation current The reverse-bias pn junction current produced by thc thermal generation of electron-hole pairs within the space charge region.

"long" diode A pn junction diode in which both the neutral p and n regions are long compared with the respective minority carrier diffusion lengths.

recombination current The forward-bias pn junction current produced as a result of the How of electrons and holes that recombine within the space charge region.

reverse saturation current The ideal reverse-bias current in a pn junction.

"short" diode Apn junction diode in which at least one of the neutral p o r n regions is short compared to the respective minority carrier diffusion length.

storage time The time required for the excess minority carrier concentrations at thr space charge edge to go from their steady-state values to zero when the diode is switched forward to reverse bias.

CHECKPOINT

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

- Describe the mechanism of charge flow across the space charge region of a pn junction when a forward-bias voltage is applied.
- State the boundary conditions for the minority carrier concentrations at the edge of the space charge region.
- Derive the expressions for the steady-state minority carrier concentrations in the pn junction.
- Derive the ideal current–voltage relationship for a pn junction diode.
- Describe the characteristics of a "short" diode.
- Describe what is meant by diffusion resistance and diffusion capacitance.
- Describe generation and recombination currents in a pn junction.
- Describe the avalanche breakdown mechanism in a pn junction.
- Describe the turn-oil transient response in a pn junction.

REVIEW QUESTIONS

1. Why does the potential barrier decrease in a forward-biased pn junction?
2. Write the boundary conditions for the excess minority carriers in a pn junction (a) under forward bias and (b) under reverse bias.
3. Sketch the steady-state minority carrier concentrations in a forward-biased pn junction.
4. Explain the procedure that is used in deriving the ideal current–voltage relationship in a pn junction diode.
5. Sketch the electron and hole currents through a forward-biased pn junction diode.
6. What is meant by a "short" diode?
7. (a) Explain the physical mechanism of diffusion capacitance. (b) What is diffusion resistance?
8. Explain the physical mechanism of the (a) generation current and (b) recombination current.
9. Why docs the breakdown voltage of a pn junction decrease as the doping concentration increases'?
10. Explain what is meant by storage time.

PROBLEMS

Section 8.1 pn Junction Current

8.1 (*a*) Consider an ideal pn junction diode at $T = 300$ K operating in the forward-bias region. Calculate the change in diode voltage that will cause a factor of 10 increase in current. (*b*) Repeat part (*a*) for a factor of 100 increase in current.

8.2 Calculate the applied reverse-bias voltage at which the ideal reverse current in a pn junction diode at T = 300 K reaches 90 percent of its reverse saturation current value.

8.3 An ideal silicon pn junction at T = 300 K is under forward bias. The minority carrier lifetimes are $\tau_{n0} = 10^{-6}$ s and $\tau_{p0} = 10^{-7}$ s. The doping concentration in the n region is $N_d = 10^{16}$ cm^{-3}. Plot the ratio of hole current to the total current crossing the space charge region as the p-region doping concentration varies over the range $10^{15} \leq N_a \leq 10^{18}$ cm^{-3}. (Use a log scale for the doping concentrations.)

A silicon pn junction diode is to be designed to operate at T = 300 K such that the diode current is I = 10 mA at a diode voltage of $V_D = 0.65$ V. The ratio of electron current to total current is to be 0.10 and the maximum current density is to be no more than 20 A/cm^2. Use the semiconductor parameters given in Example 8.2.

8.5 For a silicon pn junction at T = 300 K, assume $\tau_{p0} = 0.1\tau_{n0}$ and $\mu_n = 2.4\,\mu_p$. The ratio of electron current crossing the depletion region to the total current is defined as the electron injection efficiency. Determine the expression for the electron injection efficiency as a function of (a) N_d/N_a and (b) the ratio of n-type conductivity to p-type conductivity.

8.6 Consider a p$^+$n silicon diode at T = 300 K with doping concentrations of $N_a = 10^{18}$ cm^{-3} and $N_d = 10^{16}$ cm^{-3}. The minority carrier hole diffusion coefficient is $D_p = 12$ cm^2/s and the minority carrier hole lifetime is $\tau_{p0} = 10^{-7}$ s. The cross-sectional area is $A = 10^{-4}$ cm^2. Calculate the reverse saturation current and the diode current at a forward-bias voltage of 0.50 V.

Consider an ideal silicon pn junction diode with the following parameters: $\tau_{n0} = \tau_{p0} = 0.1 \times 10^{-6}$ s, $D_n = 25$ cm^2/s, $D_p = 10$ cm^2/s. What must be the ratio of N_a/N_d so that 95 percent of the current in the depletion region is carried by electrons?

A silicon pn junction with a cross-sectional area of 10^{-4} cm^2 has the following properties at T = 300 K:

n region	p region
$N_d = 10^{17}$ cm^{-3}	$N_a = 5 \times 10^{15}$ cm^{-3}
$\tau_{p0} = 10^{-7}$ s	$\tau_{n0} = 10^{-6}$ s
$\mu_n = 850$ cm^2/V-s	$\mu_n = 1250$ cm^2/V-s
$\mu_p = 320$ cm^2/V-s	$\mu_p = 420$ cm^2/V-s

(a) Sketch the thermal equilibrium energy-band diagram of the pn junction, including the values of the Fermi level with respect to the intrinsic level on each side of the junction. (b) Calculate the reverse saturation current I_s and determine the forward-bias current I at a forward-bias voltage of 0.5 V. (c) Determine the ratio of hole current to total current at the space charge edge x_n.

A germanium p$^+$n diode at T = 300 K has the following parameters: $N_a = 10^{18}$ cm^{-3}, $N_d = 10^{16}$ cm^{-3}, $D_p = 49$ cm^2/s, $D_n = 100$ cm^2/s, $\tau_{p0} = \tau_{n0} = 5\,\mu$s, and $A = 10^{-4}$ cm^2. Determine the diode current for (a) a forward-bias voltage of 0.2 V and (b) a reverse-bias voltage of 0.2 V.

8.10 An n$^+$p silicon diode at T = 300 K has the following parameters: $N_d = 10^{18}$ cm^{-3}, $N_a = 10^{16}$ cm^{-3}, $D_n = 25$ cm^2/s, $D_p = 10$ cm^2/s, $\tau_{n0} = \tau_{p0} = 1\,\mu$s, and $A = 10^{-4}$ cm^2. Determine the diode current for (a) a forward-bias voltage of 0.5 V and (b) a reverse-bias voltage of 0.5 V.

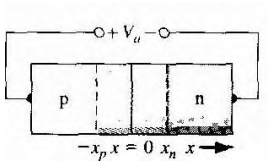


Figure 8.33 | Figure for Problem 8.11.

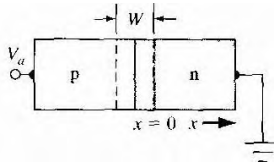


Figure 8.34 | Figure for Problem 8.12.

8.11 A silicon step junction has uniform impurity doping concentrations of $N_a = 5 \times 10^{15}$ cm^{-3} and $N_d = 1 \times 10^{15}$ cm^{-3}, and a cross-sectional area of $A = 10^{-4}$ cm^2. Let $\tau_{n0} = 0.4\ \mu$s and $\tau_{p0} = 0.1\ \mu$s. Consider the geometry in Figure 8.33. Calculate (*a*) the ideal reverse saturation current due to holes, (*b*) the ideal reverse saturation current due to electrons, (*c*) the hole concentration at x_n if $V_a = \frac{1}{2}V_{bi}$, and (*d*) the electron current at $x = x_n + \frac{1}{2}L_p$ for $V_a = \frac{1}{2}V_{bi}$

8.12 Consider the ideal long silicon pn junction shown in Figure 8.34. $T = 300$ K. The n region is doped with 10^{17} donor atoms per cm^3 and the p region is doped with 5×10^{16} acceptor atoms per cm^3. The minority carrier lifetimes are $\tau_{n0} = 0.05\ \mu$s and $\tau_{p0} = 0.01\ \mu$s. The minority carrier diffusion coefficients are $D_n = 23$ cm^2/s and $D_p = 8$ cm^2/s. The forward-bias voltage is $V_a = 0.610$ V. Calculate (a) the excess hole concentration as a function of x for $x \geq 0$, (b) the hole diffusion current density at $x = 3 \times 10^{-4}$ cm. and (r) the electron current density at $x = 3 \times 10^{-4}$ cm.

8.13 The limit of low injection is normally defined to he when the minority carrier concentration at the edge of the space charge region in the low-doped region becomes equal to one-tenth the majority carrier concentration in this region. Determine the value of the forward-bias voltage at which the limit of low injection is reached for the diode described in (*a*) Problem 8.9 and (*b*) Problem 8.10.

8.14 The cross-sectional area of a silicon pn junction is 10^{-3} cm^2. The temperature of the diode is $T = 300$ K, and the doping concentrations are $N_d = 10^{16}$ cm^{-3} and $N_a = 8 \times 10^{15}$ cm^{-3}. Assume minority carrier lifetimes of $\tau_{n0} = 10^{-6}$ s and $\tau_{p0} = 10^{-7}$ s. Calculate the total number of excess electrons in the p region and the total number of excess hales in the n region for (*a*) $V_a = 0.3$ V, (*h*) $V_a = 0.4$ V, and (*c*) $V_a = 0.5$ V.



8.15 Consider two ideal pn junctions at $T = 300$ K. having exactly the same electrical and physical parameters except for the bandgap energy of the semiconductor materials. The tint pn junction has a bandgap energy of 0.525 eV and a forward-bias current of 10 mA with $V_a = 0.255$ V. For the second pn junction, "design" the bandgap energy so that a forward-bias voltage of $V_a = 0.32$ V will produce a current of 10 μA.

8.16 The reverse-bias saturation current is a function of temperature. (*a*) Assuming that I_s varies with temperature only from the intrinsic carrier concentration, show that we can write $I_s = CT^3 \exp(-E_g/kT)$ where C is a constant and a function only of the diode parameters. (*b*) Determine the increase in I_s as the temperature increases from $T = 300$ K to $T = 400$ K for a (*i*) germanium diode and (*ii*) silicon diode.



8.17 Assume that the mobilities, diffusion coefficients, and minority carrier lifetime parameters are independent of temperature (use the $T = 300$ K values). Assume that $\tau_{n0} = 10^{-6}$ s, $\tau_{p0} = 10^{-7}$ s, $N_d = 5 \times 10^{15}$ cm^{-3}, and $N_a = 5 \times 10^{16}$ cm^{-3}. Plot the ideal reverse saturation current density from $T = 200$ K to $T = 500$ K for (*a*) silicon,

(b) germanium, and *(c)* gallium arsenide ideal pn junctions. (Use a log scale for the current density.)

8.18 An ideal uniformly doped silicon pn junction diode has a cross-sectional area of 10^{-4} cm^2. The p region is doped with 5×10^{18} acceptor atoms per cm^3 and the n region is doped with 10^{15} donor atoms per cm^3. Assume that the following parameter values are independent of temperature: $E_g = 1.10$ eV, $\tau_{n0} = \tau_{p0} = 10^{-7}$ s. $D_n = 25$ cm^2/s, $D_p = 10$ cm^2/s, $N_c = 2.8 \times 10^{19}$ cm^{-3}, and $N_v = 1.04 \times 10^{19}$ cm^{-3}. The ratio of the forward to reverse current is to be no less than 10^4 with forward- and reverse-bias voltages of 0.50 V. Also, the reverse saturation current is to be no larger than 1 μA. What is the maximum temperature at which the diode will meet these specifications?

8.19 A p^+n silicon diode is fabricated with a narrow n region as shown in Figure 8.10, in which $W_n < L_n$. Assume the boundary condition of $p_n = p_{n0}$ at $x = x_n + W_n$ *(a)* Derive the expression for the excess hole concentration $\delta p_n(x)$ as given by Equation (8.27). *(b)* Using the results of part (a), show that the current density in the diode is given by

$$J = \frac{eD_p p_{n0}}{L_p} \coth\left(\frac{W_n}{L_p}\right)\left[\exp\left(\frac{eV}{kT}\right) - 1\right]$$

8.20 A silicon diode can be used to measure temperature by operating the diode at a fixed forward-bias current. The forward-bias voltage is then a function of temperature. At T = 300 K, the diode voltage is found to be 0.60 V. Determine the diode voltage at *(a)* T = 310 K and *(b)* T = 320 K.

8.21 A forward-biased silicon diode is to be used as a temperature sensor. The diode is forward biased with a constant current source and V_a is measured as a function of temperature. *(a)* Derive an expression for $V_a(T)$ assuming that D/L for electrons and holes and E_g are independent of temperature. *(b)* If the diode is biased at $I_D = 0.1$ mA and if $I_s = 10^{-15}$ A at $T = 300$ K, plot V_a versus T for 20°C < T < 200°C. *(c)* Repeat part *(b)* if $I_D = 1$ mA. (d) Determine any changes in the results of parts *(a)* through *(c)* if the change in bandgap energy with temperature is taken into account.

Section 8.2 Small-Signal Model of the pn Junction

8.22 Calculate the small-signal ac admittance of a pn junction biased at $V_a = 0.72$ V and $I_{DQ} = 2.0$ mA. Assume the minority carrier lifetime is 1 μs in both the n and p regions. T = 300 K.

8.23 Consider a p^+n silicon diode at $T = 300$ K. The diode is forward biased at a current of 1 mA. The hole lifetime in the n region is 10^{-7} s. Neglecting the depletion capacitance, calculate the diode impedance at frequencies of 10 kHz, 100 kHz, 1 MHz, and 10 MHz.

8.24 Consider a silicon pn junction with parameters as described in Problem 8.8. *(a)* Calculate and plot the depletion capacitance and diffusion capacitance over the voltage range $-10 \le V_a \le 0.75$ V. *(b)* Determine the voltage at which the two capacitances are equal.

8.25 Consider a p^+n silicon diode at $T = 300$ K. The slope of the diffusion capacitance versus forward-bias current is 2.5×10^{-6} F/A. Determine the hole lifetime and the diffusion capacitance at a forward-bias current of 1 mA.

8.26 A one-sided n^+p silicon diode at $T = 300$ K with a cross-sectional area of 10^{-3} cm^2 is operated under forward bias. The doping levels are $N_d = 10^{18}$ cm^{-3} and

$N_a = 10^{16}$ cm^{-3}, and the minority carrier parameters are $\tau_{p0} = 10^{-8}$ s. $\tau_{n0} = 10^{-7}$ s, $D_p = 10$ cm^2/s, and $D_n = 25$ cm^2/s. The maximum diffusion capacitance is to be 1 n[illegible] Determine *(a)* the maximum current through the diode, *(b)* the maximum forward-bias voltage, and *(c)* the diffusion resistance.

8.27 A silicon pn junction diode at T = 300 K has a cross-sectional area of 10^{-2} cm^2. length of the p region is 0.2 cm and the length of the n region is 0.1 cm. The doping concentrations are $N_d = 10^{15}$ cm^{-3} and $N_a = 10^{16}$ cm^{-3}. Determine *(a)* approximately the series resistance of the diode and (b) the current through the diode that will produce a 0.1 V drop across this series resistance.

8.28 We want to consider the effect of a series resistance on the forward bias voltage required to achieve a particular diode current. (a) Assume the reverse saturation current in a diode is $I_s = 10^{-10}$ A at $T = 300$ K. The resistivity of the n region is 0.2 Ω-cm and the resistivity of the p region is 0.1 R-cm. Assume the length of each neutral region is 10^{-2} cm and the cross-sectional area is 2×10^{-5} cm^2. Determine the required applied voltage to achieve a current of (i) 1 mA and *(ii)* 10 mA. *(b)* Repeat part **(a)** neglecting the series resistance.

8.29 The minimum mall-signal diffusion resistance of an ideal forward-biased silicon pn junction diode at $T = 300$ K is to be $r_d = 48\ \Omega$. The reverse saturation current is $I_s = 2 \times 10^{-11}$ A. Calculate the maximum applied forward-bias voltage that can be applied to meet this specification.

8.30 *(a)* An ideal silicon pn junction diode at $T = 300$ K is forward biased at $V_a = +20$ mV. The reverse-saturation current is $I_s = 10^{-13}$ A. Calculate the small-signal diffusion resistance. *(b)* Repeat pan *(a)* for an applied reverse-bias voltage of $V_a = -20$ mV.

Section 8.3 Generation-Recombination Currents

8.31 Consider a reverse-biased gallium arsenide pn junction at T = 300 K. Assume that a reverse-bias voltage, $V_R = 5$ V, is applied. Assume parameter values of: $N_a = N_d = 10^{16}$ cm^{-3}, $D_p = 6$ cm^2/s, $D_n = 200$ cm^2/s, and $\tau_{p0} = \tau_{n0} = \tau_0 = 10^{-8}$ s. Calculate the ideal reverse saturation current density and the reverse-biased generation current density. How does the relative value of these two currents compare to those of the silicon pn junction?

***8.32** (a) Consider Example 8.7. Assume that all parameters except n_i are independent of temperature. Determine the temperature at which J_s and J_{gen} will be equal. What are the values of J_s and J_{gen} at this temperature? *(b)* Using the results of Example 8.7, calculate the forward-bias voltage at which the ideal diffusion current is equal to the recombination current.

8.33 Consider a GaAs pn diode at $T = 300$ K with $N_a = N_d = 10^{17}$ cm^{-3} and with a cross-sectional area of 10^{-3} cm^2. The minority carrier mobilities are $\mu_n = 3000$ cm^2/V-s and $\mu_p = 200$ cm^2/V-s. The lifetimes are $\tau_{p0} = \tau_{n0} = \tau_0 = 10^{-8}$ s. As a first approximation, assume the electron-hole generation and recombination rates are constant across the space charge region. *(a)* Calculate the total diode current at a reverse-bias voltage of 5 V and at forward-bias voltages of 0.3 V and 0.5 V. *(b)* Compare the results of pan *(a)* to an ideal diode at the same applied voltages.

8.34 Consider the pn junction diode described in Problem 8.33. Plot the diode recombination current and the ideal diode current (on a log scale) versus forward bias voltage over the range $0.1 \le V_a \le 1.0$ V.

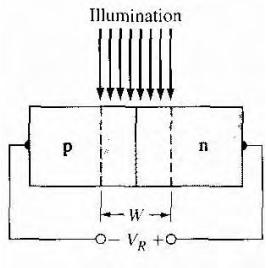


Figure 8.35 | Figure for Problem 8.38 and 8.39.

8.35 A silicon pn junction diode at T = 300 K has the following parameters: $N_a = N_d = 10^{16}$ cm^{-3}, $\tau_{p0} = \tau_{n0} = \tau_0 = 5 \times 10^{-7}$ s, $D_p = 10$ cm^2/s, $D_n = 25$ cm^2/s, and a cross-sectional area of 10^{-4} cm^2. Plot the diode recombination current and the ideal diode current (on a log scale) versus forward bias voltage over the range $0.1 \leq V_a \leq 0.6$ V.

8.36 Consider a GaAs pn diode at T = 300 K with $N_a = N_d = 10^{17}$ cm^{-3} and with a cross-sectional area of 5×10^{-3} cm^2. The minority carrier mobilities are $\mu_n = 3500$ cm^2/V-s and $\mu_p = 220$ cm^2/V-s. The electron-hole lifetimes are $\tau_{n0} = \tau_{p0} = \tau_0 = 10^{-8}$ s. Plot the diode forward-bias current including recombination current between diode voltages of $0.1 \leq V_D \leq 1.0$ V. Compare this plot to that for an ideal diode.

***8.37** Starting with Equation (8.83) and using the suitable approximations, show that the maximum recombination rate in a forward-biased pn junction is given by Equation (8.91i.

8.38 Consider. as shown in Figure 8.35, a uniformly doped silicon pn junction at $T = 300$ K with impurity doping concentrations of $N_a = N_d = 5 \times 10^{15}$ cm^{-3} and minority carrier lifetimes of $\tau_{n0} = \tau_{p0} = \tau_0 = 10^{-7}$ s. A reverse-bias voltageof $V_R = 10$ V is applied. A light source is incident only on the space charge region. producing an excess carrier generation rate of $g' = 4 \times 10^{19}$ cm^{-3} s^{-1}. Calculate the generation current density.

8.39 Along silicon pn junction diode has the following parameters: $N_d = 10^{18}$ cm^{-3}, $N_a = 3 \times 10^{17}$ cm^{-3}, $\tau_{n0} = \tau_{p0} = \tau_0 = 10^{-7}$ s, $D_n = 18$ cm^2/s, and $D_p = 6$ cm^2/s. A light source is incident on the space charge region such as shown in Figure 8.35, producing a generation current density of $J_G = 25$ mA/cm^2. The diode is open circuited. The generation current density forward biases the junction, inducing a forward-bias current in the opposite direction to the generation current. A steady-state condition is reached when the generation current density and forward-bias current density are equal in magnitude. What is the induced forward-bias voltage at this steady-state condition?

Section 8.4 Junction Breakdown

8.40 The critical electric field for breakdown in silicon is approximately $E_{crit} = 4 \times 10^5$ V/cm. Determine the maximum n-type doping concentration in an abrupt p^+n junction such that the breakdown voltage is 30 V.

8.41 Design an abrupt silicon p^+n junction diode that has a reverse breakdown voltage of 120 V and has a forward-bias current of 2 mA at V = 0.65 V. Assume that $\tau_{p0} = 10^{-7}$ s, and find μ_p from Figure 5.3.

8.42 Consider an abrupt n+p GaAs junction with a p-type doping concentration of $N_a = 10^{16}$ cm^{-3}. Determine the breakdown voltage.

8.43 A symmetrically doped silicon pn junction has doping concentrations of $N_a = N_d = 5 \times 10^{16}$ cm^{-3}. If the peakelectric field in the junction at breakdown is E = 4 x 10^5 Vlcm. determine thc breakdown voltage of this junction.

8.44 An abrupt silicon p^+n junction has an n-region doping concentration of $N_d = 5 \times 10^{15}$ cm^{-3}. What must be the minimum n-region width such that avalanche breakdown occurs before the depletion region reaches an ohmic contact (punchthrough)?

8.45 A silicon pn junction diode is doped with $N_a = N_d = 10^{18}$ cm^{-3}. Zener breakdown occurs when the peak electric field reaches 10^6 V/cm. Determine the reverse-bias breakdown voltage.

8.46 Adiode will very often have the doping profile shown in Figure 7.19. which is known as an n^+pp^+ diode. Under reverse bias, the depletion region must remain within the p region to avoid premature breakdown. Assume the p region doping is 10^{15} cm^{-3}. Determine the reverse-bias voltage such that the depletion region remains within the p region and does not reach breakdown if the p region width is *(a)* 75 μm and *(b)* 150 μm. For each case, state whether the maximum depletion width or the breakdown voltage is reached first.

8.47 Consider a silicon pn junction at T = 300 K whose doping profile varies linearly from $N_a = 10^{18}$ c m ^{3} to $N_d = 10^{18}$ cm^{-3} over a distance of 2 μm. Estimate the breakdown voltage.

Section 8.5 Charge Storage and Diode Transients

8.48 *(a)* In switching a pn junction from forward to reverse bias, assume that the ratio of reverse current, I_R, to forward current, I_F, is 0.2. Determine the ratio of storage time to minority carrier lifetime, t_s/τ_{p0}. *(b)* Repeat part (a) if the ratio of I_R to I_F is 1.0.

8.49 A pn junction is switched from forward to reverse bias. We want to specify that $t_s = 0.2\tau_{p0}$. Determine the required ratio of I_R to I_F to achieve this requirement. In this case, determine t_2/τ_{p0}.

8.50 Consider a diode with a junction capacitance of 18 pF at zero bias and 4.2 pF at a reverse bias voltage of $V_R = 10$ V. The minority carrier lifetimes are 10^{-7} s. The diode is switched from a forward bias with a current of 2 mA to a reverse bias voltage of 10 V applied through a 10 kΩ resistur. Estimate the tum-off time.

Section 8.7 The Tunnel Diode

8.51 Consider a silicon pn junction at T = 300 K with doping concentration of $N_d = N_a = 5 \times 10^{19}$ cm^{-3}. Assuming the abrupt junction approximation is valid, determine the space charge width at a forward-bias voltage of V, = 0.40 V.

8.52 Sketch the energy-band diagram of an abrupt pn junction under zero bias in which the p region is degenerately doped and $E_C = E_F$ in the n region. Sketch the forward- and reverse-bias current–voltage characteristics. This diode is sometimes called a *back-*

Summary and Review

8.53 *(n)* Explain physically why the diffusion capacitance is not important in a reverse-biased pn junction. (*b*) Consider a silicon. a germanium, and gallium arsenide pn junction. If the total current density is the same in each diode under forward bids, discuss the expected relative values of electron and hole current densities.

***8.54** A silicon pn junction diode at $T = 300$ K is to be designed to have a reverse-bias breakdown voltage of at least 50 V and to handle a forward-bias current of $I_D = 100$ mA while still operating under low injection. The minority carrier diffusion coefficients and lifetimes are $D_n = 25$ cm^2/s, $D_p = 10$ cm^2/s, and $\tau_{n0} = \tau_{p0} = 5 \times 10^{-7}$ s. The diode is to be designed for minimum cross-sectional area.

***8.55** The donor and acceptor concentrations on either side of a silicon step junction are equal. *(a)* Derive an expression for the breakdown voltage in terms of the critical electric field and doping concentration. (*b*) If the breakdown voltage is to be $V_B = 50$ V, specify the range of allowed doping concentrations.

READING LIST

1. Dimitrijev, S. *Understanding Semiconductor* ***Devices.*** New York: Oxford University Press, 2000.
2. Kano, K. ***Semiconductor Devices.*** Upper Saddle River, NJ: Prentice Hall, 1998.
3. Muller, R. S., and T. I. Kamins. ***Device Electronics*** for *Integrated* ***Circuits.*** 2nd ed. New York: Wiley. 1986.
4. Neudeck, G. W. ***The PN Junction Diode.*** Vol. 2 of the ***Modular Series on*** *Solid State* ***Devices.*** 2nd ed. Reading, MA: Addison-Wesley, 1989.

*5. Ng, K. K. ***Complete Guide*** *to Semiconductor Devices.* New York: McGraw-Hill, 1995.

6. Pierret, R. F. ***Semiconductor Device*** *Fundamentals.* Reading, MA: Addison-Wesley Publishing Co., 1996.
7. Roulston, D. J. ***An*** *Introduction to* ***the Physics of*** *Semiconductor* ***Devices.*** New York: Oxford University Press, 1999.
8. Shur, M. *Introduction to Electronic* ***Devices.*** New York: John Wiley and Sons, Inc., 1996.

*9. Shur, M. *Physics of Semiconductor Devices.* Englewood Cliffs, NJ: Prentice Hall, 1990.

10. Streetman, B. G., and S. Banerjee. *Solid State* ***Electronic Devices.*** 5th ed. Upper Saddle River, NJ: Prentice Hall, 2000.
11. Sze, S. M. ***Physics of*** *Semiconductor* ***Devices.*** 2nd ed. New York: John Wiley & Sons, 1981.
12. Sze, S. M. ***Semiconductor Devices: Physics*** *and* ***Technology.*** 2nd ed. New York: John Wiley and Sons, 2001.

*13. Wang. S. *Fundamentals of* ***Semiconductor*** *Theory and Device* ***Physics.*** Englewood Cliffs, NJ: Prentice Hall, 1989.

14. Yang, *E.* S. *Microelectronic* ***Devices.*** New York: McGraw-Hill, 1988.

CHAPTER 9

Metal–Semiconductor and Semiconductor Heterojunctions

PREVIEW

In the preceding two chapters, we considered the pn junction and assumed that the semiconductor material was the same throughout the entire 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 will 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 will examine the conditions that yield metal–semiconductor ohmic contacts..

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 setniconductor 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 will 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 will 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 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.

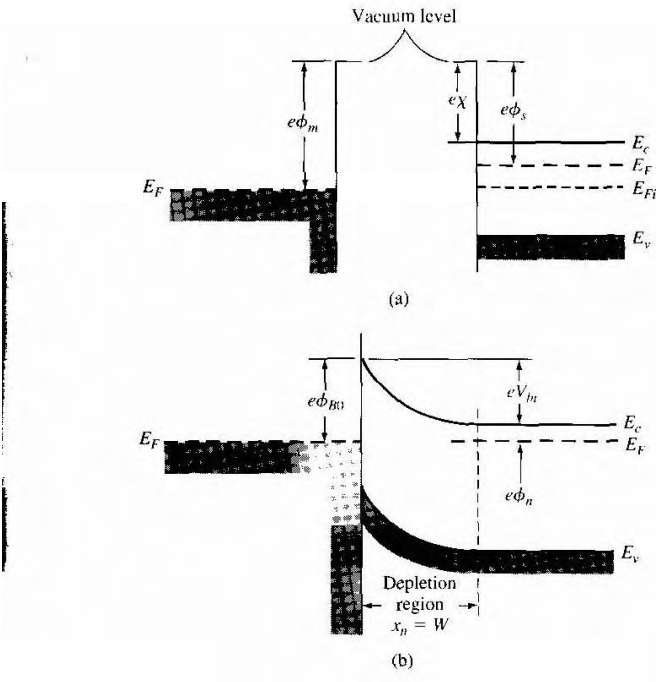


Figure 9.1 | (a) Energy-band diagram of a metal and semiconductor before contact; (b) ideal energy-band diagram of a metal-n-semiconductor junction for $\phi_m > \phi_s$.

Table **9.1** | Work functions of some elements

Element	Wnrk 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 92 | 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

This barrier is known as the *Schottky* burrier and is given. ideally, by

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

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

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

which makes V_{bi} a slight function of the semiconductor doping, as was 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 and 9.2b, where V_R is the magnitude of the reverse-bias voltage and V_a is the magnitude of the forward-bias voltage.

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 last chapter. Because of the 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 How 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-bins voltage V,.

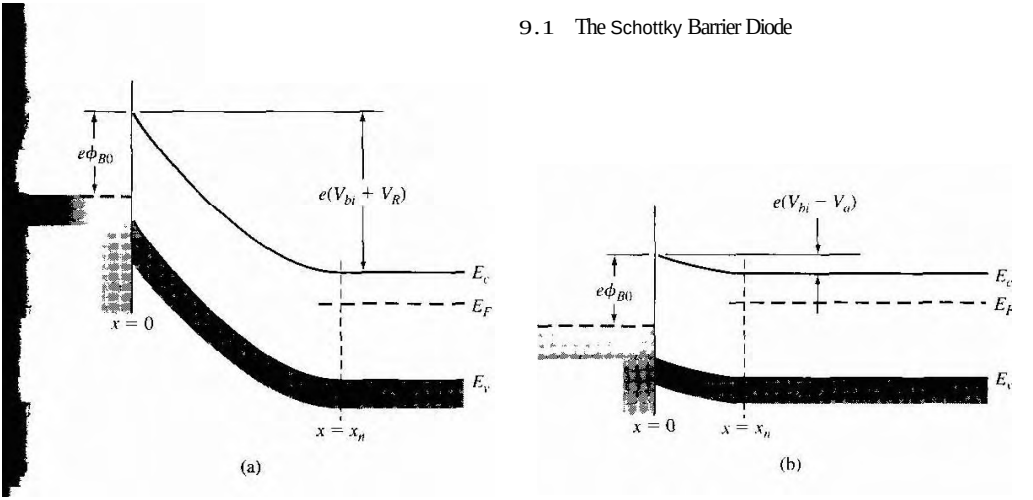


Figure 9.2 | Ideal energy-band diagram of a metal–semiconductor junction (a) under reverse bias and (b) under forward bias.

9.1.2 Ideal Junction Properties

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

$$\frac{dE}{dx} = \frac{\rho(x)}{\epsilon_s} \tag{9.3}$$

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

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

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 he found as

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

The electric field can then be written as

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

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-held 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 did for the pn junction. The result is identical to that of a one-sided p^+n junction. For the uniformly doped semiconductor, we have

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

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

EXAMPLE 9.1

Objective

To calculate the theoretical harrier height, built-in potential barrier, and maximum electric field in a meVal-semiconductor diode for zero applied bias.

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

■ **Solution**

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

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

where ϕ_{B0} is the ideal Schottky barrier height. We can calculate @„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

$$V_{bi} = \phi_{B0} - \phi_n = 0.54 - 0.206 = 0.33 \text{ V}$$

The space charge width at zero bias is

$$x_n = \left[\frac{2\epsilon_s V_{bi}}{eN_d}\right]^{1/2} = \left[\frac{2(11.7)(8.85 \times 10^{-14})(0.33)}{(1.6\times 10^{-19})(10^{16})}\right]^{1/2}$$

or

$$x_{,,} = 0.207 \times 10^{-4} \text{ cm}$$

Then the maximum electric field is

$$|E_{max}| = \frac{eN_d x_n}{\epsilon_s} = \frac{(1.6 \times 10^{-19})(10^{16})(0.207 \times 10^{-4})}{(11.7)(8.85 \times}$$

or finally

$$|E_{max}| = 3.2 \times 10^4 \text{ V/cm}$$

■ **Comment**

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

A junction capacitance can also be determined in the same way as we did 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 seniiconductor doping N_d. We can calculate the potential ϕ_n, and then determine the Schottky barrier ϕ_{B0} from Equation (9.2).

TEST YOUR UNDERSTANDING

E9.1 Consider an ideal chromium-to-n-type silicon Schottky diode at T = 300 K. Assume the semiconductor is doped at a cuncentration of $N_d = 3 \times 10^{\bullet\bullet}$ cm^{-3}. Determine the *(a)* ideal Schottky barrier height, *(b)* built-in potential harrier, *(c)* peak electric field with an applied reverse-bias voltage of $V_R = 5$ V, and *(d)* junction capacitance per unit area for $V_R = 5$ V. [Ans. (a) ϕ_{B0} (n) = 0.490 V, (b) ϕ_n = 0.253 V, (c) $|E_{max}|$ = 1 × 10^4 V/cm, C' (p) = 6.88 × 10^{-9} F/cm^2]

E9.2 Repeat E9.1 for an ideal palladium-to-n-type GaAs Schottky diode with the same impurity concentration. [Ans. (a) ϕ_{B0} (n) = 1.05 V, (b) ϕ_n = 0.619 V, (c) $|E_{max}|$ = 1 × 10^4 V/cm, C' (p) = 6.88 × 10^{-9} F/cm^2]

Objective EXAMPLE **9.2**

To calculate the semiconductor doping and Schottky barrier height from the silicon diode experimental data shown in Figure 9.3. 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

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

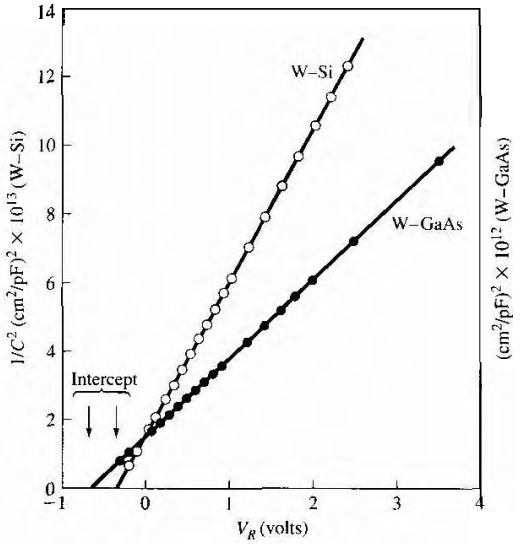


Figure 9.3 | $1/C^2$ versus V_R for W–Si and W–GaAs Schottky barrier diodes.
(From *Sze [14].*)

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,

■ Comment

The experimental value of 0.52 V can be compared with the ideal barrier height of ϕ_{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.

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.

9.1.3 Nonideal Effects on the Barrier Height

Several effects will alter the actual Schottky barrier height from the theoretical value given hy Equation (9.1). The first effect we will consider is the ***Schottky*** effect, or image-force-induced lowering of the potential barrier.

An electron in a dielectric at a distance x from the metal will create an electric field. The field lines must be perpendicular to the metal surface and will be the same as if an image charge, $+e$. is located at the same distance from the metal surface, but inside the metal. This image effect is shown in Figure 9.4a. The force on the electron, due to the coulomb attraction with the image charge, is

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

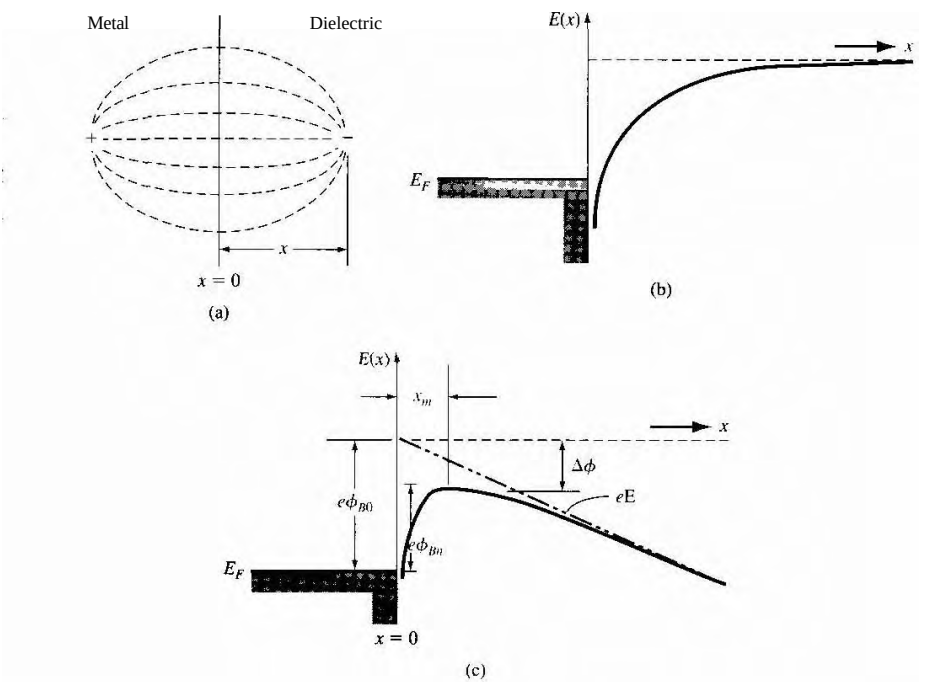


Figure 9.4 | (a) Image charge and electric field lines at a metal–dielectric interface. (b) Distortion of the potential barrier due to image forces with zero electric field and (c) with a constant electric field.

The potential can then be found as

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

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

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

$$-\phi(x) = \frac{-e}{16\pi\epsilon_s x} - \mathrm{E}x \tag{9.12}$$

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 lowerinpof the potential barrier is the Schottky effect, or image-force-induced lowering.

We can find the Schottky barrier lowering, **A@,** and the position of the maximum barrier, x,,,, from the condition that

$$\frac{d(e\phi(x))}{dx} = 0 \tag{9.13}$$

We find that

$$x_m = \sqrt{\frac{e}{16\pi\epsilon_s \mathrm{E}}} \tag{9.14}$$

and

$$\boxed{\Delta\phi = \sqrt{\frac{e\mathrm{E}}{4\pi\epsilon_s}}} \tag{9.15}$$

EXAMPLE 9.3 Objective

To 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 x 10⁴ V/cm.

Solution

The Schottky barrier lowering is given by Equation (9.15). which in this case yields

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

The position of the maximum barrier height is

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

or

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

■ **Comment**

Although the Schottky bamer 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 Schouky harrier diode.

TEST YOUR UNDERSTANDING

E9.3 Determine the Schottky barrier lowering and the position of the maximum barrier height for the junction described in E9.1. Use the value of the electric field found in this exercise. (Ans. $\Delta\phi = 0.0293$ eV, $x_m = 21$ Å)

E9.4 Repeat E9.3 for the junction described in E9.2. (Ans. $\Delta\phi = 0.0278$ V, $x_m = 19.8$ Å)

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 hanier 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

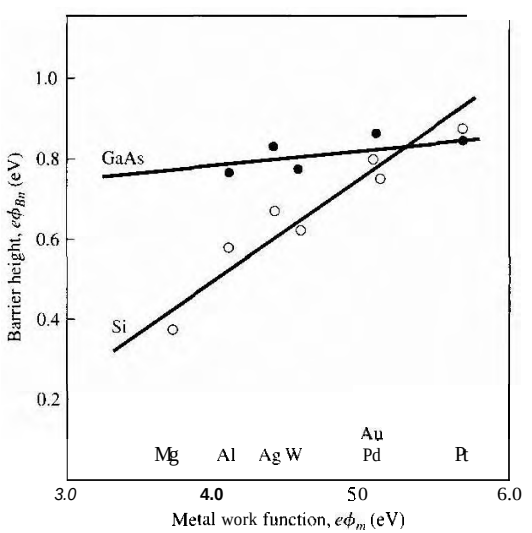


Figure 9.5 | Experimental banier heights as a function of metal work functions for GaAs and Si.
(From *Crowley and Sze [2].*)

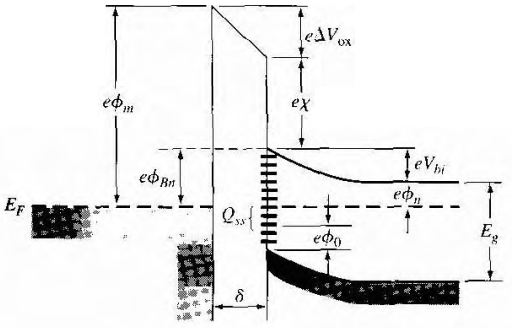


Figure 9.6 | Energy-band diagram of a metal–semiconductor junction with an interfacial layer and interface states.

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 will 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 will assume that all states below the surface potential ϕ_0 are donor states, which will be neutral if the state contains an electron and positively charged if the slate does not contain an electron. We will 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 will tend to contain electrons and will be negatively charged. We may assume that the surface state density is constant and equal to D_{it} states/cm^2-eV. The relation between the surface potential, surface state density, and other semiconductor parameters is found to be

$$(E_g - e\phi_0 - e\phi_{Bn}) = \frac{1}{eD_{it}}\sqrt{2e\epsilon_s N_d(\phi_{Bn} - \phi_n)} - \frac{\epsilon_i}{eD_{it}\delta}[\phi_m - (\chi + \phi_{Bn})] \tag{9.16}$$

We will 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

$$\phi_{Bn} = \frac{1}{e}(E_g - e\phi_0) \tag{9.17}$$

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 $D_{it}\delta \rightarrow 0$. 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.

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 arc derived by using the assumptions 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 shows two electron current density components. The current $J_{s\rightarrow m}$ is the electron current density due to the flow of electrons from the semiconductor into the metal, and the current $J_{m\rightarrow s}$ 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 How. The conventional current direction is opposite to electron flow.

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

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

where E_c' is the minimum energy required for thertnionic 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–Boltrmann approximation applies, we may write

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

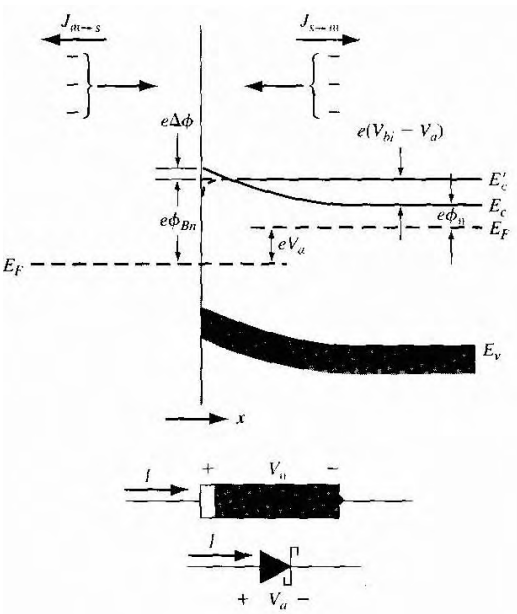


Figure 9.7 | Energy-band diagram of a forward-biased metal-semiconductor junction including the image lowering effect.

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

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

where

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

The parameter A* is called the effective Richardson constant for therrnionic 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

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

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

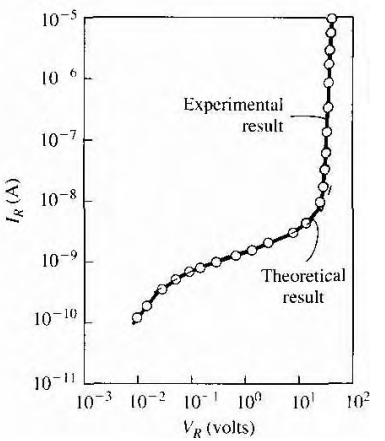


Figure 9.8 | Experimental and theoretical reverse-bias currents in a PtSi–Si diode. *(From SZE [14].)*

EXAMPLE 9.4

Objective

To calculate the effective Richardson constant from the *I–V* characteristics.

Consider the tungsten–silicon diode curve in Figure 9.9 and assume a harrier height of $\phi_{Bn} = 0.67$ V. From the figure, $J_{sT} \approx 6 \times 10^{-5}$ A/cm^2.

■ 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.

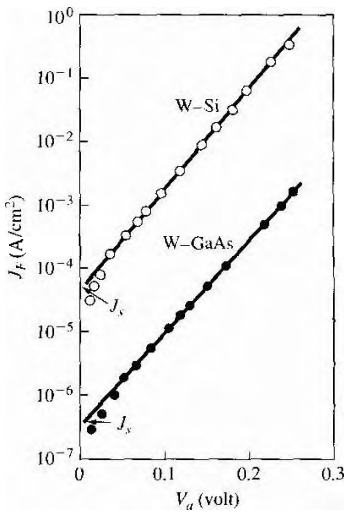


Figure 9.9 | Forward-bias current density J_F versus V_a for Wi–Si and W–GaAs diodes.
(From Sze [14].)

TEST YOUR UNDERSTANDING

E9.5 The Schottky barrier height of a silicon Schottky junction is $\phi_{Bn} = 0.59$ V, the effective Richardson constant is A' = 114 A/K^2-cm^2, and the cross-sectional area is $A = 10^{-4}$ cm^2. For T = 300 K, calculate (*a*) the ideal reverse-saturation current and (b) the diode current for $V_a = 0.30$ V. [Ans. (a) $I_s = 1.31 \times 10^{-7}$ A, (b) $I = 1.41$ mA]

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 2 orders of magnitude. This 2 order of magnitude difference will be reflected in the effective Richardson constant, assuming the harrier 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.

Objective

EXAMPLE 9.5

To calculate the 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

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

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

$$N_a = 10^{18}\ \text{cm}^{-3} \qquad N_d = 10^{16}\ \text{cm}^{-3}$$
$$D_{n} = 10\ \text{cm}^2/\text{s} \qquad D_{n} = 25\ \text{cm}^2/\text{s}$$
$$\tau_{po} = 10^{-7} \qquad \tau_{no} = 10^{-7}$$

We can then calculate the following parameters:

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

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.

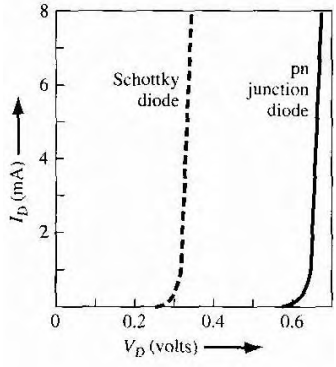


Figure 9.10 | Comparison of forward-bias *I–V* characteristics between a Schottky diode and a pn junction diode.

Recall that the reverse-bias current in a silicon pn junction diode is dominated by the generation current. A typical generation current density is approximately 10^{-7} A/cm^2. which is still 2 to 3 orders of magnitude less than the reverse-saturation current density of the Schottky bamer diode. A generation current also exists in the reverse-biased Schottky bamer 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.

EXAMPLE 9.6

Objective

To calculate the forward-bias voltage required to generate a forward-bias current density of 10 A/cm^2 in a Schottky barrier diode and a pn junction diode.

Consider diodes with the paramctrrs given in Example9.5.We can assume that thepn 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 thc two forward-bias voltages shows that the Schottky barrier diode has a turn-on voltage that ın this case, is approximately 0.37 V smaller than the tom-on voltage of the pn junction diode:

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 the next chapter, in what is referred to as a ***Schottky clamped transistor.***

TEST YOUR UNDERSTANDING

E9.6 (*a*) The reverse saturation currents of a pn junction and a Schottky diode arc 10^{-15} 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 pan (a) for forward bias currents of 1 mA.
[Ans. (*a*) 0.596 V, 0.298 V (*b*) 0.656 V, 0.358 V]

E9.7 A pn junction diode and a Schottky diode have equal cross-sectional areas and have forward-biased 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 second major difference between a Schottky barrier diode and a pn junctio diode is in the frequency response, or switching characteristics. In our discussion, w have considered the current in a Schottky diode as being due to the injection of major ity carriers over apotential barrier. The energy-band diagram of Figure 9.1. forexam ple, showed that there can he electrons in the metal directly adjacent to empty sta es i the semiconductor. If an electron from the valence hand of the semiconductor were t flow into the metal, this effect would bc equivalent to holes being injected into th semiconductor. This injection of holes would create excess minority carrier hole : i then 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 harrier diode, then, is a majority carrier device. This fact mean that there is no diffusion capacitance associated with a forward-hiased Schottky diode. The elimination of the diffusion capacitance makes the Schottky diode a higher-frequency device than the pnjunction diode. Also, when switching a Schottky diode from forward to reverse bias, there is no minority carrier stored charge to remove, as was the case in the pn junction diode. Since there is no minority camer 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.

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 will define a specific contact resistance that is used to characterize ohmic contacts.

9.2.1 Ideal Nonrectifying Barriers

We considered an ideal metal-to-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 will Row 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 harrier 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-hand diagram when a positive voltage is applied to the metal with respect to the semiconductor. Electrons can easily flow "downhill" 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 Row over the barrier from the metal into the semiconductor. This junction, then, is an ohmic contact.

Figure 9 1 3 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 will 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

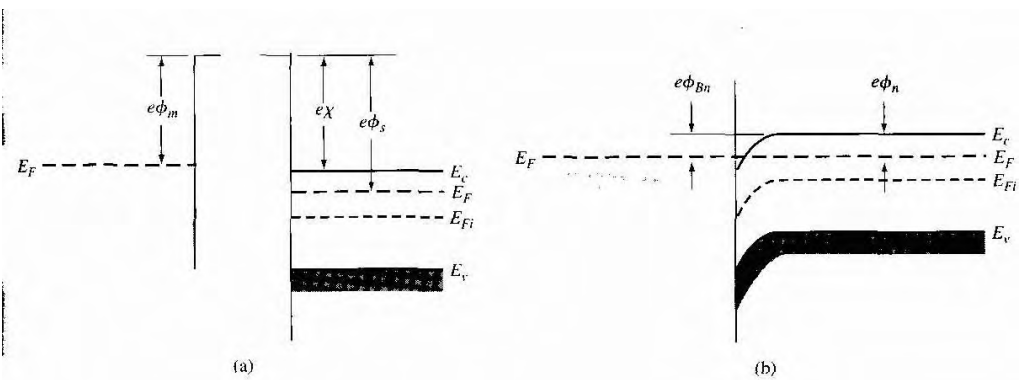


Figure 9.11 | Ideal energy-band diagram (a) before contact and (h) after contact for a metal–n-semiconductor junction for $\phi_m < \phi_s$

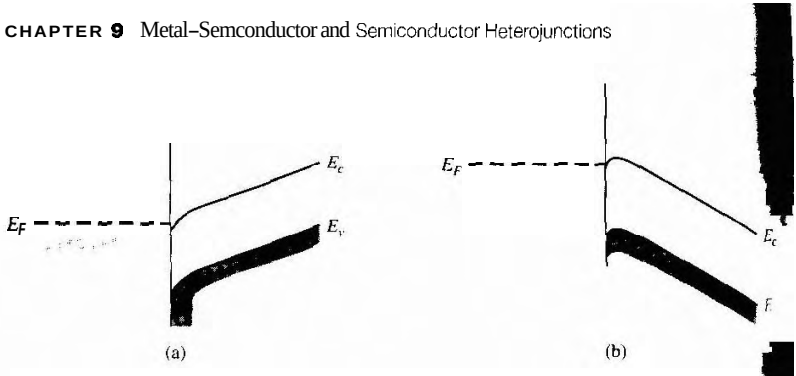


Figure 9.12 | Ideal energy-band diagram of a metal–n-semiconductor ohmic contact (a) with a positive voltage applied to the metal and (b) with a positive voltage applied to the serniconductur.

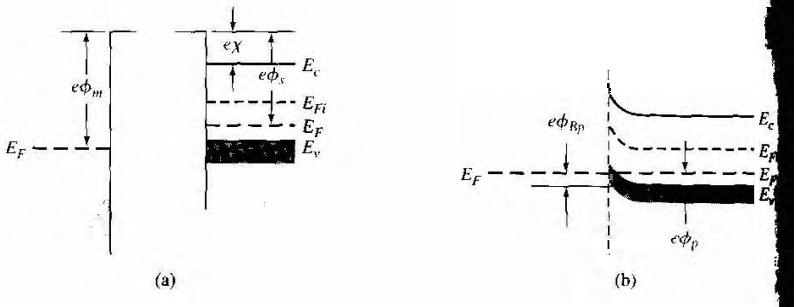


Figure **9.13** | Ideal energy-band diagram (a) before contact and (b) after contact for a metal–p-semiconductor junction for $\phi_m > \phi_s$

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 hands 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 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 he 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

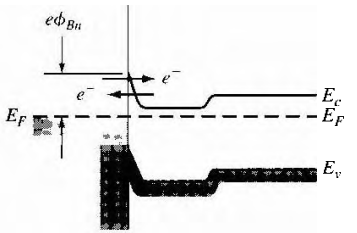


Figure 9.14 | Energy-band diagram of a heavily doped n-semiconductor-to-metal Junction.

region decreases as the doping concentration in the semiconductor increases; thus, as thedoping concentration increases, the probability of tunneling through the barrier in-,creases. Figure 9.14 shows ajunction in which the metal is in contact with a heavily doped n-type epitaxial layer.

EXAMPLE 9.7

Objective

%calculate the space charge width for a Schottky barrier on a heavily doped semiconductor.

Consider silicon at T = 300K 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 lor zero applied bias

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

$$x_n = 1.1 \times 10^{-6} \text{ cm} = 110 \text{ Å}$$

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.

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

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

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 was 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

$$R_c \propto \exp\left[\frac{+2\sqrt{\epsilon_s m_n^*}}{\hbar} \cdot \frac{\phi_{Bn}}{\sqrt{N_d}}\right] \tag{9.33}$$

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 $10^{\prime\prime}\ \text{cm}^{-3}$, 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-to-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

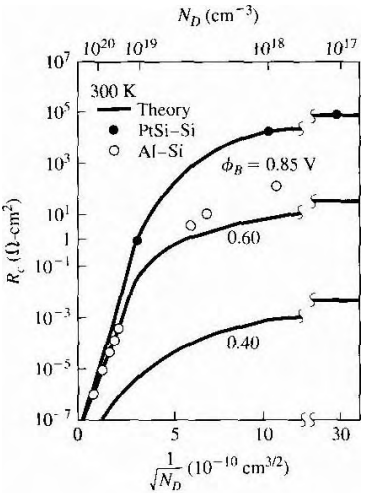


Figure 9.15 | Theoretical and experimental specific contact resistance as a function of doping.
(From Sze [14].)

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 was homogeneous throughout the entire 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–$Al_xGa_{1-x}As$ 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 $Al_xGa_{1-x}As$ 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 will consider only this case here. The other possibilities are called ***staggered*** and *broken gap* and are shown in Figures 9.16b and 9.16c.

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. Heterojuncrions with the same dopant type on either side of the junction are called *isotype*. We can form nN and 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

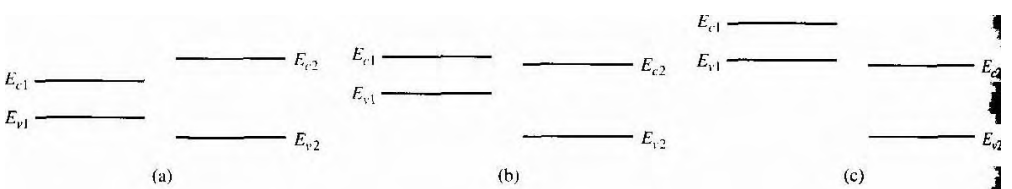


Figure 9.16 | Relation between narrow-bandgapand wide-bandgap energies: (a) straddling, (b) staggered, and (c) broken gap.

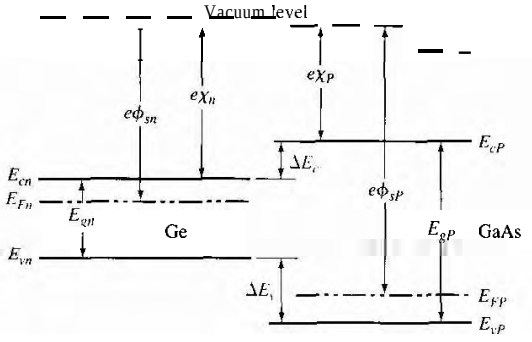


Figure 9.17 | Energy-band diagrams of a narrow-bandgep and a wide-bandgap material before contact.

wide-bandgap material is less than that of the narrow-bandgap material. The difference between the two conduction band energies is denoted by $AE_{,,}$ and the difference between the two valence band energies is denoted by AE,. From Figure 9.17. we can see that

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

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

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 AE, and AE, discontinuities will exist at the 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,, and the space charge width into the P-type region is denoted 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 he moderetely to heavily doped n type, while the GaAs can be more lightly doped or

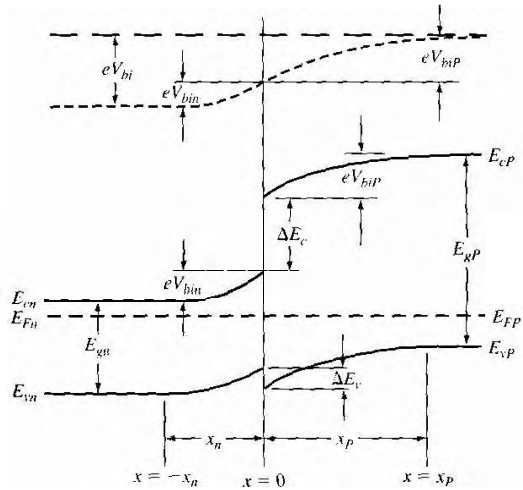


Figure 9.18 | Ideal energy-band diagram of an nP heterojunction in thermal equilibrium.

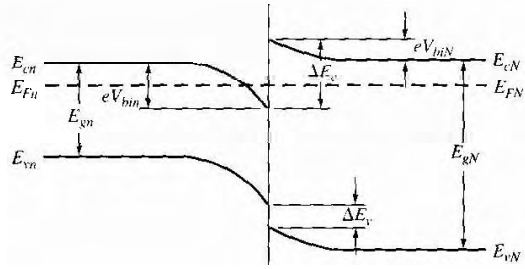


Figure 9.19 | Ideal energy-hand diagram of an nN heterojunction in thermal equilibrium.

even intrinsic. As mentioned previously. to achieve thermal equilibrium, electron from the wide-bandgap AlGaAs Row into the GaAs. forming an accumulation laye 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 con. tained in a potential well is quantized. The phrase *two-dimensional electron* gas 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

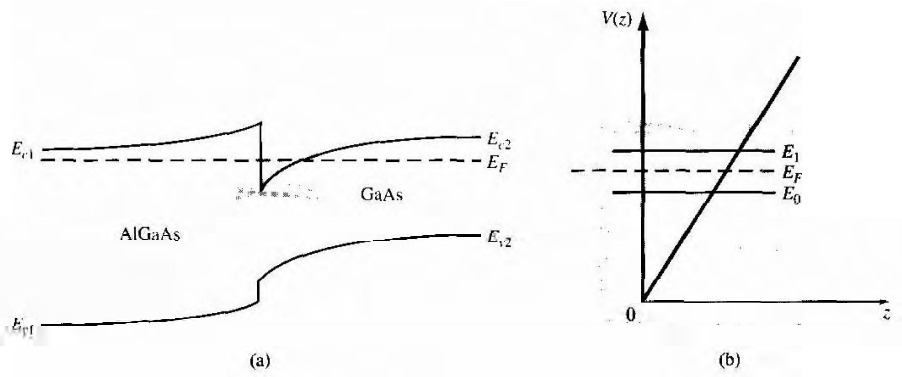


Figure 9.20 | (a) Conduction-band edge at N-AlGaAs, n-GaAs heterojunction; (b) triangular well approximation with discrete electron energies.

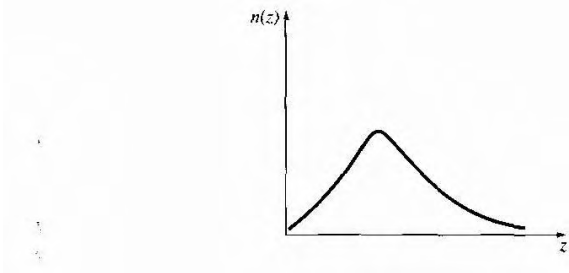


Figure 9.21 | Electron density in triangular potential well.

interface and Figure 9.20b shows the approximation of the triangular potential well. We can write

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

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

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.

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.

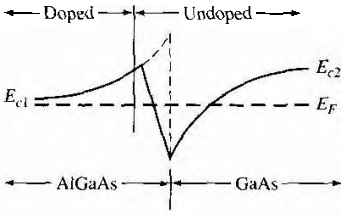


Figure 9.22 / Conduction-hand edge at a graded heterojunction.

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 will now consider the electrostatics of the nP heterojunction that was shown in Figure 9.18. As in the case of the homojunction, potential differences exist across the space charge regions in both then region and the Pregion. 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 wntten as

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

or

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

which can be expressed as

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

Finally, we can write Equation (9.38) as

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

where p_{po} and $p_{,,}$ are the hole concentrations in the P and n materials, respectively, and N,,and N_{vP} are the effective density of states functions in then and P materials, respectively. We could also obtain an expression for the built-in potential barrier in terms of the conduction band shift as

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

Objective EXAMPLE 9.8

To determine $\boldsymbol{AE,.\ At,.}$ and V_{bi} for an n–Ge to P–GaAs heterojunction using the electron affinity rule.

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

■ Solution

From Equation (9.34a), we have

$$\boldsymbol{AE,} \simeq e(\chi_n - \chi_P) = e(4.13 - 4.07) = 0.06 \text{ eV}$$

and from Equation (9.34b), we have

$$\boldsymbol{AE,} = \Delta E_g - \boldsymbol{AE,} = (1.43 - 0.67) - 0.06 = 0.70 \text{ eV}$$

Todetermine V_{bi} using Equation (9.39), wc need to determine p_{no} in Ge. or

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

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

or, finally.

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

■ Comment

There is a nonsymmetry in the AE, and $\boldsymbol{AE,}$ values that will tend to make the potential harriers seen by electrons and holes different. This nonsymmetry does not occur in homojunctions.

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

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

and in the P region

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

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

$$\epsilon_n \mathrm{E}_n(x = 0) = \epsilon_P \mathrm{E}_P(x = 0) \tag{9.42a}$$

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 then 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 baniers 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 cp 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-bias 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 before, V_R is the magnitude of the reverse-bias 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 showed 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 he 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 that

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

for the straddling type of heterojunction, although the ΔE_c and ΔE_g values may differ from those determined from the electron affinity rule.

We may consider the general characteristics of the energy-hand diagrams of the other types of heterojunction. Figure 9.23 shows the energy-band diagram of an Np

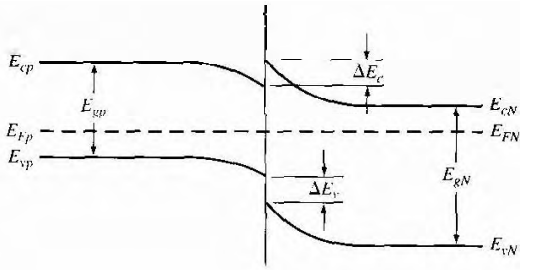


Figure 9.23 | Ideal energy-band diagram of an Np heterojunction in thermal equilibrium.

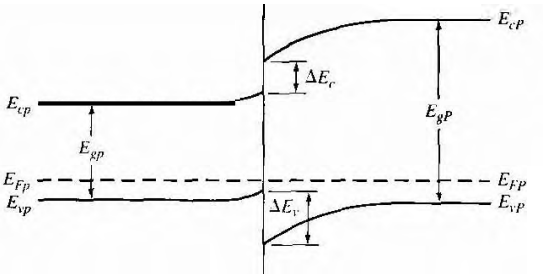


Figure 9.21 | Ideal energy-bend diagram of a pP heterojunction in thermal cquilihrium.

heterojunction. The same AE, 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-hand diagram of the nN junction was shown in Figure 9.19. To achieve thermal equilihrium. electrons from the wide-bandgap material will How into the narrow-bandgap material. A positive space charge region exists in the widc-gap material and an accumulation layer of electrons now exists at the interface in the narrow-gap 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 Row into the narrow-bandgap material, creating an accumulation layer of holes in the narrow-handgap material at the interface. These types of isotype heterojunctions are obviously not possible in a homojunction.

*9.3.5 Current–Voltage Characteristics

The ideal current–voltage characteristics of a pn homojunction were 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 harrier 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 harrier heights seen by electrons and holes are not the same. The energy-band diagrams in Figures 9.18 and 9.23 demonstrated 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 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 for holes, the electron current will he 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 did in the 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 barrierjunction. The heterojunction 1–Vcharacteristics, however, may need to he 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 banier 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 How 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 thennionic 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 encrgies 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 heterojunctian, 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 harrier 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 harrier 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 ina potential well at a heterojunction interface that 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 How 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 interface states on the characteristics of a Schottky harrier diode.
- Describe one effect of a larger reverse saturation current in a Schottky banier 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 harrier 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. Sketch the ideal energy band diagram of a metal–semiconductorjunction in which $\phi_m < \phi_s$. Explain why this is an ohmic contact.
6. Sketch the energy band diagram of a tunneling junction. Why is this an ohmic contact?
7. What is a hetcrojunction'?
8. 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 fur 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 pan (*b*). (*d*) Repeat parts (*b*) and (*c*) using the data in Figure 9.5.

9.2 An ideal rectifying contact is formed by depositing gold on n-type silicon doped at 10^{15} cm^{-3}. At $T = 300$ K, determine (*a*) ϕ_{B0}, (*b*) V_{bi}, (*c*) W, and (*d*) E_{max}, all under equilibrium conditions.

9.3 Consider a gold Schottky diode at T = 300 K formed on n-type GaAs doped at $N_d = 5 \times 10^{16}$ cm^{-3}. Determine (*a*) the theoretical barrier height. ϕ_{B0}, (*b*) ϕ_n, (*c*) V_{bi},

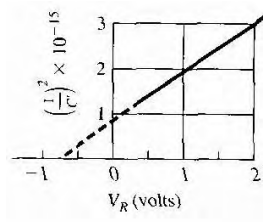


Figure 9.25 | Ftgure for Problem 9 6

(*d*) the space charge width, x_n, for V_R = 5 V. and (*e*) the electric field at the metal junction for $V_R = 5$ V.

9.4 Repeat problem 9.3, parts (b) through (*e*), if the experimentally determined harrier height is found to be $\phi_{Bn} = 0.86$ V.

9.5 An Au–n-Si junction with $N_d = 5 \times 10^{15}$ cm^{-3} has a crocs-sectional area of $A = 5 \times 10^{-4}$ cm^2. T = 300 K. Use the data in Figure 9.5. ***(a)*** Determine the junction capacitance when $V_R = 4$ V. ***(b)*** Repeat part (a) if the doping is increased to $N_d = 5 \times 10^{16}$ cm^{-3}.

9.6 A Schottky diode with n-type GaAs at $T = 300$ K yields the $1/C'^2$ versus V_R plot shown in Figure 9.25. where C' is the capacitance per cm'. Determine (a) V_{bi}, (b) N_d, (*c*) ϕ_n, and (*d*) ϕ_{B0}

9.7 Consider an Al–n-Si Schottky barrier at $T = 300$ K with $N_d = 10^{\prime\prime\prime}$ cm^{-3}. Use the data in Figure 9.5 to determine the barrier height. (o) Determine V_{bi}, x_d, and E_{max} at zero bias. (*b*) Using the value of E_{max} from pan (*a*), determine $\Delta\phi$ and x_m for the Schottky barrier lowering. (*c*) Repeat part (*b*) for the case when a reverse bias of V_R = 4 V is npplicd.

9.8 Starting with Equation (9.12), derive Equations (9.14) and (9.15).

9.9 An Au–n-GaAs Schottky diode is at T = 300 K with $N_d = 5 \times 10^{16}$ cm^{-3}. Use the data in Figure 9.5 to determine the harrier height. (*a*) Determine V_{bi}, x_d, and E_{max} at zero bias. (b) Determine the reverse-bias voltage at which the Schottky barrier lowering, $\Delta\phi$, will he 7 percent of ϕ_{Bn}. (Use the value of E_{max} in the space charge region.)

9.10 Consider n-type silicon doped at $N_d = 10^{16}$ cm^{-3} with a gold contact to form a Schottky diode. Investigate the effect of Schottky barrier lowering. ***(a)*** Plot the Schottky barrier lowering $\Delta\phi$ versus reverse-bias voltage over the range $0 \leq V_R \leq 50$ V. (b) Plot the ratio $J_{sT}(V_R)/J_{sT}(V_R = 0)$ over the same range of reverse-bias voltage.

***9.11** The energy-band diagram of a Schottky diode is shown in Figure 9.6. Assume the following parameters:

$\phi_m = 5.2$ V	$\phi_n = 0.10$ V	$\phi_0 = 0.60$ V
$E_g = 1.43$ eV	$\delta = 25$ Å	$\epsilon_i = \epsilon_0$
$\epsilon_s = (13.1)\epsilon_0$	$\chi = 4.07$ V	$N_d = 10^{16}$ cm^{-3}
		$D_{it} = 10^{13}$ eV^{-1} cm^{-2}

(*a*) Determine the theoretical barrier height ϕ_{B0}, without interface states. (*b*) Determine the barrier height with interface states. (*c*) Repeat pans (a) and (*h*) if Q is changed to $\phi_m = 4.5$ V.

***9.12** A Schottky barrier diode contains interface states and an interfacial layer Assume the following parameters:

$\phi_m = 4.75$ V $\quad \phi_n = 0.164$ V $\quad \phi_0 = 0.230$ V

$E_g = 1.12$ eV $\quad \delta = 20$ Å $\quad \epsilon_i = \epsilon_0$

, $= (11.7)\epsilon_0 \quad \chi = 4.01$ V $\quad N_d = 5 \times 10^{16}$ cm^{-3}

$\phi_{B0} = 0.60$ V

Determine the interface state density, D_{it}, in units of eV^{-1} cm^{-2}.

9.13 A PtSi Schottky diode at T = 300 K is fabricated on n-type silicon with a doping of $N_d = 10^{16}$ cm^{-3}. From Figure 9.5, the barrier height is 0.89 V. Determine (*a*) ϕ_n, (b) V_{bi}, (*c*) J_{sT}, when the harrier lowering is neglected, and (*d*) V_a so that $J_n = 2$ A/cm^2.

9.14 (*a*) Consider a Schottky diode at $T = 300$ K formed with tungsten on n-type silicon. Let $N_d = 5 \times 10^{15}$ cm^{-3} and assume a cross-sectional area of $A = 5 \times 10^{-4}$ cm^2. Determine the forward-bias voltage required to obtain a current of 1 mA, 10 mA, and 100 mA. (*b*) Repeat pan *(n)* if the temperature is increased to T = 400 K. (Neglect Schottky barrier lowering.)

9.15 A Schottky diode is formed by depositing Au on n-type GaAs doped at $N_d = 5 \times 10^{16}$ cm^{-3}. T = 300 K. (*a*) Determine the forward-bias voltage required to obtain $J_n = 5$ A/cm^2. (b) What is the change in forward-hias voltage necessary to double the current? (Neglect Schottky barrier lowering.)

9.16 ***(a)***Consider an Au n-type GaAs Schottky diode with a cross-sectional area of 10^{-4} cm^2. Plot the forward-bias current–voltage characteristics over a voltage range of $0 \leq V_D \leq 0.5$ V. Plot the current an a log scale. (*b*) Repeat part (*a*) for an Au n-type silicon Schottky diode. (*c*) What conclusions can he drawn from these results?



9.17 A Schottky diode at T = 300 K is formed between tungsten and n-type silicon doped at $N_d = 10^{16}$ c m^{-3}. The cross-sectional area is A = 10^{-4} cm^2. Determine the reverse-bias saturation current at (*a*) $V_R = 2$ V and (*b*) $V_R = 4$ V. (Take into account the Schottky barrier lowering.)

***9.18** Starting with the basic current equation given by Equation (9.18), derive the relation given by Equation (9.23).

9.19 A Schottky diode and a pn junction diode have cross-sectional arras of A = 5×10^{-4} cm^2. The reverse saturation current density of the Schottky diode is 3×10^{-8} A/cm^2 and the reverse saturation current density of the pn junction diode is 3×10^{-12} A/cm^2. The temperature is 300 K. Determine the forward-bias voltage in each diode required to yield diode currents of 1 mA.

9.20 The reverse saturation current densities in a pn junction diode and a Schottky diode are 5×10^{-12} A/cm^2 and 7×10^{-8} A/cm^2, respectively, at $T = 300$ K. The cross-sectional area of the pn junction diode is A = 8×10^{-4} cm^2. Determine the cross-sectional area of the Schottky diode so that the difference in forward-bias voltages to achieve 1.2 mA is 0.265 V.

9.21 ***(a)***The reverse-saturation currents of a Schotlky diode and a pn junction diode at T = 300 K are 5×10^{-8} A and 10^{-12} A, respectively. The diodes are connected in

parallel and arc driven by a constant current of 0.5 mA. (*i*) Determine the current in each diode. (ii) Determine the voltage across each diode. (*b*) Repeat part (*a*) if the diodes are connected in series.

9.22 A Schottky diode and a pn junction diode have cross-sectional areas of $A = 7 \times 10^{-4}$ cm^2. The reverse-saturation current densities at $T = 300$ K of the Schottky diode and pn junction are 4×10^{-8} A/cm^2 and 3×10^{-12} A/cm^2, respectively. A forward-bias current of 0.8 mA is required in each diode. (*a*) Determine the forward-bias voltage required across each diode. (*b*) If the voltage from part (a) is maintained across each diode, determine the current in each diode if the temperature is increased to 400 K. (Take into account the temperature dependence of the reverse-saturtion currents. Assume $E_g = 1.12$ eV for the pn junction diode and $\phi_{B0} = 0.82$ V. for the Schottky diode.)

9.23 Compare the current-voltage characteristics of a Schottky barrier diode and a pn junction diode. Use the results of Example 9.5, and assume diode areas of 5×10^{-4} cm^2. Plot the current–voltage characteristics on a linear scale over a current range of $0 \leq I_D \leq 10$ mA.

Section 9.2 Metal-Semiconductor Ohmic Contacts

9.24 It is possible, theoretically, to form an ohmic contact between a metal and silicon that has a very low barrier height. Considering the specific contact resistance, determine the value of ϕ_{Bn} that will give a value of $R_c = 10^{-5}$ Ω-cm^2 at T = 300 K.

9.25 A metal, with a work function $\phi_m = 4.2$ V, is deposited on an n-type silicon semiconductor with $\chi_s = 4.0$ V and $E_g = 1.12$ eV. Assume no interface states exist at the junction. Let T = 300 K. (*a*) Sketch the energy-hand diagram for zero bias fur the case when no space charge region exists at the junction. (*b*) Determine N_d so that the condition in pan (*a*) is satisfied. (c) What is the potential barrier height seen by electrons in the metal moving into the semiconductor?

9.26 Consider the energy-hand diagram of a silicon Schottky junction under zero bias shown in Figure 9.26. Let $\phi_{B0} = 0.7$ V and $T = 300$ K. Determine the doping required so that $x_d = 50$ A at the point where the potential is $\phi_{B0}/2$ below the peak value. (Neglect the barrier lowering effect.)

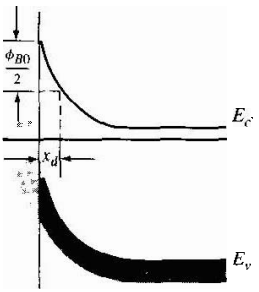


Figure 9.26 | Figure for Problem 9.26.

9.27 A metal–semiconductor junction is formed between a metal with a work function of 4.3 eV and p-type silicon with an electron affinity of 4.0 eV. The acceptor doping concentration in the silicon is $N_a = 5 \times 10^{16}\ cm^{-3}$. Assume T = 300 K. (a) Sketch the thermal equilibrium energy band diagram. (*h*) Determine the height of the Schottky barrier. (*c*) Sketch the energy band diagram with an applied reverse-bias voltage of $V_R = 3$ V. (*d*) Sketch the energy band diagram with an applied forward-bias voltage of $V_a = 0.25$ V.

9.28 (*a*) Consider a metal–semicanductor junction formed between a metal with a work function of 4.65 eV and Ge with an electron affinity of 4.13 eV. The dopingconcentration in the Ge material is $N_d = 6 \times 10^{13}\ cm^{-3}$ and $N_a = 3 \times 10^{13}\ cm^{-3}$. Assume T = 300 K. Sketch the zero bias energy-band diagram and determine the Schottky barrier height. (b) Repeat pan *(a)* if the metal work function is 4.35 eV.

Section 9.3 Heterojunctions

9.29 Sketch the energy-band diagrams of an abrupt $Al_{0.3}Ga_{0.7}As$–GaAs heterojunction for: (a) N^+–AlGaAs, intrinsic GaAs, (*b*) N^+–AlGaAs, p–GaAs, and (c) P^+–AlGaAs, n^+–GaAs. Assume $E_g = 1.85$ eV for $Al_{0.3}Ga_{0.7}As$ and assume $AE, = \frac{2}{3}\Delta E_g$.

9.30 Repeat Problem 9.29 assuming the ideal electron affinity rule. Determine $AE,$ and $AE,$

***9.31** Starting with Poisson's equation, derive Equation (9.48) for an abrupt heterojunction.

Summary and Review

***9.32** **(a)** Derive an expression for dV_a/dT as a function of current density in a Schottky diode. Assume the minority carrier current is negligible. (*b*) Compare dV_a/dT for a GaAs Schottky diode to that for a Si Schottky diode. (*c*) Compare dV_a/dT for a Si Schottky diode to that for a Si pn junction diode.

9.33 The $(1/C_j)^2$ versus V_R data are measured for two Schottky diodes with equal areas. One diode is fabricated with 1 Ω-cm silicon and the other diode with 5 Ω-cm silicon. The plots intersect the voltage axis as $V_R = -0.5$ V for diode A and at $V_R = -1.0$ V for diode B. The slope of the plot for diode A is $1.5 \times 10^{18}\ (F^2\text{-}V)^{-1}$ and that for diode B is $1.5 \times 10^{17}\ (F^2\text{-}V)^{-1}$. Determine which diode has the higher metal work function and which diode has the lower resistivity silicon.

***9.34** Both Schottky barrier diodes and ohmic contacts are to be fabricated by depositing a particular metal on a silicon integrated circuit. The work function of the metal is 4.5 V. Considering the ideal metal–semiconductor contact, determine the allowable range of doping concentrations for each type of contact. Consider both p- and n-type silicon regions.

9.30 Consider an n-GaAs–p-AlGaAs heterojunction in which the bandgap offsets are $AE, = 0.3$ eV and $\Delta E_v = 0.15$ eV. Discuss the difference in the expected electron and hole currents when the junction is forward biased.

READING LIST

1. Anderson, R. L. "Experimenls on Ge–GaAs Heterojunctions." *Solid-State* Electronics 5, no. 5 (September–October 1962), pp. 341–351

2. Crowley, A. M., and S. M. Sze. "Surface States and Barrier Height of Metal Semiconductor Systems." *Journal of Applied Physics* 36 (1965), p. 3212.
3. MacMillan, H. F.; H. C. Hamaker; G. F. Virshup; and J. G. Werthrn. "Multijunction III-V Solar Cells: Recent and Projected Results." *Twentieth IEEE Photovoltaic Specialists Conference* (1988), pp. *48–54.*
4. Michaelson, H. B. "Relation between an Atomic Electronegativity Scale and the Work Function." *IBM Journal of Research and Development 22. no.* 1 (January 1978), pp. 72–80.
5. Pierret, R. F. *Semiconductor Device Fundamentals.* Reading, MA: Addison-Wesley, 1996.
6. Rideout, V. L. "A Review of the Theory, Technology and Applications of Metal-Semiconductor Rectifiers." *Thin Solid Films* 48, no. 3 (February 1, 1978), pp. 261–291.
7. Roulston, D. J. *Bipolar Semiconductor Devices.* New York: McGraw-Hill, 1990.
8. Shur, M. *Introduction to Electronic Devices. New* York: John Wiley and Sons, 1996.
*9. Shur, M. *GaAs Devices und Circuits.* New York: Plenum Press, 1987.
*10. ——. *Physics of Semiconductor Devices.* Englewood Cliffs. NJ: Prentice Hall. 1990.
*11. Singh. J. *Physics of Semiconductors and Their Heterostructures.* New York: McGraw-Hill, 1993.
12. Singh, J. *Semiconductor Devices: Basic Principles.* New York: John Wiley and Sons, 2001.
13. Streetman, B. G., and S. Banerjee. *Solid State Electronic Devices.* 5th cd. Upper Saddle River. NJ: Prentice Hall, 2000.
14. Sze, S. M. *Physics of Semiconductor Devices.* 2nd ed. New York: Wiley, 1981.
*15. Wang, S. *Fundamentals of Semiconductor Theory and Device* **Physics.** Englewood Cliffs, NJ: Prentice Hall. 1989.
*16. Wolfe, C. M., N. Holonyak, Jr., and G. E. Stillman. *Physical Properties of Semiconductors.* Englewood Cliffs, *NJ:* Prentice Hall, 1989.
17. Yang, E. S. *Microelectronic Devices.* New York: McGraw-Hill, 1988.
*18. Yuan. J. S. *SiGe, GaAs, and InP Heterojunction Bipolar Transistors.* New York: John Wiley and Sons. 1999.

CHAPTER



The Bipolar Transistor

PREVIEW

The single-junction devices we have considered, including the pn homojunction diode, can be used to obtain rectifying current–voltage characteristics, and to form electronic switching circuits. The transistor is a multifunction semiconductor device that, in conjunction with other circuit elements, is capable of current gain, voltage gain, and signal-power gain. The transistor is therefore referred to as an active device whereas the diode is passive. The basic transistor action is the control of current at one terminal by voltage applied across two other terminals of the device.

The three basic transistor types are the bipolar transistor, the metal-oxide-semiconductor field-effect transistor (MOSFET), and the junction field-effect transistor (JFET). The bipolar transistor is covered in this chapter, the MOSFET is treated in Chapters 11 and 12, and the JFET is discussed in Chapter 13. The chapters dealing with each of the transistor types are written to stand alone, so that each type of transistor may be covered in any order desired.

The bipolar transistor has three separately doped regions and two pn junctions, sufficiently close together so that interactions occur between the two junctions. We will use much of the theory developed for the pn junction in the analysis of the bipolar transistor. Since the flows of both electrons and holes are involved in this device, it is called a bipolar transistor.

We will first discuss the basic geometry and operation of the transistor. Since there is more than one pn junction in the bipolar transistor, several combinations of reverse- and forward-bias junction voltages are possible, leading to different operating modes in the device. As with the pn junction diode, minority carrier distributions in the bipolar transistor are an important part of the physics of the device—minority carrier gradients produce diffusion currents. We will determine the minority carrier distribution in each region of the transistor, and the corresponding currents.

The bipolar transistor is a voltage-controlled current source. We will consider the various factors that determine the current gain and derive its mathematical expression.

As with anv semiconductor device. nonideal effects influence device characteristics: a few of these effects, such as breakdown voltage, will be described.

In order to analyze or design a transistor circuit. especially using computer simulations, one needs a mathematical model or equivalent circuit of the transistor. We will develop two equivalent circuits. The first equivalent circuit, the Ebers–Moll model, can be used for a transistor biased in any of its operating modes and is especially used for transistors in switching circuits. The second equivalent circuit, the hybrid-pi model, is applied when transistors are operated in a small signal linear amplifier and takes into account frequency effects within the transistor.

Various physical factors affect the frequency response of the bipolar transistor. There are several time-delay factors within the device that determine the limiting frequency response. We will define these time delays and develop expressions for each factor. The limiting frequency is given in terms of a cutoff frequency, a figure of merit for the transistor. The frequency response generally applies to the small signal, steady-state characteristics of the device. The switching characteristics, in contrast, determine the transient behavior of the transistor to large changes in the input signal..

10.1 | THE BIPOLAR TRANSISTOR ACTION

The bipolar transistor has three separately doped regions and two pn junctions. Figure 10.1 shows the basic structure of an npn bipolar transistor and a pnp bipolar transistor, along with the circuit symbols. The three terminal connections are called the emitter, base, and collector. The width of the base region is small compared to the minority carrier diffusion length. The (++)and (+) notation indicates the relative magnitudes of the impurity doping concentrations normally used in the bipolar transistor, with (++) meaning very heavily doped and (+) meaning moderately doped. The emitter region has the largest doping concentration: the collector region has the smallest. The reasons for using these relative impurity concentrations, and for the narrow base width. will become clear as we develop the theory of the bipolar transistor. The concepts developed for the pn junction apply directly to the bipolar transistor.

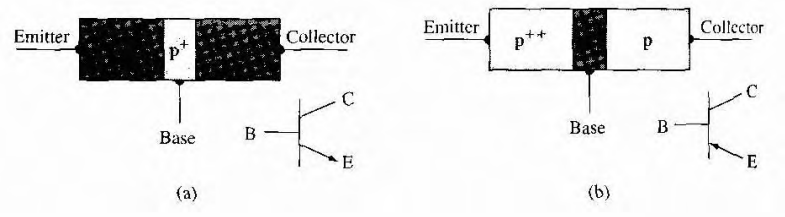


Figure 10.1 | Simplified block diagrams and circuit symbols **of** (a) npn and (b) pnp bipolar transistors.

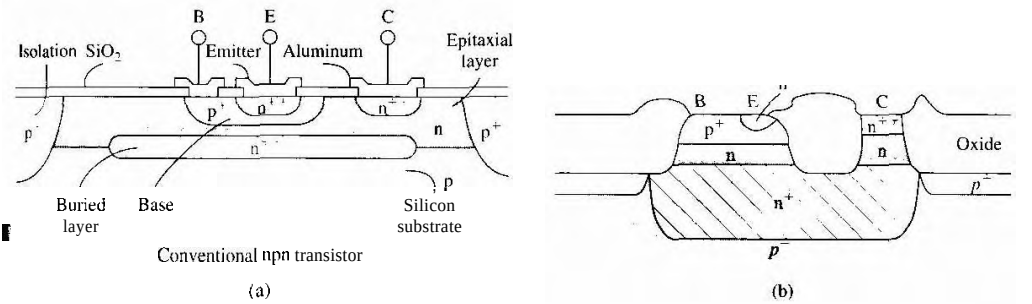


Figure 10.2 | Cross section of (a) a conventional integrated-circuit npn bipolar transistor and (b) an oxide-isolated npn bipolar transistor
(From Muller and Kamins [3].)

The block diagrams of Figure 10.1 show the basic structure of the transistor. but in very simplified sketches. Figure 10.2a shows a cross section of a classic npn bipolar transistor fabricated in an integrated circuit configuration, and Figure 10.2b shows the cross section of an npn bipolar transistor fabricated by a more modem technology. One can immediately observe that the actual structure of the bipolar transistor is not nearly as simple as the block diagrams of Figure 10.1 might suggest. A reason for the complexity is that terminal connections are made at the surface: in order to minimize semiconductor resistances, heavily doped n^+ buried layers must be included. Another reason for complexity arises out of the desire to fabricate more than one bipolar transistor on a single piece of semiconductor material. Individual transistors must be isolated from each other since all collectors, for example, will not be at the same potential. This isolation is accomplished by adding p^+ regions so that devices are separated by reverse-biased pn junctions as shown in Figure 10.2a. or they are isolated by large oxide regions as shown in Figure 10.2b.

An important point to note from the devices shown in Figure 10.2 is that the bipolar transistor is not a symmetrical device. Although the transistor may contain two n regions or two p regions, the impurity doping concentrations in the emitter and collector are different and the geometry of these regions can be vastly different. The block diagrams of Figure 10.1 are highly simplified, but useful, concepts in the development of the basic transistor theory.

10.1.1 The Basic Principle of Operation

The npn and pnp transistors are complementary devices. We will develop the bipolar transistor theory using the npn transistor, but the same basic principles and equations also apply to the pnp device. Figure 10.3 shows an idealized impurity doping profile in an npn bipolar transistor for the case when each region is uniformly doped. Typical impurity doping concentrations in the emitter, base, and collector may be on the order of 10^{19}, 10^{17}, and 10^{15} cm^{-3}, respectively.

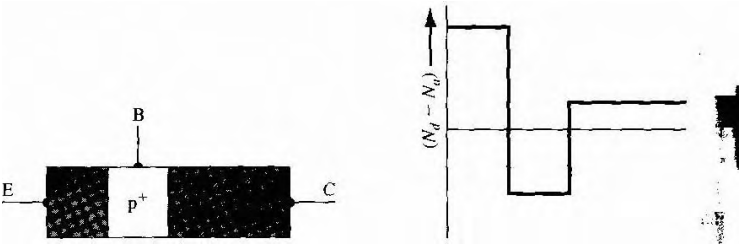


Figure 10.3 | Idealized doping profile of a uniformly doped npn bipolar transistor.

The base–emitter (B-E) pn junction is forward-biased, and the base–collector (B-C) pn junction is reverse-biased in the normal bias configuration as shown in Figure 10.4a. This configuration is called the *forward-active* operating mode: The B-E junction is forward-biased so electrons from the emitter are injected across the B-E junction into the base. These injected electrons create an excess concentration of minority carriers in the base. The B-C junction is reverse biased, so the minority carrier electron concentration at the edge of the B-C junction is ideally zero. We expect the electron concentration in the base to be like that shown in Figure 10.4b. The large gradient in the electron concentration means that electrons injected from the emitter will diffuse across the base region into the B-C space charge region. where the electric field will sweep the electrons into the collector. We want as many electrons as possible to reach the collector without recombining with any majority carrier holes in the base. For this reason, the width of the base needs to be small compared with the minority carrier diffusion length. If the base width is small, then the minority carrier electron concentration is a function of both the B-E and B-C junction voltages. The two junctions are close enough to be called *interacting* pn junctions.

Figure 10.5 shows a cross section of an npn transistor with the injection of electrons from the n-type emitter (hence the name emitter) and the collection of the electrons in the collector (hence the name collector).

10.1.2 Simplified Transistor Current Relations

We can gain a basic understanding of the operation of the transistor and the relations between the various currents and voltages by considering a simplified analysis. After this discussion, we will then delve into a more detailed analysis of the physics of the bipolar transistor.

The minority carrier concentrations are again shown in Figure 10.6 for an npn bipolar transistor biased in the forward active mode. Ideally. the minority carrier electron concentration in the base is a linear function of distance, which implies no recombination. The electrons diffuse across the base and are swept into the collector by the electric field in the B-C space charge region.

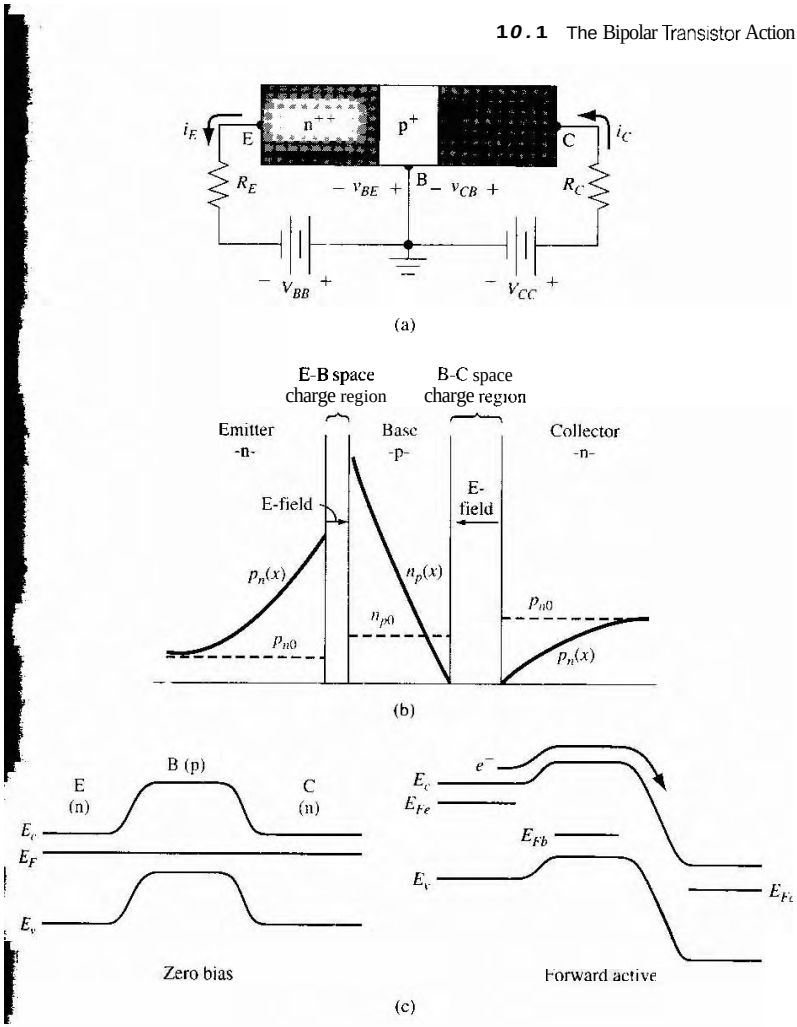


Figure 10.4 | (a) Biasing of an npn bipolar transistor in the forwad-active mode, (b) minority carrier distribution in an npn bipolar transistor operating in the forward-active mode. and (c) energy band diagram of the npn bipolar transistor under zero bias **and** under a forward-active mode bias.

Collector Current Assuming the ideal linear electron distribution in the base, the collector current can be written as a diffusion current given by

$$i_C = eD_n A_{BE}\frac{dn(x)}{dx} = eD_n A_{BE}\left[\frac{n_B(0)-0}{0-x_B}\right] = \frac{-eD_n A_{BE}}{x_B}\cdot n_{B0}\exp\left(\frac{v_{BE}}{V_t}\right)$$

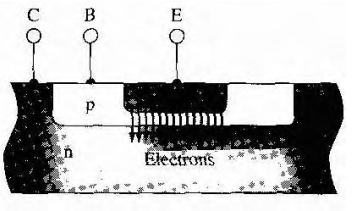


Figure 10.5 | Cross section of an npn bipolar transistor showing the injection and collection of electrons in the forward-active mode.

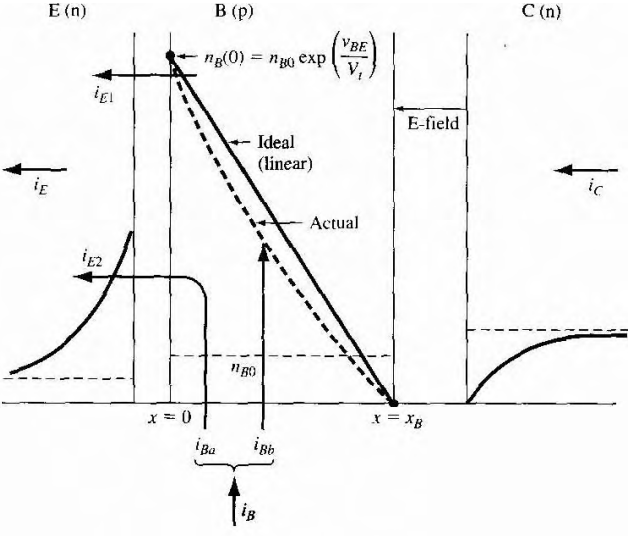


Figure 10.6 | Minority carrier distributions and basic currents in a forward-biased npn bipolar transistor.

where A_{BE} is the cross-sectional area of the B-E junction, n_{B0} is the thermal equilibrium electron concentration in the base, and V_t is the thermal voltage. The diffusion of electrons is in the $+x$ direction so that the conventional current is in the $-x$ direction. Considering magnitudes only, Equation (10.1) can be written as

$$i_C = I_S \exp\left(\frac{v_{BE}}{V_t}\right) \tag{10.2}$$

The collector current is controlled by the base–emitter voltage; that is, the current at one terminal of the device is controlled by the voltage applied to the other two terminals of the device. As we have mentioned, this is the basic transistor action.

mitter Current One component of emitter current, i_{E1}, shown in Figure 10.6 is ue to the flow of elections injected from the emitter into the hase. This current, then, equal to the collector current given by Equation (10.1).

Since the base–emitter junction is forward biased, majority canier holes in the se are injected across the B-E junction into the emitter. These injected holes pro- ce a pn junction current i_{E2} as indicated in Figure 10.6. This current is only a B-E nction current so this component of emitter current is not part of the collector cur- nt. Since i_{E2} is a forward-biased pn junction current, we can write (considering agnitude only)

$$i_{E2} = I_{S2} \exp\left(\frac{v_{BE}}{V_t}\right) \tag{10.3}$$

where I_{S2} involves the minority carrier hole parameters in the emitter. The total emitter current is the sum of the two components, or

$$i_E = i_{E1} + i_{E2} = i_C + i_{E2} = I_{SE} \exp\left(\frac{v_{BE}}{V_t}\right) \tag{10.4}$$

Since all current components in Equation (10.4) are functions of exp (v_{BE}/ V,), the ratio of collector current to emitter current is a constant. We can write

$$\frac{i_C}{i_E} \equiv \alpha \tag{10.5}$$

where a is called the *common-base current gain*. By considering Equation (10.4), we see that $i_C < i_E$ or $\alpha < 1$. Since i_{E2} is not part of the basic transistor action, we would like this component of current to be as small as possible. We would then like the common base current gain to be as close to unity as possible.

Referring to Figure 10.4a and Equation (10.4), note that the emitter current is an exponential function of the base–emitter voltage and the collector current is $i_C = \alpha i_E$. To a first approximation, the collector current is independent of the hase–collector voltage as long as the B-C junction is reverse biased. We can sketch the common-base transistor characteristics as shown in Figure 10.7. The bipolar transistor acts like a constant current source.

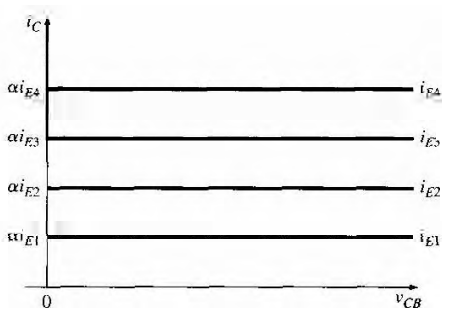


Figure 10.7 Ideal bipolar transistor common-base current–voltage characteristics.

Base Current **As** shown in Figure 10.6, the component of emitter current i_{E2} is a B-E junction current so that this current is also a component of base current sh as i_{Ba}. This component of base current is proportional to $\exp(v_{BE}/V_t)$.

There is also a second component of base current. We have considered the ideal case in which there is no recombination of minority carrier electrons with ma carrier holes in the base. However, in reality, there will be some recombi Since majority carrier holes in the base are disappearing, they must he resuppli a flow of positive charge into the base terminal. This flow of charge is indicated as a current i_{Bb} in Figure 10.6. The number of holes per unit time recombining in the base is directly related to the number of minority carrier electrons in the base (see Equation (6.13)). Therefore, the current i_{Bb} is also proportional to $\exp(v_{BE}/V_t)$. The total base current is the sum of i_{Ba} and i_{Bb}, and is proportional to $\exp(v_{BE}/V_t)$.

The ratio of collector current to hase current is a constant since both current, are directly proportional to $\exp(v_{BE}/V_t)$. We can then write

$$\frac{i_C}{i_B} \equiv \beta \qquad (10.$$

where β is called the common-emitter current ***gain.*** Normally. the hase current will be relatively small so that, in general, the common-emitter current gain is much larger than unity (on the order of 100 or larger).

10.1.3 The **Modes of Operation**

Figure 10.8 shows the npn transistor in a simple circuit. In this configuration, the transistor may be biased in one of three modes of operation. If the B-E voltage is zero or reverse biased ($V_{BE} \leq 0$), then majority carrier electrons from the emitter will not be injected into the hase. The B-C junction is also reverse biased; thus. the emitter and collector currents will be zero for this case. This condition is referred to as *cutoff*—all currents in the transistor are zero.

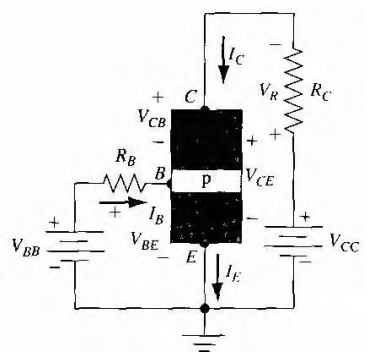


Figure 10.8 | An npn bipolar transistor in a common-emittercircuit configuration.

When the B-E junction becomes forward biased, an emitter current will be generated as we have discussed, and the injection of electrons into the base results in a collector current. We may write the KVL equations around the collector–emitter loop as

$$V_{CC} = I_C R_C + V_{CB} + V_{BE} = V_R + V_{CE} \tag{10.7}$$

If V_{CC} is large enough and if V_R is small enough, then $V_{CB} > 0$, which means that the B-C junction is reverse biased for this npn transistor. Again, this condition is the forward-active region of operation.

As the forward-biased B-E voltage increases, the collector current and hence V_R will also increase. The increase in V_R means that the reverse-biased C-B voltage decreases, or $|V_{CB}|$ decreases. At some point, the collector current may become large enough that the combination of V_R and V_{CC} produces zero voltage across the B-C junction. A slight increase in I_C beyond this point will cause a slight increase in V_R and the B-C junction will become forward biased ($V_{CB} < 0$). This condition is called saturation. In the saturation mode of operation, both B-E and B-C junctions are forward biased and the collector current is no longer controlled by the B-E voltage.

Figure 10.9 shows the transistor current characteristics, I_C versus V_{CE}, for constant base currents when the transistor is connected in the common-emitter configuration (Figure 10.8). When the collector–emitter voltage is large enough so that the base–collector junction is reverse biased, the collector current is a constant in this first-order theory. For small values of C-E voltage, the base–collector junction becomes forward biased and the collector current decreases to zero for a constant base current.

Writing a Kirchhoff's voltage equation around the C-E loop, we find

$$V_{CE} = V_{CC} - I_C R_C \tag{10.8}$$

Equation (10.8) shows a linear relation between collector current and collector–emitter voltage. This linear relation is called a ***load*** line and is plotted in Figure 10.9. The load line. superimposed on the transistor characteristics, can be used to visualize the bias condition and operating mode of the transistor. The cutoff mode occurs when

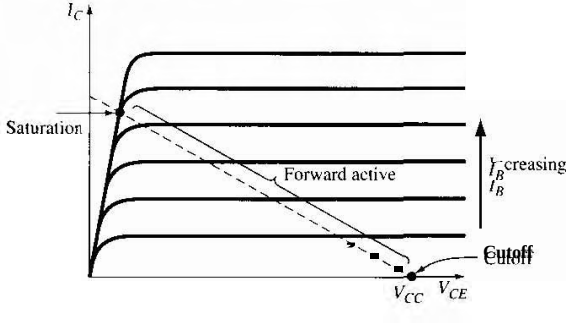


Figure 10.9 | Bipolar transistor common-emitter current–voltage characteristics with load line superimposed.

$I_C = 0$, saturation occurs when there is no longer a change in collector current for change in base current, and the forward-active mode occurs when the relation $I_C = \beta I_B$ is valid. These three operating modes are indicated on the figure.

A fourth mode of operation for the bipolar transistor is possible, although not with the circuit configuration shown in Figure 10.8. This fourth mode, known as inverse *active*, occurs when the B-E junction is reverse biased and the B-C junction is forward biased. In this case the transistor is operating "upside down," and the roles of the emitter and collector are reversed. We have argued that the transistor is not a symmetrical device; therefore, the inverse-active characteristics will not be the same as the forward-active characteristics.

The junction voltage conditions for the four operating modes are shown in Figure 10.10.

10.1.4 Amplification with Bipolar Transistors

Voltages and currents can he amplified by bipolar transistors in conjunction with other elements. We will demonstrate this amplification qualitatively in the following discussion. Figure 10.11 shows an npn bipolar transistor in a common-emitter configuration. The dc voltage sources, V_{BB} and V_{CC}, are used to bias the transistor in the forward-active mode. The voltage source v_i represents a time-varying input voltage (such as a signal from a satellite) that needs to be amplified.

Figure 10.12 shows the various voltages and currents that are generated in the circuit assuming that v_i is a sinusoidal voltage. The sinusoidal voltage v_i induces a sinusoidal component of base current superimposed on a dc quiescent value. Since $i_C = \beta i_B$, then a relatively large sinusoidal collector current is superimposed on a dc value of collector current. The tirne-varying collector current induces a time-varying voltage across the R_C resistor which, by Kirchhoff's voltage law, means that a sinusoidal vollage, superimposed on a dc value, exists between the collector and emitter of the bipolar transistor The sinusoidal voltages in the collector-emitter

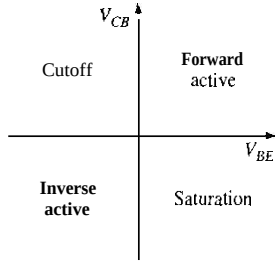


Figure 10.10 | Junction voltage conditions for the four operating modes of a bipolar transistor.

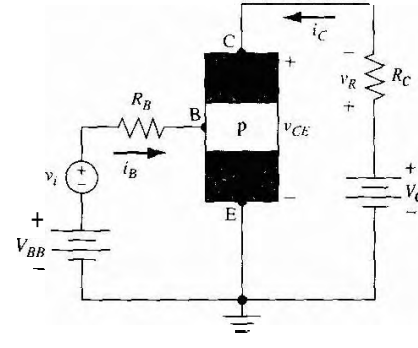


Figure 10.11 | Common-emitter npn bipolar circuit configuration with a time-varying signal voltage v_i included in the basc-emitter loop.

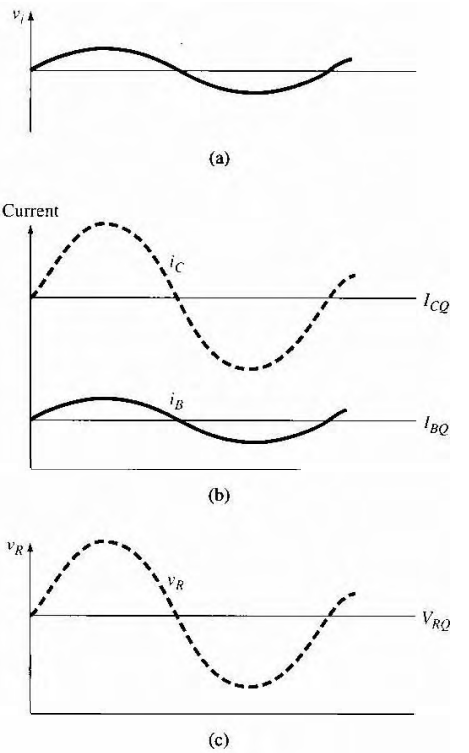


Figure 10.12 | Currents and voltages existing in the circuit shown in Figure 10.11. (a) Input sinusoidal signal voltage. (b) Sinusoidal base and collector currents superimposed on the quiescent dc values. (c) Sinusoidal voltage across the R_C resistor superimposed on the quiescent dc value.

portion of the circuit are larger than the signal input voltage v_i, so that the circuit has produced a ***voltage gain*** in the time-varying signals. Hence, the circuit is known as a ***voltage*** *amplifier.*

In the remainder of the chapter, we will consider the operation and characteristics of the bipolar transistor in more detail.

10.2 | MINORITY CARRIER DISTRIBUTION

We are interested in calculating currents in the bipolar transistor which, as in the simple pn junction, are determined by minority carrier diffusion. Since diffusion currents are produced by minority carrier gradients, we must determine the steady-state

Table 10.1 | Notation used in the analysis of the bipolar transistor

Notation	Definition
For **both** the npn and pnp transistors	
N_E, N_B, N_C	Doping concentrations in the emitter, base, and collector
x_E, x_B, x_C	Widths of neutral emitter, base. and collector regions
D_E, D_B, D_C	***Minority carrier*** diffusion coefficients in emitter, base. and collector regions
L_E, L_B, L_C	***Minority carrier*** diffusion lengths in emitter, base, and collector regions
$\tau_{E0}, \tau_{B0}, \tau_{C0}$	***Minority carrier*** lifetimes in emitter, base, and collector regions
For the npn	
p_{E0}, n_{B0}, p_{C0}	Thermal equilibrium *minority carrier* hole, electron, and hole concentrations in the emitter, base, and collector
$p_E(x'), n_B(x), p_C(x'')$	Total ***minority carrier*** hole, electron. and hole concentrations in the emitter base, and collector
$\delta p_E(x'), \delta n_B(x), \delta p_C(x'')$	Excess *minority carrier* hole, electron. and hole concentrations in the emitter, base, and collector
For the pnp	
n_{E0}, p_{B0}, n_{C0}	Thermal equilibrium ***minority carrier*** electron, hole, and electron concentrations in the emitter, base. and collector
$n_E(x'), p_B(x), n_C(x'')$	Total ***minority carrier*** electron. hole, and electron concentrations in the emitter, base, and collector
$\delta n_E(x'), \delta p_B(x), \delta n_C(x'')$	Excess ***minority carrier*** electron, hole, and electron concentrations in the emitter, base, **and** collector

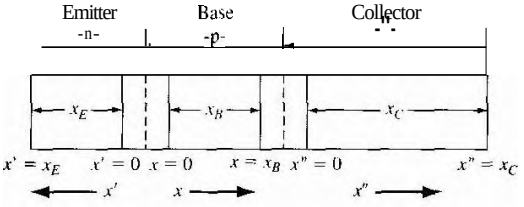


Figure 10.13 | Geometry of the npn bipolar transistor used to calculate the minority carrier distribution.

minority carrier distribution in each of the three transistor regions. Let us first consider the forward-active mode. and then the other modes of operation. Table 10.1 summarizes the notation used in the following analysis.

10.2.1 Forward-Active Mode

Consider a uniformly doped npn bipolar transistor with the geometry shown in Figure 10.13. When we consider the individual emitter, base, and collector regions, we

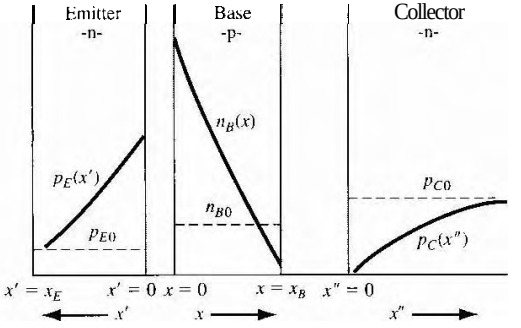


Figure 10.14 | Minority carrier distribution in an npn bipolar transistor operating in the forward-active mode.

will shift the origin to the edge of the space charge region and consider a positive x, x', or x" coordinate as shown in the figure.

In the forward-active mode, the B-E junction is forward biased and the B-C is reverse biased. We expect the minority carrier distributions to look like those shown in Figure 10.14. As there are two n regions, we will have minority carrier holes in both emitter and collector. To distinguish between these two minority carrier hole distributions, we will use the notation shown in the figure. Keep in mind that we will be dealing only with minority carriers. The parameters p_{E0}, n_{B0}, and p_{C0} denote the thermal-equilibrium minority carrier concentrations in the emitter, base, and collector, respectively. The functions $p_E(x')$, $n_B(x)$, and $p_C(x'')$ denote the steady-state minority carrier concentrations in the emitter, base, and collector, respectively. We will assume that the neutral collector length x_C is long compared to the minority carrier diffusion length L_C in the collector, but we will take into account a finite emitter length x_E. If we assume that the surface recombination velocity at x' = x_E is infinite, then the excess minority carrier concentration at x' = x_E is zero, or $p_E(x' = x_E) = p_{E0}$. An infinite surface recombination velocity is a good approximation when an ohmic contact is fabricated at x' = x_E.

Base Region The steady-state excess minority carrier electron concentration is found from the ambipolar transport equation, which we discussed in detail in Chapter 6. For a zero electric field in the neutral base region, the ambipolar transport equation in steady state reduces to

$$D_B \frac{\partial^2(\delta n_B(x))}{\partial x^2} - \frac{\delta n_B(x)}{\tau_{B0}} = 0$$

where δn_B is the excess minority carrier electron concentration, and D_B and τ_{B0} are the minority carrier diffusion coefficient and lifetime in the base region, respectively. The excess electron concentration is defined as

$$\delta n_B(x) = n_B(x) - n_{B0} \tag{10.10}$$

The general solution to Equation (10.9) can be written as

$$\delta n_B(x) = A \exp\left(\frac{+x}{L_B}\right) + B \exp\left(\frac{-x}{L_B}\right) \tag{10.11}$$

where L_B is the minority carrier diffusion length in the base, given by $L_B = \sqrt{D_B \tau_{B0}}$. The base is of finite width so both exponential terms in Equation (10.11) must be retained.

The excess minority carrier electron concentrations at the two boundaries become

$$\delta n_B(x = 0) \equiv \delta n_B(0) = A + B \tag{10.12a}$$

and

$$\delta n_B(x = x_B) \equiv \delta n_B(x_B) = A \exp\left(\frac{+x_B}{L_B}\right) + B \exp\left(\frac{-x_B}{L_B}\right) \tag{10.12b}$$

The B-E junction is forward biased, so the boundary condition at $x = 0$ is

$$\delta n_B(0) = n_B(x = 0) - n_{B0} = n_{B0}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \tag{10.13a}$$

The B-C junction is reverse biased, so the second boundary condition at $x = x_B$ is

$$\delta n_B(x_B) = n_B(x = x_B) - n_{B0} = 0 - n_{B0} = -n_{B0} \tag{10.13b}$$

From the boundary conditions given by Equations (10.13a) and (10.13b), the coefficients A and B from Equations (10.12a) and (10.12b) can be determined. The results are

$$A = \frac{-n_{B0} - n_{B0}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right]\exp\left(\frac{-x_B}{L_B}\right)}{2\sinh\left(\frac{x_B}{L_B}\right)} \tag{10.14a}$$

and

$$B = \frac{n_{B0}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right]\exp\left(\frac{x_B}{L_B}\right) + n_{B0}}{2\sinh\left(\frac{x_B}{L_B}\right)} \tag{10.14b}$$

Then, substituting Equations (10.14a) and (10.14b) into Equation (10.9), we can write the excess minority carrier electron concentration in the base region as

$$\delta n_B(x) = \frac{n_{B0}\left\{\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right]\sinh\left(\frac{x_B - x}{L_B}\right) - \sinh\left(\frac{x}{L_B}\right)\right\}}{\sinh\left(\frac{x_B}{L_B}\right)} \tag{10.15a}$$

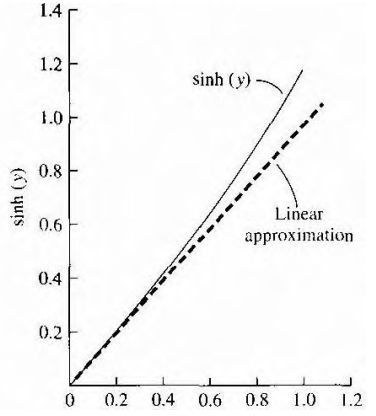


Figure 10.15 | Hyperbolic sine function **and** its linear approximation.

Equation (10.15a) may look formidable with the sinh functions. We have stressed that we want the base width x_B to be small compared to the minority carrier diffusion length L_B. This condition may seem somewhat arbitrary at this point, but the reason will become clear as we proceed through all of the calculations. Since we want $x_B < L_B$, the argument in the sinh functions is always less than unity and in most cases will he much less than unity. Figure 10.15 shows a plot of sinh (y) for $0 \leq y \leq 1$ and also shows the linear approximation for small values of y. If $y < 0.4$, the sinh (y) function differs from its linear approximation by less than 3 percent. All of this leads to the ***conclusion that the excess electron concentration δn_B in Equation (10.15a) is approximately a lineur function of x through the neutral base region.*** Using the approximation that sinh $(x) \approx x$ for $x \ll 1$. the excess electron concentration in the base is given by

$$\delta n_B(x) \approx \frac{n_{B0}}{x_B}\left\{\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right](x_B - x) - x\right\} \qquad (10.15b)$$

We will use this linear approximation later in some of the example calculations. The difference in the excess canier concentrations determined from Equations (10.15a) and (10.15b) is demonstrated in the following exercise.

TEST YOUR UNDERSTANDING

E10.1 The emitter and base of a silicon npn bipolar transistor are uniformly doped at impurity concentrations of 10^{18} cm^{-3} and 10^{16} cm^{-3}, respectively. A forward-bias B-E voltage of $V_{BE} = 0.610$ V is applied. The neutral base width is $x_B = 2\,\mu$m and

the minority carrier diffusion length in the base is $L_B = 10\,\mu$m. Calculate the excess minority carrier concentration in the base at (*a*) $x = 0$ and (*b*) $x = x_B/2$. (*c*) Determine the ratio of the actual minority carrier concentration at $x = x_B/2$ (Equation (10.15a)) to that in the ideal case of a linear minority carrier distribution (Equation (10.15b)).

[Ans. (*a*) $\delta n_B(0) = 3.81 \times 10^{14}$ cm^{-3}, (*b*) $\delta n_B(x_B/2) \approx n_B(x_B/2) = 1.8947 \times 10^{14}$ cm^{-3}, (*c*) Ratio $= (1.8947 \times 10^{14})/1.9042 \times 10^{14}) = 0.99503$]

Tahle 10.2 shows the Taylor expansions of some of the hyperbolic functions that will he encountered in this section of the chapter. In most cases. we will consider only the linear terms when expanding these functions.

Emitter Region Consider, now, the minority carrier hole concentration in the emitter. The steady-state excess hole concentration is determined from the equation

$$D_E \frac{\partial^2(\delta p_E(x'))}{\partial x'^2} - \frac{\delta p_E(x')}{\tau_{E0}} = 0 \tag{10.16}$$

where D_E and τ_{E0} are the minority carrier diffusion coefficient and minority carrier lifetime, respectively. in the emitter. The excess hole concentration is given by

$$\delta p_E(x') = p_E(x') - p_{E0} \tag{10.17}$$

The general solution to Equation (10.16) can he written as

$$\delta p_E(x') = C \exp\left(\frac{+x'}{L_E}\right) + D \exp\left(\frac{-x'}{L_E}\right) \tag{10.18}$$

where $L_E = \sqrt{D_E \tau_{E0}}$. If we assume the neutral emitter length x_E is not necessarily long compared to L_E. then both exponential terms in Equation (10.18) must be retained.

The excess minority carrier hole concentrations at the two boundaries are

$$\delta p_E(x' \geq 0) \equiv \delta p_E(0) = C + D \tag{10.19a}$$

and

$$\delta p_E(x' = x_E) \equiv \delta p_E(x_E) = C \exp\left(\frac{x_E}{L_E}\right) + D \exp\left(\frac{-x_E}{L_E}\right) \tag{10.19b}$$

Table 10.2 | Taylor expansions of hyperholic functions

Function	Taylor expansion
$\sinh(x)$	$x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$
$\cosh(x)$	$1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$
$\tanh(x)$	$x - \frac{x^3}{3} + \frac{2x^5}{15} + \cdots$

Again, the B-E junction is forward biased so

$$\delta p_E(0) = p_E(x' = 0) - p_{E0} = p_{E0}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \tag{10.20a}$$

An infinite surface recombination velocity at $x' = x_E$ implies that

$$\delta p_E(x_E) = 0 \tag{10.20b}$$

Solving for C and D using Equations (10.19) and (10.20) yields the excess minority carrier hole concentration in Equation (10.18):

$$\boxed{\delta p_E(x') = \frac{p_{E0}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right]\sinh\left(\frac{x_E - x'}{L_E}\right)}{\sinh\left(\frac{x_E}{L_E}\right)}} \tag{10.21a}$$

This exress *concentration* will also vary *approximately* linearly with distance *if* x_E is small. We find

$$\boxed{\delta p_E(x') \approx \frac{p_{E0}}{x_E}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right](x_E - x')} \tag{10.21b}$$

If x_E is comparable to L_E, then $\delta p_E(x')$ shows an exponential dependence on x_E

TEST YOUR UNDERSTANDING

E10.2 Consider a silicon npn bipolar transistor with emitter and base regions uniformly doped at concentrations of 10^{18} cm^{-3} and 10^{16} cm^{-3}, respectively. A forward bias B-E voltage of $V_{BE} = 0.610$ V is applied. The neutral emitter width is $x_E = 4\,\mu$m and the minority carrier diffusion length in the emitter is $L_E = 4\,\mu$m. Calculate the excess minority carrier concentration in the emitter at (a) $x' = 0$ and (b) $x' = x_E/2$

[Ans. (a) 3.808×10^{12} cm^{-3}, (b) 1.689×10^{12} cm^{-3}]

Collector **Region** The excess minority carrier hole concentration in the collector can he determined from the equation

$$D_C\frac{\partial^2(\delta p_C(x''))}{\partial x''^2} - \frac{\delta p_C(x'')}{\tau_{C0}} = 0 \tag{10.22}$$

where D_C and τ_{C0} are the minority carrier diffusion coefficient and minority carrier lifetime, respectively, in the collector. We can express the excess minority carrier hole concentration in the collector as

$$\delta p_C(x'') = p_C(x'') - p_{C0} \tag{10.23}$$

The general solution to Equation (10.22) can he written as

$$\delta p_C(x'') = G \exp\left(\frac{x''}{L_C}\right) + H \exp\left(\frac{-x''}{L_C}\right) \quad (10.2$$

where $L_C = \sqrt{D_C \tau_{C0}}$. If we assume that the collector is long. then the coefficient must be zero since the excess concentration must remain finite. The second boundar condition gives

$$\delta p_C(x'' = 0) \equiv \delta p_C(0) = p_C(x'' = 0) - p_{C0} = 0 - \text{pco} = -p_{C0} \quad (10.2$$

The excess minority carrier hole concentration in the collector is then given as

$$\boxed{\delta p_C(x'') = -p_{C0} \exp\left(\frac{-x''}{L_C}\right)} \quad (10.26$$

This result is exactly what we expect from the results of a reverse-biased pn junction

TEST YOUR UNDERSTANDING

E10.3 Consider the collector region **of** an npn bipolar transistor biased in the forward active region. At whet value of x'', compared to L_C, does the magnitude of the minority carrier concentration reach 95 percent of the thermal equilibrium value.
$(\xi \approx {}^{\circ}L/{}_{''}x \cdot \text{su}\forall)$

10.2.2 Other Modes of Operation

The bipolar transistor can also operate in the cutoff, saturation, or inverse-active mode. We will qualitatively discuss the minority carrier distrihutions for these operating conditions and treat the actual calculations as problems at the end of the chapter.

Figure 10.16a shows the minority carrier distribution in an npn bipolar transistor in cutoff. In cutoff, both the B-E and B-C junctions are reverse biased: thus, the,

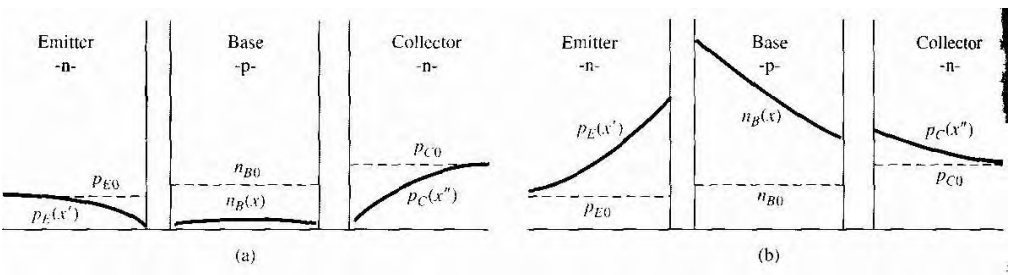


Figure 10.16 | Minority carrier distribution in an npn bipolar transistor operating in (a) cutoff and (b) saturation

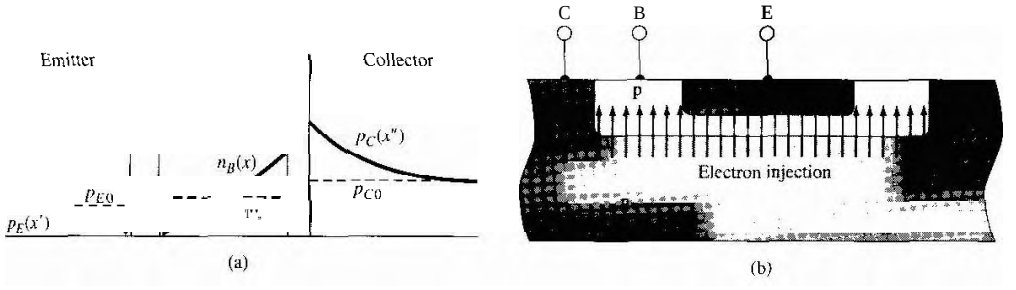


Figure 10.17 | (a) Minorily carrier distribution in an npn bipolar transistor operating in the inverse-active mode. (b) Cross section of an npn bipolar transistor showing the injection and collection of electrons in the inverse-active mode.

minority carrier concentrations are zero at each space charge edge. The emitter and collector regions are assumed to be "long" in this figure, while the base is narrow compared with the minority carrier diffusion length. Since $x_B \ll L_B$, essentially all minority carriers are swept out of the base region.

Figure 10.16b shows the minority carrier distribution in the npn bipolar transistor operating in saturation. Both the B-E and B-C junctions are forward biased; thus, excess minority carriers exist at the edge of each space charge region. However, since a collector current still exists when the transistor is in saturation, a gradient will still exist in the minority carrier electron concentration in the base.

Finally, Figure 10.17a shows the minority carrier distribution in the npn transistor for the inverse-active mode. In this case, the B-E is reverse biased and the B-C is forward biased. Electrons from the collector are now injected into the base. The gradient in the minority carrier electron concentration in the base is in the opposite direction compared with the forward-active mode, so the emitter and collector currents will change direction. Figure 10.17b shows the injection of electrons from the collector into the base. Since the B-C area is normally much larger than the B-E area. not all of the injected electrons will be collected by the emitter. The relative doping concentrations in the base and collector are also different compared with those in the base and emitter; thus, we see that the transistor is not symmetrical. We then expect the characteristics to be significantly different between the forward-active and inverse-active modes of operation.

10.3 | LOW-FREQUENCY COMMON-BASE CURRENT GAIN

The basic principle of operation of the bipolar transistor is the control of the collector current by the B-E voltage. The collector current is a function of the number of majority carriers reaching the collector after being injected from the emitter across the B-E junction. The *common-base current gain* is defined as the ratio of collector current to emitter current. The flow of various charged carriers leads to definitions of

particular currents in the device. We can use these definitions to define the current gain of the transistor in terms of several factors.

10.3.1 Contributing Factors

Figure 10.18 shows the various particle flux components in the npn bipolar transistor. We will define the various flux components and then consider the resulting currents. Although there seems to be a large number of flux components, we may help clarify the situation by correlating each factor with the minority carrier distributions shown in Figure 10.14.

The factor J_{nE}^{-} is the electron flux injected from the emitter into the base. As the electrons diffuse across the base, a few will recombine with majority carrier holes. The majority carrier holes that are lost by recomhination must be replenished from the base terminal. This replacement hole flux is denoted by J_{RB}^{+}. The electron flux that reaches the collector is J_{nC}^{-}. The majority camer holes from the base that are injected back into the emitter result in a hole flux denoted by J_{pE}^{+}. Some electrons and holes that are injected into the forward-biased B-E space charge region will recombine in this region. This recombination leads to the electron flux J_{R}^{-}. Generation of electrons and holes occurs in the reverse-biased B-C junction. This generation yields a hole flux J_{G}^{+}. Finally, the ideal reverse-saturation current in the B-C junction is denoted by the hole flux J_{pc0}^{+}.

The corresponding electric current density components in the npn transistor are shown in Figure 10.19 along with the minority carrier distributions for the forward active mode. The curves are the same as in Figure 10.14. As in the pn junction, the currents in the bipolar transistor are defined in terms of minority carrier diffusion currents. The current densities are defined as follows:

J_{nE}: Due to the diffusion of minority carrier electrons in the base at $x = 0$.

J_{nC}: Due to the diffusion of minority carrier electrons in the base at $x = x_B$.

J_{RB}: The difference between J_{nE} and J_{nC}, which is due to the recomhination of excess minority carrier electrons with majority carrier holes in the base. The J_{RB} current is the flow of holes into the base to replace the holes lost by recombination.

J_{pE}: Due to the diffusion of minority carrier holes in the emitter at $x' = 0$.

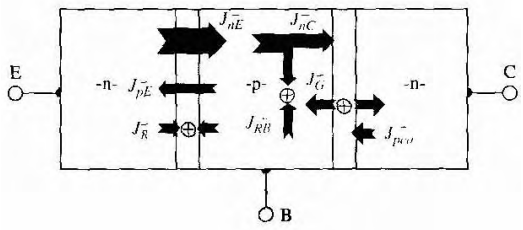


Figure 10.18 | Particle current density or flux components in an npn bipolar transistor operating in the fanvard-activemode

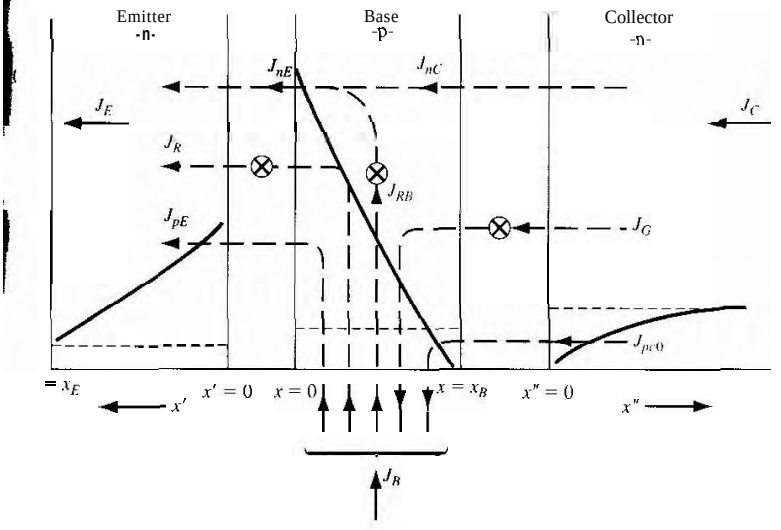


Figure 10.19 | Current density components in an npn bipolar transistor operating in the forward-active mode.

J_R: Due to the recombination of carriers in the forward-biased B-E junction.

J_{pc0}: Due to the diffusion of minority carrier holes in the collector at $x'' = 0$.

J_G: Due to the generation of carriers in the reverse-biased **B-C** junction.

The currents J_{RB}, J_{pE}, and J_R are B-E junction currents only and do not contribute to the collector current. The currents J_{pc0} and J_G are **B-C** junction currents only. These current components do not contribute to the transistor action or the current gain.

The dc common-base current gain is defined as

$$\alpha_0 = \frac{I_C}{I_E} \tag{10.27}$$

If we assume that the active cross-sectional area is the same for the collector and emitter, then we can write the current gain in terms of the current densities, or

$$\alpha_0 = \frac{J_C}{J_E} = \frac{J_{nC} + J_G + J_{pc0}}{J_{nE} + J_R + J_{pE}} \tag{10.28}$$

We are primarily interested in determining how the collector current will change with a change in emitter current. The small-signal, or sinusoidal, common-base current gain is defined as

$$\alpha = \frac{\partial J_C}{\partial J_E} = \frac{J_{nC}}{J_{nE} + J_R + J_{pE}} \tag{10.29}$$

The reverse-bias **B-C** currents, J_G and J_{pc0}, are not functions of the emitter current.

We can rewrite Equation (10.29) in the form

$$\alpha = \left(\frac{J_{nE}}{J_{nE}+J_{pE}}\right)\left(\frac{J_{nC}}{J_{nE}}\right)\left(\frac{J_{nE}+J_{pE}}{J_{nE}+J_R+J_{pE}}\right) \tag{10.30a}$$

or

$$a = \gamma\alpha_T\delta \tag{10.30b}$$

The factors in Equation (10.30b) are defined as:

$$\gamma = \left(\frac{J_{nE}}{J_{nE}+J_{pE}}\right) \equiv \text{emitter injection efficiency factor} \tag{10.31a}$$

$$\alpha_T = \left(\frac{J_{nC}}{J_{nE}}\right) \equiv \text{base transport factor} \tag{10.31b}$$

$$\delta = \frac{J_{nE}+J_{pE}}{J_{nE}+J_R+J_{pE}} \equiv \text{recombination factor} \tag{10.31c}$$

We would like to have the change in collector current be exactly the same as th change in emitter current or. ideally, to have $a = 1$. However, a consideration o Equation (10.29) shows that α will always be less than unity. The goal is to make as close to one as possible. To achieve this goal, we must make each term in Equ tion (10.30b) as close to one as possible, since each factor is less than unity.

The emitter injection *efficiency* factor γ takes into account the minority came hole diffusion current in the emitter. This current is part of the emitter current, bu does not contribute to the transistor action in that J_{pE} is not part of the collector cur rent. The *base transport factor* α_T takes into account any recombination of exce minority carrier electrons in the base. Ideally, we want no recombination in the base The recombinution factor δ takes into account the recombination in the forward biased B-E junction. The current J_R contributes to the emitter current, but does no contribute to collector current.

10.3.2 Mathematical Derivation of Current Gain Factors

We now wish to determine each of the gain factors in terms of the electrical and geometrical parameters of the transistor. The results of these derivations will show how the various parameters in the transistor influence the electrical properties of the device and will point the way to the design of a "good bipolar transistor.

Emitter Injection Efficiency Factor Consider, initially, the emitter injection effi- ciency factor. We have from Equation (10.31a)

$$\gamma = \left(\frac{J_{nE}}{J_{nE}+J_{pE}}\right) = \frac{1}{\left(1+\dfrac{J_{pE}}{J_{nE}}\right)} \tag{10.32}$$

We derived the minority carrier distribution functions for the forward-active mode in Section 10.2.1. Noting that J_{nE}, as defined in Figure 10.19, is in the negative

x direction, we can write the current densities as

$$J_{pE} = -eD_E \left.\frac{d(\delta p_E(x'))}{dx'}\right|_{x'=0} \tag{10.33a}$$

and

$$J_{nE} = (-)eD_B \left.\frac{d(\delta n_B(x))}{dx}\right|_{x=0} \tag{10.33b}$$

where $\delta p_E(x')$ and $\delta n_B(x)$ are given by Equations (10.21) and (10.15), respectively.

Taking the appropriate derivatives of $\delta p_E(x')$ and $\delta n_B(x)$, we obtain

$$J_{pE} = \frac{eD_E p_{E0}}{L_E}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \cdot \frac{1}{\tanh(x_E/L_E)} \tag{10.34a}$$

and

$$J_{nE} = \frac{eD_B n_{B0}}{L_B}\left\{\frac{1}{\sinh(x_B/L_B)} + \frac{[\exp(eV_{BE}/kT) - 1]}{\tanh(x_B/L_B)}\right\}$$

Positive J_{pE} and J_{nE} values imply that the currents are in the directions shown in Figure 10.19. If we assume that the B-E junction is biased sufficiently far in the forward bias so that $V_{BE} \gg kT/e$, then

$$\exp\left(\frac{eV_{BE}}{kT}\right) \gg 1$$

and also

$$\frac{\exp(eV_{BE}/kT)}{\tanh(x_B/L_B)} \gg \frac{1}{\sinh(x_B/L_B)}$$

The emitter injection efficiency, from Equation (10.32), then becomes

$$\boxed{\gamma = \frac{1}{1 + \dfrac{p_{E0}D_E L_B}{n_{B0}D_B L_E} \cdot \dfrac{\tanh(x_B/L_B)}{\tanh(x_E/L_E)}}} \tag{10.35a}$$

If we assume that all the parameters in Equation (10.35a) except p_{E0} and n_{B0} are fixed, then in order for $\gamma \approx 1$, we must have $p_{E0} \ll n_{B0}$. We can write

$$p_{E0} = \frac{n_i^2}{N_E} \qquad \text{and} \qquad n_{B0} = \frac{n_i^2}{N_B}$$

where N_E and N_B are the impurity doping concentrations in the emitter and base, respectively. Then the condition that $p_{E0} \ll n_{B0}$ implies that $N_E \gg N_B$. For the emitter injection efficiency to be close to unity, the emitter doping must be large compared to the base doping. This condition means that many more electrons from the n-type emitter than holes from the p-type base will be injected across the B-E space

charge region. If both $x_B \ll L_B$ and $x_E \ll L_E$, then the emitter injection efficiency can be written as

$$\gamma \approx \frac{1}{1 + \frac{N_B}{N_E} \cdot \frac{D_E}{D_B} \cdot \frac{x_B}{x_E}} \tag{10.35b}$$

Base Transport Factor The next term to consider is the base transport factor, given by Equation (10.31b) as $\alpha_T = J_{nC}/J_{nE}$. From the definitions of the current directions shown in Figure 10.19, we can write

$$J_{nC} = (-)eD_B \frac{d(\delta n_B(x))}{dx}\bigg|_{x=x_B} \tag{10.36a}$$

and

$$J_{nE} = (-)eD_B \frac{d(\delta n_B(x))}{dx}\bigg|_{x=0} \tag{10.36b}$$

Using the expression for $\delta n_B(x)$ given in Equation (10.15), we find that

$$J_{nC} = \frac{eD_B n_{B0}}{L_B} \left\{ [\exp(eV_{BE}/kT) - 1] + \frac{1}{\sinh(x_B/L_B)} - \frac{1}{\tanh(x_B/L_B)} \right\} \tag{10.37}$$

The expression for J_{nE} was given in Equation (10.34a).

If we again assume that the B-E junction is biased sufficiently far in the forward bias so that $V_{BE} \gg kT/e$, then $\exp(eV_{BE}/kT) \gg 1$. Substituting Equations (10.37) and (10.34b) into Equation (10.31b), we have

$$\alpha_T = \frac{J_{nC}}{J_{nE}} \approx \frac{\exp(eV_{BE}/kT) + \cosh(x_B/L_B)}{1 + \exp(eV_{BE}/kT)\cosh(x_B/L_B)} \tag{10.38}$$

In order for α_T to be close to unity, the neutral base width x_B must be much smaller than the minority carrier diffusion length in the base L_B. If $x_B \ll L_B$, then $\cosh(x_B/L_B)$ will be just slightly greater than unity. In addition, if $\exp(eV_{BE}/kT) \gg 1$, then the base transport factor is approximately

$$\alpha_T \approx \frac{1}{\cosh(x_B/L_B)} \tag{10.39a}$$

For $x_B \ll L_B$, we may expand the cosh function in a Taylor series, so that

$$\alpha_T = \frac{1}{\cosh(x_B/L_B)} \sim \frac{1}{1 + \frac{1}{2}(x_B/L_B)^2} \approx 1 - \frac{1}{2}(x_B/L_B)^2 \tag{10.39b}$$

The base transport factor α_T will be close to one if $x_B \ll L_B$. We can now see why we indicated earlier that the neutral base width x_B would be less than L_B.

Recombination Factor The recombination factor was given by Equation (10.31c). We can write

$$\delta = \frac{J_{nE} + J_{pE}}{J_{nE} + J_R + J_{pE}} \approx \frac{J_{nE}}{J_{nE} + J_R} = \frac{1}{1 + J_R/J_{nE}} \tag{10.40}$$

We have assumed in Equation (10.40) that $J_{pE} \ll J_{nE}$. The recombination current density, due to the recombination in a forward-biased pn junction, was discussed in Chapter 8 and can be written as

$$J_R = \frac{ex_{BE}n_i}{2\tau_0} \exp\left(\frac{eV_{BE}}{2kT}\right) = J_{r0} \exp\left(\frac{eV_{BE}}{2kT}\right) \tag{10.41}$$

where x_{BE} is the B-E space charge width.

The current J_{nE} from Equation (10 34b) can be approximated as

$$J_{nE} = J_{s0} \exp\left(\frac{eV_{BE}}{kT}\right) \tag{10.42}$$

where

$$J_{s0} = \frac{eD_B n_{B0}}{L_B \tanh(x_B/L_B)} \tag{10.41}$$

The recombination factor, from Equation (10.40), can then be written as

$$\boxed{\delta = \frac{1}{1 + \dfrac{J_{r0}}{J_{s0}} \exp\left(\dfrac{-eV_{BE}}{2kT}\right)}} \tag{10.44}$$

The recombination factor is a function of the B-E voltage. As V_{BE} increases, the recombination current becomes less dominant and the recombination factor approaches unity.

The recombination factor must also include surface effects. The surface effects can be described by the surface recombination velocity as we discussed in Chapter 6. Figure 10.20a shows the B-E junction of an npn transistor near the semiconductor surface. We will assume that the B-E junction is forward biased. Figure 10.20b shows the excess minority carrier electron concentration in the base along the cross section A-A'. This curve is the usual forward-biased junction minority carrier concentration. Figure 10.20c shows the excess minority carrier electron concentration along the cross section C-C' from the surface. We showed earlier that the excess concentration at a surface is smaller than the excess concentration in the bulk material. With this electron distribution, there is a diffusion of electrons from the bulk toward the surface where the electrons recombine with the majority carrier holes. Figure 10.20d shows the injection of electrons from the emitter into the base and the diffusion of electrons toward the surface. This diffusion generates another component of recombination current and this component of recombination current must be included in the recombination factor δ. Although the actual calculation is difficult because of the two-dimensional analysis required, the form of the recombination current is the same as that of Equation (10.41).

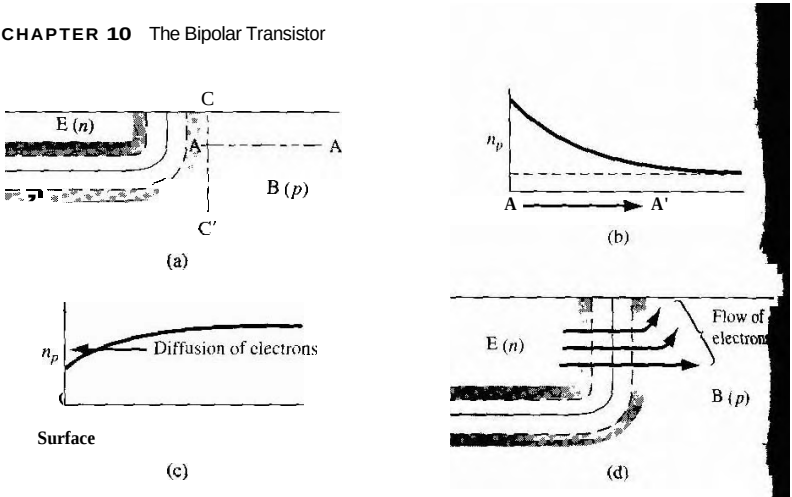


Figure 10.20 | The surface at the E-B junction showing the diffusion of carriers toward the surface.

10.3.3 Summary

Although we have considered an npn transistor in all of the derivations, exactly the same analysis applies to a pnp transistor; the same minority carrier distributions will be obtained except that the electron concentrations will become hole concentrations and vice versa. The current directions and voltage polarities will also change.

We have been considering the common-base current gain, defined in Equation (10.27) as $\alpha_0 = 1$. The common-emitter current gain is defined as $\beta_0 = I_C/I_B$. From Figure 10.8 we see that $I_E = I_B + I_C$. We can determine the relation between common-emitter and common-base current gains from the KCL equation. We can write

$$\frac{I_E}{I_C} = \frac{I_B}{I_C} + 1$$

Substituting the definitions of current gains, we have

$$\frac{1}{\alpha_0} = \frac{1}{\beta_0} + 1$$

Since this relation actually holds for hoth dc and small-signal conditions, we can drop the subscript. The common-emitter current gain can now be written in terms of the common-base current gain as

$$\beta = \frac{\alpha}{1 - \alpha}$$

The common-base current gain, in terms of the common-emitter current gain, is found to be

$$\alpha = \frac{\beta}{1 + \beta}$$

Table 10.3 | Summary of limiting factors

Emitter injection efficiency

$$\gamma \approx \frac{1}{1 + \frac{N_B}{N_E} \cdot \frac{D_E}{D_B} \cdot \frac{x_B}{x_E}} \qquad (x_B \ll L_B), (x_E \ll L_E)$$

Base transport factor

$$\alpha_T \approx \frac{1}{1 + \frac{1}{2}\left(\frac{x_B}{L_B}\right)^2} \qquad (x_B \ll L_B)$$

Recombination factor

$$\delta = \frac{1}{1 + \frac{J_{r0}}{J_{s0}} \exp\left(\frac{-eV_{BE}}{2kT}\right)}$$

Common-base current gain

$$\alpha = \gamma \alpha_T \delta \approx \frac{1}{1 + \frac{N_B}{N_E} \cdot \frac{D_E}{D_B} \cdot \frac{x_B}{x_E} + \frac{1}{2}\left(\frac{x_B}{L_B}\right)^2 + \frac{J_{r0}}{J_{s0}} \exp\left(\frac{-eV_{BE}}{2kT}\right)}$$

Common-emitter current gain

$$\beta = \frac{\alpha}{1-\alpha} \approx \frac{1}{\frac{N_B}{N_E} \cdot \frac{D_E}{D_B} \cdot \frac{x_B}{x_E} + \frac{1}{2}\left(\frac{x_B}{L_B}\right)^2 + \frac{J_{r0}}{J_{s0}} \exp\left(\frac{-eV_{BE}}{2kT}\right)}$$

Table 10.3 summarizes the expressions for the limiting factors in the common base current gain assuming that $x_B \ll L_B$ and $x_E \ll L_E$. Also given are the approximate expressions for the common-base current gain and the common-emitter current gain.

10.3.4 Example Calculations of the Gain Factors

If we assume a typical value of β to be 100, then $\alpha = 0.99$. If we also assume that $\gamma = \alpha_T = \delta$, then each factor would have to be equal to 0.9967 in order that $\beta = 100$. This calculation gives an indication of how close to unity each factor must be in order to achieve a reasonable current gain.

DESIGN EXAMPLE 10.1

Objective

To design the ratio of emitter doping to base doping in order to achieve an emitter injection efficiency factor equal to $\gamma = 0.9967$.

Consider an npn bipolar transistor. Assume, for simplicity, that $D_E = D_B$, $L_E = L_B$, and



Solution

Equation (10.35a) reduces to

$$\gamma = \frac{1}{1 + \dfrac{p_{E0}}{n_{B0}}} = \frac{1}{1 + \dfrac{n_i^2/N_E}{n_i^2/N_B}}$$

so

$$\gamma = \frac{1}{1 + \dfrac{N_B}{N_E}} = 0.9967$$

Then

$$\frac{N_B}{N_E} = 0.00331 \qquad \text{or} \qquad \frac{N_E}{N_A} = 302$$

■ Comment

The emitter doping concentration must be much larger than the base doping concentration to achieve a high emitter injection efficiency.

DESIGN EXAMPLE 10.2

Objective

To design the base width required to achieve a base transport factor equal to $\alpha_T = 0.9967$.

Consider a pnp bipolar transistor. Assume that $D_B = 10\ \text{cm}^2/\text{s}$ and $\tau_{B0} = 10^{-7}$ s.

Solution

The base transport factor applies to both pnp and npn transistors and is given by

$$\alpha_T = \frac{1}{\cosh(x_B/L_B)} = 0.9967$$

Then

$$x_B/L_B = 0.0814$$

We have

$$L_B = \sqrt{D_B \tau_{B0}} = \sqrt{(10)(10^{-7})} = 10^{-3}\ \text{cm}$$

so that the base width must then be

$$x_B = 0.814 \times 10^{-4}\ \text{cm} = 0.813\ \mu\text{m}$$

■ Comment

If the base width is less than approximately $0.8\ \mu$m, then the required base transport factor will be achieved. In most cases, the base transport factor will not be the limiting factor in the bipolar transistor current gain.

Objective | EXAMPLE 10.3

To calculate the forward-biased B-E voltage required to achieve a recombination factor equal to6 = 0.9967.

Consider an npn bipolar transistor at $T = 300$ K. Assume that $J_{r0} = 10^{-8}$ A/cm^2 and that $J_{s0} = 10^{-11}$ A/cm^2.

■ **Solution**

The recombination factor, from Equation (10.44), is

$$\delta = \frac{1}{1 + \frac{J_{r0}}{J_{s0}} \exp\left(\frac{-eV_{BE}}{2kT}\right)}$$

We then have

$$0.9967 = \frac{1}{1 + \frac{10^{-8}}{10^{-11}} \exp\left(\frac{-eV_{BE}}{2kT}\right)}$$

We can rearrange this equation and write

$$\exp\left(\frac{+eV_{BE}}{2kT}\right) = \frac{0.9967 \times 10^3}{1 - 0.9967} = 3.02 \times 10^5$$

$$V_{BE} = 2(0.0259) \ln(3.02 \times 10^5) = 0.654 \text{ V}$$

■ **Comment**

This example demonstrates that the recombination factor may be an important limiting factor in the bipolar current gain. In this example, if V_{BE} is smaller than 0.654 V, then the recombination factor S will fall below the desired 0.9967 value.

Objective | EXAMPLE 10.4

To calculate the common-emitter current gain of a silicon npn bipolar transistor at $T = 300$ K given a set of parameters.

Assume the following parameters:

$D_E = 10$ cm^2/s	$x_B = 0.70\ \mu$m
$D_B = 25$ cm^2/s	$x_E = 0.50\ \mu$m
$\tau_{E0} = 1 \times 10^{-7}$ s	$N_E = 1 \times 10^{18}$ cm^{-3}
$\tau_{B0} = 5 \times 10^{-7}$ s	$N_B = 1 \times 10^{16}$ cm^{-3}
$J_{r0} = 5 \times 10^{-8}$ A/cm^2	$V_{BE} = 0.65$ V

The following parameters are calculated:

$$p_{E0} = \frac{(1.5 \times 10^{10})^2}{1 \times 10^{18}} = 2.25 \times 10^2 \text{ cm}^{-3}$$

$$n_{B0} = \frac{(1.5 \times 10^{10})^2}{1 \times 10^{16}} = 2.25 \times 10^4 \text{ cm}^{-3}$$

$$L_E = \sqrt{D_E \tau_{E0}} = 10^{-3} \text{ cm}$$

$$L_B = \sqrt{D_B \tau_{B0}} = 3.54 \times 10^{-3} \text{ cm}$$

■ Solution

The emitter injection efficiency factor, from Equation (10.35a), is

$$\gamma = \frac{1}{1 + \dfrac{(2.25 \times 10^2)(10)(3.54 \times 10^{-3})}{(2.25 \times 10^4)(25)(10^{-3})} \cdot \dfrac{\tanh(0.0198)}{\tanh(0.050)}} = 0.9944$$

The base transport factor, from Equation (10.39a) is

$$\alpha_T = \frac{1}{\cosh\left(\dfrac{0.70 \times 10^{-4}}{3.54 \times 10^{-3}}\right)} = 0.9998$$

The recombination factor, from Equation (10.44), is

$$\delta = \frac{1}{1 + \dfrac{5 \times 10^{-8}}{J_{s0}} \exp\left(\dfrac{-0.65}{2(0.0259)}\right)}$$

where

$$J_{s0} = \frac{eD_B n_{B0}}{L_B \tanh\left(\dfrac{x_B}{L_B}\right)} = \frac{(1.6 \times 10^{-19})(25)(2.25 \times 10^4)}{3.54 \times 10^{-3} \tanh(1.977 \times 10^{-2})} = 1.29 \times 10^{-9} \text{ A/cm}^2$$

We can now calculate $\delta = 0.99986$. The common-base current gain is then

$$\alpha = \gamma \alpha_T \delta = (0.9944)(0.9998)(0.99986) = 0.99406$$

which gives a common-emitter current gain of

$$\beta = \frac{\alpha}{1 - \alpha} = \frac{0.99406}{1 - 0.99406} = 167$$

■ Comment

In this example, the emitter injection efficiency is the limiting factor in the current gain.

TEST YOUR UNDERSTANDING

NOTE: In Exercises E10.4 through E10.9, assume a silicon npn bipolar transistor at T = 300 K has the following minority carrier parameters: $D_E = 8 \text{ cm}^2/\text{s}$, $D_B = 20 \text{ cm}^2/\text{s}$, $D_C = 12 \text{ cm}^2/\text{s}$, $\tau_{E0} = 10^{-8}$ s, $\tau_{B0} = 10^{-7}$ s, $\tau_{C0} = 10^{-6}$ s.

E10.4 If the emitter doping concentration is $N_E = 5 \times 10^{18}$ cm^{-3}, find the base doping concentration such that the emitter injection efficiency is $\gamma = 0.9950$. Assume $x_E = 2x_B = 2\ \mu$m. ($N_B = 1.03 \times 10^{16}$ cm^{-3})

E10.5 Assume that $\alpha_T = \delta = 0.9967, x_B = x_E = 1\ \mu$m, $N_B = 5 \times 10^{16}$ cm^{-3}, and $N_E = 5 \times 10^{18}$ cm^{-3}. Determine the common emitter current gain β. ($\beta = 276$)

E10.6 Determine the minimum base width x_B such that the base transport factor is $\alpha_T = 0.9980$. ($x_B = 0.896\ \mu$m)

E10.7 Assume that $\gamma = \delta = 0.9967$ and $x_B = 0.80\ \mu$m. Determine the common-emitter current gain β. ($\beta = 121$)

E10.8 If $J_{r0} = 10^{-8}$ A/cm^2 and $J_{s0} = 10^{-11}$ A/cm^2, find the value of V_{BE} such that $\delta = 0.9960$. ($V_{BE} = 0.649$ V)

E10.9 Assume that $\gamma = \alpha_T = 0.9967$. $J_{r0} = 5 \times 10^{-9}$ A/cm^2, $J_{s0} = 10^{-11}$ A/cm^2, and $V_{BE} = 0.585$ V. Determine the common-emitter current gain β. ($\beta = 77.4$)

10.4 | NONIDEAL EFFECTS

In all previous discussions, we have considered a transistor with uniformly doped regions, low injection, constant emitter and base widths, an ideal constant energy bandgap, uniform current densities, and junctions which are not in breakdown. If any of these ideal conditions are not present, then the transistor properties will deviate from the ideal characteristics we have derived.

10.4.1 Base Width Modulation

We have implicitly assumed that the neutral base width x_B was constant. This base width, however, is a function of the **B-C** voltage, since the width of the space charge region extending into the base region varies with B-C voltage. As the **B-C** reverse-bias voltage increases, the B-C space charge region width increases, which reduces x_B. A change in the neutral base width will change the collector current as can be observed in Figure 10.21. A reduction in base width will cause the gradient in the minority carrier concentration to increase, which in turn causes an increase in the diffusion current. This effect is known as ***base width modulation;*** it is also called the ***Early effect.***

The Early effect can be seen in the current-voltage characteristics shown in Figure 10.22. In most cases, a constant base current is equivalent to a constant B-E voltage. Ideally the collector current is independent of the **B-C** voltage so that the slope of the curves would be zero; thus the output conductance of the transistor would be zero. However, the base width modulation, or Early effect, produces a nonzero slope and gives rise to a finite output conductance. If the collector current characteristics are extrapolated to zero collector current, the curves intersect the voltage axis at a point that is defined as the Early voltage. The Early voltage is considered to be a positive value. It is a common parameter given in transistor specifications; typical values of Early voltage are in the 100- to 300-volt range.

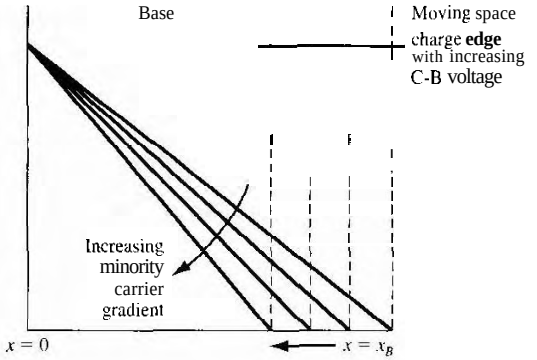


Figure 10.21 | The change in the base width and the change in the minority carrier gradient as the B-C space charge width changes.

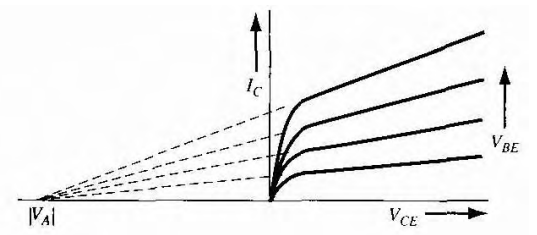


Figure 10.22 | The collector current versus collector–emitter voltage showing the Early effect and Early voltage.

From Figure 10.22, we can write that

$$\frac{dI_C}{dV_{CE}} \equiv g_0 = \frac{I_C}{V_{CE} + V_A} \tag{10.45a}$$

where V_A and V_{CE} are defined as positive quantities and g_0 is defined as the output conductance. Equation (10.45a) can be rewritten in the form

$$I_C = g_0(V_{CE} + V_A) \tag{10.45b}$$

showing explicitly that the collector current is now a function of the C-E voltage or the C-B voltage.

EXAMPLE 10.5 | Objective

To calculate the change in the neutral base width with a change in C-B voltage.

Consider a uniformly doped silicon bipolar transistor at T = 300 K with a base doping of $N_B = 5 \times 10^{16}$ cm^{-3} and a collector doping of $N_C = 2 \times 10^{15}$ cm^{-3}. Assume the metallurgical

base width is 0.70 μm. Calculate the change in the neutral base width as the C-B voltage changes from 2 to 10 V.

■ Solution

The space charge width extending into the base region can be written as

$$x_{dB} = \left\{\frac{2\epsilon_s(V_{bi}+V_{CB})}{e}\left[\frac{N_C}{N_B}\cdot\frac{1}{(N_B+N_C)}\right]\right\}^{1/2}$$

$$x_{dB} = \left\{\frac{2(11.7)(8.85\times10^{-14})(V_{bi}+V_{CB})}{1.6\times10^{-19}}\times\left[\frac{2\times10^{15}}{5\times10^{16}}\cdot\frac{1}{(5\times10^{16}+2\times10^{15}}\right]\right\}^{1/2}$$

which becomes

$$x_{dB} = ((9.96\times10^{-12})(V_{bi}+V_{CB}))^{1/2}$$

The built-in potential is

$$V_{bi} = \frac{kT}{e}\ln\left[\frac{N_B N_C}{n_i^2}\right] = 0.718 \text{ V}$$

For $V_{CB} = 2$ V, we find $x_{dB} = 0.052$ μm, and for $V_{CB} = 10$ V, we find $x_{dB} = 0.103$ μm. If we neglect the B-E space charge region, which will be small because of the forward-biased junction, then we can calculate the neutral base width. For $V_{CB} = 2$ V.

$$x_B = 0.70 - 0.052 = 0.648 \ \mu\text{m}$$

and for $V_{CB} = 10$ V,

$$x_B = 0.70 - 0.103 = 0.597 \ \mu\text{m}$$

■ Comment

This example shows that the neutral base width can easily change by approximately 8 percent as the C-B voltage changes from 2 to 10 V.

Objective

EXAMPLE 10.6

To calculate the change in collector current with a change in neutral base width, and to estimate the Early voltage.

Consider a uniformly doped silicon npn bipolar transistor with parameters described in Example 10.5. Assume $D_B = 25$ cm^2/s, and $V_{BE} = 0.60$ V, and also assume that $x_B \ll L_B$.

■ Solution

The excess minority carrier electron concentration in the base is given by Equation (10.15) as

$$\delta n_B(x) = \frac{n_{B0}\left\{\left[\exp\left(\frac{eV_{BE}}{kT}\right)-1\right]\sinh\left(\frac{x_B-x}{L_B}\right)-\sinh\left(\frac{x}{L_B}\right)\right\}}{\sinh\left(\frac{x_B}{L_B}\right)}$$

If $x_B \ll L_B$, then $(x_B - x) \ll L_B$ so we can write the approximations

$$\sinh\left(\frac{x_B}{L_B}\right) \quad \left(\frac{x_B}{L_B}\right) \qquad \text{and} \qquad \sinh\left(\frac{x_B - x}{L_B}\right) \approx \left(\frac{x_B - x}{L_B}\right)$$

The expression for $\delta n_B(x)$ can then he approximated as

$$\delta n_B(x) \approx \frac{n_{B0}}{x_B}\left\{\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right](x_B - x) - x\right\}$$

The collector current is now

$$|J_C| = eD_B\frac{d(\delta n_B(x))}{dx} \approx \frac{eD_B n_{B0}}{x_B}\exp\left[\frac{eV_{BE}}{kT}\right]$$

The value of n_{B0} is calculated as

$$n_{B0} = \frac{n_i^2}{N_B} = \frac{(1.5 \times 10^{10})^2}{5 \times 10^{16}} = 4.5 \times 10^3 \text{ cm}^{-3}$$

If we let $x_B = 0.648\,\mu\text{m}$ when $V_{CB} = 2\text{ V}$ ($V_{CE} = 2.6\text{ V}$), then

$$|J_C| = \frac{(1.6 \times 10^{-19})(25)(4.5 \times 10^3)}{0.648 \times 10^{-4}}\exp\left(\frac{0.60}{0.0259}\right) = 3.20\text{ A/cm}^2$$

Now let $x_B = 0.597\,\mu\text{m}$ when $V_{CB} = 10\text{ V}$ ($V_{CE} = 10.6\text{ V}$). In this case we have $|J_C| = 3.47\text{ A/cm}^2$. From Equation (10.45a), we can write

$$\frac{dJ_C}{dV_{CE}} = \frac{J_C}{V_{CE} + V_A} = \frac{\Delta J_C}{\Delta V_{CE}}$$

Using the calculated values of current and voltage, we have

$$\frac{\Delta J_C}{\Delta V_{CE}} = \frac{3.47 - 3.20}{10.6 - 2.6} = \frac{J_C}{V_{CE} + V_A} \approx \frac{3.20}{2.6 + V_A}$$

The Early voltage is then determined to be

$$V_A \approx 92\text{ V}$$

■ Comment

This example indicates how much the collector current can change as the neutral base width changes with a change in the B-C space charge width. and it also indicates the magnitude of the Early voltage.

The example demonstrates, too, that we can expect variations in transistor properties due to tolerances in transistor-fabrication processes. There will be variations, in particular, in the base width of narrow-base transistors that will cause variations in the collector current characteristics simply due to the tolerances in processing.

TEST YOUR UNDERSTANDING

E10.10 A particular transistor has an output resistance of 200 kΩ and an Early voltage of $V_A = 125$ V. Determine the change in collector current when V_{CE} increases from 2 V to 8 V. (Ans. 0.03 mA)

E10.11 (*a*) If, because of fabrication tolerances, the neutral base width for a set of transistors varies over the range of $0.800 \leq x_B \leq 1.00\ \mu$m, determine the variation in the base transport factor α_T. Assume $L_B = 1.414 \times 10^{-3}$ cm. (*b*) Using the results of part (*a*) and assuming $\gamma = \delta = 0.9967$, what is the variation in common emitter current gain. [Ans. (*a*) $0.9975 \leq \alpha_T \leq 0.9984$, (*b*) $109 \leq \beta \leq 121$]

10.4.2 High Injection

The ambipolar transport equation that we have used to determine the minority carrier distributions assumed low injection. As V_{BE} increases, the injected minority carrier concentration may approach, or even become larger than, the majority carrier concentration. If we assume quasi-charge neutrality, then the majority carrier hole concentration in the p-type base at $x = 0$ will increase as shown in Figure 10.23 because of the excess holes.

Two effects occur in the transistor at high injection. The first effect is a reduction in emitter injection efficiency. Since the majority carrier hole concentration at $x = 0$ increases with high injection, more holes are injected back into the emitter because of the forward-biased B-E voltage. An increase in the hole injection causes an increase in the J_{pE} current, and an increase in J_{pE} reduces the emitter injection efficiency. The common, emitter current gain decreases, then, with high injection. Figure 10.24 shows a typical common-emitter current gain versus collector current

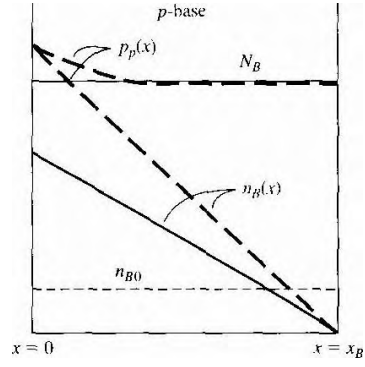


Figure 10.23 | Minority and majority carrier concentrations in the base under low and high injection (solid line: low injection: dashed line: high injection).

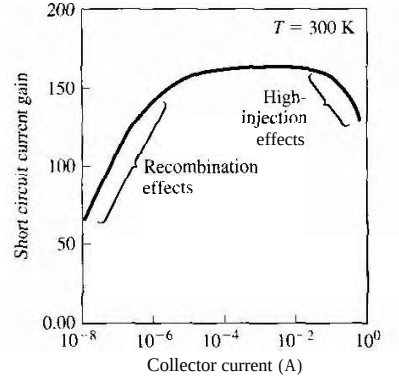


Figure 10.24 | Common-emitter current gain versus collector current. *(From Shur [13].)*

curve. The low gain at low currents is due to the small recombination factor and the drop-off at the high current is due to the high-injection effect.

We will now consider the second high-injection effect. At low injection, the majority carrier hole concentration at $x = 0$ for the npn transistor is

$$p_p(0) = p_{p0} = N_a \tag{10.46a}$$

and the minority carrier electron concentration is

$$n_p(0) = n_{p0} \exp\left(\frac{eV_{BE}}{kT}\right) \tag{10.46b}$$

The pn product is

$$p_p(0)n_p(0) = p_{p0}n_{p0} \exp\left(\frac{eV_{BE}}{kT}\right) \tag{10.46c}$$

At high injection, Equation (10.46c) still applies. However, $p_p(0)$ will also increase, and for very high injection it will increase at nearly the same rate as $n_p(0)$. The increase in $n_p(0)$ will asymptotically approach the function

$$n_p(0) \approx n_{p0} \exp\left(\frac{eV_{BE}}{2kT}\right) \tag{10.47}$$

The excess minority carrier concentration in the base, and hence the collector current, will increase at a slower rate with B-E voltage in high injection than low injection. This effect is shown in Figure *10.25*. The high-injection effect is very similar to the effect of a series resistance in a pn junction diode.

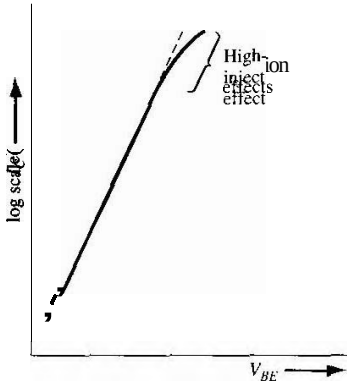


Figure 10.25 | Collector current versus base–emitter voltage showing high-injection effects.

10.4.3 Emitter Bandgap Narrowing

Another phenomenon affecting the emitter injection efficiency is bandgap narrowing. We implied from our previous discussion that the emitter injection efficiency (actor would continue to increase and approach unity as the ratio of emitter doping to base doping continued to increase. As silicon becomes heavily doped, the discrete donor energy level in an n-type emitter splits into a band of energies. The distance between donor atoms decreases as the concentration of impurity donor atoms increases and the splitting of the donor level is caused by the interaction of donor atoms with each other. As the doping continues to increase, the donor hand widens, becomes skewed, and moves up toward the conduction hand, eventually merging with it. At this point, the effective bandgap energy has decreased. Figure 10.26 shows a plot of the change in the bandgap energy with impurity doping concentration.

A reduction in the bandgap energy increases the intrinsic carrier concentration. The intrinsic carrier concentration is given by

$$n_i^2 = \mathrm{N}, N_v \exp\left(\frac{-E_g}{kT}\right) \tag{10.48}$$

In a heavily doped emitter, the intrinsic carrier concentration can he written as

$$n_{iE}^2 = N_c \mathrm{N}, \exp\left[\frac{-(E_{g0} - \Delta E_g)}{kT}\right] = n_i^2 \exp\left(\frac{\Delta E_g}{kT}\right) \tag{10.49}$$

where E_{g0} is the bandgap energy at a low doping concentration and $\boldsymbol{AE}$, is the bandgap narrowing factor.

The emitter injection efficiency factor was given by Equation (10.35) as

$$\gamma = \frac{1}{1 + \dfrac{p_{E0} D_E L_B}{n_{B0} D_B L_E} \cdot \dfrac{\tanh\,(x_B/L_B)}{\tanh\,(x_E/L_E)}}$$

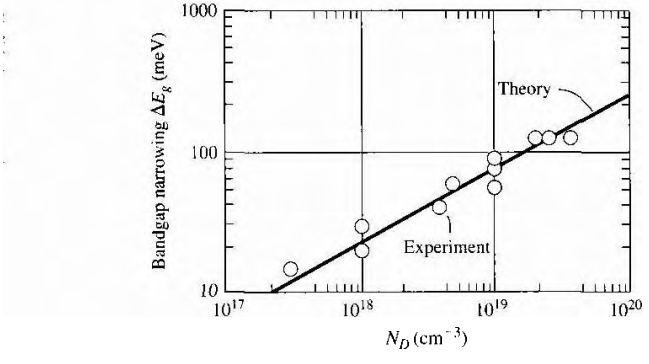


Figure 10.26 | Bandgap-narrowing factor **versus** donor impurity concentration in silicon.
(From Sze [18].)

The term p_{E0} is the thermal-equilibrium minority carrier concentration in the emitter and must be written as

$$p_{E0} = \frac{n_{iE}^2}{N_E} = \frac{n_i^2}{N_E} \exp\left(\frac{\Delta E_g}{kT}\right) \quad (10$$

As the emitter doping increases, ΔE_g increases; thus, p_{E0} does not continue to decrease with increased emitter doping. If p_{E0} starts to increase because of the bandgap narrowing, the emitter injection efficiency begins to fall off instead of continuing increase with increased emitter doping.

EXAMPLE 10.7

Objective

To determine the increase in p_{E0} in the emitter due to bandgap narrowing.

Consider a silicon emitter at T = 300 K. Assume the emitter doping increases fr 10^{18} cm^{-3} to 10^{19} cm^{-3}. Calculate the change in the p_{E0} value.

Solution

For emitter dopings of $N_E = 10^{18}$ cm^{-3} and 10^{19} cm^{-3}, we have, neglecting bandgap narrowing,

$$p_{E0} = \frac{n_i^2}{N_E} = \frac{(1.5 \times 10^{10})^2}{10^{18}} = 2.25 \times 10^2 \text{ cm}^{-3}$$

and

$$p_{E0} = \frac{(1.5 \times 10^{10})^2}{10^{19}} = 2.25 \times 10^1 \text{ cm}^{-3}$$

Taking into account the bandgap narrowing, we obtain, respectively, for $N_E = 10^{18}$ cm^{-3} and $N_E = 10^{19}$ cm^{-3}

$$p_{E0} = \frac{(1.5 \times 10^{10})^2}{10^{18}} \exp\left(\frac{0.030}{0.0259}\right) = 7.16 \times 10^2 \text{ cm}^{-3}$$

and

$$p_{E0} = \frac{(1.5 \times 10^{10})^2}{10^{19}} \exp\left(\frac{0.1}{0.0259}\right) = 1.07 \times 10^3 \text{ cm}^{-3}$$

■ Comment

If the emitter doping increases from 10^{18} to 10^{19} cm^{-3}, the thermal-equilibrium minority carrier concentration actually increases by a factor of 1.5 instead of decreasing by the expected factor of 10. This effect is due to bandgap narrowing.

As the emitter doping increases, the bandgap narrowing factor, AE,, will increase; this can actually cause p_{E0} to increase. As p_{E0} increases, the emitter injection efficiency decreases; this then causes the transistor gain to decrease, as in Figure 10.24.

A very high emitter doping may result in a smaller current gain than we anticipate because of the handgap-narrowing effect.

10.4.4 Current Crowding

It is tempting to minimize the effects of base current in a transistor since the base current is usually much smaller than either the collector or the emitter current. Figure 10.27 is a cross section of an npn transistor showing the lateral distribution of base current. The base region is typically less than a micrometer thick, so there can be a sizable base resistance. The nonzero base resistance results in a lateral potential difference under the emitter region. For the npn transistor, the potential decreases from the edge of the emitter toward the center. The emitter is highly doped, so as a first approximation the emitter can be considered an equipotential region.

The number of electrons from the emitter injected into the base is exponentially dependent on the B-E voltage. With the lateral voltage drop in the base between the edge and center of the emitter, more electrons will be injected near the emitter edges than in the center, causing the emitter current to be crowded toward the edges. This current-crowding effect is schematically shown in Figure 10.28. The larger current density near the emitter edge may cause localized heatingeffects as well as localized high-injection effects. The nonuniform emitter current also results in a nonuniform lateral base current under the emitter. A two-dimensional analysis would he required to calculate the actual potential drop versus distance because of the nonuniform base current. Another approach is to slice the transistor into a number of smaller parallel transistors and to lump the resistance of each base section into an equivalent external resistance.

Power transistors, designed to handle large currents, require large emitter areas to maintain reasonable current densities. To avoid the current-crowding effect, these transistors are usually designed with narrow emitter widths and fabricated with an interdigitated design. Figure 10.29 shows the basic geometry. In effect, many narrow emitters are connected in parallel to achieve the required emitter area.

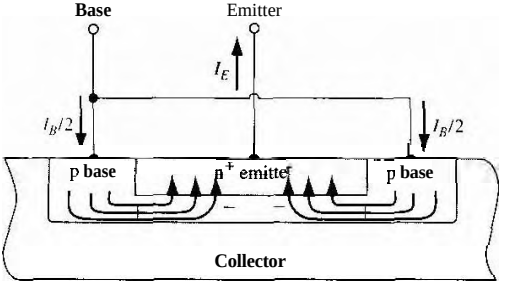


Figure 10.27 | Cross section of an npn bipolar transistor showing the base current distribution and the lateral potential drop in the base region.

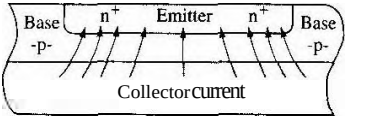


Figure 10.28 | Cross section of an npn bipolar transistor showing the emitter current-crowding effect.

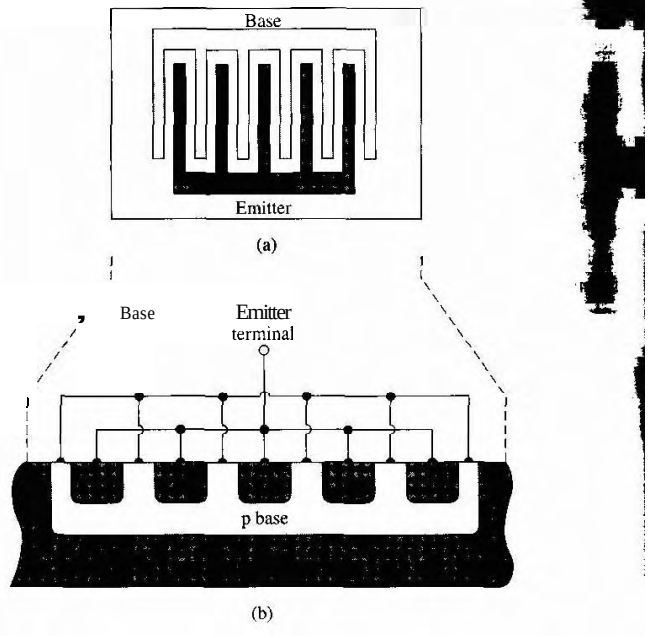


Figure 10.29 | (a) Top view and (b) cross section of an interdigitated npn bipolar transistor structure.

TEST YOUR UNDERSTANDING

E10.12 Consider the geometry shown in Figure 10.30. The base doping concentration is $N_B = 10^{16}$ cm^{-3}, the neutral base width is $x_B = 0.80\ \mu$m, the emitter width is $S = 10\ \mu$m, and the emitter length is L $= 10\ \mu$m. (a) Determine the resistance of the base between x $= 0$ and $x = S/2$. Assume a hole mobility of $\mu_p = 400$ cm^2/V-s. (b) If the base current in this region is uniform and given by $I_B/2 = 5\ \mu$A, determine the potential difference between $x = 0$ and x $= S/2$. (c) Using the results of part (b), what is the ratio of emitter current density at x $= 0$ to that at x $= S/2$?

[Ans. (a) 9.77 kΩ, (b) 48.85 mV, (c) 6.59]

*10.4.5 Nonuniform Base Doping

In the analysis of the bipolar transistor, we assumed uniformly doped regions. However, uniform doping rarely occurs. Figure 10.31 shows a doping profile in a doubly diffused npn transistor. We can start with a uniformly doped n-type substrate, diffuse acceptor atoms from the surface to form a compensated p-type base, and then diffuse donor atoms from the surface to form a doubly compensated n-type emitter. The diffusion process results in a nonuniform doping profile.

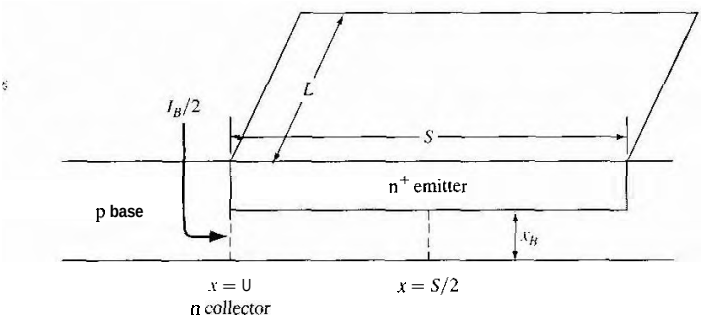


Figure 10.30 | Figure for E10.12.

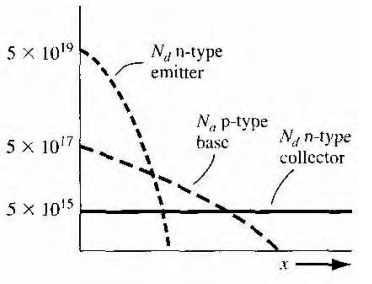


Figure 10.31 | Impurity concentration profiles of a double-diffused npn bipolar transistor.

We determined in Chapter 5 that a graded impurity concentration leads to an induced electric field. For the p-type base region in thermal equilibrium, we can write

$$J_p = e\mu_p N_a \mathrm{E} - eD_p \frac{dN_a}{dx} = 0 \tag{10.51}$$

Then

$$\mathrm{E} = +\left(\frac{kT}{e}\right)\frac{1}{N_a}\frac{dN_a}{dx} \tag{10.52}$$

According to the example of Figure 10.31, dN_a/dx is negative; hence, the induced electric field is in the negative x direction.

Electrons are injected from the n-type emitter into the base and the minority carrier base electrons begin diffusing toward the collector region. The induced electric field in the base, because of the nonuniform doping, produces a force on the electrons in the direction toward the collector. The induced electric field, then, aids the flow of minority carriers across the base region. This electric field is called an ***accelerating field.***

The accelerating field will produce a drift component of current that is in addition to the existing diffusion current. Since the minority carrier electron concentration varies across the base, the drift current density will not he constant. The total current across the base, however, is nearly constant. The induced electric field in the base due to nonuniform base doping will alter the minority carrier distribution through the base so that the sum of drift current and diffusion current will be a constant. Calculations have shown that the uniformly doped base theory is very useful for estimating the base characteristics.

10.4.6 Breakdown Voltage

There are two breakdown mechanisms to consider in a bipolar transistor. The first is called punch-through. As the reverse-bias B-C voltage increases, the B-C space charge region widens and extends farther into the neutral base. It is possible for the B-C depletion region to penetrate completely through the base and reach the B-E space charge region, the effect called *punch-through*. Figure 10.32a shows the energy-band diagram of an npn bipolar transistor in thermal equilibrium and Figure 10.32b shows the energy-band diagram for two values of reverse-bias B-C junction voltage. When a small C-B voltage, V_{R1}, is applied, the B-E potential barrier is not affected; thus, the transistor current is still essentially zero. When a large reverse-bias voltage, V_{R2}, is applied, the depletion region extends through the base region and the B-E potential barrier is lowered because of the C-B voltage. The lowering of the potential banier at the B-E junction produces a large increase in current with a very small increase in C-B voltage. This effect is the punch-through breakdown phenomenon.

Figure 10.33 shows the geometry for calculating the punch-through voltage. Assume that N_B and N_C are the uniform impurity doping concentrations in the base and collector, respectively. Let W_B be the metallurgical width of the base and let x_{dB} be the space charge width extending into the base from the B-C junction. If we neglect the

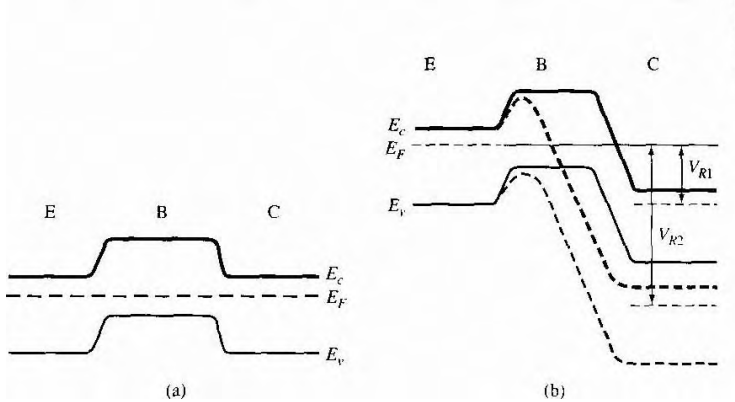


Figure 10.32 | Energy-band diagram of an npn bipolar transistor (a) in thermal equilibrium. and (b) with a reverse-bias B-C voltage before punch-through, V_{R1} and after punch-through, V_{R2}.

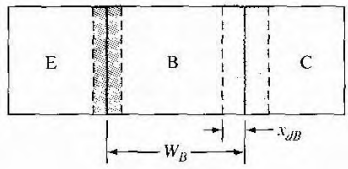


Figure 10.33 | Geometry of a bipolar transistor to calculate the punch-through voltage.

narrow space charge width of a zero-biased or forward-biased B-E junction. then punch-through. assuming the abrupt junction approximation, occurs when $x_{dB} = W_B$. We can write that

$$x_{dB} = W_B = \left\{ \frac{2\epsilon_s(V_{bi} + V_{pt})}{e} \cdot \frac{N_C}{N_B} \cdot \frac{1}{N_C + N_B} \right\}^{1/2} \tag{10.53}$$

where V_{pt} is the reverse-biased B-C voltage at punch-through. Neglecting V_{bi} compared to V_{pt}, we can solve for V_{pt} as

$$\boxed{V_{pt} = \frac{eW_B^2}{2\epsilon_s} \cdot \frac{N_B(N_C + N_B)}{N_C}} \tag{10.54}$$

Objective

DESIGN EXAMPLE 10.8



To design the collector doping and collector width to meet a punch-through voltage specification.

Consider a uniformly doped silicon bipolar transistor with a metallurgical base width of 0.5 μm and a base doping of $N_B = 10^{16}$ c m ^{3}. The punch-through voltage is to be $V_{pt} = 25$ V.

Solution

The maximum collector doping concentration can be determined from Equation (10.54) as

$$25 = \frac{(1.6 \times 10^{-19})(0.5 \times 10^{-4})^2(10^{16})(N_C + 10^{16})}{2(11.7)(8.85 \times 10^{-14})N_C}$$

$$12.94 = 1 + \frac{10^{16}}{N_C}$$

which yields

$$N_C = 8\ 38 \times 10^{14}\ \text{cm}^{-3}$$

This n-type doping concentration in the collector must extend at least as far as the depletion width extends into the collector to avoid breakdown in the collector region. We have, using

results from Chapter 7,

$$x_n = \left[\frac{2\epsilon_s(V_{bi} + V_R)}{e}\left(\frac{N_B}{N_C}\cdot\frac{1}{N_B + N_C}\right)\right]^{1/2}$$

Neglecting V_{bi} compared to $V_R = V_{pt}$, we obtain

$$x_n = \left[\frac{2(11.7)(8.85 \times 10^{-14})(25)}{1.6 \times 10^{-19}}\left(\frac{10^{16}}{8.38 \times 10^{14}}\right)\left(\frac{1}{10^{16} + 8.38 \times 10^{14}}\right)\right]^{1/2}$$

or

$$x_n = 5.97\ \mu\text{m}$$

■ Comment

From Figure 8.25, the expected avalanche breakdown voltage for this junction is greater than 300 volts. Obviously punch-through will occur before the normal breakdown voltage in this case. For a larger punch-through voltage, a larger metallurgical hase width will be required since a lower collector doping concentration is becoming impractical. A larger punch-through voltage will also require a larger collector width in order to avoid premature breakdown in this region.

TEST YOUR UNDERSTANDING

E10.13 The metallurgical hase width of a silicon npn bipolar transistor is $W_B = 0.80\ \mu$m. The base and collector doping concentrations are $N_B = 2 \times 10^{16}$ cm^{-3} and $N_C = 10^{\prime\prime}$ cm^{-3}. Find the punch-through breakdown voltage. (Ans. 80Z 'suV)

E10.14 The base impurity doping concentration is $N_B = 3 \times 10^{16}$ cm^{-3} and the metallurgical base width is $W_B = 0.70\ \mu$m. The minimum required punch-through breakdown voltage is specified to be $V_{pt} = 70$ V. What is the maximum allowed collectordoping concentration? (Ans. $N_C = 5.81 \times 10^{15}$ cm^{-3} 'suV)

The second breakdown mechanism to consider is avalanche breakdown, but taking into account the gain of the transistor.[1] Figure 10.34a is an npn transistor with a reverse-bias voltage applied to the B-C junction and with the emitter left open. The current I_{CBO} is the reverse-biased junction current. Figure 10.34b shows the transistor with an applied C-E voltage and with the base terminal left open. This bias condition also makes the B-C junction reverse biased. The current in the transistor for this bias configuration is denoted as I_{CEO}.

The current I_{CBO} shown in Figure 10.34b is the normal reverse-biased B-C junction current. Pan of this current is due to the flow of minority carrier holes from the collector across the B-C space charge region into the base. The Row of holes into the

[1] The doping concentrations in the base and collector of the transistor we assume to be small enough that Zener breakdown is not a factor to be considered.

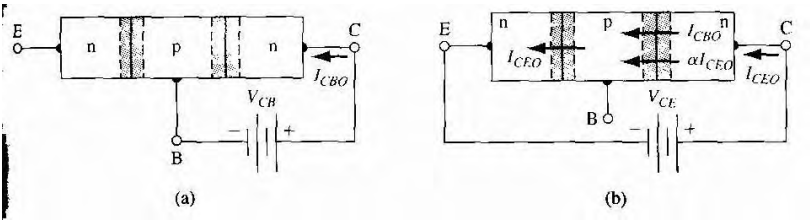


Figure 10.34 | (a) Open emitter configuration with saturation current I_{CBO}. (b) Open base configuration with saturation current I_{CEO}.

base makes the base positive with respect to the emitter, and the B-E junction becomes forward biased. The forward-biased B-E junction produces the current I_{CEO}, due primarily to the injection of electrons from the emitter into the base. The injected electrons diffuse across the base toward the B-C junction. These electrons are subject to all of the recombination processes in the bipolar transistor. When the electrons reach the B-C junction, this current component is αI_{CEO} where α is the common base current gain. We therefore have

$$I_{CEO} = \alpha I_{CEO} + I_{CBO} \tag{10.55a}$$

or

$$I_{CEO} = \frac{I_{CBO}}{1-\alpha} \approx \beta I_{CBO} \tag{10.55b}$$

where β is the common-emitter current pain. The reverse-biased junction current I_{CBO} is multiplied by the current gain β when the transistor is biased in the open-base configuration.

When the transistor is biased in the open-emitter configuration as in Figure 10.34a, thecurrent I_{CBO} at breakdown becomes $I_{CBO} \rightarrow M I_{CBO}$, where M is the multiplication factor. An empirical approximation for the multiplication factor is usually written as

$$M = \frac{1}{1-(V_{CB}/BV_{CBO})^n} \tag{10.56}$$

where n is an empirical constant, usually between 3 and 6, and BV_{CBO} is the B-C breakdown voltage with the emitter left open.

When the transistor is biased with the base open circuited as shown in Figure 10.34b, the currents in the B-C junction at breakdown are multiplied, so that

$$I_{CEO} = M(\alpha I_{CEO} + I_{CBO}) \tag{10.57}$$

Solving for I_{CEO}, we obtain

$$I_{CEO} = \frac{M I_{CBO}}{1-\alpha M} \tag{10.58}$$

The condition for breakdown corresponds to

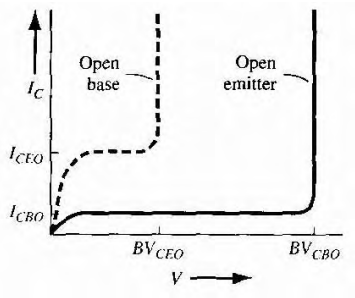


Figure 10.35 | Relative breakdown voltages and saturation currents of the open base and open emitter configurations

Using Equation (10.56) and assuming that $V_{CB} \approx V_{CE}$, Equation (10.59) becomes

$$\frac{\alpha}{1-(BV_{CEO}/BV_{CBO})^n} = 1 \qquad (10$$

where BV_{CEO} is the C-E voltage at breakdown in the open base configuration. Solving for BV_{CEO}, we find

$$BV_{CEO} = BV_{CBO}\sqrt[n]{1-\alpha} \qquad (10.6$$

where, again, α is the common-base current gain. The common-emitter and common-base current gains are related by

$$\beta = \frac{\alpha}{1-\alpha} \qquad (10.62a$$

Normally $\alpha \approx 1$, so that

$$1-\alpha \approx \frac{1}{\beta} \qquad (10.62b$$

Then Equation (10.61) can be written as

$$\boxed{BV_{CEO} = \frac{BV_{CBO}}{\sqrt[n]{\beta}}} \qquad (10.63$$

The breakdown voltage in the open-base configuration is smaller, by the factor $\sqrt[n]{\beta}$, than the actual avalanche junction breakdown voltage. This characteristic is shown in Figure 10.35.

DESIGN EXAMPLE 10.9

Objective

To design a bipolar transistor to meet a breakdown voltage specification.

Consider a silicon bipolar transistor with a common-emitter current gain of $\beta = 100$ and a base doping concentration of $N_B = 10^{17}$ cm^{-3}. The minimum open-base breakdown voltage is to be 15 volts.

■ Solution

From Equation (10.63).the minimum open-emitter junction breakdown voltage must be

$$BV_{CBO} = \sqrt[n]{\beta}\, BV_{CEO}$$

Assuming the empirical constant n is 3, we find

$$BV_{CBO} = \sqrt[3]{100}(15) = 69.6 \text{ V}$$

From Figure 8.25, the maximum collector doping concentration should be approximately 7×10^{15} cm^{-3} to achieve this breakdown voltage.

■ Comment

In a transistor circuit, the transistor must be designed to operate under a worst-case situation. In this example, the transistor must be able to operate in an open-base configuration without going into breakdown. As we determined previously, an increase in breakdown voltage can be achieved by decreasing the collector doping concentration.

TEST YOUR UNDERSTANDING

E10.15 A uniformly doped silicon transistor has base and collector doping concentrations of 5×10^{16} cm^{-3} and 5×10^{15} cm^{-3}, respectively. The common emitter current gain is $\beta = 85$. Assuming an empirical constant value of $n = 3$, determine BV_{CEO} (Ans. 21.6 V)

E10.16 The minimum required breakdown voltage of a uniformly doped silicon npn bipolar transistor is to be $BV_{CEO} = 70\,V$. The base impurity doping concentration is $N_B = 3 \times 10^{16}$ cm^{-3}, the common-emitter current gain is $\beta = 85$, **and** the empirical constant value is $n = 3$. Determine the maximum collector impurity doping concentration. (1×10^{15} cm^{-3} Ans. $N_C \approx$)

10.5 | EQUIVALENT CIRCUIT MODELS

In order to analyze a transistor circuit either by hand calculations or using computer codes, one needs a mathematical model, or equivalent circuit, of the transistor. There are several possible models, each one having certain advantages and disadvantages. Adetailed study of all possible models is beyond the scope of this text. However, we will consider three equivalent circuit models. Each of these follows directly from the work we have done on the pn junction diode and on the bipolar transistor. Computer analysis of electronic circuits is more commonly used than hand calculations, but it is instructive to consider the types of transistor model used in computer codes.

It is useful to divide bipolar transistors into two categories — switching and amplification — definedby their use in electronic circuits. Switching usually involves turning a transistor from its "off" state, or cutoff, to its "on" state, either forward-active or saturation, and then back to its "off" state. Amplification usually involves superimposing sinusoidal signals on dc values so that bias voltages and currents are only perturbed. The *Ebers–Moll* ***model*** is used in switching applications; the *hybrid*-pi model is used in amplification applications.

*10.5.1 Ebers–Moll Model

The Ebers–Moll model, or equivalent circuit, is one of the classic models of the bipolar transistor. This particular model is based on the interacting diode junctions and is applicable in any of the transistor operating modes. Figure 10.36 shows the current directions and voltage polarities used in the Ebers–Moll model. The currents are defined as all entering the terminals so that

$$I_E + I_B + I_C = 0 \tag{10.6}$$

The direction of the emitter current is opposite to what we have considered up to this point, but as long as we are consistent in the analysis, the defined direction does not matter.

The collector current can be written in general as

$$I_C = \alpha_F I_F - I_R \tag{10.65}$$

where α_F is the common base current gain in the forward-active mode. In this mode, Equation (10.65a) becomes

$$I_C = \alpha_F I_F + I_{CS} \tag{10.65b}$$

where the current I_{CS} is the reverse-bias B-C junction current. The current I_F is given by

$$I_F = I_{ES}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \tag{10.66}$$

If the B-C junction becomes forward biased, such as in saturation, then we can write the current I_R as

$$I_R = I_{CS}\left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right] \tag{10.67}$$

Using Equations (10.66) and (10.67), the collector current from Equation (10.65a) can be written as

$$I_C = \alpha_F I_{ES}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] - I_{CS}\left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right] \tag{10.68}$$

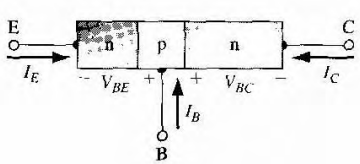


Figure 10.36 | Current direction and voltage polarity definitions for the Ebers–Moll model.

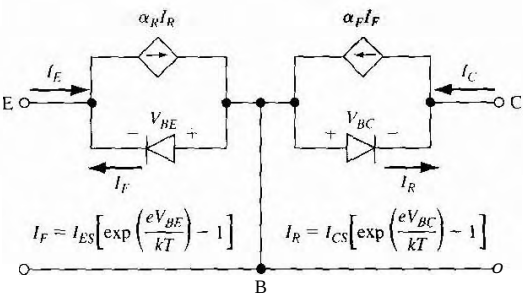


Figure 10.37 | Basic Ebers–Moll equivalent circuit.

We can also write the emitter current as

$$I_E = \alpha_R I_R - I_F \tag{10.69}$$

or

$$I_E = \alpha_R I_{CS}\left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right] - I_{ES}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \tag{10.70}$$

The current I_{ES} is the reverse-bias B-E junction current and α_R is the common base current gain for the inverse-active mode. Equations (10.68) and (10.70) are the classic Ebers–Moll equations.

Figure 10.37 shows the equivalent circuit corresponding to Equations (10.68) and (10.70). The current sources in the equivalent circuit represent current components that depend on voltages across otherjunctions. The Ebers–Moll model has four parameters: α_F, α_R, I_{ES}, and I_{CS}. However, only three parameters are independent. The reciprocity relationship states that

$$\alpha_F I_{ES} = \alpha_R I_{CS} \tag{10.71}$$

Since the Ebers–Moll model is valid in each of the four operating modes, we can, for example, use the model for the transistor in saturation. In the saturation mode, both B-E and B-C junctions are forward biased, so that $V_{BE} > 0$ and $V_{BC} > 0$. The B-E voltage will be a known parameter since we will apply a voltage across this junction. The forward-biased B-C voltage is a result of driving the transistor into saturation and is the unknown to be determined from the Ebers–Moll equations. Normally in electronic circuit applications, the collector–emitter voltage at saturation is of interest. We can define the C-E saturation voltage as

$$V_{CE}(\text{sat}) = V_{BE} - V_{BC} \tag{10.72}$$

We will find an expression for $V_{CE}(\text{sat})$ by combining the Ebers–Moll equations. In the following example we see how the Ebers–Moll equations can be used in a hand calculation, and we may also see how a computer analysis would make the calculations easier.

Combining Equations (10.64) and (10.70), we have

$$-(I_B + I_C) = \alpha_R I_{CS}\left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right] - I_{ES}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \quad (10.7$$

If we solve for $[\exp(eV_{BC}/kT) - 1]$ from Equation (10.73). and substitute the re ing expression into Equation (10.68), we can then find V_{BE} as

$$V_{BE} = V_t \ln\left[\frac{I_C(1-\alpha_R) + I_B + I_{ES}(1-\alpha_F\alpha_R)}{I_{ES}(1-\alpha_F\alpha_R)}\right] \quad (10.74)$$

where V_t is the thermal voltage. Similarly, if we solve for $[\exp(eV_{BE}/kT) - 1]$ from Equation (10.68), and substitute this expression into Equation (10.73), we can V_{BC} as

$$V_{BC} = V_t \ln\left[\frac{\alpha_F I_B - (1-\alpha_F)I_C + I_{CS}(1-\alpha_F\alpha_R)}{I_{CS}(1-\alpha_F\alpha_R)}\right] \quad (10.75)$$

We may neglect the I_{ES} and I_{CS} terms in the numerators of Equations (10.74) and (10.75). Solving for $V_{CE}(\text{sat})$, we have

$$V_{CE}(\text{sat}) = V_{BE} - V_{CB} = V_t \ln\left[\frac{I_C(1-\alpha_R) + I_B}{\alpha_F I_B - (1-\alpha_F)I_C} \cdot \frac{I_{CS}}{I_{ES}}\right] \quad (10.76)$$

The ratio of I_{CS} to I_{ES} can be written in terms of α_F and α_R from Equation (10.71). We can finally write

$$V_{CE}(\text{sat}) = V_t \ln\left[\frac{I_C(1-\alpha_R) + I_B}{\alpha_F I_B - (1-\alpha_F)I_C} \cdot \frac{\alpha_F}{\alpha_R}\right] \quad (10.77)$$

EXAMPLE 10.10

Objective

To calculate the collector–emitter saturation voltage of a bipolar transistor at T = 300 K.

Assume that $\alpha_F = 0.99$. $\alpha_R = 0.20$. $I_C = 1$ mA, and $I_B = 50\ \mu$A.

■ Solution

Substituting the parameters into Equation (10.77). we have

$$V_{CE}(\text{sat}) = (0.0259)\ln\left[\frac{(1)(1-0.2) + (0.05)}{(0.99)(0.05) - (1-0.99)(1)}\left(\frac{0.99}{0.20}\right)\right] = 0.121\ \text{V}$$

■ Comment

This $V_{CE}(\text{sat})$ value is typical of collector-zmitter saturation voltages. Because of the log function, $V_{CE}(\text{sat})$ is not a strong function of I_C or I_B.

10.5.2 Gummel–Poon Model

The Gummel–Poon model of the BJT considers more physics of the transistor thar the Ehers–Moll model. This model can he used if, for example, there is a nonunifom doping concentration in the base.

The electron current density in the base of an npn transistor can be written as

$$J_n = e\mu_n n(x)\mathrm{E} + eD_n \frac{dn(x)}{\mathrm{d}x} \tag{10.78}$$

An electric field will occur in the base if nonuniform doping exists in the base. This was discussed in Section 10.4.5. The electric field, from Equation (10.52), can be written in the form

$$\mathrm{E} = \frac{kT}{e} \cdot \frac{1}{p(x)} \cdot \frac{dp(x)}{dx} \tag{10.79}$$

where $p(x)$ is the majority carrier hole concentration in the base. Under low injection, the hole concentration is just the acceptor impurity concentration. With the doping profile shown in Figure 10.31, the electric field is negative (from the collector to the emitter). The direction of this electric field aids the flow of electrons across the base.

Substituting Equation (10.79) into Equation (10.78), we obtain

$$J_n = e\mu_n n(x) \cdot \frac{kT}{e} \cdot \frac{1}{p(x)} \cdot \frac{dp(x)}{dx} + eD_n \frac{dn(x)}{dx} \tag{10.80}$$

Using Einstein's relation, we can write Equation (10.80) in the form

$$J_n = \frac{eD_n}{p(x)} \left(n(x)\frac{dp(x)}{dx} + p(x)\frac{dn(x)}{\mathrm{d}x} \right) = \frac{eD_n}{p(x)} \cdot \frac{d(pn)}{dx} \tag{10.81}$$

Equation (10.81) can be written in the form

$$\frac{J_n p(x)}{eD_n} = \frac{d(pn)}{dx} \tag{10.82}$$

Integrating Equation (10.82) through the base region while assuming that the electron current density is essentially a constant and the diffusion coefficient is a constant, we find

$$\frac{J_n}{eD_n} \int_0^{x_B} p(x)\,dx = \int_0^{x_B} \frac{dp(x)}{dx}\,dx = p(x_B)n(x_B) - p(0)n(0) \tag{10.83}$$

Assuming the B-E junction is forward biased and the B-C junction is reverse biased, we have $n(0) = n_{B0}\exp(V_{BE}/V_t)$ and $n(x_B) = 0$. We may note that $n_{B0}p = n_i^2$ so that Equation (10.83) can be written as

$$J_n = \frac{-eD_n n_i^2 \exp(V_{BE}/V_t)}{\int_0^{x_B} p(x)\,dx} \tag{10.84}$$

The integral in the denominator is the total majority carrier charge in the base and is known as the *base Gummel* number; defined as Q_B.

If we perform the same analysis in the emitter, we find that the hole current density in the emitter of an npn transistor can be expressed as

$$J_p = \frac{-eD_p n_i^2 \exp(V_{BE}/V_t)}{\int_0^{x_E} n(x')\,\mathrm{dx}'} \tag{10.85}$$

The integral in the denominator is the total majority carrier charge in the emitter and is known as the emitter Gummel number, defined as Q_E

Since the currents in the Gummel–Poon model are functions of the total integrated charges in the base and emitter. these currents can easily be determined fa nonuniformly doped transistors.

The Gummel–Poon model can also take into account nonideal effects, such a the Early effect and high-level injection. 4 s the B-C voltage changes, the neutr base width changes so that the base Gummel number Q_B changes. The change in Q with B-C voltage then makes the electron current density given by Equation (10.84) a function of the B-C voltage. This is the base width modulation effect or Early effect as discussed previously in Section 10.4.1.

If the B-E voltage becomes too large. low injection no longer applies, which leads to high-level injection. In this case, the total hole concentration in the base increases because of the increased excess hole concentration. This means that the base Gummel number will increase. The change in base Gummel number implies, from Equation (10.84). that the electron current density will also change. High-level injection was also previously discussed in Section 10.4.2.

The Gummel–Poon model can then he used to describe the basic operation of the transistor as well as to describe nonideal effects.

10.5.3 Hybrid-Pi Model

Bipolar transistors are commonly used in circuits that amplify time-varying or sinusoidal signals. In these linear amplifier circuits, the transistor is biased in the forward-active region and small sinusoidal voltages and currents are superimposed on dc voltages and currents. In these applications, the sinusoidal parameters are of interest, so it is convenient to develop a small-signal equivalent circuit of the bipolar transistor using the small-signal admittance parameters of the pn junction developed in Chapter 8.

Figure 10.38a shows an npn bipolar transistor in a common emitter configuration with the small-signal terminal voltages and currents. Figure 10.38b shows the cross section of the npn transistor. The C, B, and E terminals are the external connections to the transistor, while the C', B', and E' points are the idealized internal collector, hase, and emitter regions.

We can begin constructing the equivalent circuit of the transistor by considering the various terminals individually. Figure 10.39a shows the equivalent circuit between the external input base terminal and the external emitter terminal. The resistance r_b is the series resistance in the base between the external base terminal B and the internal base region B'. The B′-E′ junction is forward biased, so C_π is the junction diffusion capacitance and r_π is the junction diffusion resistance. The diffusion capacitance C_π is the same as the diffusion capacitance C_d given by Equation (8.72), and the diffusion resistance r_π is the same as the diffusion resistance r_d given by Equation (8.35).The values of both parameters are functions of the junction current. These two elements are in parallel with the junction capacitance, which is C_{je}. Finally, r_{ex} is the series resistance between the external emitter terminal and the

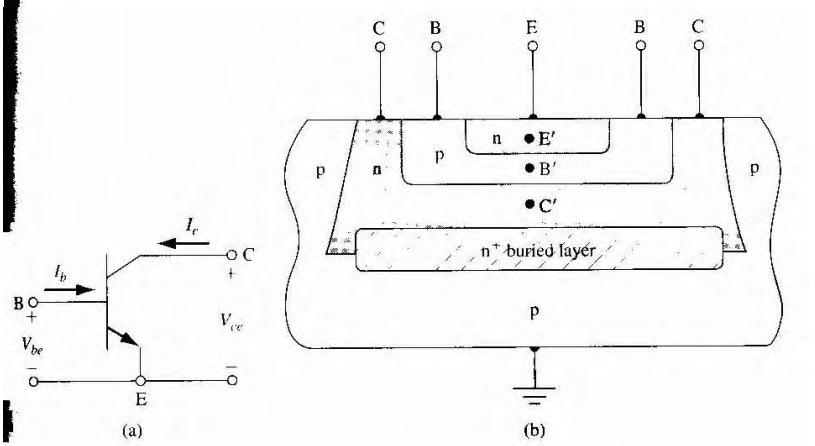
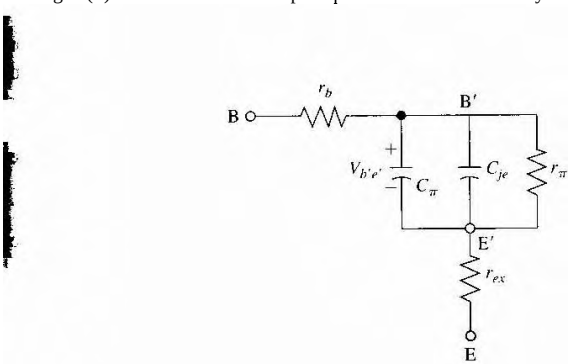


Figure 10.38 | **(a)** Common emitter npn bipolar transistor with small-signal current and voltages. (b) Cross section of an npn bipolar transistor for the hybrid-pi model.



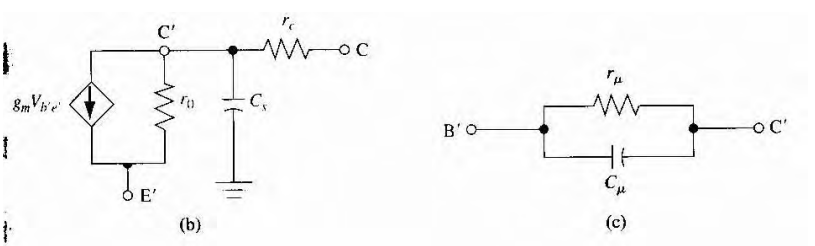


Figure 10.39 | Components of the hybrid-pi equivalent circuit between (a) the base and emitter, (h) the collector and emitter. and (c) the base and collector.

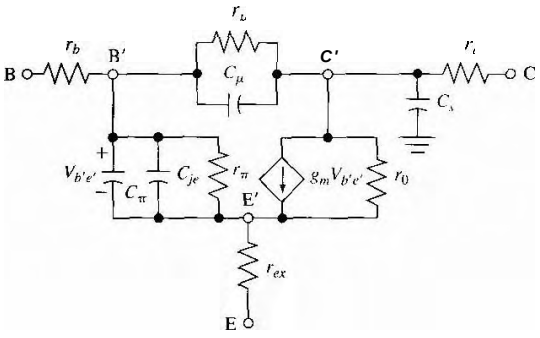


Figure 10.40 | Hybrid-pi equivalent circuit.

internal emitter region. This resistance is usually very small and may he on the order of 1 to 2 Ω.

Figure 10.39b shows the equivalent circuit looking into the collector terminal. The r_c resistance is the series resistance between the external and internal collector connections and the capacitance C_s is the junction capacitance of the reverse-biased collector-substrate junction. The dependent current source. $g_m V_{b'e'}$, is the collector current in the transistor, which is controlled by the internal base-emitter voltage. The resistance r_0 is the inverse of the output conductance g_0 and is primarily due to the Early effect.

Finally, Figure 10.39c shows the equivalent circuit of the reverse-biased B'-C' junction. The **C,** parameter is the reverse-biased junction capacitance and r_μ is the reverse-biased diffusion resistance. Normally, r_μ is on the order of megohms and can be neglected. The value of C_μ is usually much smaller than C_π but, because of the feedback effect which leads to the Miller effect and Miller capacitance, C_μ cannot be ignored in most cases. The Miller capacitance is the equivalent capacitance between B' and E′ due to C_μ and the feedback effect, which includes the gain of the transistor. The Miller effect also reflects C, between the C' and E' terminals at the output. However, the effect on the output characteristics can usually he ignored.

Figure 10.40 shows the complete hybrid-pi equivalent circuit. A computer simulation is usually required for this complete model because of the large number of elements. However, some simplifications can be made in order to gain an appreciation for the frequency effects of the bipolar transistor. The capacitances lead to frequency effects in the transistor, which means that the gain, for example, is a function of the input signal frequency.

EXAMPLE 10.11 | Objective

To determine, to a first approximation, the frequency at which the small-signal current gain decreases to $1/\sqrt{2}$ of its low frequency value.

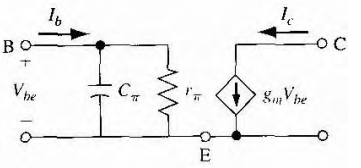


Figure 10.41 | Simplified hybrid-pi equivalent circuit.

Consider the simplified hybrid-pi circuit shown in Figure 10.41. We are ignoring C_{μ}, C_s, r_μ, C_{μ}, r_0, and the series resistances. We must emphasize that this is a first order calculation and that C_μ normally cannot be neglected.

■ Solution

At very low frequency, we may neglect C_π so that

$$V_{be} = I_b r_\pi \qquad \text{and} \qquad I_c = g_m V_{be} = g_m r_\pi I_b$$

We can then write

$$h_{fe0} = \frac{I_c}{I_b} = g_m r_\pi$$

where h_{fe0} is the low-frequency, small-signal common emitter current gain

Taking into account C_π, we have

$$V_{be} = I_b \left[\frac{r_\pi}{1 + j\omega r_\pi C_\pi} \right]$$

Then

$$I_c = g_m V_{be} = I_b \left[\frac{h_{fe0}}{1 + j\omega r_\pi C_\pi} \right]$$

or the small-signal current gain can be written as

$$A_i = \frac{I_c}{I_b} = \left[\frac{h_{fe0}}{1 + j\omega r_\pi C_\pi} \right]$$

The magnitude of the current gain drops to $1/\sqrt{2}$ of its low-frequency value at $f = 1/2\pi r_\pi C_\pi$.

If, for example, $r_\pi = 2.6$ kΩ and $C_\pi = 4$ pF, then

$$f = 15.3 \text{ MHz}$$

■ Comment

High-frequency transistors must have small diffusion capacitances, implying the use of small devices.

10.6 | FREQUENCY LIMITATIONS

The hybrid-pi equivalent circuit, developed in the last section, introduces frequency effects through the capacitor-resistor circuits. We will now discuss the various physical factors in the bipolar transistor affecting the frequency limitations of the device, and then define the transistor cutoff frequency, which is a figure of merit for a transistor.

10.6.1 Time-Delay Factors

The bipolar transistor is a transit-time device. When the voltage across the B-E junction increases, for example, additional carriers from the emitter are injected into the base, diffuse across the base, and are collected in the collector region. As the frequency increases, this transit time can become comparable to the period of the input signal. At this point, the output response will no longer be in phase with the input and the magnitude of the current gain will decrease.

The total emitter-to-collector time constant or delay time is composed of four separate time constants. We can write

$$\tau_{ec} = \tau_e + \tau_b + \tau_d + \tau_c \tag{10.86}$$

where

τ_{ec} = emitter-to-collector time delay
τ_e = emitter–base junction capacitance charging time
τ_b = base transit time
τ_d = collector depletion region transit time
τ_c = collector capacitance charging time

The equivalent circuit of the forward-biased B-E junction was given in Figure 10.39a. The capacitance C_{je} is the junction capacitance. If we ignore the series resistance, then the emitter–base junction capacitance charging time is

$$\boxed{\tau_e = r_e'(C_{je} + C_p)} \tag{10.87}$$

where r_e' is the emitter junction or diffusion resistance. The capacitance C_p, includes any parasitic capacitance between the base and emitter. The resistance r_e' is found as the inverse of the slope of the I_E versus V_{BE} curve. We obtain

$$r' = \frac{kT}{e} \cdot \frac{1}{I_E} \tag{10.88}$$

where I_E is the dc emitter current.

The second term, τ_b, is the base transit time, the time required for the minority carriers to diffuse across the neutral base region. The base transit time is related to the diffusion capacitance C_π of the B-E junction. For the npn transistor, the electron current density in the base can he written as

$$J_n = -en_B(x)v(x) \tag{10.89}$$

where $v(x)$ is an average velocity. We can write

$$v(x) = dx/dt \qquad \text{or} \qquad dt = dx/v(x) \tag{10.90}$$

The transit time can then be found by integrating, or

$$\tau_b = \int_0^{x_B} dt = \int_0^{x_B} \frac{dx}{v(x)} = \int_0^{x_B} \frac{en_B(x)\,dx}{(-J_n)} \tag{10.91}$$

The electron concentration in the base is approximately linear (see Example 10.6) so we can write

$$n_B(x) \approx n_{B0}\left[\exp\left(\frac{eV_{BE}}{kT}\right)\right]\left(1 - \frac{x}{x_B}\right) \tag{10.92}$$

and the electron current density is given by

$$J_n = eD_n \frac{dn_B(x)}{dx} \tag{10.93}$$

The base transit time is then found by combining Equations (10.92) and (10.93) with Equation (10.91). We find that

$$\boxed{\tau_b = \frac{x_B^2}{2D_n}} \tag{10.94}$$

The third time-delay factor is τ_d, the collector depletion region transit time. Assuming that the electrons in the npn device travel across the B-C space charge region at their saturation velocity, we have

$$\boxed{\tau_d = \frac{x_{dc}}{v_s}} \tag{10.95}$$

where x_{dc} is the B-C space charge width and v_s is the electron saturation velocity.

The fourth time-delay factor, τ_c, is the collector capacitance charging time. The B-C is reverse biased so that the diffusion resistance in parallel with the junction capacitance is very large. The charging time constant is then a function of the collector series resistance r_c. We can write

$$\boxed{\tau_c = r_c(C_\mu + C_s)} \tag{10.96}$$

where C_μ is the B-C junction capacitance and C_s is the collector-to-substrate capacitance. The series resistance in small epitaxial transistors is usually small; thus the time delay τ_c may be neglected in some cases.

Example calculations of the various time-delay factors will he given in the next section as part of the cutoff frequency discussion.

10.6.2 Transistor Cutoff Frequency

The current gain as a function of frequency was developed in Example 10.11 so that we can also write the common base current gain as

$$\alpha = \frac{\alpha_0}{1 + j\dfrac{f}{f_\alpha}} \tag{10.97}$$

where α_0 is the low-frequency common base current gain and f_α is defined as the *alpha cutoff frequency*. The frequency f_α is related to the emitter-to-collector time delay τ_{ec} as

$$f_\alpha = \frac{1}{2\pi\tau_{ec}} \tag{10.98}$$

When the frequency is equal to the alpha cutoff frequency, the magnitude of the common base current gain is $1/\sqrt{2}$ of its low-frequency value.

We can relate the alpha cutoff frequency to the common emitter current gain by considering

$$\beta = \frac{\alpha}{1 - \alpha} \tag{10.99}$$

We may replace u in Equation (10.99) with the expression given by Equation (10.97). When the frequency f is of the same order of magnitude as f_α, then

$$|\beta| = \left|\frac{\alpha}{1 - \alpha}\right| \approx \frac{f_\alpha}{f} \tag{10.100}$$

where we have assumed that $\alpha_0 \approx 1$. When the signal frequency is equal to the alpha cutoff frequency. the magnitude of the common emitter current gain is equal to unity. The usual notation is to define this *cutoff frequency* as f_T, so we have

$$\boxed{f_T = \frac{1}{2\pi\tau_{ec}}} \tag{10.101}$$

From the analysis in Example 10.11, we may also write the common-emitter current gain as

$$\beta = \frac{\beta_0}{1 + j(f/f_\beta)} \tag{10.102}$$

where f_β is called the *beta cutoff frequency* and is the frequency at which the magnitude of the common-emitter current gain β drops to $1/\sqrt{2}$ of its low-frequency value.

Combining Equations (10.99) and (10.97). we can write

$$\beta = \frac{\alpha}{1 - \alpha} = \frac{\dfrac{\alpha_0}{1 + j(f/f_T)}}{1 - \dfrac{\alpha_0}{1 + j(f/f_T)}} = \frac{\alpha_0}{1 - \alpha_0 + j(f/f_T)} \tag{10.103}$$

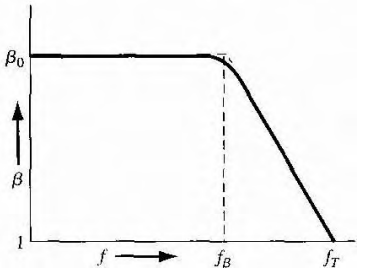


Figure 10.42 | Bode plot of common emitter current gain versus frequency.

$$\beta = \frac{\alpha_0}{(1-\alpha_0)\left[1 + j\dfrac{f}{(1-\alpha_0)f_T}\right]} \approx \frac{\beta_0}{1 + j\dfrac{\beta_0 f}{f_T}} \tag{10.104}$$

where

$$\beta_0 = \frac{\alpha_0}{1-\alpha_0} \approx \frac{1}{1-\alpha_0}$$

Comparing Equations (10.104) and (10.102), the beta cutoff frequency is related to the cutoff frequency by

$$\boxed{f_\beta \approx \frac{f_T}{\beta_0}} \tag{10.105}$$

Figure 10.42 shows a Bode plot of the common emitter current gain as a function of frequency and shows the relative values of the beta and cutoff frequencies. Keep in mind that the frequency is plotted on a log scale, so f_β and f_T usually have significantly different values.

EXAMPLE 10.12 Objective

To calculate the emitter-to-collector transit time and the cutoff frequency of a bipolar transistor, given the transistor parameters.

Consider a silicon npn transistor at T = 300 K. Assume the following parameters:

$I_E = 1$ mA $\quad C_{je} = 1$ pF

$x_B = 0.5\,\mu$m $\quad D_n = 25$ cm^2/s

$x_{dc} = 2.4\,\mu$m $\quad r_c = 20\,\Omega$

$C_\mu = 0.1$ pF $\quad C_s = 0.1$ pF

■ Solution

We will initially calculate the various time-delay factors. If we neglect the parasitic capacitance, the emitter-base junction charging time is

$$\tau_e = r_e' C_{je}$$

where

$$r_e' = \frac{kT}{e} \cdot \frac{1}{I_E} = \frac{0.0259}{1 \times 10^{-3}} = 25.9\ \Omega$$

Then

$$\tau_e = (25.9)(10^{-12}) = 25.9\ \text{ps}$$

The base transit time is

$$\tau_b = \frac{x_B^2}{2D_n} = \frac{(0.5 \times 10^{-4})^2}{2(25)} = 50\ \text{ps}$$

The collector depletion region transit time is

$$\tau_b = \frac{x_{dc}}{v_s} = \frac{2.4 \times 10^{-4}}{10^7} = 24\ \text{ps}$$

The collector capacitance charging time is

$$\tau_c = r_c(C_\mu + C_s) = (20)(0.2 \times 10^{-12}) = 4\ \text{ps}$$

The total emitter-to-collector time delay is then

$$\tau_{ec} = 25.9 + 50 + 24 + 4 = 103.9\ \text{ps}$$

so that the cutoff frequency is calculated as

$$f_T = \frac{1}{2\pi \tau_{ec}} = \frac{1}{2\pi(103.9 \times 10^{-12})} = 1.53\ \text{GHz}$$

If we assume a low-frequency common-emitter current gain of $\beta = 100$, then the beta cutoff frequency is

$$f_\beta = \frac{f_T}{\beta_0} = \frac{1.53 \times 10^9}{100} = 15.3\ \text{MHz}$$

■ Comment

The design of high-frequency transistors requires small device geometries in order to reduce capacitances, and narrow base widths in order to reduce the base transit time.

TEST YOUR UNDERSTANDING

E10.17 A silicon npn bipolar transistor is biased at $I_E = 0.5$ mA and has a junction capacitance of $C_{je} = 2$ pF. All other parameters are the same as listed in Example 10.12. Find the emitter-to-collector transit time, the cutoff frequency, and the beta cutoff frequency. (Ans. $\tau_{ec} = 181.6$ ps, $f_T = 1.14$ GHz, $f_\beta = 11.4$ MHz)

10.7 | LARGE-SIGNAL SWITCHING

Switching a transistor from one state to another is strongly related to the frequency characteristics just discussed. However, switching is considered to be a large-signal change whereas the frequency effects assumed only small changes in the magnitude of the signal.

10.7.1 Switching Characteristics

Consider an npn transistor in the circuit shown in Figure 10.43a switching from cutoff to saturation, and then switching back from saturation to cutoff. We will describe the physical processes taking place in the transistor during the switching cycle.

Consider, initially, the case of switching from cutoff to saturation. Assume that in cutoff $V_{BE} \approx V_{BB} < 0$, thus the B-E junction is reverse biased. At $t = 0$, assume that V_{BB} switches to a value of V_{BB0} as shown in Figure 10.43b. We will assume that V_{BB0} is sufficiently positive to eventually drive the transistor into saturation. For $0 \leq t \leq t_1$, the base current supplies charge to bring the B-E junction from reverse bias to a slight forward bias. The space charge width of the B-E junction is narrowing, and ionized donors and acceptors are being neutralized. A small amount of charge is also injected into the base during this time. The collector current increases from zero to 10 percent of its final value during this time period, referred to as the delay time.

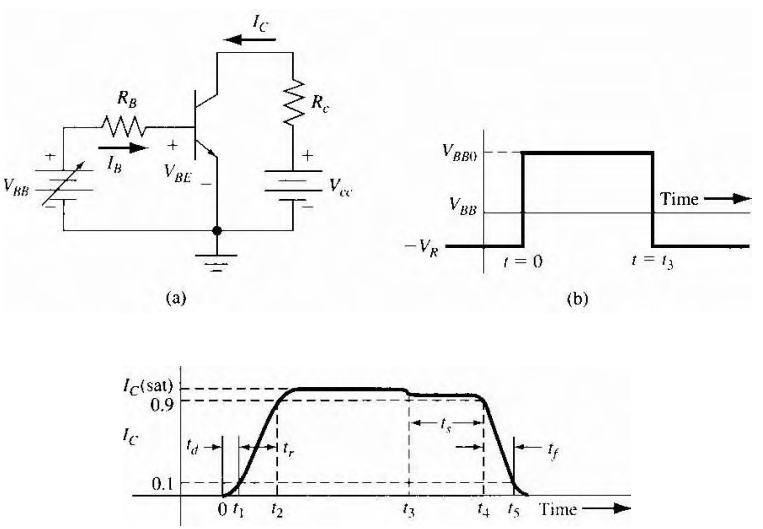


Figure 10.43 | (a) Circuit used for transistor switching. (b) Input base drive for transistor switching. *(c)* Collector current versus time during transistor switching.

During the next time period, $t_1 \leq t \leq t_2$, the base current is supplying charge which increases the B-E junction voltage from near cutoff to near saturation. During this time, additional carriers are being injected into the base so that the gradient of the minority carrier electron concentration in the base increases, causing the collector current to increase. We refer to this time period as the rise time, during which the collector current increases from 10 percent to 90 percent of the final value. For $t > t_2$ the base drive continues to supply base current, driving the transistor into saturation and establishing the final minority carrier distribution in the device.

The switching of the transistor from saturation to cutoff involves removing all of the excess minority carriers stored in the emitter, base, and collector regions. Figure 10.44 shows the charge storage in the base and collector when the transistor is in saturation. The charge Q_B is the excess charge stored in a forward-active transistor and Q_{BX} and Q_C are the extra charges stored when the transistor is biased in saturation. At $t = t_3$, the base voltage V_{BB} switches to a negative value of $(-V_R)$. The base current in the transistor reverses direction as was the case in switching a pn junction diode from forward to reverse bias. The reverse base current pulls the excess stored carriers from the emitter and base regions. Initially, the collector current does not change significantly, since the gradient of the minority carrier concentration in the base does not change instantaneously. Recall that when the transistor is biased in saturation, both the B-E and B-C junctions are forward biased. The charge Q_{BX} in the base must be removed to reduce the forward-biased **B-C** voltage to zero volts before the collector current can change. This time delay is called the *storage time* and is denoted by t_s. The storage time is the time between the point at which V_{BB} switches to the time when the collector current is reduced to 90 percent of its maximum saturation value. The storage time is usually the most important parameter in the switching speed of the bipolar transistor.

The final switching delay time is the fall time t_f during which the collector current decreases from the 90 percent to the 10 percent value. During this time, the B-C

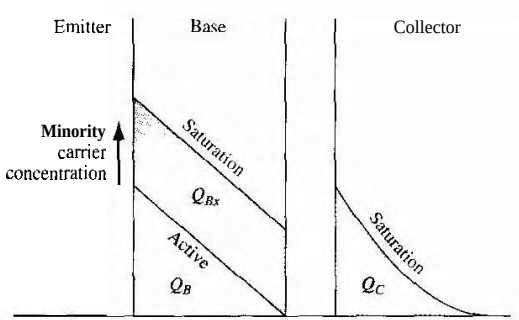


Figure 10.44 | Charge storage in the base and collector at saturation and in **the** active mode.

junction is reverse biased but excess carriers in the base are still being removed, and the **B-E** junction voltage is decreasing.

The switching-time response of the transistor can be determined by using the Ebers–Moll model. The frequency-dependent gain parameters must be used, and normally the Laplace transform technique is used to obtain the time response. The details of this analysis are quite tedious and will not be presented here.

10.7.2 The Schottky-Clamped Transistor

One method frequently employed to reduce the storage time and increase the switching speed is the use of a Schottky-clamped transistor. This is a normal npn bipolar device with a Schottky diode connected between base and collector, as shown in Figure 10.45a. The circuit symbol for the Schottky-clamped transistor is shown in Figure 10.45b. When the transistor is biased in the forward-active mode, the **B-C** junction is reverse biased; hence, the Schottky diode is reverse biased and effectively out of the circuit. The characteristics of the Schottky-clamped transistor—or simply the Schottky transistor—are those of the normal npn bipolar device.

When the transistor is driven into saturation, the **B-C** junction becomes forward biased; hence the Schottky diode also becomes forward biased. We may recall from our discussion in the previous chapter that the effective turn-on voltage of the Schottky diode is approximately half that of the pn junction. The difference in turn-on voltage means that most of the excess base current will be shunted through the Schottky diode and away from the base so that the amount of excess stored charge in the base and collector is drastically reduced. The excess minority carrier concentration in the base and collector at the **B-C** junction is an exponential function of V_{BC}. If V_{BC} is reduced from 0.5 volt to 0.3 volt, for example, the excess minority carrier concentration is reduced by over 3 orders of magnitude. The reduced excess stored charge in the base of the Schottky transistor greatly reduces the storage time—storage times on the order of 1 ns or less are common in Schottky transistors.

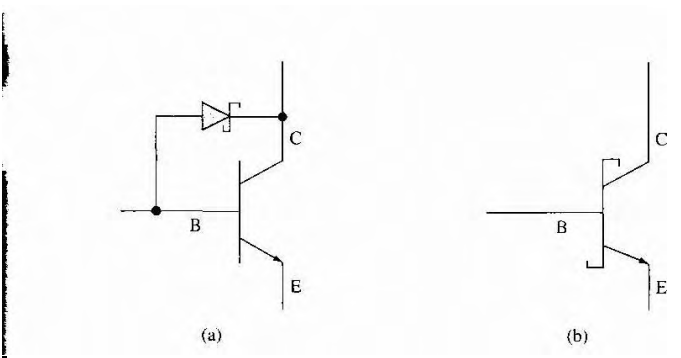


Figure 10.45 | (a) The Schottky-clamped transistor. (b) Circuit symbol of the Schottky-clamped transistor.

*10.8 | OTHER BIPOLAR TRANSISTOR STRUCTURES

This section is intended to briefly introduce three specialized bipolar transistor structures. The first structure is the polysilicon emitter bipolar junction transistor (BJT), the second is the SiGe-base transistor, and the third is the heterojunction bipolar transistor (HBT). The polysilicon emitter BJT is being used in some recent integrated circuits, and the SiGe-base transistor and HBT are intended for high-frequency/high-speed applications.

10.8.1 Polysilicon Emitter BJT

The emitter injection efficiency is degraded by the carriers injected from the base back into the emitter. The emitter width, in general, is thin, which increases speed and reduces parasitic resistance. However. a thin emitter increases the gradient in the minority carrier concentration, as indicated in Figure 10.19. The increase in the gradient increases the B-E junction current, which in turn decreases the emitter injection efficiency and decreases the common emitter current gain. This effect is also shown in the summary of Fable 10.3.

Figure 10.46 shows the idealized cross section of an npn bipolar transistor with a polysilicon emitter. As shown in the figure, there is a very thin n^+ single crystal silicon region between the p-type base and the n-type polysilicon. As a first approximation to the analysis, we may treat the polysilicon portion of the emitter as low-mobility silicon, which means that the corresponding diffusion coefficient is small.

Assuming that the neutral widths of both the polysilicon and single-crystal portions of the emitter are much smaller than the respective diffusion lengths, then the minority carrier distribution functions will be linear in each region. Both the minority carrier concentration and diffusion current must be continuous across the polysilicon/silicon interface. We can therefore write

$$eD_{E(\text{poly})}\frac{d(\delta p_{E(\text{poly})})}{dx} = eD_{E(n^+)}\frac{d(\delta p_{E(n^+)})}{dx} \qquad (10.106)$$

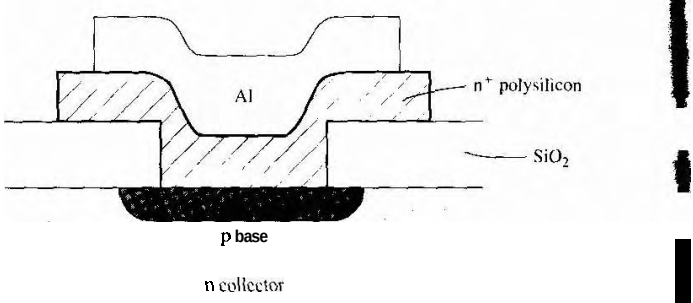


Figure 10.46 | Simplified cross section of an npn polysilicon emitter BJT.

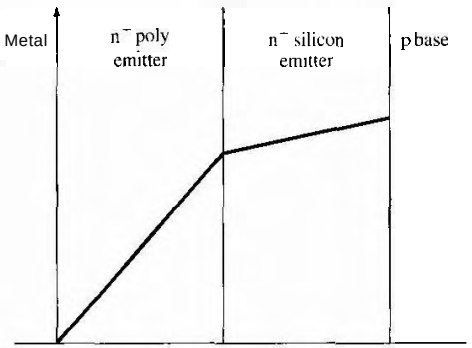


Figure 10.47 | Excess minority carrier hole concentrations in n^+ polysilicon and n^+ silicon emitter.

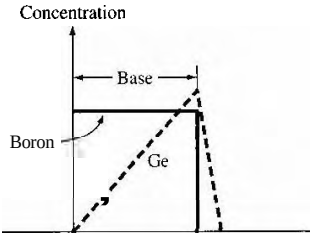
$$\frac{d\left(\delta p_{E(n^-)}\right)}{dx} = \frac{D_{E(\text{poly})}}{D_{E(n^+)}} \cdot \frac{d\left(\delta p_{E(\text{poly})}\right)}{dx} \tag{10.106b}$$

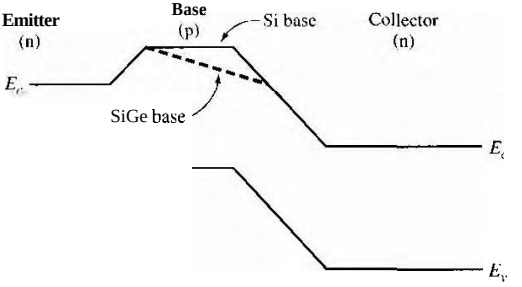
Since $D_{E(\text{poly})} < D_{E(n^+)}$, then the gradient of the minority carrier concentration at the emitter edge of the B-E depletion region in the n^+ region is reduced as Figure 10.47 shows. This implies that the current back-injected from the base into the emitter is reduced so that the common-emitter current gain is increased.

10.8.2 Silicon–Germanium Base Transistor

The bandgap energy of Ge (−0.67 eV) is significantly smaller than the bandgap energy of Si (−1.12 eV). By incorporating Ge into Si, the bandgap energy will decrease compared to pure Si. If Ge is incorporated into the base region of a Si bipolar transistor, the decrease in bandgap energy will influence the device characteristics. The desired Ge concentration profile is to have the largest amount of Ge near the base–collector junction and the least amount of Ge near the base–emitter junction. Figure 10.48a shows an ideal uniform boron doping concentration in the p-type base and a linear Ge concentration profile.

The energy bands of a SiGe-base npn transistor compared to a Si-base npn transistor, assuming the boron and Ge concentrations given in Figure 10.48a, are shown in Figure 10.48b. The emitter–base junctions of the two transistors are essentially identical, since the Ge concentration is very small in this region. However, the bandgap energy of the SiGe-base transistor near the base–collector junction is smaller than that of the Si-base transistor. The base current is determined by the base-emitter junction parameters and hence will be essentially the same in the two transistors. This change in bandgap energy will influence the collector current.





(b)

Figure 10.48 | (a) Assumed boron and germanium concentrations in the base of the SiGe-base transistor. (b) Energy band diagram of the Si- and SiGe-base transistors.

Collector Current and Current Gain Effects Figure 10.49 shows the thermal equilibrium minority carrier electron concentration through the base region of the SiGe and Si transistors. This concentration is given by

$$n_{B0} = \frac{n_i^2}{N_B} \tag{10.107}$$

where N_B is assumed to be constant. The intrinsic concentration, however, is a function of the bandgap energy. We may write

$$\frac{n_i^2(\text{SiGe})}{n_i^2(\text{Si})} = \exp\left(\frac{\Delta E_g}{kT}\right) \tag{10.108}$$

where $n_i(\text{SiGe})$ is the intrinsic carrier concentration in the SiGe material, $n_i(\text{Si})$ is the intrinsic carrier concentration in the Si material, and AE, is the change in the bandgap energy of the SiGe material compared to that of Si

The collector current in a SiGe-base transistor will increase. As a first approximation, we can see this from the previous analysis. The collector current was found from Equation (10.36a), in which the derivative was evaluated at the base–collector

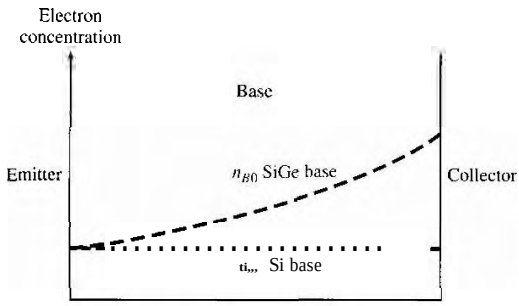


Figure **10.49** | Thermal equilibrium minority carrier electron concentration through the base of the Si- and SiGe-base transistors.

junction. This means that the value of n_{B_0} in the collector current expression in Equation (10.37) is the value at the base–collector junction. Since this value is larger for the SiGe-base transistor (Figure 10.49), the collector current will be larger compared to the Si-base transistor. Since the base currents are the same in the two transistors, the increase in collector current then implies that the current gain in the SiGe-base transistor is larger. If the bandgap narrowing is 100 meV, then the increase in the collector current and current gain will be approximately a factor of four.

Early Voltage Effects The Early voltage in a SiGe-base transistor is larger than that of the Si-base transistor. The explanation for this effect is less obvious than the explanation for the increase in collector current and current gain. For a bandgap narrowing of 100 meV, the Early voltage is increased by approximately a factor of 12. Incorporating Ge into the base region can increase the Early voltage by a large factor.

Base Transit Time and Emitter–Base Charging Time Effects The decrease in bandgap energy from the base–emitter junction to the base–collector junction induces an electric field in the base that helps accelerate electrons across the p-type base region. For a bandgap narrowing of 100 meV, the induced electric field can be on the order of 10^3 to 10^4 V/cm. This electric field reduces the base-transit time by approximately a factor of 2.5.

The emitter–base junction charging time constant, given by Equation (10.87), is directly proportional to the emitter diffusion resistance r_e'. This parameter is inversely proportional to the emitter current, as seen in Equation (10.88). For a given base current, the emitter current in the SiGe-base transistor is larger, since the current gain is larger. The emitter–base junction charging time is then smaller in a SiGe-base transistor than that in a Si-base transistor.

The reduction in both the base-transit time and the emitter–base charging time increases the cutoff frequency of the SiGe-base transistor. The cutoff frequency of these devices can be substantially higher than that of the Si-base device.

10.8.3 Heterojunction Bipolar Transistors

As mentioned previously, one of the basic limitations of the current gain in the bipolar transistor is the emitter injection efficiency. The emitter injection efficiency γ can be increased by reducing the value of the thermal-equilibrium minority carrier concentration p_{E0} in the emitter. However, as the emitter doping increases, the bandgap-narrowing effect offsets any improvement in the emitter injection efficiency. One possible solution is to use a wide-bandgap material for the emitter, which will minimize the injection of carriers from the base back into the emitter.

Figure 10.50a shows a discrete aluminum gallium arsenide/gallium arsenide heterojunction bipolar transistor, and Figure 10.50b shows the band diagram of the

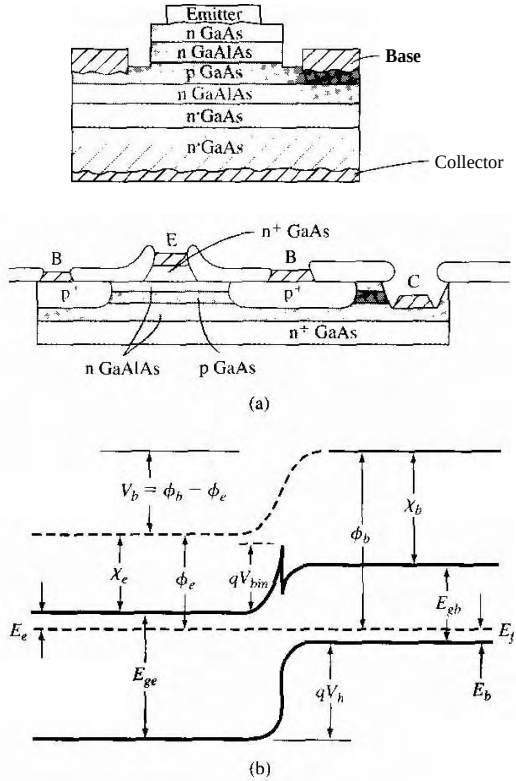


Figure 10.50 | (a) Cross section of AlGaAs/GaAs heterojunction bipolar transistor showing a discrete and integrated structure. (b) Energy-band diagram of the n AlGaAs emitter and p GaAs base junction.
(From Tiwari et al. [19].)

n-AlGaAs emitter to p-GaAs base junction. The large potential barrier V_h limits the number of holes that will be injected back from the base into the emitter.

The intrinsic carrier concentration is a function of bandgap energy as

$$n_i^2 \propto \exp\left(\frac{-E_g}{kT}\right)$$

For a given emitter doping. the numher of minority carrier holes injected into the emitter is reduced by a factor of

$$\exp\left(\frac{\Delta E_g}{kT}\right)$$

inchanging from a narrow- to wide-handgapemitter. If $\Delta E_g = 0.30$ eV, for example, n_i^2 would he reduced by approximately 10^5 at T = 300 K. The drastic reduction in n_i^2 for the wide-bandgap emitter mean? that the requirements of a very high emitter doping can be relaxed and a high emitter injection efficiency can still be obtained. A lower emitter doping reduces the handgap-narrowing effect.

The heterojunction GaAs bipolar transistor has the potential of being a very high frequency devicc. A lower emitter doping in the wide-bandgap emitter leads to a smaller junction capacitance, increasing the speed of the device. Also, for the GaAs npn device. the minority carriers in the base arc electrons with a high mobility. The electron mobility in GaAs is approximately 5 times that in silicon; thus, the base transit time in the GaAs base is very short. Experimental AlGaAs/GaAs heterojunction transistors with hase widths on the order of 0.1 μm have shown cutoff frequencies on the order of 40 GHz.

One disadvantage of GaAs is the low minority carrier lifetime. The small lifetime is not a factor in the base of a narrow-base device, but results in a larger B-E recombination current. which decreases the recombination factor and reduces the current gain. A current gain of 150 has been reported.

10.9 | SUMMARY

- There are two complementary bipolar transistors—npn and pnp. Each transistor has three separately doped regions and two pn junctions. The center region (base) is very narrow. so the two pn junctions are said to he interacting junctions.
- In the forward-active mode. the B-E junction is forward hiascd and the B-C junction is reverse biased. Majority carriers from the emitter arc injected into the hase whcrc they become minority carriers. These minority carriers diffuse across the base into thr B-C space charge region where they are swept into the collector.
- When a transistor is hiased in the forward-active mode of operation, the current at one terminal of the transistor (collector current) is controlled hy the voltage applied across the other two terminals of the transistor (base–emitter voltage). This is the basic transistor action
- The minority carrier concentrations were determined in each region of the transistor. The principal currents in the device are determined hy the diffusion of these minority carriers.

The common-base current gain, which leads to the common-emitter current gain, is a function of three factors—emitter injection efficiency, base transport factor, and recombination factor The emitter injection efficiency takes into account carriers from the base that are injected back into the emitter, the base transport factor takes into account recombination in the base region. and the recombination factor takes into account carriers that recombine within the forward-biased B-E junction.

- Several nonideal effects were considered:
  1. Base width modulation, or Early effect — thechangc in the neutral base width with a change in B-C voltage, producing a change in collector current with a change in B-C or C-E voltage.
  2. High-injection effects that cause the collector current to increase at a slower rate with base-emitter voltage.
  3. Emitter bandgap narrowing that produces a smaller emitter injection efficiency because of a very large emitter region doping concentration.
  4. Current crowding effects that produce a larger current density at the emitter edge than in the center of the emitter
  5. A nonuniform base doping concentration that induces an electric field in the base region, which aids the flow of minority carriers across the base.
  6. Two breakdown voltage mechanisms — punch-through and avalanche.
- Three equivalent circuits or mathematical models of the transistor were considered. The Ebers–Moll model and equivalent circuit are applicable in any of the transistor operating modes. The Gummel–Poon model is convenient to use when nonuniform doping exists in the transistor. The small-signal hybrid-pi model applies to transistors operating in the forward-active mode in linear amplifier circuits.
  The cutoff frequency of a transistor, a figure of merit for the transistor, is the frequency at which the magnitude of the common-emitter current gain hccomes equal to unity. The frequency responsc is a function of the emitter–base junction capacitance charging time, the base transit time, the collector depletion region transit time, and the collector capacitance charging time.
- The switching characteristics are closely related to the frequency limitations although switching involves large changes in currents and voltages. An important parameter in switching is the charge storage time. which applies to a transistor switching from saturation to cutoff.

GLOSSARY OF IMPORTANT TERMS

alpha cutoff frequency The frequency at which the magnitude of the common base current is $1/\sqrt{2}$ of its low-Srequcncy value; also equal to the cutoff frequency.

bandgap narrowing The reduction in the forbidden energy bandgap with high emitter doping concentration.

hase transit time The time that it takes a minority carrier to cross the neutral base region.

base transport factor The factor in the common base current gain that accounts for recombination in the neutral base width.

base width modulation The changc in the neutral base width with C-E or C-B voltage.

beta cutoff frequency The frequency at which the magnitude of the common emitter current gain is $1/\sqrt{2}$ of its low frequency value.

collector capacitance charging time The time constant that describes the time required for the **B-C** and collector–substrate space charge widths to change with a change in emitter current.

collector depletion region transit time The time that it takes a carrier to be swept across the B-C space charge region.

common-base current gain The ratio of collector current to emitter current.

rommon-emitter current gain The ratio of collector current to base current.

current **crowding** The nonuniform current density across the emitter junction areacreated by a lateral voltage drop in the base region due to a finite base current and base resistance.

cutoff The bias condition in which zero- or reverse-bias voltages are applied to both transistor junctions, resulting in zero transistor currents.

cutofffrequency The frequency at which the magnitude of the common emitter current gain is unity.

early effect Another term fur base width modulation.

early voltage The value of voltage (magnitude) at the intercept on the voltage axis obtained by extrapolating the I_C versus V_{CE} curves to zero current.

emitter–base junction capacitance charging time The time constant describing the time for the B-E space charge width to change with a change in emitter current.

emitter injection efficiency factor The factor in the common-base current gain that takes into account the injection of carriers from the base into the emitter.

forward active The bias condition in which the **B-E** junction is forward biased and ***the*** **B-C** junction is reverse biased.

inverse active The bias condition in which the **B-E** junction is reverse biased and the **B-C** junction is forward biased.

output conductance The ratio of a differential change in collector current to the corresponding differential change in C-E voltage.

CHECKPOINT

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

- Describe the basic operation of the transistor.
- Sketch the energy bands of the transistor in thermal equilibrium and when biased in the various operating modes.
- Calculate, to a good hrst approximation, the collector current as a function of base–emitter voltage.
- Sketch the minority carrier concentrations throughout the transistor under the various operating modes.
- Define the various diffusion and other current components in the transistor from the minority carrier distribution curves.
- Explain the physical mechanisms of the current gain limiting factors. Define the current-limiting factors from the current components in the transistor.
- Describe the physical mechanism of base width modulation and its effect on the current–voltage characteristics of the transistor
- Describe the voltage breakdown mechanisms in a bipolar transistor
- Sketch the simplified small-signal hybrid-pi equivalent circuit of the transistor biased in the forward-active mode.

■ Describe qualitatively the four time-delay or time-constant components in the frequen response of the bipolar transistor.

REVIEW QUESTIONS

1. Describe the charge flow in an npn bipolar transistor biased in the forward-active mode. Is the current by drift or diffusion?
2. Define the common-emitter current gain and explain why, to a first approximation, the current gain is a constant.
3. Explain the conditions of the cutoff, saturation, and inverse-active modes.
4. Sketch the minority carrier concentrations in a pnp bipolar transistor biased in the forward-active mode.
5. Define and describe the three limiting factors in the common-base current gain.
6. What is meant by base width modulation? What is another term used for this effect?
7. What is meant by high injection?
8. Explain emitter current crowding.
9. Define I_{CBO} and I_{CEO}, and explain why $I_{CEO} > I_{CBO}$
10. Sketch a simplified hybrid-pi model for an npn bipolar transistor and explain when this equivalent circuit is used.
11. Describe the time-delay factors in the frequency limitation o i the bipolar transistor.
12. What is the cutoff frequency of a bipolar transistor?
13. Describe the response of a bipolar transistor when it is switching between saturation and cutoff.

PROBLEMS

(Note: In the following problems, use the transistor geometry shown in Figure 10.13. Assume T = 300 K unless otherwise stated.)

Section 10.1 The Bipolar Transistor Action

10.1 For a uniformly doped $n^{++}p^{+}n$ bipolar transistor in thermal equilibrium, (a) sketch the energy-band diagram, (*b*) sketch the electric field through the device, and (c) repeat parts (*a*) and (h) for the transistor biased in the forward-active region.

10.2 Consider a $p^{++}n+p$ bipolar transistor, uniformly doped in each region. Sketch the energy-band diagram for the case when the transistor is (a) in thermal equilibrium, (*b*) biased in the forward-active mode, (*c*) biased in the inverse-active region, and (*d*) biased in cutoff with both the B-E and B-C junctions reverse biased.

10.3 The parameters in the base region of an npn bipolar transistor are $D_n = 20\,\text{cm}^2/\text{s}$, $n_{B0} = 10^4\,\text{cm}^{-3}$, $x_B = 1\ \mu\text{m}$, and $A_{BE} = 10^{-4}\,\text{cm}^2$. (*a*) Comparing Equations (10.1) and (10.2), calculate the magnitude of 4. (*b*) Determine the collector current for (*i*) $v_{BE} = 0.5$ V, (*ii*) $v_{BE} = 0.6$ V, and (iii) $v_{BE} = 0.7$ V.

10.4 Assume the common-base current gain for the transistor described in Problem 10.3 is $a = 0.9920$. (*a*) What is the common-emitter current gain β? [Note that $\beta = \alpha/(1 - \alpha)$.] (b) Determine the emitter and base currents corresponding to the collector currents determined in Problem 10.3b.

10.5 (*a*) In a bipolar transistor biased in the forward-active region, the base current is $i_B = 6.0\,\mu A$ and the collector current is $i_C = 510\,\mu A$. Determine β, α, and i_E. (Note that $i_E = i_C + i_B$.) (b) Repeat part (a) if $i_B = 50\,\mu A$ and $i_C = 2.65$ mA.

10.6 Assume that an npn bipolar transistor has a common-emitter current gain of $\beta = 100$. (*a*) Sketch the ideal current–voltage characteristics (i_C versus v_{CE}), like those in Figure 10.9, as i_B varies from zero to 0.1 mA in 0.01-mA increments. Let v_{CE} vary over the range $0 \leq v_{CE} \leq 10$ V. (*b*) Assuming $V_{CC} = 10$ V and $R_C = 1\,k\Omega$ in the circuit in Figure 10.8, superimpose the load line on the transistor characteristics in part (*a*).(c) Plot, on the resulting graph, the value of i_C and v_{CE} corresponding to $i_B = 0.05$ mA.

10.7 Consider Equation (10.7). Assume $V_{CC} = 10$ V, $R_C = 2\,k\Omega$, and $V_{BE} = 0.6$ V. (*a*) Plot V_{CB} versus I_C over the range $0 \leq I_C \leq 5$ mA. (*b*)At what value of I_C does $V_{CB} = 0$?

Section 10.2 Minority Carrier Distribution

10.8 A uniformly doped silicon npn bipolar transistor is to he biased in the forward-active mode with the B-C junction reverse hiased by 3 V The metallurgical base width is 1.10 μm. The transistor dopings are $N_E = 10^{17}$ cm^{-3}, $N_B = 10^{16}$ cm^{-3}, and $N_C = 10^{15}$ cm^{-3}. (*a*) For T = 300 K, calculate the B-E voltage at which the minority carrier electron concentration at $x = 0$ is 10 percent of the majority carrier hole concentration. (b) At this bias, determine the minority carrier hole concentration at x' = 0. (*c*) Determine the neutral base width for this bias.

10.9 A silicon npn bipolar transistor is uniformly doped and biased in the forward-active region. The neutral base width is $x_B = 0.8\,\mu$m. The transistor doping concentrations are $N_E = 5 \times 10^{17}$ cm^{-3}, $N_B = 10^{16}$ cm^{-3}, and $N_C = 10^{15}$ cm^{-3}. (*n*) Calculate the values of p_{E0}, n_{B0}, and p_{C0}. (*b*) For $V_{BE} = 0.625$ V, determine n_B at x = 0 and p_E at x' = 0. (c) Sketch the minority carrier concentrations through the device and label each curve.

10.10 A uniformly doped silicon pnp transistor is biased in the forward-active mode. The doping concentrations are $N_E = 10^{18}$ cm^{-3}, $N_B = 5 \times 10^{16}$ cm^{-3}, and $N_C = 10^{15}$ cm^{-3}. (*a*) Calculate the values of n_{E0}, p_{B0}, and n_{C0}. (*b*) For $V_{EB} = 0.650$ V, determine p_B at $x = 0$ and n_E at $x' = 0$. (*c*) Sketch the minority carrier concentrations through the device and label each curve.

10.11 Consider the minority carrier electron concentration in the base of an npn bipolar transistor as given by Equation (10.15a). In this problem, we want to compare the gradient of the electron concentration evaluated at the B-C junction to that evaluated at the B-E junction. In particular, calculate the ratio of $d(\delta n_B)/dx$ at $x = x_B$ to $d(\delta n_B)/dx$ at $x = 0$ for (*a*) $x_B/L_B = 0.1$, (*b*) $x_B/L_B = 1.0$, and (*c*) $x_B/L_B = 10$.

10.12 Derive the expressions for the coefficients given by Equations (10.14a) and (10.14b).

***10.13** Derive the expression for the excess minority carrier hole concentration in the base region of a uniformly doped pnp bipolar transistor operating in the forward-active region.

10.14 The excess electron concentration in the base of an npn bipolar transistor is given by Equation (10.15a). The linear approximation is given by Equation (10.15b). If $\delta n_{B0}(x)$ is the linear approximation given by Equation (10.15b) and $\delta n_B(x)$ is the

actual distribution given by Equation (10.15a), determine

$$\frac{\delta n_{B0}(x) - \delta n_B(x)}{\delta n_{B0}(x)} \times 100\%$$

at $x = x_B/2$ for (a) $x_B/L_B = 0.1$ and (b) $x_B/L_B = 1.0$. Assume $V_{BE} \gg kT/e$.

10.15 Consider a pnp bipolar transistor. Assume that the excess minority carrier hole concentrations at the edges of the B-E and B-C space charge regions are $\delta p_B(0) = 8 \times 10^{14}$ cm^{-3} and $\delta p_B(x_B) = -2.25 \times 10^{4}$ cm^{-3}, respectively. Plot, on the same graph, $\delta p_B(x)$ for (*a*) the ideal case when no recombination occurs in the base, and (*b*) the case when $x_B = L_B = 10\,\mu$m. (*c*) Assuming $D_B = 10$ cm^2/s, calculate the diffusion current density at $x = 0$ and $x = x_B$ for the conditions in parts (*a*) and (*b*) Determine the ratio $J(x = x_B)/J(x = 0)$ for the two cases.

***10.16** (*a*) A uniformly doped npn bipolar transistor at $T = 300$ K is biased in saturation. Starting with the continuity equation for minority carriers, show that the excess electron concentration in the base region can be expressed as

$$\delta n_B(x) = n_{B0}\left\{\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right]\left[1 - \frac{x}{x_B}\right] + \left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right]\left[\frac{x}{x_B}\right]\right\}$$

for $x_B/L_B \ll 1$ where x_B is the neutral base width. (*b*) Show that the minority carrier diffusion current in the base is then given by

$$J_n = -\frac{eD_B n_{B0}}{x_B}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - \exp\left(\frac{eV_{BC}}{kT}\right)\right]$$

(*c*) Show that the total excess minority carrier charge (C/cm^2) in the base region is given by

$$\delta Q_{nB} = \frac{-en_{B0}x_B}{2}\left\{\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] + \left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right]\right\}$$

***10.17** Consider a silicon pnp bipolar transistor at $T = 300$ K with uniform dopings of $N_E = 5 \times 10^{18}$ cm^{-3}, $N_B = 10^{17}$ cm^{-3}, and $N_C = 5 \times 10^{15}$ cm^{-3}. Let $D_B = 10$ cm^2/s, $x_B = 0.7\,\mu$m, and assume $x_B \ll L_B$. The transistor is operating in saturation with $J_p = 165$ A/cm^2 and $V_{EB} = 0.75$ V. Determine (*a*) V_{CB}. (*b*) V_{EC}(sat), (*c*) the #/cm^2 of excess minority carrier holes in the base, and (*d*) the #/cm^2 of excess minority carrier electrons in the long collector. Let $L_C = 35\,\mu$m.

10.18 An npn silicon bipolar transistor at $T = 300$ K has uniform dopings of $N_E = 10^{19}$ cm^{-3}, $N_B = 10^{17}$ cm^{-3}, and $N_C = 7 \times 10^{15}$ cm^{-3}. The transistor is operating in the inverse-active mode with $V_{BE} = -2$ V and $V_{BC} = 0.565$ V. (a) Sketch the minority carrier distribution through the device. (*b*) Determine the minority carrier concentrations at $x = x_B$ and $x = 0$. (*c*) If the metallurgical base width is $1.2\,\mu$m, determine the neutral base width.

10.19 A uniformly doped silicon pnp bipolar transistor at $T = 300$ K with dopings of $N_E = 5 \times 10^{17}$ cm^{-3}, $N_B = 10^{16}$ cm^{-3}, and $N_C = 5 \times 10^{14}$ cm^{-3} is biased in the inverse-active mode. What is the maximum B-C voltage so that the low-injection condition applies?

Section 10.3 Low-Frequency Common-Base Current Gain

10.20 The following currents are measured in a uniformly doped npn bipolar transistor:

$$I_{nE} = 1.20\ \text{mA} \qquad I_{pE} = 0.10\ \text{mA}$$
$$I_{nC} = 1.18\ \text{mA} \qquad I_R = 0.20\ \text{mA}$$
$$I_G = 0.001\ \text{mA} \qquad I_{rG} = 0.001\ \text{mA}$$

Determine *(a)* α. *(b)* γ, *(c)* α_T, *(d)* δ, and *(e)* β.

10.21 A silicon npn transistor at T = 300 K has an area of 10^{-3} cm^2, neutral base width of 1 μm, and doping concentrations of $N_E = 10^{18}$ cm^{-3}, $N_B = 10^{17}$ cm^{-3}, $N_C = 10^{16}$ cm^{-3}. Other semiconductor parameters are $D_B = 20$ cm^2/s, $\tau_{E0} = \tau_{B0} = 10^{-7}$ s, and $\tau_{C0} = 10^{-6}$ s. Assuming the transistor is biased in the active region and the recombination factor is unity, calculate the collector current for: *(a)* $V_{BE} = 0.5$ V, *(b)* $I_E = 1.5$ mA, and (c) $I_B = 2\ \mu$A.

10.22 Consider a uniformly doped npn bipolar transistor at $T = 300$ K with the following parameters:

$N_E = 10^{18}$ cm^{-3}	$N_B = 5 \times 10^{16}$ cm^{-3}	$N_C = 10^{15}$ cm^{-3}
$D_E = 8$ cm^2/s	$D_B = 15$ cm^2/s	$D_C = 12$ cm^2/s
$\tau_{E0} = 10^{-8}$ s	$\tau_{B0} = 5 \times 10^{-8}$ s	$\tau_{C0} = 10^{-7}$ s
$x_E = 0.8\ \mu$m	$x_B = 0.7\ \mu$m	$J_{r0} = 3 \times 10^{-8}$ A/cm^2

For $V_{BE} = 0.60$ V and $V_{CE} = 5$ V, calculate *(a)* the currents J_{nE}, J_{pE}, J_{nC}, and J_R and *(b)* the current gain factors γ, α_T, δ, α, and β.

10.23 Three npn bipolar transistors have identical parameters except for the bass doping concentrations and neutral base widths. The base parameters for the three devices are as follows:

Device	Base duping	Base width
A	$N_B = N_{B0}$	$x_B = x_{B0}$
B	$N_B = 2N_{B0}$	$x_B = x_{B0}$
C	$N_B = N_{B0}$	$x_B = x_{B0}/2$

(The base doping concentration for the *B* device is twice that of A and C, and the neutral base width for the C device is half that of *A* and *B.)*

(a) Determine the ratio of the emitter injection efficiency of *(i)* device B to device A and *(ii)* device C to device A.

(h) Repeat part *(a)* for the base lranspon factor.

(c) Repeat pan (a)for the recombination factor.

(d) Which device has the largest common-emitter current gain β?

10.24 Repeat Problem 10.23 for three devices in which the emitter paramelers **vary.** The emitter parameters for the three devices are as follows:

Device	Emitter doping	Emitter width
A	$N_E = N_{E0}$	$x_E = x_{E0}$
B	$N_E = 2N_{E0}$	$x_E = x_{E0}$
C	$N_E = N_{E0}$	$x_E = x_{E0}/2$

10.25 An npn silicon transistor is biased in the inverse active mode with $V_{BE} = -3$ V and $V_{BC} = 0.6$ V The doping concentrations are $N_E = 10^{18}$ cm^{-3}, $N_B = 10^{17}$ cm^{-3}, and $N_C = 10^{16}$ cm^{-3}. Other parameters are $x_B = 1$ μm, $\tau_{E0} = \tau_{B0} = \tau_{C0} = 2 \times 10^{-7}$ s, $D_E = 10$ cm^2/s, $D_B = 20$ cm^2/s, D_C =15 cm^2/s, and $A = 10^{-3}$ cm'. (*a*) Calculate and plot the minority carrier distribution in the device. (*b*) Calculate the collector and emitter currents. (Neglect geometry factors and assume the recombination factor is unity.)

10.26 (a) Calculate the base transport factor, α_T, for $x_B/L_B = 0.01, 0.10, 1.0$, and 10. Assuming that γ and δ are unity, determine β for each case. (*b*) Calculate the emitter injection efficiency, γ, for $N_B/N_E = 0.01, 0.10, 1.0$, and 10. Assuming that α_T and δ are unity, determine β for each case. (*c*) Considering the results of pans (a)and (*b*), what conclusionr can be made concerning when the base transport factor or when the emitter injection efficiency are the limiting factors for the common-emitter current gain?

10.27 *Ia)* Calculate the recombination factor for $V_{BE} = 0.2, 0.4$, and 0.6 V. Assume the following parameters:

$$D_B = 25 \text{ cm}^2/\text{s} \qquad D_E = 10 \text{ cm}^2/\text{s}$$
$$N_E = 5 \times 10^{18} \text{ cm}^{-3} \qquad N_B = 1 \times 10^{17} \text{ cm}^{-3}$$
$$N_C = 5 \times 10^{15} \text{ cm}^{-3} \qquad x_B = 0.7 \ \mu\text{m}$$
$$\tau_{B0} = \tau_{E0} = 10^{-7} \text{ s} \qquad J_{r0} = 2 \times 10^{-9} \text{ A/cm}^2$$
$$n_i = 1.5 \times 10^{10} \text{ cm}^{-3}$$

(*b*) Assuming the base transport and emitter injection efficiency factors are unity, calculate the common-emitter current gain for the conditions in part (*a*). (*c*) Considering the results of part (*b*), what can be said about the recombination factor being the limiting factor in the common emitter current gain.

10.28 Consider an npn silicon bipolar transistor at T = 300 K with the following parameters:

$$D_B = 25 \text{ cm}^2\text{1s} \qquad D_E = 10 \text{ cm}^2/\text{s}$$
$$\tau_{B0} = \qquad \tau_{E0} = 5 \times$$
$$N_B = 10^{16} \text{ cm}^{-3} \qquad x_E = 0.5 \ \mu\text{m}$$

The recombination factor, δ, has been determined to he $\delta = 0.998$. We need a common-emitter current gain of $\beta = 120$. Assuming that $\alpha_T = \gamma$, determine the maximum base width, x_B, and the minimum emitter doping, N_E, to achieve this specification.

***10.29** (*a*) The recombination current density, J_{r0}, in an npn silicon bipolar transistor at $T = 300$ K is $J_{r0} = 5 \times 10^{-8}$ A/cm^2. The uniform dopings are $N_E = 10^{18}$ cm^{-3}, $N_B = 5 \times 10^{16}$ cm^{-3}, and $N_C = 10^{15}$ cm^{-3}. Other parameters are $D_E = 10$ cm^2/s, $D_B = 25$ cm^2/s. $\tau_{E0} = 10^{-8}$ s, and $\tau_{B0} = 10^{-7}$ s. Determine the neutral base width so that the recornbination factor is $\delta = 0.995$ when $V_{BE} = 0.55$ V (*b*) If J_{r0} remains constant with temperature, what is the value of δ when $V_{BE} = 0.55$ V for the case when the temperature is T = 400 K? Use the value of x_B determined in pan (a).



***10.30** (a)Plot, for a bipolar transistor, the base transport factor. α_T, as a function of (x_B/L_B) over the range $0.01 \leq (x_B/L_B) \leq 10$. (Use a log scale an the horizontal axis.)

(*b*) Assuming that the emitter injection efficiency and recombination factors are unity, plot the common emitter gain for the conditions in part ***(a).(c)*** Considering the results of pan ***(b),*** what can be said about the base transport factor being the limiting factor in the common emitter current gain?

1031 (a) Plot the emitter injection efficiency as a function of the doping ratio, N_B/N_E, over therange $0.01 \leq N_B/N_E \leq 10$. Assume that $D_E = D_B$, $L_B = L_E$, and $x_B = x_E$. (Use a log scale on the horizontal axis.) Neglect bandgap narrowing effects. (*b*) Assuming that the base transport factor and recombination factors are unity, plot the common emitter current gain for the conditions in part (a). (*c*) Considering the results of part (*b*), what can be said about the emitler injection efficiency being the limiting factor in the common emitter current gain.

10.32 **(a)** Plot the recombination factor as a function of the forward-bias B-E voltage for $0.1 \leq V_{BE} \leq 0.6$. Assume the following parameters:

$$D_B = 25\ \text{cm}^2/\text{s} \qquad D_E = 10\ \text{cm}^2/\text{s}$$
$$N_E = 5 \times 10^{18}\ \text{cm}^{-3} \qquad N_B = 1 \times 10^{17}\ \text{cm}^{-3}$$
$$N_C = 5 \times 10^{15}\ \text{cm}^{-3} \qquad x_B = 0.7\ \mu\text{m}$$
$$\tau_{B0} = \tau_{E0} = 10^{-7}\ \text{s} \qquad J_{r0} = 2 \times 10^{-9}\ \text{A/cm}^2$$
$$n_i = 1.5 \times 10^{10}\ \text{cm}^{-3}$$

(*b*) Assuming the base transport and emitter injection efficiency factors are unity, plot the common emitter current gain for the conditions in part *(a).(c)* Considering the results of part (*b*), what can be said about the recombination factor being the limiting factor in the common emitter current gain.

10.33 The emitter in a BJT is often made very thin to achieve high operating speed. In this problem. we investigate the effect of emitter width an current gain. Consider the emitter injection efficiency given by Equation (10.35a). Assume that $N_E = 100N_B$, $D_E = D_B$, and $L_E = L_B$. Also let $x_B = 0.1L_B$. Plot the emitter injection efficiency for $0.01L_E \leq x_E \leq 10L_E$. From these results, discuss the effect of emitter width on the current gain.

Section 10.4 Nonideal Effects

10.34 A silicon pnp bipolar transistor at T = 300 K has uniform dopings of $N_E = 10^{18}\ \text{cm}^{-3}$, $N_B = 10^{16}\ \text{cm}^{-3}$, and $N_C = 10^{15}\ \text{cm}^{-3}$. The metallurgical hase width is 1.2 μm. Let $D_B = 10\ \text{cm}^2/\text{s}$ and $\tau_{B0} = 5 \times 10^{-7}$ s. Assume that the minority carrier hole concentration in the hase can be approximated by a linear distribution. Let $V_{EB} = 0.625$ V (*a*) Determine the hole diffusion current density in the base for $V_{BC} = 5$ V, $V_{BC} = 10$ V, and $V_{BC} = 15$ V (*b*) Estimate the Early voltage.

***10.35** The base width of a bipolar transistor is normally small to provide a large current gain and increased speed. The hase width also affects the Early voltage. In a silicon npn bipolar transistor at T = 300 K, the doping concentrations arc $N_E = 10^{18}\ \text{cm}^{-3}$, $N_B = 3 \times 10^{16}\ \text{cm}^{-3}$, and $N_C = 5 \times 10^{15}\ \text{cm}^{-3}$. Assume $D_B = 20\ \text{cm}^2/\text{s}$ and $\tau_{B0} = 5 \times 10^{-7}$ s, and let $V_{BE} = 0.70$ V. Using voltages $V_{CB} = 5$ V and $V_{CB} = 10$ V as two data paints. estimate the Early voltage for metallurgical base widths of (a) 1.0 μm, (*b*) 0.80 μm, and (*c*) 0.60 μm.

10.36 An npn silicon bipolar transistor has a base doping concentration of $N_B = 10^{17}$ cm^{-} a collector doping concentration of $N_C = 10^{16}$ cm^{-3}, a metallurgical base width of 1.1 μm, and a base minority carrier diffusion coefficient of $D_B = 20$ cm^2/s. The transistor is biased in the forward-active region with $V_{BE} = 0.60$ V. Determine ***(a)*** the change in the neutral base width as V_{CB} changes from 1 V to 5 V, and *(b)* the corresponding change in the collector current.

1037 Consider a uniformly doped silicon npn bipolar transistor in which $x_E = x_B$, $L_E = L_B$, and $D_E = D_B$. Assume that $\alpha_T = \delta = 0.995$ and let $N_B = 10^{17}$ cm^{-3}. Calculate and plot the common emitter current gain β for $N_E = 10^{17}, 10^{18}, 10^{19}$, and 10^{20} cm^{-3}, and for the case ***(a)*** when the bandgap narrowing effect is neglected. and (b) when the bandgap narrowing effect is taken into account.

10.38 A silicon pnp bipolar transistor at T = 300 K is to be designed so that the emitter injection efficiency is $\gamma = 0.996$. Assume that $x_E = x_B$, $L_E = L_B$, $D_E = D_B$, and let $N_E = 10^{19}$ cm^{-3}. *(a)* Determine the maximum base doping, taking into account bandgap narrowing. ***(h)*** If bandgap narrowing were neglected, what would he the maximum base doping required?

10.39 A first-approximation type calculation of the current crowding effect can be made using the geometry shown in Figure 10.51. Assume that one-half of the base current enters from each side of the emitter strip and flows uniformly to the center of the emitter. Assume the base is p type with the following parameters:

$N_B = 10^{16}$ cm^{-3}	$x_B = 0.70\ \mu$m
$\mu_p = 400$ cm^2/V-s	$S = 8\ \mu$m
Emitter length	L = 100 μm

(a) Calculate the resistance between $x = 0$ and $x = S/2$. *(b)* If $\frac{1}{2} I_B = 10\ \mu$A, calculate the voltage drop between $x = 0$ and $x = S/2$. *(c)* If $V_{BE} = 0.6$ V at $x = 0$, estimate in percent the number of electrons being injected into the base at $x = S/2$ compared to $x = 0$.

10.40 Consider the geometry shown in Figure 10.51 and the device parameters in Problem 10.39 except the emitter width S. The emitter width S is to be changed so that the number of electrons injected into the hase at $x = Sf2$ is no more than 10 percent less than the number of electrons injected into the base at $x = 0$. Calculate S.

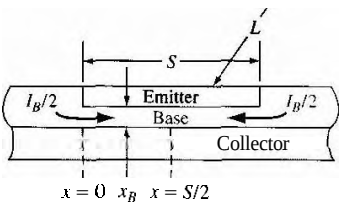


Figure 10.51 | Figure for Problems 10.39 and 10.40.

***10.41** The base doping in a diffused n^+pn bipolar transistor can be approximated by an exponential as

$$N_B = N_B(0)\exp\left(\frac{-ax}{x_B}\right)$$

where a is a constant and is given by

$$a = \ln\left(\frac{N_B(0)}{N_B(x_B)}\right)$$

(*a*) Show that, in thermal equilibrium, the electric field in the neutral base region is a constant. **(*b*)** Indicate the direction of the electric field. Does this electric field aid or retard the flow of minority carrier electrons across the base? (*c*) Derive an expression for the steady-state minority carrier electron concentration in the base under forward bias. Assume no recombination occurs in the base. (Express the electron concentration in terms of the electron current density.)

10.42 Consider a silicon npn bipolar transistor with uniform dopings of $N_E = 5 \times 10^{18}$ cm^{-3}, $N_B = 10^{17}$ cm^{-3}, and $N_C = 5 \times 10^{15}$ cm^{-3}. Assume the common-base current gain is $\alpha = 0.9920$. Determine (*a*) BV_{CBO}, (*b*) BV_{CEO}, and (*c*) the base–emitter breakdown voltage. (Assume $n = 3$ for the empirical constant.)

10.43 A high-voltage silicon npn bipolar transistor is to be designed such that the uniform base doping is $N_B = 10^{16}$ cm^{-3} and the common-emitter current grain is $\beta = 50$. The breakdown voltage BV_{CEO} is to he at least 60 V. Determine the maximum collector doping and the minimum collector length to support this voltage. (Assume $n = 3$.)

10.44 A uniformly doped silicon epitaxial npn bipolar transistor is fabricated with a base doping of $N_B = 3 \times 10^{16}$ cm^{-3} and a heavily doped collector region with $N_C = 5 \times 10^{17}$ cm^{-3}. The neutral base width is $x_B = 0.70\ \mu$m when $V_{BE} = V_{BC} = 0$. Determine V_{BC} at which punch-through occurs. Compare this value to the expected avalanche breakdown voltage of the junction.

10.45 A silicon npn bipolar transistor has a base doping concentration of $N_B = 10^{17}$ cm^{-3}, a collector duping concentration of $N_C = 7 \times 10^{15}$ cm^{-3}, and a metallurgical base width of 0.50 μm. Let $V_{BE} = 0.60$ V. (*a*) Determine V_{CE} at punch-through. (*b*) Determine the peak electric field in the B-C space charge region at punch-through.

10.46 A uniformly doped silicon pnp bipolar transistor is to be designed with $N_E = 10^{19}$ c m^{-3} and $N_C = 10^{16}$ cm^{-3}. The metallurgical base width is 0.75 μm. Determine the minimum base doping so that the punch-through voltage is no less than $V_{pt} = 25$ V.

Section 10.5 Equivalent Circuit Models

10.47 The V_{CE}(sat) voltage of an npn transistor in saturation continues to decrease slowly as the base current increases. In the Ehers–Moll model. assume $\alpha_F = 0.99$. $\alpha_R = 0.20$. and $I_C = 1$ mA. For T = 300 K, determine the base current, I_B, necessary to give (*a*) V_{CE}(sat) = 0.30 V, **(*b*)** V_{CE}(sat) = 0.20 V, and (*c*) V_{CE}(sat) = 0.10 V.

10.48 Consider an npn bipolar transistor biased in the active mode. Using the Ebers–Moll model, derive the equation for the base current, I_B, in terms of α_F, α_R, I_{ES}, I_{CS}, and V_{BE}.

10.49 Consider the Ebers–Moll model and let the base terminal be open so $I_B = 0$. Sho that. when a collector-emitter voltage is applied. we have

$$I_C \equiv I_{CEO} = I_{CS}\frac{(1-\alpha_F\alpha_R)}{(1-\alpha_F)}$$

10.50 In the Ebers–Moll model. let $\alpha_F = 0.98$. $I_{}$, = 10^{-13} A, and I, , = 5 x 10^{-13} A. $T = 300$ K. Plot I_C versus V_{C} , for $-V_{BE} < V_{CB} < 3$ V and for $V_{BE} = 0.2, 0.4$. and 0.6 V. (Note that $V_{CB} = -V_{BC}$.) What can be said about the base width modu lation effect using this model?



10.51 The collector–emitter saturation voltage. from the Ebers–Mull model, is given by Equation (10.77). Consider a power BJT in which $\alpha_F = 0.98$. $\alpha_R = 0.20$. and 4 I A. Plot V_{CE}(sat) versus I_B over the range $0.03 \leq I_B \leq 1.0$ A.

Section 10.6 Frequency Limitations

10.52 Consider a silicon npn transistor st $T = 300$ K. Assume thr following parameters:

$I_E = 0.5$ mA	C_{je} = 0.8 pF
$x_B = 0.7\ \mu$m	D_n = 25 cm^2/s
$x_{dc} = 2.0\ \mu$m	$r_c = 30\ \Omega$
$C_{} = C_{} = 0.08$ pF	$\beta = 50$

(a) Calculate the transit time factors. (*b*) Calculate the cutoff and beta cutoff frequencies. f_t and f_β. respectively.

10.53 In a particular bipolar transistor, the base transit time is 20 percent of the total delay time. The base width is 0.5 μm and the base diffusion coefficient is $D_B = 20$ cm^2/s. Determine the cutoff frequency.

10.54 Assume the base transit time of a BJT is 100 ps and carriers cross the 1.2 μm B-C space charge region at a speed of 10^7 cm/s. The emitter–base junction charging time is 25 ps and the collector capacitance and resistance **are** 0.10 pF and 10 Ω. respectivcly. Determine the cutoff frequency.

Summary and Review



***10.55** (*a*) A silicon npn bipolar transistor at $T = 300$ K is to be designed with an Early voltage of at least 200 V and a current gain of at least $\beta = 80$. (b) Repeat pan (*a*) for s pnp bipolar transistor,



***10.56** Design a uniformly doped silicon npn bipolar transistor so that $\beta = 100$ at $T = 300$ K. The maximum CE voltage is to be 15 V and any breakdown voltage is to be at least 3 times this value. Assume the recombination factor is constant at $\delta = 0.995$. The transistor is to be operated in low injection with a maximum collector current **of** $I_C = 5$ mA. Bandgap narrowing effects and base width modulation effects are to be minimized. Let $D_E = 6$ cm^2/s, $D_B = 25$ cm^2/s, $\tau_{E0} = 10^{-8}$ s, and $\tau_{B0} = 10^{-7}$ s. Determine duping concentrations, the metallurgical base width, the active area, and the maximum allowable V_{BE}.



***10.57** Deign a pair of complementary npn and pnp bipolar transistors. The transistors are to have the same metallurgical base and emitter widths of $W_B = 0.75\ \mu$m and

$x_E = 0.5\ \mu m$. Assume that the following minority carrier parameters apply to each device.

$$D_n = 23\ cm^2/s \qquad \tau_{n0} = 10^{-7}\ s$$
$$D_p = 8\ cm^2/s \qquad \tau_{p0} = 5 \times 10^{-8}\ s$$

The collector doping concentration in each device is $5 \times 10^{15}\ cm^{-3}$ and the recombination factor in each device is constant at $\delta = 0.9950$. *(a)* Design, if possible, fhe devices so that $\beta = 100$ in each device. If this is not possible. how close a match can be obtained'! (*b*) With equal forward-bias base–emitter voltages applied. the collector currents are to he $I_C = 5$ mA with each device operating in low-injection. Determine the active cross-sectional areas.

READING LIST

1. 'Dimitrijev, S. *Understanding Semiconductor* ***Devices.*** New York: Oxford University Press, 2000.
2. Kano, K. ***Semiconductor*** *Devices.* Upper Saddle River, NJ: Prentice Hall, 1998.
3. Muller, R. S., and T. I. Kamins. ***Device*** *Electronics for* ***Integrated Circuits.*** 2nd ed. New York: Wiley, 1986.
4. Navon, D. H. ***Semiconductor*** *Microdevices and Materials.* New York: Holt, Rinehart, & Winston, 1986.
5. Neudeck. G. W. *The Bipolar* ***Junction*** *Transistor.* Vol. 3 of the ***Modular*** *Series on Solid State Devices.* 2nd ed. Reading, MA: Addison-Wesley, 1989.
6. Ng, K. K. ***Complete Guide to Semiconductor*** *Devices.* New York: McGraw-Hill, 1995.
7. Ning, T. H., and R. D. Isaac. "Effect of Emitter Contact on Current Gain of Silicon Bipolar Devices." *Polysilicon Emitter* ***Bipolar Transistors.*** eds. ***A. K.*** Kapoor and D. J. Roulston. New York: IEEE Press, 1989.
8. Pierret. R. F. *Semiconductor* ***Device Fundamentals.*** Reading, ***MA:*** Addison-Wesley, 1996.
9. Roulston, D. J. ***Bipolar Semiconductor*** *Devices.* New York: McGraw-Hill, 1990.
10. Roulston, D. J. An ***Introduction*** *to* ***the Physics of Semiconductor Devices.*** New York Oxford University Press, 1999.
11. Shur, M. *GaAs* ***Devices and Circuits.*** New York: Plenum Press. 1987.
12. Shur, M. ***Introduction*** *to* ***Electronic Devices.*** New York: John Wiley & Sons, Inc., 1996.
13. Shur, M. ***Physics*** *of Semiconductor* ***Devices.*** Englewood Cliffs, NJ: Prentice Hall, 1990.
14. Singh, J. ***Semiconducror Devices: An*** *Introduction.* New York: McGraw-Hill, 1994.
15. Singh, J. ***Semiconducror Devices: Basic Principles.*** New York: John Wiley & Sons, Inc., 2001
16. Streetman, B. G.. and S. Banerjee. *Solid State Electronic Devices,* 5th ed. Upper Saddle River, NJ: Prentice Hall, 2000.
17. Sze. S. M. ***High-Speed*** *Semiconductor Devices.* New ***York:*** Wiley, 1990.
18. Sze, S. M. *Physics of Semiconductor Devices.* 2nd ed. New York: Wiley, 1981.

19. Tiwari, S., S. L. Wright, and A. W. Kleinsasser. "Transport and Related Properties o (Ga. Al)As/GaAs Double Heterojunction Bipolar Junction Transistors." IEEE *Transactions on Electron Devices,* ED-34 (February 1987), pp. 185–87.

***20.** Taur, Y., and T. H. Ning. *Fundamentals of Modern VLSI Devices. New* York: Cambridge University Press, 1998.

***21.** Wang, S. *Fundamentals of Semiconductor Theory and Device Physics.* Englewood Cliff*, NJ: Prentice Hall, 1989.

***22.** Warner, R. M., Jr., and B. L. Grung. *Transistors: Fundamentals for the Integrated-Circuit Engineer.* New York: Wiley. 1983.

23. Yang. E. S. *Microelectronic Devices.* New York: McGraw-Hill, 1988.

***24.** Yuan, J. S. *SiGe, GaAs, and InP Heterojunction Bipolar Transistors.* New York: John Wiley & Sons, Inc., 1999.

CHAPTER 11

Fundamentals of the Metal–Oxide–Semiconductor Field-Effect Transistor

PREVIEW

The fundamental physics of the Metal–Oxide–Semiconductor Field-Effect Transistor (MOSFET) is developed in this chapter. Although the bipolar transistor was discussed in the last chapter, the material in this chapter presumes a knowledge only of the semiconductor material properties and characteristics of the pn junction.

The MOSFET, in conjunction with other circuit elements, is capable of voltage gain and signal-power gain. The MOSFET is also used extensively in digital circuit applications where, because of its relatively small size, thousands of devices can be fabricated in a single integrated circuit. The MOSFET is, without doubt. the core of integrated circuit design at the present lime.

The MOS designation is implicity used only for the metal–silicon dioxide (SiO_2)–silicon system. The more general terminology is metal–insulator–semiconductor (MIS), where the insulator is not necessarily silicon dioxide and the semiconductor is not necessarily silicon. We will use the MOS system throughout this chapter although the same basic physics applies to the MIS system.

The heart of the MOSFET is a metal–oxide–semiconductor structure known as an MOS capacitor. The energy bands in the semiconductor near the oxide–semiconductor interface bend as a voltage is applied across the MOS capacitor. The position of the conduction and valence bands relative to the Fermi level at the oxide–semiconductor interface is a function of the MOS capacitor voltage, so that the characteristics of the semiconductor surface can be inverted from p-type to n-type, or from n-type to p-type, by applying the proper voltage. The operation and characteristics of the MOSFET are dependent on this inversion and the creation of

an inversion charge density at the semiconductor surface. The threshold voltage is defined as the applied gate voltage required to create the inversion layer charge and is one of the important parameters of the MOSFET.

The various types of MOSFETs are examined and a qualitative discussion of the current–voltage characteristics is initially presented. A mathematical derivation of the current–voltage relation is then covered in detail. The frequency response and limitations of the MOSFET are also considered.

Although we have not discussed fabrication processes in any detail in this text, there is an MOS technology that should be considered, since it directly influences the characteristics and properties of the MOS devices and circuits. We will consider the complementary MOS (CMOS) process. The discussion of this technology will be brief, but should provide a good base for further in-depth study. ■

11.1 | THE TWO-TERMINAL MOS STRUCTURE

The heart of the MOSFET is the metal–oxide–semiconductor capacitor shown in Figure 11.1. The metal may be aluminum or some other type of metal, although in many cases, it is actually a high-conductivity polycrystalline silicon that has been deposited on the oxide; however, the term metal is usually still used. The parameter t_{ox} in the figure is the thickness of the oxide and ϵ_{ox} is the permittivity of the oxide.

11.1.1 Energy-Band Diagrams

The physics of the MOS structure can be more easily explained with the aid of the simple parallel-plate capacitor. Figure 11.2a shows a parallel-plate capacitor with the top plate at a negative voltage with respect to the bottom plate. An insulator material separates the two plates. With this bias, a negative charge exists on the top plate, a positive charge exists on the bottom plate, and an electric field is induced between the two plates as shown. The capacitance per unit area for this geometry is

$$C' = \frac{\epsilon}{d} \tag{11.1}$$

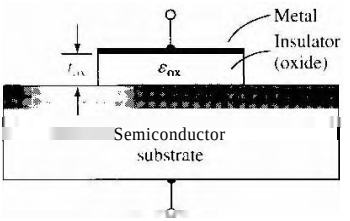


Figure 11.1 | The basic MOS capacitor structure.

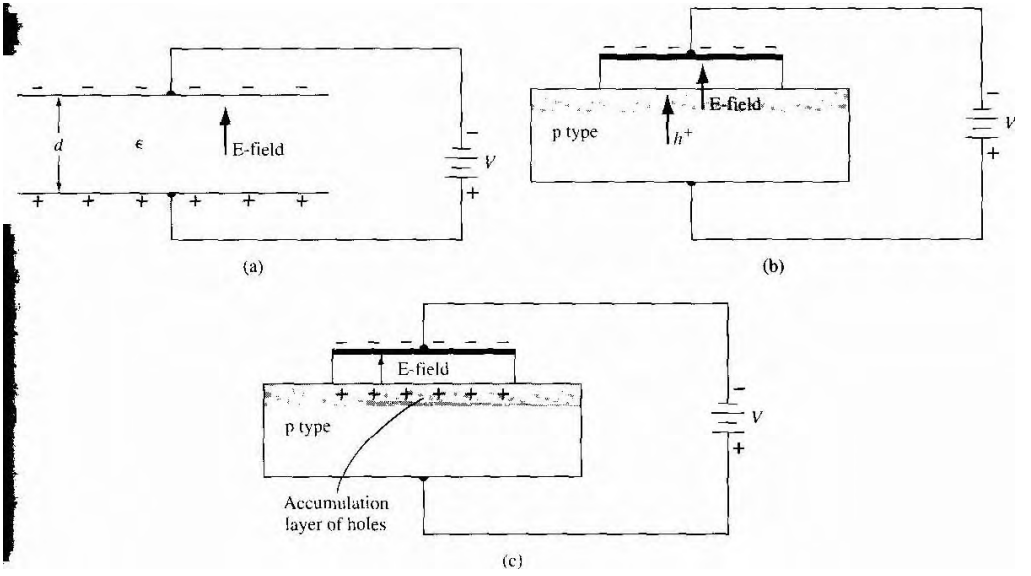


Figure 11.2 | (a) A parallel-plate capacitor showing the electric field and conductor charges. (b) A corresponding MOS capacitor with a negative gate bias showing the electric field and charge Row. (c) The MOS capacitor with an accumulation layer of holes.

where ϵ is the permittivity of the insulator and d is the distance between the two plates. The magnitude of the charge per unit area on either plate is

$$Q' = C'V \tag{11.2}$$

where the prime indicates charge or capacitance per unit area. The magnitude of the electric field is

$$E = \frac{V}{d} \tag{11.3}$$

Figure 11.2b shows an MOS capacitor with a p-type semiconductor substrate. The top metal gate is at a negative voltage with respect to the semiconductor substrate. From the example of the parallel-plate capacitor, we can see that a negative charge will exist on the top metal plate and an electric field will be induced with the direction shown in the figure. If the electric field were to penetrate into the semiconductor, the majority carrier holes would experience a force toward the oxide–semiconductor interface. Figure 11.2c shows the equilibrium distribution of charge in the MOS capacitor with this particular applied voltage. An *accumulation layer* of holes in the oxide–semiconductor junction corresponds to the positive charge on the bottom "plate" of the MOS capacitor.

Figure 11.3a shows the same MOS capacitor in which the polarity of the applied voltage is reversed. A positive charge now exists on the top metal plate and the induced electric field is in the opposite direction as shown. If the electric field penetrates the semiconductor in this case, majority carrier holes will experience a force away from the oxide–semiconductor interface. As the holes are pushed away from the interface, a negative space charge region is created because of the fixed ionized acceptor atoms. The negative charge in the induced depletion region corresponds to the negative charge on the bottom "plate" of the MOS capacitor. Figure 11.3b shows the equilibrium distribution of charge in the MOS capacitor with this applied voltage.

The energy-hand diagram of the MOS capacitor with the p-type substrate, for the case when anegative voltage is applied to the top metal gate, is shown in Figure 11.4a. The valence-band edge is closer to the Fermi level at the oxide–semiconductor interface than in the bulk material, which implies that there is an accumulation of holes. The semiconductor surface appears to he more p-type than the bulk material. The Fermi level is a constant in the semiconductor since the MOS system is in thermal equilibrium and there is no current through the oxide.

Figure 11.4b shows the energy-band diagram of the MOS system when a positive voltage is applied to the gate. The conduction and valence band edges bend as

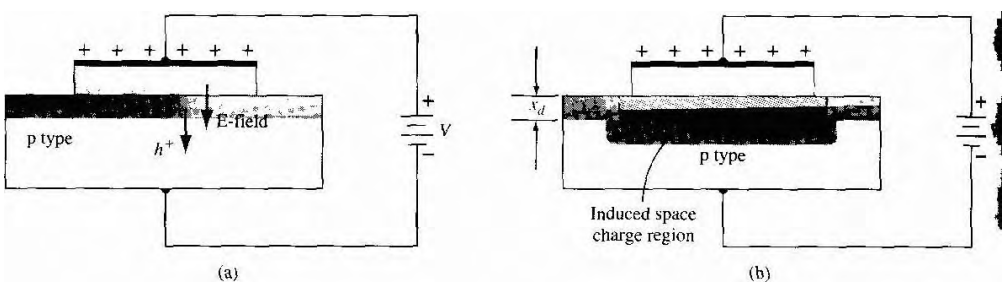


Figure **11.3** | The MOS capacitor with a moderate positive gate bias, showing (a) the electric field and charge flow and (b) the induced space charge region.

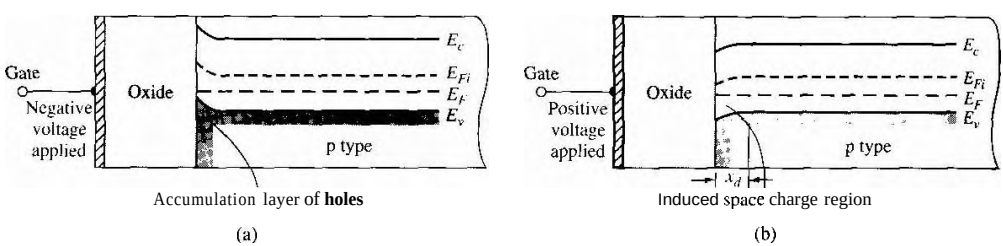


Figure 11.4 | The energy-band diagram of an MOS capacitor with a p-type substrate for (a) a negative gate bias and (b) a moderate positive gate bias.

hown in the figure, indicating a space charge region similar to that in a pn junction. The conduction hand and intrinsic Fermi levels move closer to the Fermi level. The induced space charge width is x_d.

Now consider the case when a still larger positive voltage is applied to the top metal gate of the MOS capacitor. We expect the induced electric field to increase in magnitude and the corresponding positive and negative charges on the MOS capacitor to increase. Alarger negative charge in the MOS capacitor implies a larger induced space charge region and more band bending. Figure 11.5 shows such a condition. The intrinsic Fermi level at the surface is now below the Fermi level: thus, the conduction band is closer to the Fermi level than the valence band is. This result implies that the surface in the semiconductor adjacent to the oxide-semiconductor interface is n type. By applying a sufficiently large positive gate voltage, we have inverted the surface of the semiconductor from a p-type to an n-type semiconductor. We have created an *inversion layer* of electrons at the oxide-semiconductor interface.

In the MOS capacitor structure that we have just considered, we assumed a p-type semiconductor substrate. The same type of energy-band diagrams can be constructed for an MOS capacitor with an n-type semiconductor substrate. Figure 11.6a shows the MOS capacitor structure with a positive voltage applied to the top gate terminal. Apositive charge exists on the top gate and an electric field is induced with the direction shown in the figure. An accumulation layer of electrons will be induced in

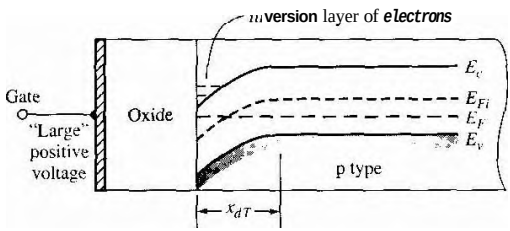


Figure 11.5 | The energy-band diagram of the MOS capacitor with a p-type substrate for a "large" positive gate bias.

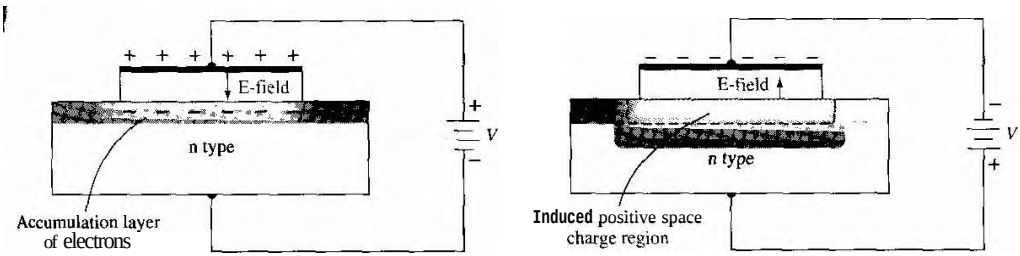


Figure 11.6 | The MOS capacitor with an n-type substrate for (a) a positive gate bias and (b) a moderate negative gate bias.

the n-type substrate. The case when a negative voltage is applied to the top gate is shown in Figure 11.6b. A positive space charge region is induced in the n-type semiconductor in this situation.

The energy-band diagrams for this MOS capacitor with the n-type suhstrate ar shown in Figure 11.7. Figure 11.7a shows the case when a positive voltage is ap plied to the gate and an accumulation layer of electrons is formed. Figure 11.7 shows the positive space charge region induced by an applied negative gate voltag in it the conduction and valence band energies bend upward. Figure 11.7c shows th energy bands when a larger negative voltage is applied to the gate. The conduction and valence bands are bent even more and the intrinsic Fermi level has moved above the Fermi level so that the valence band is closer to the Fermi level than the

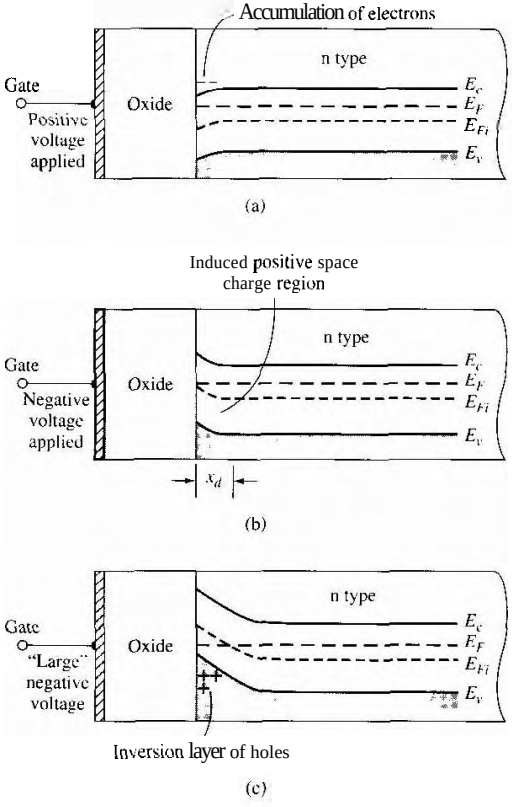


Figure 11.7 | The energy-band diagram of the MOS capacitor with an n-type substrate for (a) a positive gate bias. (b) a moderate negative bias, and (c) a "**large**" negative gate bias.

conduction band is. This result implies that the semiconductor surface adjacent to the oxide-semiconductor interface is p type. By applying a sufficiently large negative voltage to the gate of the MOS capacitor, the semiconductor surface has been inverted from n type to p type. An inversion layer of holes has been induced at the oxide-semiconductor interface.

11.1.2 Depletion Layer Thickness

We may calculate the width of the induced space charge region adjacent to the oxide-semiconductor interface. Figure 11.8 shows the space charge region in a p-type semiconductor substrate. The potential ϕ_{fp} is the difference (in volts) between E_{Fi} and E_F and is given by

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) \tag{11.4}$$

where N_a is the acceptor doping concentration and n_i is the intrinsic carrier concentration.

The potential ϕ_s is called the surface potential; it is the difference (in volts) between E_{Fi} measured in the bulk semiconductor and E_{Fi} measured at the surface. The surface potential is the potential difference across the space charge layer. The space charge width can now be written in a form similar to that of a one-sided pn junction. We can write that

$$x_d = \left(\frac{2\epsilon_s \phi_s}{eN_a}\right)^{1/2} \tag{11.5}$$

where ϵ_s is the permittivity of the semiconductor. Equation (11.5) assumes that the abrupt depletion approximation is valid.

Figure 11.9 shows the energy bands for the case in which $\phi_s = 2\phi_{fp}$. The Fermi level at the surface is as far above the intrinsic level as the Fermi level is below the

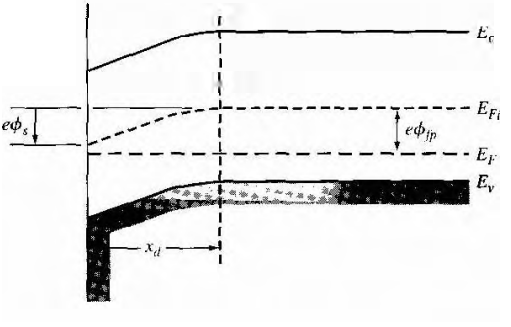


Figure 11.8 | The energy-band diagram in the p-type semiconductor, indicating surface potential.

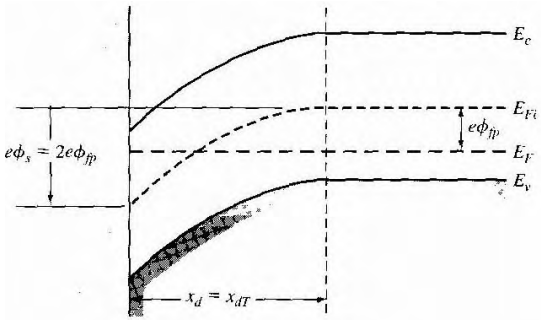


Figure 11.9 | The energy-hand diagram in the p-type semiconductor at the threshold inversion point.

intrinsic level in the hulk semiconductor. The electron concentration at the surface is the same as the hole concentration in the bulk material. This condition is known as the *threshold inversion point.* The applied gate voltage creating this condition is known as the ***threshold voltage.*** If the gate voltage increases above this threshold value, the conduction bend will hrnd slightly closer to the Fermi level. but the change in the conduction hand at the surface is now only a slight function of gate voltage. The electron concentration at the surface. however, is an exponential function of the surface potential. The surface potential may increase by a few (kT/e) volts, which will change the electron concentration by orders of magnitude, but the space charge width changes only slightly. In this case, then, the space charge region has essentially reached a maximum width.

The maximum space charge width, x_{dT}, at this inversion transition point can be calculated from Equation (11.5) by setting $\phi_s = 2\phi_{fp}$. Then

$$x_{dT} = \left(\frac{4\epsilon_s \phi_{fp}}{eN_a}\right)^{1/2} \tag{11.6}$$

EXAMPLE 11.1

Objective

To calculate the maximum space charge width given a particular semiconductor doping concentration.

Consider silicon at T = 300 K doped to $N_a = 10^{16}$ cm^{-3}. The intrinsic carrier concentration is $n_i = 1.5 \times 10^{10}$ cm^{-3}.

■ **Solution**

From Equatiun (11.4), we have

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259)\ln\left(\frac{10^{16}}{1.5 \times 10^{10}}\right) = 0.347 \text{ V}$$

Then the maximum space charge width is

$$x_{dT} = \left[\frac{4\epsilon_s\phi_{fp}}{eN_a}\right]^{1/2} = \left[\frac{4(11.7)(8.85\times10^{-14})(0.347)}{(1.6\times10^{-19})(10^{16})}\right]^{1/2}$$

or

$$x_{dT} = 0.30 \times 10^{-4}\ \text{cm} = 0.30\,\mu\text{m}$$

■ **Comment**

The maximum induced space charge width is on the same order of magnitude as pn junction space charge widths.

We have been considering a p-type semiconductor substrate. The same maximum induced space charge region width occurs in an n-type substrate. Figure 11.10 is the energy-band diagram at the threshold voltage with an n-type substrate. We can write

$$\boxed{\phi_{fn} = V_t \ln\left(\frac{N_d}{n_i}\right)} \tag{11.7}$$

$$\boxed{x_{dT} = \left(\frac{4\epsilon_s\phi_{fn}}{eN_d}\right)^{1/2}} \tag{11.8}$$

Note that we are always assuming the parameters ϕ_{fp} and ϕ_{fn} to be positive quantities.

Figure 11.11 is a plot of x_{dT} at T = 300 K as a function of doping concentration in silicon. The semiconductor doping can be either n-type or p-type.

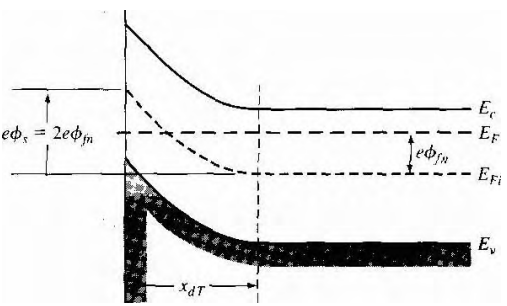


Figure 11.10 | The energy-band diagrem in the n-type semiconductor at the threshold inversion point.

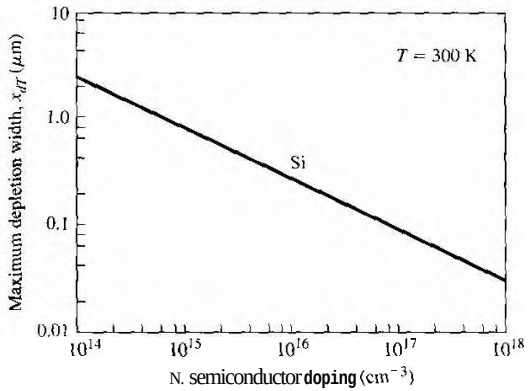


Figure 11.11 | Maximum induced space charge region width versus semiconductnr doping.

TEST YOUR UNDERSTANDING

E11.1 (*a*) Consider as oxide-to-p-type silicon junction at $\mathbf{7} = 300$ K. The impurity doping concentration in the silicon is N, $= 3 \times 10^{16}$ cm^{-3}. Calculate the maximum space-charge width in the silicon. (*b*) Repeat part *(a)* for an impurity concentration of $N_a = 10^{15}$ cm^{-3}. [ɯɹn £98'0 (*q*) 'ɯɹn 081'0 (*n*) 'sn∀]

E11.Z Consider an oxide-to-n-type silicon junction at $T = 300$ K. The impurity doping concentration in the silicon is $N_d = 8 \times 10^{15}$ cm^{-3}. Calculate the maximum space-charge width in the silicon. (ɯɹn ɛɛ'0 'sn∀)

11.1.3 Work Function Differences

We have been concerned. so far. wtth the energy-band diagrams of the semiconductor material. Figure 11.12a shows the energy levels in the metal, silicon dioxide, and silicon relative to the vacuum level. The metal work function is ϕ_m and the electron affinity is χ. The parameter χ_i is the oxide electron affinity and, for silicon dioxide $\chi_i = 0.9$ V.

Figure 11.12b shows the energy-band diagram of the entire metal–oxide–semiconductor structure with zero gate voltage applied. The Fermi level is a constant through the entire system at thermal equilibrium. We may define ϕ_m' as a modified metal work function—the potential required to inject an electron from the metal into the conduction band of the oxide. Similarly, χ' is defined as a modified electron affinity. The voltage V_{ox0} is the potential drop across the oxide for zero applied gate voltage and is not necessarily zero because of the difference between ϕ_m and χ. The potential ϕ_{s0} is the surface potential for this case.

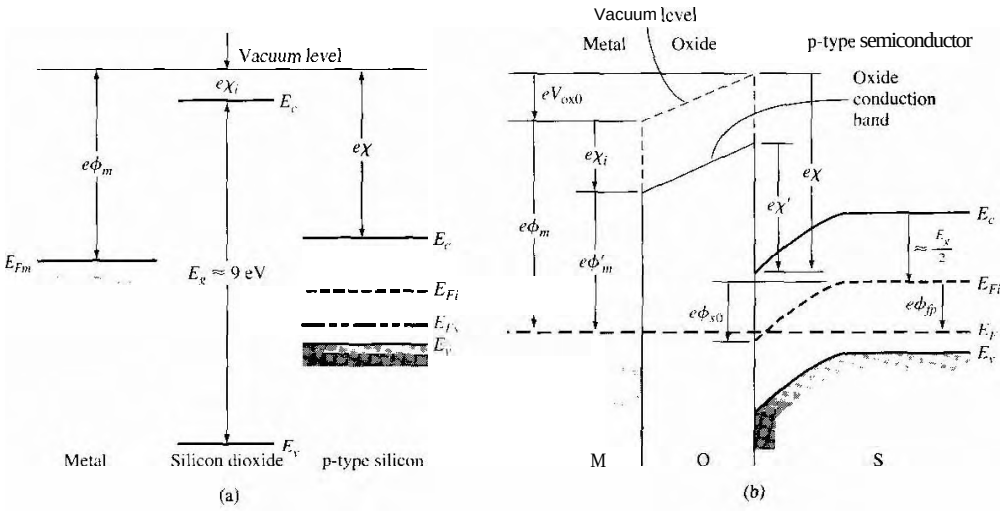


Figure 11.12 | (a) Energy levels in an MOS system pnor to contact and (b) energy-hand diagram through the MOS structure in thermal equilibrium after contact.

If we sum the energies from the Fermi level on the metal side to the Fermi level on the semiconductor side, we have

$$e\phi'_m + eV_{ox0} = e\chi' + \frac{E_g}{2} - e\phi_{s0} + e\phi_{fp} \tag{11.9}$$

Equation (11.9) can be rewritten as

$$V_{ox0} + \phi_{s0} = -\left[\phi'_m - \left(\chi' + \frac{E_g}{2e} + \phi_{fp}\right)\right] \tag{11.10}$$

We can define a potential ϕ_{ms} as

$$\boxed{\phi_{ms} \equiv \left[\phi'_m - \left(\chi' + \frac{E_g}{2e} + \phi_{fp}\right)\right]} \tag{11.11}$$

which is known as the metal–semiconductor work function difference

Objective | EXAMPLE 11.2

To calculate the metal–semiconductor work function difference ϕ_{ms} for a given MOS system and semiconductor doping.

For an aluminium–silicon dioxide junction, $\phi'_m = 3.20$ V and for a silicon–silicon dioxide junction, $\chi' = 3.25$ V. We may assume that $E_g = 1.11$ eV. Let the p-type doping be $N_a = 10^{14}$ cm^{-3}.

■ Solution

For silicon at $T = 300$ K, we may calculate ϕ_{fp} as

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259)\ln\left(\frac{10^{14}}{1.5 \times 10^{10}}\right) = 0.228 \text{ V}$$

Then the work function difference is

$$\phi_{ms} = \phi'_m - \left(\chi' + \frac{E_g}{2e} + \phi_{fp}\right) = 3.20 - (3.25 + 0.555 + 0.228)$$

or

$$\phi_{ms} = -0.83 \text{ V}$$

■ Comment

The value of ϕ_{ms} will become more negative as the doping of the p-type substrate increases.

Degenerately doped polysilicon deposited on the oxide is also often used as the metal gate. Figure 11.13a shows the energy-band diagram of an MOS capacitor with an n^+ polysilicon gate and a p-type substrate. Figure 11.13b shows the energy-band diagram for the case of a p^+ polysilicon gate and the p-type silicon substrate. In the degenerately doped polysilicon, we will initially assume that $E_F = E_c$ for the n-case and $E_F = E_v$ for the p^+ case.

For the n^+ polysilicon gate, the metal-semiconductor work function difference can be written as

$$\phi_{ms} = \left[\chi' - \left(\chi' + \frac{E_g}{2e} + \phi_{fp}\right)\right] = -\left(\frac{E_g}{2e} + \phi_{fp}\right) \tag{11.12}$$

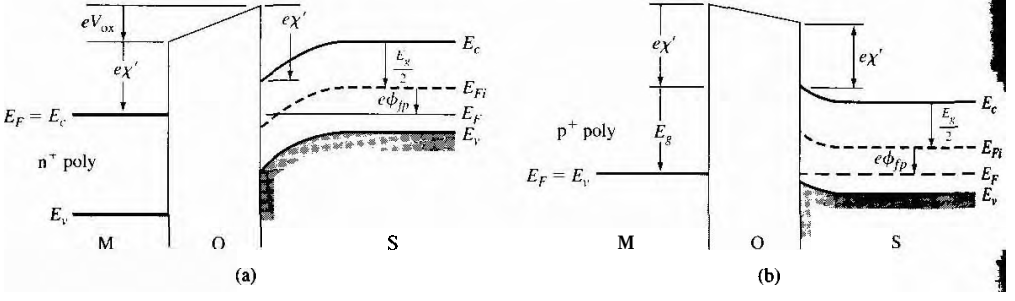


Figure 11.13 | Energy-band diagram through the MOS structure with a p-type substrate at zero gate bias for (a) an n^+ polysilicon gate and (b) a p^+ polysilicon gate.

and for the p^+ polysilicon gate, we have

$$\phi_{ms} = \left[\left(\chi' + \frac{E_g}{e}\right) - \left(\chi' + \frac{E_g}{2e} + \phi_{fp}\right)\right] = \left(\frac{E_g}{2e} - \phi_{fp}\right) \tag{11.13}$$

However, for degenerately doped n^+ polysilicon and p^+ polysilicon, the Fermi level can be above E_c and below E_v, respectively, by 0.1 to 0.2 V. The experimental ϕ_{ms} values will then be slightly different from the values calculated by using Equations (11.12) and (11.13).

We have been considering a p-type semiconductor substrate. We may also have an n-type semiconductor substrate in an MOS capacitor. Figure 11.14 shows the energy-band diagram of the MOS capacitor with a metal gate and the n-type semiconductor substrate, for the case when a negative voltage is applied to the gate. The metal–semiconductor work function difference for this case is defined as

$$\phi_{ms} = \phi'_m - \left(\chi' + \frac{E_g}{2e} - \phi_{fn}\right) \tag{11.14}$$

where ϕ_{fn} is assumed to be a positive value. We will have similar expressions for n^+ and p^+ polysilicon gates.

Figure 11.15 shows the work function differences as a function of semiconductor doping for the various types of gates. We may note that the magnitude of ϕ_{ms} for the polysilicon gates are somewhat larger than Equations (11.12) and (11.13) predict. This difference again is because the Fermi level is not equal to the conduction band

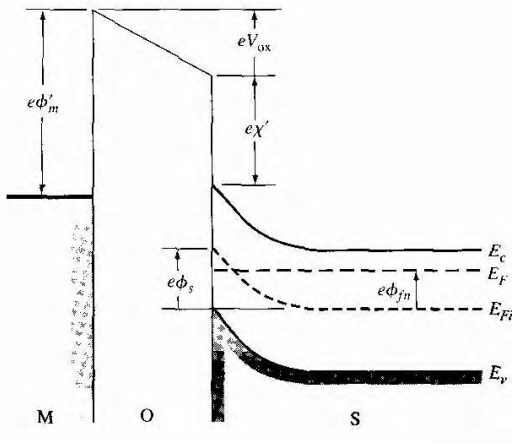


Figure 11.14 | Energy-band diagram through the MOS structure with an n-type substrate for a negative applied gate bias.

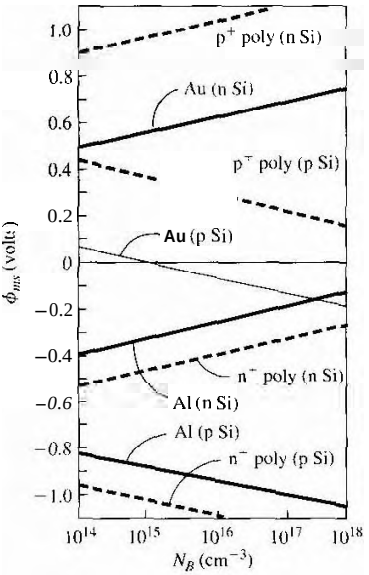


Figure **11.15** | Metal-semiconductor work function difference versus doping for aluminum, gold, and n^+ and p^+ polysilicon gates.
(From Sze 1161 *and* **Werner** *[19].)*

energy for the n^+ gate and is not equal to the valence band energy for the p^+ gate. The metal-semiconductor work function difference becomes important in the flat-band and threshold voltage parameters discussed next.

TEST YOUR UNDERSTANDING

E11.3 The silicon impurity doping concentration in an aluminum–silicon dioxide–silicon MOS structure is $N_a = 3 \times 10^{16}$ cm^{-3}. Using the parameters in Example 11.2, determine the metal-semiconductor work function difference @,,,,(Λ 186'0− = "'@suV)

E11.4 Consider an n^+ polysilicon gate in an MOS structure with a p-type silicon substrate. The doping concentration of the silicon is N,, = 3 x 10^{16} cm^{-3}. Using Equation (11.12), find the value of ϕ_{ms}. (Λ 1E6'0− = $^{ms}\phi$ 'suV)

E11.5 Repeat E11.4 for a p^+ polysilicon gate using Equation (11.13).(Λ 611'0+ = $^{ms}\phi$ 'suV)

11.1.4 Flat-Band Voltage

The *flat-band voltage* is defined as the applied gate voltage such that there is no band bending in the semiconductor and, as a result, zero net space charge in this region. Figure 11.16 shows this Rat-band condition. Because of the work function difference

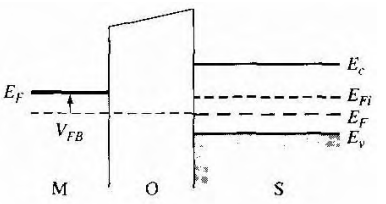


Figure 11.16 | Energy-band diagram of an MOS capacitor at flat band.

and possible trapped charge in the oxide, the voltage across the oxide for this case is not necessarily zero.

We have implicitly been assuming that there is zero net charge density in the oxide material. This assumption may not be valid—a net fixed charge density, usually positive, may exist in the insulator. The positive charge has been identified with broken or dangling covalent bonds near the oxide–semiconductor interface. During the thermal formation of SiO_2, oxygen diffuses through the oxide and reacts near the Si–SiO_2 interface to form the SiO_2. Silicon atoms may also break away from the silicon material just prior to reacting to form SiO_2. When the oxidation process is terminated, excess silicon may exist in the oxide near the interface, resulting in the dangling bonds. The magnitude of this oxide charge seems, in general, to be a strong function of the oxidizing conditions such as oxidizing ambient and temperature. The charge density can be altered to some degree by annealing the oxide in an argon or nitrogen atmosphere. However, the charge is rarely zero.

The net fixed charge in the oxide appears to be located fairly close to the oxide–semiconductor interface. We will assume in the analysis of the MOS structure that an equivalent trapped charge per unit area, Q'_{ss}, is located in the oxide directly adjacent to the oxide–semiconductor interface. For the moment, we will ignore any other oxide-type charges that may exist in the device. The parameter Q'_{ss} is usually given in terms of number of electronic charges per unit area.

Equation (11.10), for zero applied gate voltage, can be written as

$$V_{ox0} + \phi_{s0} = -\phi_{ms} \tag{11.15}$$

If a gate voltage is applied, the potential drop across the oxide and the surface potential will change. We can then write

$$V_G = \Delta V_{ox} + \Delta\phi_s = (V_{ox} - V_{ox0}) + (\phi_s - \phi_{s0}) \tag{11.16}$$

Using Equation (11.15), we have

$$V_G = V_{ox} + \phi_s + \phi_{ms} \tag{11.17}$$

Figure 11.17 shows the charge distribution in the MOS structure for the flat-band condition. There is zero net charge in the semiconductor and we can assume that an equivalent fixed surface charge density exists in the oxide. The charge density on the metal is Q'_m, and from charge neutrality we have

$$Q'_m + Q'_{ss} = 0 \tag{11.18}$$

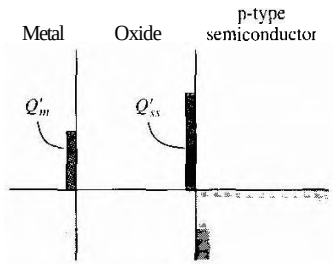


Figure **11.17** | Charge distribution in an **MOS** capacitor at flat band.

We can relate Q'_m to the voltage across the oxide by

$$V'_{} = \frac{Q'_m}{C_{ox}} \tag{11.19}$$

where C_{ox} is the oxide capacitance per unit area.[1] Substituting Equation (11.18) into Equation (11.19), we have

$$V_{ox} = \frac{-Q'_{ss}}{C_{ox}} \tag{11.20}$$

In the flat-hand condition, the surface potential is zero, or $\phi_s = 0$. Then from Equation (11.17). we have

$$V_G = \boxed{V_{FB} = \phi_{ms} - \frac{Q'_{ss}}{C_{ox}}} \tag{11.21}$$

Equation (11.21) is the flat-band voltage for this MOS device.

EXAMPLE 11.3 | Objective

To calculate the flat-band voltage for **an** MOS capacitor with a p-type semiconductor substrate.

Consider an MOS structure with a p-type semiconductor substrate doped to $N_a = 10^{16}$ cm^{-3}, a silicon dioxide insulator with a thickness of $t_{ox} = 500$ Å. and an n$^+$ polysilicon gate. Assume that $Q'_{ss} = 10^{11}$ electronic charges percm'.

Solution

The work function difference, from Figure 11.15, is $\phi_{ms} = -1.1$ V. The oxide capacitance can be found as

$$C_{,,} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{(3.9)(8.85 \times 10^{-14})}{500\ x} = 6.9 \times 10^{-8}\ \text{F/cm}^2$$

[1] Although we will. in general, use the primed notation for capacitance per unit area or charge per unit area. we will omit, fur convenience, the prime on the **oxide** capacitance per unit area parameter

The equivalent oxide surface charge density is

$$Q'_{ss} = (10^{11})(1.6 \times 10^{-19}) = 1.6 \times 10^{-8}\ \text{C/cm}^2$$

The Hat-band voltage is then calculated as

$$V_{FB} = \phi_{ms} - \frac{Q'_{ss}}{C_{ox}} = -1.1 - \left(\frac{1.6 \times 10^{-8}}{6.9 \times 10^{-8}}\right) = -1.33\ \text{V}$$

■ Comment

The applied gate voltage required to achieve the flat-band condition for this p-type substrate is negative. If the amount of fixed oxide charge increases, the Hat-band voltage becomes even more negative.

TEST YOUR UNDERSTANDING

E11.6 Consider the MOS structure described in El 1.3. Far an oxide thickness of $t_{ox} =$ 200 A and an axide charge of $Q'_{ss} = 8 \times 10^{10}\ \text{cm}^{-2}$, calculate the flat-band voltage. (A 90'1− = $_{FB}V$ 'suV)

E11.7 Repeat E11.6 for the MOS device described in El1.4. (A 10'1− = $_{FB}V$ 'suV)

E11.8 Repeat El 1.6 for the MOS device described in E11.5. (A S01'0+ = $_{FB}V$ 'suV)

11.1.5 Threshold Voltage

The threshold voltage was defined as the applied gate voltage required to achieve the threshold inversion point. The threshold inversion point, in turn, is defined as the condition when the surface potential is @, = ?@, for the p-type semiconductor and @, = $2\phi_{fn}$ for the n-type semiconductor. These conditions were shown in Figures 11.9a and 11.10. The threshold voltage will he derived in terms of rhe electrical and geometrical properties of the MOS capacitor.

Figure 11.18 shows the charge distrihution through the MOS device at the threshold inversion point for a p-type semiconductor substrate. The space charge

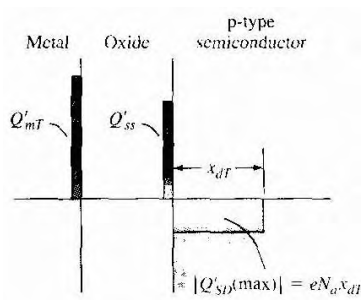


Figure 11.18 | Charge distribution in an MOS capacitor with a p-type substrate at the threshold inversion point.

width has reached its maximum value. We will assume that there is an equivalent oxide charge Q'_{ss} and the positive charge on the metal gate at threshold is Q'_{mT} .The prime on the charge terms indicates charge per unit area. Even though we are assuming that the surface has been inverted, we will neglect the inversion layer charge at this threshold inversion point. From conservation of charge, we can write

$$Q'_{mT} + Q'_{ss} = |Q'_{SD}(\text{max})| \tag{11.22}$$

where

$$\boxed{|Q'_{SD}(\text{max})| = eN_a x_{dT}} \tag{11.23}$$

and is the magnitude of the maximum space charge density per unit area of the depletion region.

The energy-band diagram of the MOS system with an applied positive gate voltage is shown in Figure 11.19. As we mentioned, an applied gate voltage will change the voltage across the oxide and will change the surface potential. We had from Equation (11.16) that

$$V_G = \Delta V_{ox} + \Delta\phi_s = V_{ox} + \phi_s + \phi_{ms}$$

At threshold, we can define $V_G = V_{TN}$, where V_{TN} is the threshold voltage that creates the electron inversion layer charge. The surface potential is $\phi_s = 2\phi_{fp}$ at threshold so Equation (11.16) can be written as

$$V_{TN} = V_{oxT} + 2\phi_{fp} + \phi_{ms} \tag{11.24}$$

where V_{oxT} is the voltage across the oxide at this threshold inversion point.

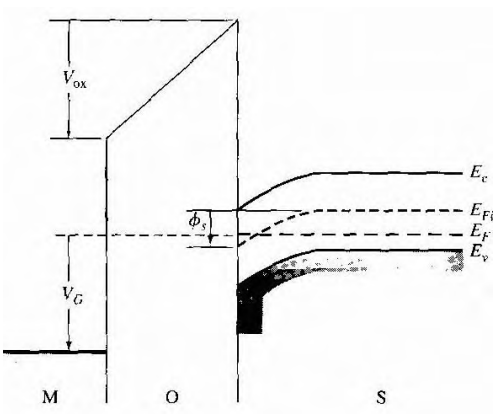


Figure 11.19 | Energy-band diagram through the MOS structure with apositive applied gate bias.

The voltage V_{oxT} can be related to the charge on the metal and to the oxide capacitance by

$$V_{oxT} = \frac{Q'_{mT}}{C_{ox}} \tag{11.25}$$

where again C_{ox} is the oxide capacitance per unit area. Using Equation (11.22), we can write

$$V_{oxT} = \frac{Q'_{mT}}{C_{ox}} = \frac{1}{C_{ox}}(|Q'_{SD}(\max)| - Q'_{ss}) \tag{11.26}$$

Finally, the threshold voltage can be written as

$$V_{TN} = \frac{|Q'_{SD}(\max)|}{C_{ox}} - \frac{Q'_{ss}}{C_{ox}} + \phi_{ms} + 2\phi_{fp} \tag{11.27a}$$

$$V_{TN} = (|Q'_{SD}(\max)| - Q'_{ss})\left(\frac{t_{ox}}{\epsilon_{ox}}\right) + \phi_{ms} + 2\phi_{fp} \tag{11.27b}$$

Using the definition of flat-band voltage from Equation (11.21), we can also express the threshold voltage as

$$V_{TN} = \frac{|Q'_{SD}(\max)|}{C_{ox}} + V_{FB} + 2\phi_{fp} \tag{11.27c}$$

For a given semiconductor material, oxide material, and gate metal, the threshold voltage is a function of semiconductor doping, oxide charge Q'_{ss}, and oxide thickness.

DESIGN EXAMPLE 11.4

Objective

To design the oxide thickness of an MOS system to yield a specified threshold voltage.

Consider an n^+ polysilicon gate and a p-type silicon substrate doped to $N_a = 3 \times 10^{16}$ cm^{-3}. Assume $Q'_{ss} = 10^{11}$ cm^{-2}. Determine the oxide thickness such that $V_{TN} = +0.65$ V.



■ Solution

From Figure 11.15, the work function difference is $\phi_{ms} = -1.13$ V. The various parameters can be calculated as

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259)\ln\left(\frac{3 \times 10^{16}}{1.5 \times 10^{10}}\right) = 0.376 \text{ V}$$

$$x_{dT} = \left\{\frac{4\epsilon_s\phi_{fp}}{eN_a}\right\}^{1/2} = \left\{\frac{4(11.7)(8.85 \times 10^{-14})(0.376)}{(1.6 \times 10^{-19})(3 \times 10^{16})}\right\}^{1/2} = 0.18\ \mu\text{m}$$

Then

$$|Q'_{SD}(\max)| = eN_a x_{dT} = (1.6 \times 10^{-19})(3 \times 10^{16})(0.18 \times 10^{-4})$$

or

$$|Q'_{SD}(\text{max})| = 8.64 \times 10^{-8}\ \text{C/cm}^2$$

The oxide thickness can be determined from the threshold voltage equation

$$V_{TN} = (|Q'_{SD}(\text{max})| - Q'_{ss})\left(\frac{t_{ox}}{\epsilon_{ox}}\right) + \phi_{ms} + 2\phi_{fp}$$

Then

$$0.65 = \frac{[(8.64 \times 10^{-8}) - (10^{11})(1.6 \times 10^{-19})]}{(3.9)(8.85 \times 10^{-14})} \cdot t_{ox} - 1.13 + 2(0.376)$$

or

$$0.65 = 2.0 \times 10^{5} t_{ox} - 0.378$$

which yields

$$t_{ox} = 504\ \text{Å}$$

■ **Comment**

The threshold voltage for chis case is a positive quantity, which means that the MOS device is an enhancement mode device: a gate voltage must he applied to create the inversion layer charge, which is zero for zero applied gate voltage.

The threshold voltage must be within the voltage range of a circuit design. Although we have not yet considered the current in an MOS transistor, the threshold voltage is the point at which the transistor turns on. If a circuit is to operate between 0 and 5 V and the threshold voltage of a MOSFET is 10 V, for example, the device and circuit cannot be turned "on" and "off." The threshold voltage, then, is one of the important parameters of the MOSFET.

EXAMPLE 11.5 Objective

To calculate the threshold voltage of an MOS system using the aluminum gate.

Consider a p-type silicon substrate at T = 300 K doped to $N_a = 10^{14}\ \text{cm}^{-3}$. Let $Q'_{ss} = 10^{10}\ \text{cm}^{-2}$, $t_{ox} = 500$ Å, and assume the oxide is silicon dioxide. From Figure 11.15, we have that $\phi_{ms} = -0.83$ V.

■ **Solution**

We can start calculating the various parameters as

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259)\ln\left(\frac{10^{14}}{1.5 \times 10^{10}}\right) = 0.228\ \text{V}$$

and

$$x_{dT} = \left\{\frac{4\epsilon_s \phi_{fp}}{eN_a}\right\}^{1/2} = \left\{\frac{4(11.7)(8.85 \times 10^{-14})(0.228)}{(1.6 \times 10^{-19})(10^{14})}\right\}^{1/2} = 2.43\ \mu\text{m}$$

Then

$$|Q'_{SD}(\text{max})| = eN_a x_{dT} = (1.6 \times 10^{-19})(10^{14})(2.43 \times 10^{-4}) = 3.89 \times 10^{-9}\ \text{C/cm}^2$$

We can now calculate the threshold voltage as

$$\begin{aligned} V_{TN} &= (|Q'_{SD}(\text{max})| - Q'_{ss})\left(\frac{t_{ox}}{\epsilon_{ox}}\right) + \phi_{ms} + 2\phi_{fp} \\ &= [(3.89 \times 10^{-9}) - (10^{10})(1.6 \times 10^{-19})]\left[\frac{500 \times 10^{-8}}{(3.9)(8.85 \times 10^{-14})}\right] \\ &\quad - 0.83 + 2(0.228) \\ &= -0.341\ \text{V} \end{aligned}$$

■ Comment

In this example, the semiconductor is very lightly doped which, in conjunction with the positive charge in the oxide and the work function potential difference, is sufficient to induce an electron inversion layer charge even with zero applied gate voltage. This condition makes the threshold voltage negative.

A negative threshold voltage for a p-type substrate implies a depletion mode device. A negative voltage must be applied to the gate in order to make the inversion layer charge equal to zero, whereas a positive gate voltage will induce a larger inversion layer charge.

Figure 11.20 is a plot of the threshold voltage V_{TN} as a function of the acceptor doping concentration for various positive oxide charge values. We may note that the p-type semiconductor must be somewhat heavily doped in order to obtain an enhancement mode device.

The previous derivation of the threshold voltage assumed a p-type semiconductor substrate. The same type of derivation can be done with an n-type semiconductor substrate, where a negative gate voltage can induce an inversion layer of holes at the oxide-semiconductor interface.

Figure 11.14 showed the energy-band diagram of the MOS structure with an n-type substrate and with an applied negative gate voltage. The threshold voltage for this case can be derived and is given by

$$V_{TP} = (-|Q'_{SD}(\text{max})| - Q'_{ss})\left(\frac{t_{ox}}{\epsilon_{ox}}\right) + \phi_{ms} - 2\phi_{fn} \tag{11.28}$$

where

$$\phi_{ms} = \phi'_m - \left(\chi' + \frac{E_g}{2e} - \phi_{fn}\right) \tag{11.29a}$$

$$|Q'_{SD}(\text{max})| = eN_d x_{dT} \tag{11.29b}$$

$$x_{dT} = \left\{\frac{4\epsilon_s \phi_{fn}}{eN_d}\right\}^{1/2} \tag{11.29c}$$

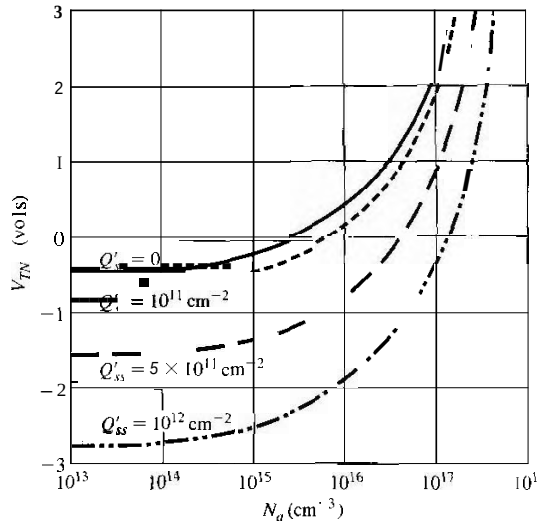


Figure 11.20 | Threshold voltage of an n-channel MOSFET versus the p-type substrate doping concentration for various values of oxide trapped charge (t_{ox} = 500 Å, aluminum gate).

and

$$\phi_{fn} = V_t \ln\left(\frac{N_d}{n_i}\right) \tag{11.29d}$$

We may note that x_{dT} and ϕ_{fn} are defined as positive quantities. We may also note that the notation of V_{TP} is the threshold voltage that will induce an inversion layer of holes. We will later drop the N and P subscript notation on the threshold voltage, but, for the moment, the notation may be useful for clarity.

DESIGN EXAMPLE 11.6

Objective

To design the semiconductor doping concentration to yield a specified threshold voltage.

Consider an aluminum–silicon dioxide–silicon MOS structure. The silicon is n type, the oxide thickness is t_{ox} = 650 Å, and the trapped charge density is $Q'_{ss} = 10^{10}$ cm^{-2}. Determine the doping concentration such that $V_{TP} = -1.0$ V.

■ Solution

The solution to this design problem is not straightforward, since the doping concentration appears in the terms ϕ_{fn}, x_{dT}, Q'_{SD}(max) and ϕ_{ms}. The threshold voltage, then, is a nonlinear function of N_d. Without a computer-generated solution, we resort to trial and error.

For $N_d = 2.5 \times 10^{14}$ cm^{-3}, we find

$$\phi_{fn} = V_t \ln\left(\frac{N_d}{n_i}\right) = 0.253 \text{ V}$$

and

$$x_{dT} = \left\{\frac{4\epsilon_s \phi_{fn}}{eN_d}\right\}^{1/2} = 1.62 \ \mu\text{m}$$

Then

$$|Q'_{SD}(\text{max})| = eN_d x_{dT} = 6.48 \times 10^{-9} \text{ C/cm}^2$$

From Figure 11.15,

$$\phi_{ms} = -0.35 \text{ V}$$

The threshold voltage is

$$V_{TP} = (-|Q'_{SD}(\text{max})| - Q'_{ss})\left(\frac{t_{ox}}{\epsilon_{ox}}\right) + \phi_{ms} - 2\phi_{fn}$$

$$= \frac{[-(6.48 \times 10^{-9}) - (10^{10})(1.6 \times 10^{-19})](650 \times 10^{-8})}{(3.9)(8.85 \times 10^{-14})} - 0.35 - 2(0.253)$$

which yields

$$V_{TP} = -1.008 \text{ V}$$

and is essentially equal to the desired result.

■ **Comment**

The threshold voltage is negative, implying that this **MOS** capacitor, with the n-type substrate, is an enhancement mode device. The inversion layer charge is zero with zero gate voltage, and anegative gate voltage must he applied to induce the hole inversion layer.

Figure 11.21 is a plot of V_{TP} versus doping concentration for several values of Q'_{ss}. We may note that, for all values of positive oxide charge, this MOS capacitor is always an enhancement mode device. As the Q'_{ss} charge increases, the threshold voltage becomes more negative, which means that it takes a larger applied gate voltage to create the inversion layer of holes at the oxide-semiconductor interface.

11.1.6 Charge Distribution

We've discussed the various charges in the MOS structure. We may gain a better understanding by considering the following figures. The electron concentration in the inversion layer (p-type substrate) at the oxide interface is given by $n_s = (n_i^2/N_a) \exp(\phi_s/V_t)$. For silicon at T = 300 K with an impurity doping concentration of $N_a = 1 \times 10^{16}$ cm^{-3}, the surface potential at the threshold inversion point is $\phi_s = 2\phi_{fp} = 0.695$ V. The electron concentration at the oxide interface at this surface potential is just $n_s = 1 \times 10^{16}$ cm^{-3} as we have discussed before. Figure 11.22

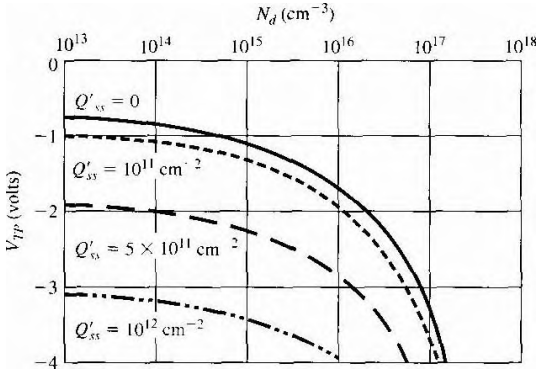


Figure 11.21 | Threshold voltage of a p-channel MOSFET versus the n-type substrate doping concentration for various values of oxide trapped charge ($t_{ox} = 500$ A. aluminum gate).

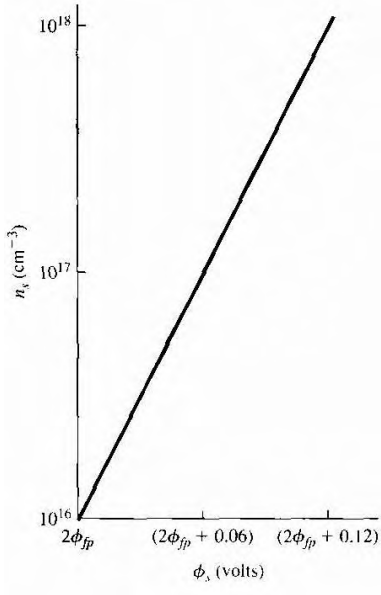


Figure 11.22 | Electron inversion charge density as a function of surface potential.

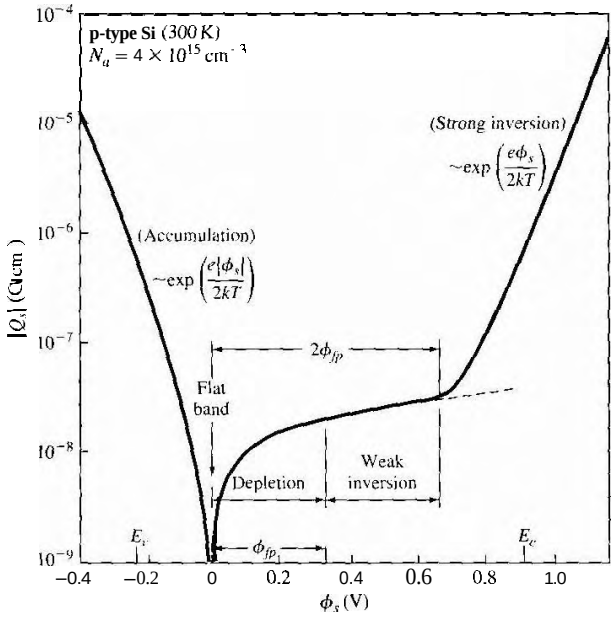


Figure 11.23 | Variation of surface charge density (accumulation charge and inversion charge) as a function of surfece potential.
(From Sze [16].)

shows the increase in electron concentration at the surface with an increase in surface potential. As discussed previously, since the electron concentration increases rapidly with very small changes in surface potential, the space charge width has essentially reached a maximum value.

Figure 11.23 shows the total charge density (C/cm^2) in the silicon as a function of the surface potential. At flat band. the total charge is zero. For $0 \leq \phi_s \leq \phi_{fp}$, we are operating in the depletion mode since the inversion charge has not yet been formed. For $\phi_{fp} \leq \phi_s < 2\phi_{fp}$, the Fermi energy at the surface is in the upper half of the band diagram, which implies an n-type material, but we have not yet reached the threshold inversion point. This condition is referred to as weak inversion. The condition for $\phi_s > 2\phi_{fp}$ is called strong inversion, since the inversion charge density increases rapidly with an increase in surface potential. as we have seen.

TEST YOUR UNDERSTANDING

E11.9 An MOS device has the following parameters: aluminum gate, p-type substrate with $N_a = 3 \times 10^{16}$ cm^{-3}, $t_{ox} = 250$ A. and $Q'_{ss} = 10^{11}$ c m^{-2}. Determine the threshold voltage. (Ans. $V_T = +0.782$)

E11.10 Consider an MOS device with the following parameters: p^+ polysilicon gate, n-type substrate with $N_d = 10^{15}$ cm^{-3}, $t_{ox} = 220$ Å, and $Q'_{ss} = 8 \times 10^{10}$ cm^{-2}. (Use Figure 11.15). Determine the threshold voltage. (Ans. $V_{TP} = +0.224$ V)

***E11.11** The device described in E11.10 is to be redesigned by changing the n-type doping concentration such that the threshold voltage is in the range $-0.50 \le V_{TP} \le -0.30$ V. (Ans. By trial and error, $V_{TP} = -0.405$ V for $N_d = 4 \times 10^{16}$ cm^{-3})

11.2 | CAPACITANCE–VOLTAGE CHARACTERISTICS

The MOS capacitor structure is the heart of the MOSFET. A great deal of information about the MOS device and the oxide–semiconductor interface can be obtained from the capacitance versus voltage or *C-V* characteristics of the device. The capacitance of a device is defined as

$$C = \frac{dQ}{dV} \tag{11.30}$$

where dQ is the magnitude of the differential change in charge on one plate as a function of the differential change in voltage dV across the capacitor. The capacitance is a small-signal or ac parameter and is measured by superimposing a small ac voltage on an applied dc gate voltage. The capacitance, then, is measured as a function of the applied dc gate voltage.

11.2.1 Ideal C-V Characteristics

First we will consider the ideal *C-V* characteristics of the MOS capacitor and then discuss some of the deviations that occur from these idealized results. We will initially assume that there is zero charge trapped in the oxide and also that there is no charge trapped at the oxide–semiconductor interface.

There are three operating conditions of interest in the MOS capacitor: accumulation, depletion, and inversion. Figure 11.24a shows the energy-band diagram of an MOS capacitor with a p-type substrate for the case when a negative voltage is applied to the gate, inducing an accumulation layer of holes in the semiconductor at the oxide-semiconductor interface. A small differential change in voltage across the MOS structure will cause a differential change in charge on the metal gate and also in the hole accumulation charge, as shown in Figure 11.24b. The differential changes in charge density occur at the edges of the oxide, as in a parallel-plate capacitor. The capacitance C' per unit area of the MOS capacitor for this accumulation mode is just the oxide capacitance, or

$$C'(\text{acc}) = C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \tag{11.31}$$

Figure 11.25a shows the energy-band diagram of the MOS device when a small positive voltage is applied to the gate, inducing a space charge region in the semiconductor; Figure 11.25b shows the charge distribution through the device for

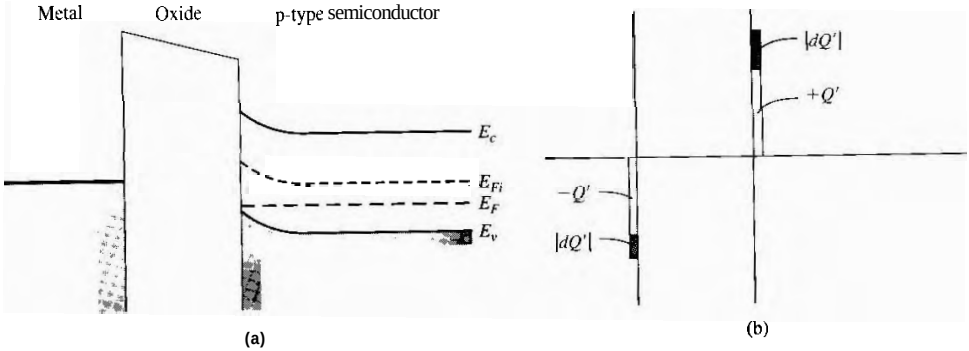


Figure 11.24 | (a) Energy-band diagram through an MOS capacitor for the accumulation mode. (b) Differential charge distribution at accumulation for a differential change in gate voltage.

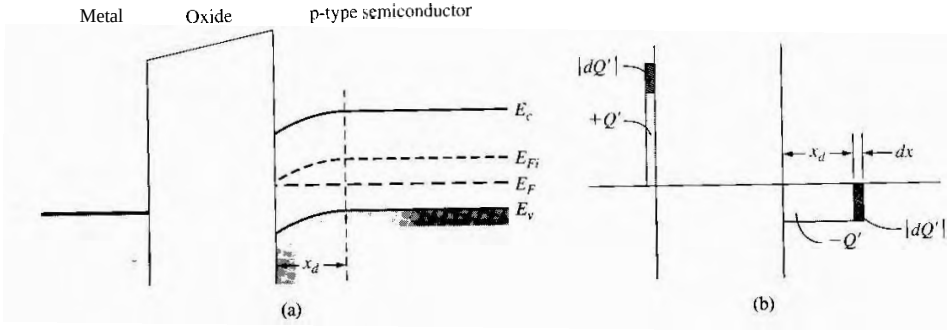


Figure 11.25 | (a) Energy-band diagram through an MOS capacitor for the depletion mode. (b) Differential charge distribution at depletion for a differential change in gate voltage.

this condition. The oxide capacitance and the capacitance of the depletion region are in series. A small differential change in voltage across the capacitor will cause a differential change in the space charge width. The corresponding differential changes in charge densities are shown in the figure. The total capacitance of the series combination is

$$\frac{1}{C'(\text{depl})} = \frac{1}{C_{ox}} + \frac{1}{C'_{SD}} \tag{11.32a}$$

$$C'(\text{depl}) = \frac{C_{ox}C'_{SD}}{C_{ox} + C'_{SD}} \tag{11.32b}$$

Since $C_{ox} = \epsilon_{ox}/t_{ox}$ and $C'_{SD} = \epsilon_s/x_d$, Equation (11.32b) can be written as

$$C'(\text{depl}) = \frac{C_{ox}}{1 + \dfrac{C_{ox}}{C'_{SD}}} = \frac{\epsilon_{ox}}{t_{ox} + \left(\dfrac{\epsilon_{ox}}{\epsilon_s}\right) x_d} \tag{11.33}$$

As the space charge width increases, the total capacitance $C'(\text{depl})$ decreases.

We had defined the threshold inversion point to he the condition when the maximum depletion width is reached but there is essentially zero inversion charge density. This condition will yield a minimum capacitance C'_{min} which is given by

$$C'_{min} = \frac{\epsilon_{ox}}{t_{ox} + \left(\dfrac{\epsilon_{ox}}{\epsilon_s}\right) x_{dT}} \tag{11.34}$$

Figure 11.26a shows the energy-band diagram of this MOS device for the inversion condition. In the ideal case, a small incremental change in the voltage across the MOS capacitor will cause a differential change in the inversion layer charge density The space charge width does not change. If the inversion charge can respond to the change in capacitor voltage as indicated in Figure 11.26b, then the capacitance is again just the oxide capacitance, or

$$C'(\text{inv}) = C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \tag{11.3}$$

Figure 11.27 shows the ideal capacitance versus gate voltage, or C-V, characteristics of the MOS capacitor with a p-type substrate. The three dashed segments correspond to the three components C_{ox}, C'_{SD}, and C'_{min}. The solid curve is the ideal net capacitance of the MOS capacitor. Moderate inversion, which is indicated in the figure, is the transition region between the point when only the space charge density

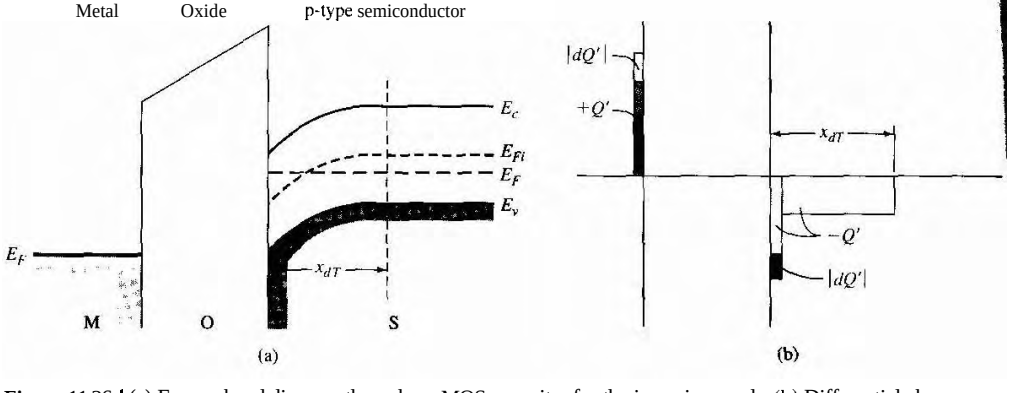


Figure 11.26 | (a) Energy-band diagram through an MOS capacitor for the inversion mode. (b) Differential charge distribution at inversion for a low-frequency differential change in gate voltage.

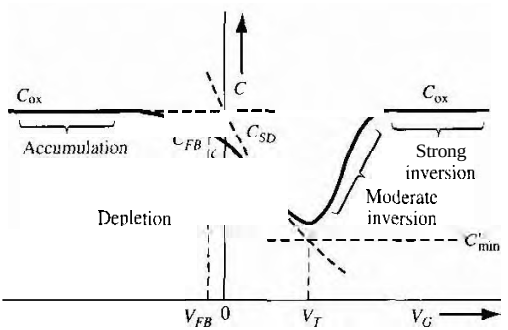


Figure 11.27 | Ideal low-frequency capacitance versus gate voltage of an MOS capacitor with a p-type substrate. Individual capacitance components are also shown.

changes with gate voltage and when only the inversion charge density changes with gate voltage.

The point on the curve that corresponds to the flat-band condition is of interest. The flat-band condition occurs between the accumulation and depletion conditions. The capacitance at flat band is given by

$$C'_{FB} = \frac{\epsilon_{ox}}{t_{ox} + \left(\frac{\epsilon_{ox}}{\epsilon_s}\right)\sqrt{\left(\frac{kT}{e}\right)\left(\frac{\epsilon_s}{eN_a}\right)}} \tag{11.36}$$

We may note that the flat-band capacitance is a function of oxide thickness as well as semiconductor doping. The general location of this point on the C-Vplot is shown in Figure 11.27.

Objective | **EXAMPLE 11.7**

To calculate C_{ox}, C'_{min}, and C'_{FB} for an MOS capacitor.

Consider a p-type silicon substrate at T = 300 K doped to $N_a = 10^{16}$ cm^{-3}. The oxide is silicon dioxide with a thickness of 550 Å and the gate is aluminum.

■ **Solution**

The oxide capacitance is

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{(3.9)(8.85 \times 10^{-14})}{550 \times 10^{-8}} = 6.28 \times 10^{-8} \text{ F/cm}^2$$

To find the minimum capacitance, we need to calculate

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259) \ln\left(\frac{10^{16}}{1.5 \times 10^{10}}\right) = 0.347 \text{ V}$$

and

$$x_{dT} = \left\{\frac{4\epsilon_s \phi_{fp}}{eN_a}\right\}^{1/2} = \left\{\frac{4(11.7)(8.85 \times 10^{-14})(0.347)}{(1.6 \times 10^{-19})(10^{16})}\right\}^{1/2} = 0.30 \times 10^{-4} \text{ cm}$$

Then

$$C'_{\text{min}} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}} + \left(\frac{\epsilon_{\text{ox}}}{\epsilon_s}\right) x_{dT}} = \frac{(3.9)(8.85 \times 10^{-14})}{(550 \times 10^{-8}) + \left(\frac{3.9}{11.7}\right)(0.3 \times 10^{-4})} = 2.23 \times 10^{-8} \text{ F/cm}^2$$

We may note that

$$\frac{C'_{\text{min}}}{C_{\text{ox}}} = \frac{2.23 \times 10^{-8}}{6.28 \times 10^{-8}} = 0.355$$

The Rat-hand capacitance is

$$C'_{FB} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}} + \left(\frac{\epsilon_{\text{ox}}}{\epsilon_s}\right)\sqrt{\left(\frac{kT}{e}\right)\left(\frac{\epsilon_s}{eN_a}\right)}}$$

$$= \frac{(3.9)(8.85 \times 10^{-14})}{(550 \times 10^{-8}) + \left(\frac{3.9}{11.7}\right)\sqrt{(0.0259)\frac{(11.7)(8.85 \times 10^{-14})}{(1.6 \times 10^{-19})(10^{16})}}}$$

$$= 5.03 \times 10^{-8} \text{ F/cm}^2$$

We may also note that

$$\frac{C'_{FB}}{C_{\text{ox}}} = \frac{5.03 \times 10^{-8}}{6.28 \times 10^{-8}} = 0.80$$

■ **Comment**

The ratios of C'_{min} to C_{ox} and of C'_{FB} to C_{ox} are typical values obtained in C-V plots

TEST YOUR UNDERSTANDING

E11.12 For the device described in E11.9, determine $C'_{\text{min}}/C_{\text{ox}}$ and C'_{FB}/C_{ox}.

(Ans. $C'_{\text{min}}/C_{\text{ox}} = 0.294$, $C'_{FB}/C_{\text{ox}} = 0.736$)

Typical values of channel length and width are 2 μm and 20 μm, respectively. The total gate oxide capacitance for this example is then

$$C_{\text{oxT}} = (6.28 \times 10^{-8})(2 \times 10^{-4})(20 \times 10^{-4}) = 0.025 \times 10^{-12} \text{ F} = 0.025 \text{ pF}$$

The total oxide capacitance in a typical MOS device is quite small.

The same type of ideal C-V characteristics are obtained for an MOS capacitor with an n-type subsrrate by changing the sign of the voltage axis. The accumulation

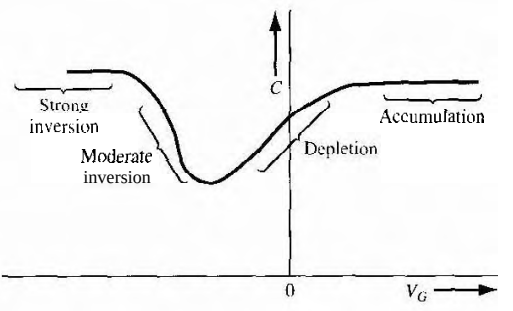


Figure 11.28 | Ideal low-frequency capacitance versus gate voltage of an MOS capacitor with an n-type substrate.

condition is obtained for a positive gate bias and the inversion condition is obtained for a negative gate bias. This ideal curve is shown in Figure 11.28.

11.2.2 Frequency Effects

Figure 11.26a showed the MOS capacitor with a p-type substrate and biased in the inversion condition. We have argued that a differential change in the capacitor voltage in the ideal case causes a differential change in the inversion layer charge density. However, we must consider the source of electrons that produces a change in the inversion charge density.

There are two sources of electrons that can change the charge density of the inversion layer. The first source is by diffusion of minority carrier electrons **from** the p-type substrate across the space charge region. This diffusion process is the same as that in a reverse-biased pn junction that generates the ideal reverse saturation current. The second source of electrons is by thermal generation of electron-hole pairs within the space charge region. This process is again the same as that in a reverse-biased pn junction generating the reverse-biased generation current. Both of these processes generate electrons at **a** particular rate. The electron concentration in the inversion layer. then, cannot change instantaneously. If the ac voltage across the MOS capacitor changes rapidly, the change in the inversion layer charge will not be able to respond. The *C-V* characteristics will then be a function of the frequency of the ac signal used to measure the capacitance.

In the limit of a very high frequency, the inversion layer charge will not respond to a differential change in capacitor voltage. Figure 11.29 shows the charge distribution in the MOS capacitor with ap-type substrate. At a high-signal frequency, the differential change in charge occurs at the metal and in the space charge width in the semiconductor. The capacitance of the MOS capacitor is then C'_{min}, which we discussed earlier.

The high-frequency and low-frequency limits of the C-V characteristics are shown in Figure 11.30. In general, high frequency corresponds to a value on the order

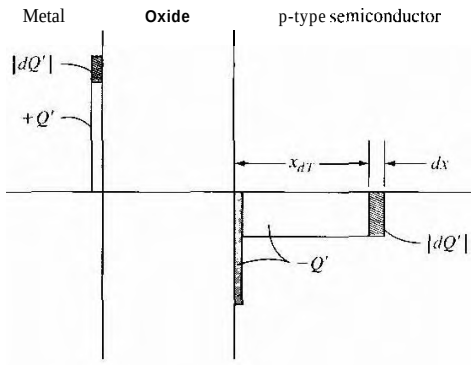


Figure 11.29 | Differential charge distribution at inversion for a high-frequency differential change in gate voltage.

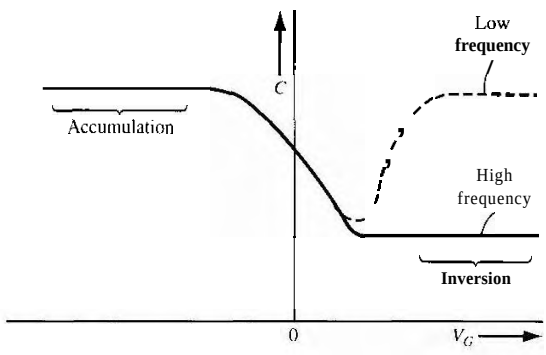


Figure **11.30** | Low-frequency and high-frequency capacitance versus gate voltage of an MOS capacitor with a p-type substrate.

of 1 MHz and low frequency correspond5 to values in the range of 5 to 100 Hz. Typically, the high-frequency characteristics of the MOS capacitor are measured.

11.2.3 Fixed Oxide and Interface Charge Effects

In all of the discussion concerning C-Vcharacteristics so far, we have assumed an ideal oxide in which there are no fixed oxide or oxide-semiconductor interface charges. These two types of charges will change the C-Vcharacteristics.

We previously discussed how the fixed oxide charge affects the threshold voltage. This charge will also affect the flat-band voltage. The fiat-band voltage from

Equation (11.21) was given by

$$V_{FB} = \phi_{ms} - \frac{Q'_{ss}}{C_{ox}}$$

where Q'_{ss} is the equivalent fixed oxide charge and ϕ_{ms} is the metal–semiconductor work function difference. The flat-band voltage shifts to more negative voltages for a positive lixed oxide charge. Since the oxide charge is not a function of gate voltage, the curves show a parallel shift with oxide charge. and the shape of the *C-V* curves remains the same as the ideal characteristics. Figure 11.31 shows the high-frequency characteristics of an MOS capacitor with a p-type substrate for several values of fixed positive oxide charge.

The *C-V* characteristics can be used to determine the equivalent fixed oxide charge. For agiven MOS structure, ϕ_{ms} and C_{ox} are known, so the ideal flat-band voltage and flat-band capacitance can be calculated. The experimental value of flat-band voltage can be measured from the *C-V* curve and the value of fixed oxide charge can then be determined. The *C-V* measurements are a valuable diagnostic tool to characterize an MOS device. This characterization is especially useful in the study of radiation effects on MOS devices, for example, which we will discuss in the next chapter.

We first encountered oxide–semiconductor interface states in Chapter 9 in the discussion of Schottky barrier diodes. Figure 11.32 shows the energy-band diagram of a semiconductor at the oxide-semiconductor interface. The periodic nature of the semiconductor is abruptly terminated at the interface so that allowed electronic energy levels will exist within the forbidden bandgap. These allowed energy states are referred to as interface states. Charge can flow between the semiconductor and interface states, in contrast to the fixed oxide charge. The net charge in these interface states is a function of the position of the Fermi level in the bandgap.

In general, acceptor states exist in the upper half of the bandgap and donor states exist in the lower half of the bandgap. An acceptor state is neutral if the Fermi level

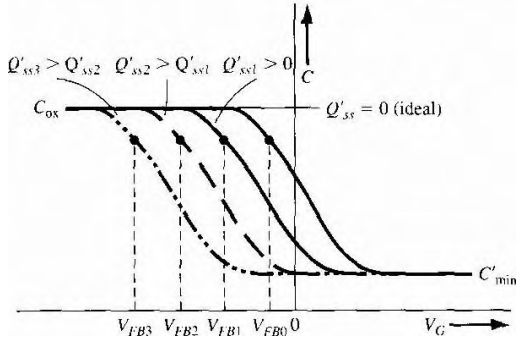


Figure 11.31 | High-frequency capacitance versus gate voltage of an MOS capacitor with a p-type substrate for several values of effective trapped oxide charge.

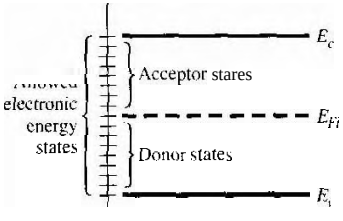


Figure 11.32 | Schematic diagram showing interface states at the oxide-semiconductor interface.

is below the state, and becomes negatively charged if the Fermi level is above the state. A donor state is neutral if the Fermi level is above the state and becomes positively charged if the Fermi level is below the state. The charge of the interface states is then a function of the gate voltage applied across the MOS capacitor.

Figure 11.33a shows the energy-band diagram in a p-type semiconductor of an MOS capacitor biased in the accumulation condition. In this case, there is a net positive charge trapped in the donor states. Now let the gate voltage change to produce the energy-band diagram shown in Figure 11.33b. The Fermi level corresponds to the intrinsic Fermi level at the surface; thus, all interface states are neutral. This

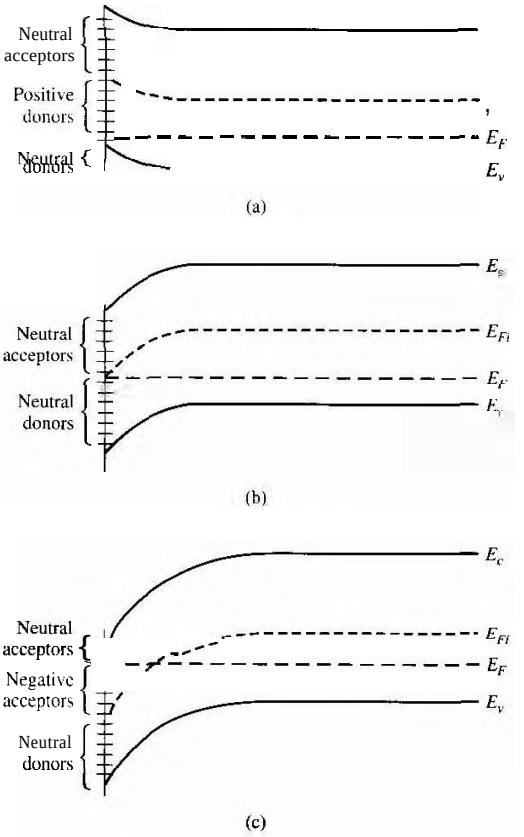


Figure 11.33 | Energy-band diagram in a p-type semiconductor showing the charge trapped in the interface states when the MOS capacitor is biased (a) in accumulation, (b) at midgap, and (c) at inversion.

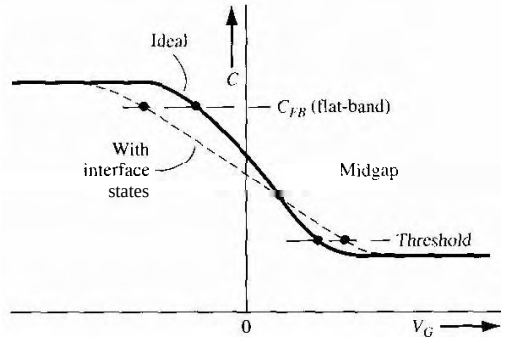


Figure 11.34 | High-frequency *C-V* characteristics of an MOS capacitor showing effects of interface states.

particular bias condition is known as *midgap*. Figure 11.33c shows the condition at inversion in which there is now a net negative charge in the acceptor states.

The net charge in the interface states changes from positive to negative as the gate voltage sweeps from the accumulation, depletion, to the inversion condition. We noted that the C-Vcurves shifted in the negative gate voltage direction due to positive fixed oxide charge. When interface states are present, the amount and direction of the shift changes as we sweep through the gate voltage, since the amount and sign of the interface trapped charge changes. The *C-V* curves now become "smeared out" as shown in Figure 11.34.

Again, the *C-V* measurements can be used as a diagnostic tool in semiconductor device process control. For a given MOS device, the ideal C-V curve can be determined. Any "smearing out" in the experimental curve indicates the presence of interface states and any parallel shift indicates the presence of fixed oxide charge. The amount of smearing out can be used to determine the density of interface states. These types of measurement are extremely useful in the study of radiation effects on MOS devices, which we will consider in the next chapter.

11.3 | THE BASIC MOSFET OPERATION

The current in an MOS field-effect transistor is due to the flow of charge in the inversion layer or channel region adjacent to the oxide-semiconductor interface. We have discussed the creation of the inversion layer charge in enhancement-type MOS capacitors. We may also have depletion-type devices in which a channel already exists at zero gate voltage.

11.3.1 MOSFET Structures

There are four basic MOSFET device types. Figure 11.35 shows an n-channel enhancement mode MOSFET. Implicit in the enhancement mode notation is the idea

that the semiconductor substrate is not inverted directly under the oxide with zero gate voltage. A positive gate voltage induces the electron inversion layer, which then "connects" the n-type source and the n-type drain regions. The source terminal is the source of carriers that flow through the channel to the drain terminal. For this n-channel device, electrons flow from the source to the drain so the conventional current will enter the drain and leave the source. The conventional circuit symbol for this n-channel enhancement mode device is also shown in this figure.

Figure 11.36 shows an n-channel depletion mode MOSFET. An n-channel region exists under the oxide with zero volts applied to the gate. However, we have shown that the threshold voltage of an MOS device with a p-type substrate may be:

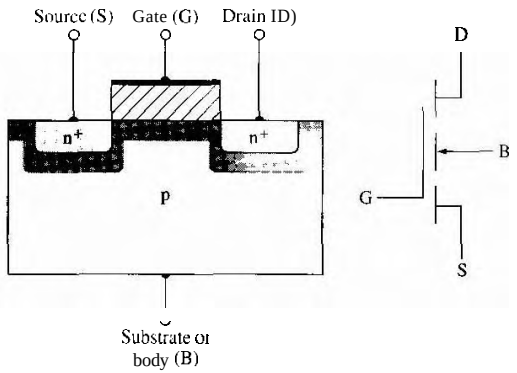


Figure 11.35 | Cross section and circuit symbol for an n-channel enhancement-mode MOSFET.

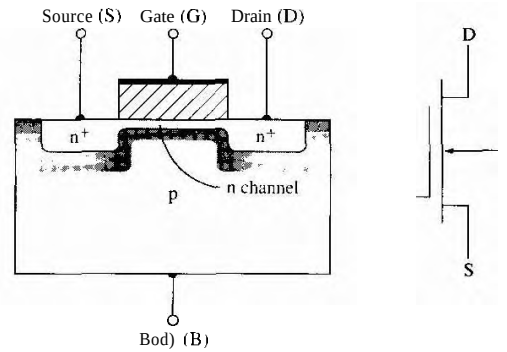


Figure 11.36 | Cross sectiun and circuit symbol for an n-channel depletion-mode MOSFET.

negative; this means that an electron inversion layer already exists with zero gate voltage applied. Such a device is also considered to be a depletion mode device. The n-channel shown in this figure can be an electron inversion layer or an intentionally doped n-region. The conventional circuit symbol for the n-channel depletion mode MOSFET is also shown in the tigure.

Figures 11.37a and 11.37b show a p-channel enhancement mode MOSFET and a p-channel depletion mode MOSFET. In the p-channel enhancement mode device, a negative gate voltage must he applied to create an inversion layer of holes that will "connect" thep-type source and drain regions. Holes flow from the source to the drain, so the conventional current will enter the source and leave the drain. A p-channel region exists in the depletion mode device even with zero gate voltage. The conventional circuit symbols are shown in the figure.

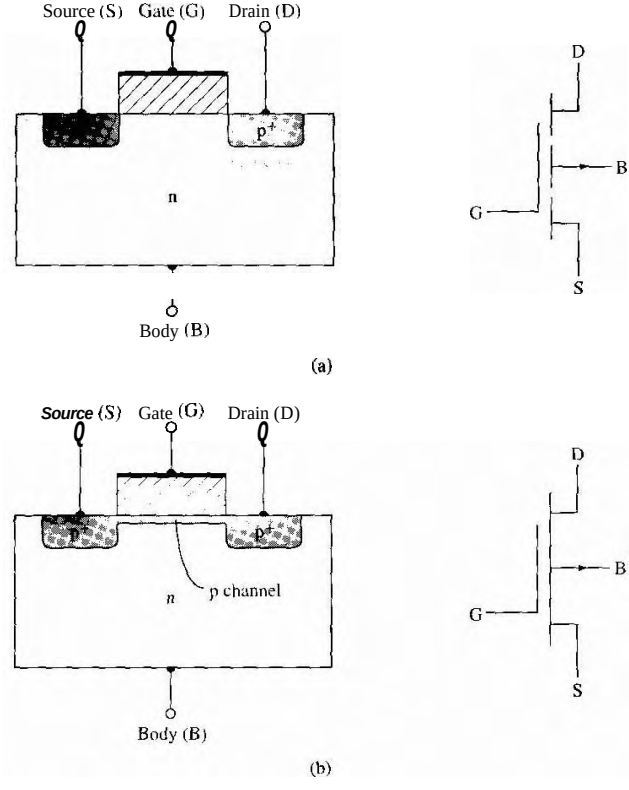


Figure 11.37 | Cross section and circuit symbol for (a) a p-channel enhancement mode MOSFET and (b) a p-channel depletion mode MOSFET

11.3.2 Current–Voltage Relationship—Concepts

Figure 11.38a shows an n-channel enhancement mode MOSFET with a gate-to-source voltage that is less than the threshold voltage and with only a very small drain-to-source voltage. The source and substrate, or body, terminals are held at ground potential. With this bias configuration, there is no electron inversion layer, the drain-to-substrate pn junction is reverse biased, and the drain current is zero (disregarding pn junction leakage currents).

Figure 11.38b shows the same MOSFET with an applied gate voltage such that $V_{GS} > V_T$. An electron inversion layer has been created so that, when a small drain voltage is applied, the electrons in the inversion layer will flow from the source to the positive drain terminal. The conventional current enters the drain terminal and leaves the source terminal. In this ideal case, there is no current through the oxide to the gate terminal.

For small V_{DS} values, the channel region has the characteristics of a resistor, so we can write

$$I_D = g_d V_{DS} \tag{11.37}$$

where g_d is defined as the channel conductance in the limit as $V_{DS} \to 0$. The channel conductance is given by

$$g_d = \frac{W}{L} \cdot \mu_n |Q_n'| \tag{11.38}$$

where μ_n is the mobility of the electrons in the inversion layer and $|Q_n'|$ is the magnitude of the inversion layer charge per unit area. The inversion layer charge is a function of the gate voltage; thus, the basic MOS transistor action is the modulation of the channel conductance by the gate voltage. The channel conductance, in turn, determines the drain current. We will initially assume that the mobility is a constant; we will discuss mobility effects and variations in the next chapter.

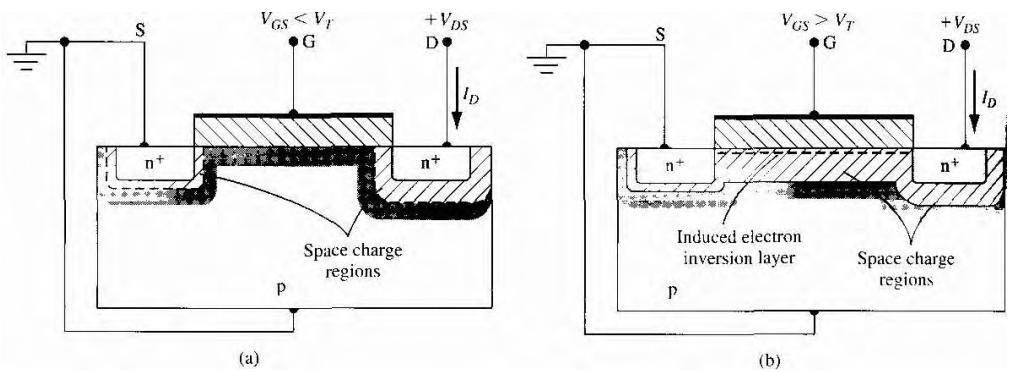


Figure 11.38 | The n-channel enhancement mode MOSFET (a) with an applied gate voltage $V_{GS} < V_T$, and (b) with an applied gate voltage $V_{GS} > V_T$.

The I_D versus V_{DS} characteristics, for small values of V_{DS}, are shown in Figure 11.39. When $V_{GS} < V_T$, the drain current is zero. As V_{GS} becomes larger than V_T, channel inversion charge density increases, which increases the channel conductance. A larger value of g_d produces a larger initial slope of the I_D versus V_{DS} characteristic as shown in the figure.

Figure 11.40a shows the basic MOS structure for the case when $V_{GS} > V_T$ and the applied V_{DS} voltage is small. The thickness of the inversion channel layer in the figure qualitatively indicates the relative charge density, which is essentially constant along the entire channel length for this case. The corresponding I_D versus V_{DS} curve is shown in the figure.

Figure 11.40b shows the situation when the V_{DS} value increases. As the drain voltage increases, the voltage drop across the oxide near the drain terminal decreases, which means that the induced inversion charge density near the drain also decreases. The incremental conductance of the channel at the drain decreases, which then means that the slope of the I_D versus V_{DS} curve will decrease. This effect is shown in the I_D versus V_{DS} curve in the figure.

When V_{DS} increases to the point where the potential drop across the oxide at the drain terminal is equal to V_T, the induced inversion charge density is zero at the drain terminal. This effect is schematically shown in Figure 11.40c. At this point, the incremental conductance at the drain is zero, which means that the slope of the I_D versus V_{DS} curve is zero. We can write

$$V_{GS} - V_{DS}(\text{sat}) = V_T \tag{11.39a}$$

or

$$\boxed{V_{DS}(\text{sat}) = V_{GS} - V_T} \tag{11.39b}$$

where $V_{DS}(\text{sat})$ is the drain-to-source voltage producing zero inversion charge density at the drain terminal.

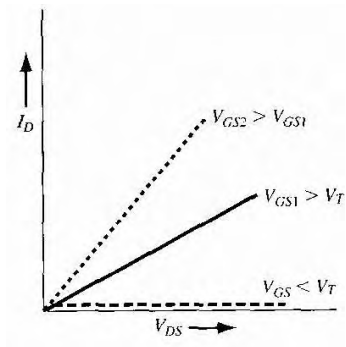


Figure 11.39 I_D versus V_{DS} characteristics for small values of V_{DS} at three V_{GS} voltages.

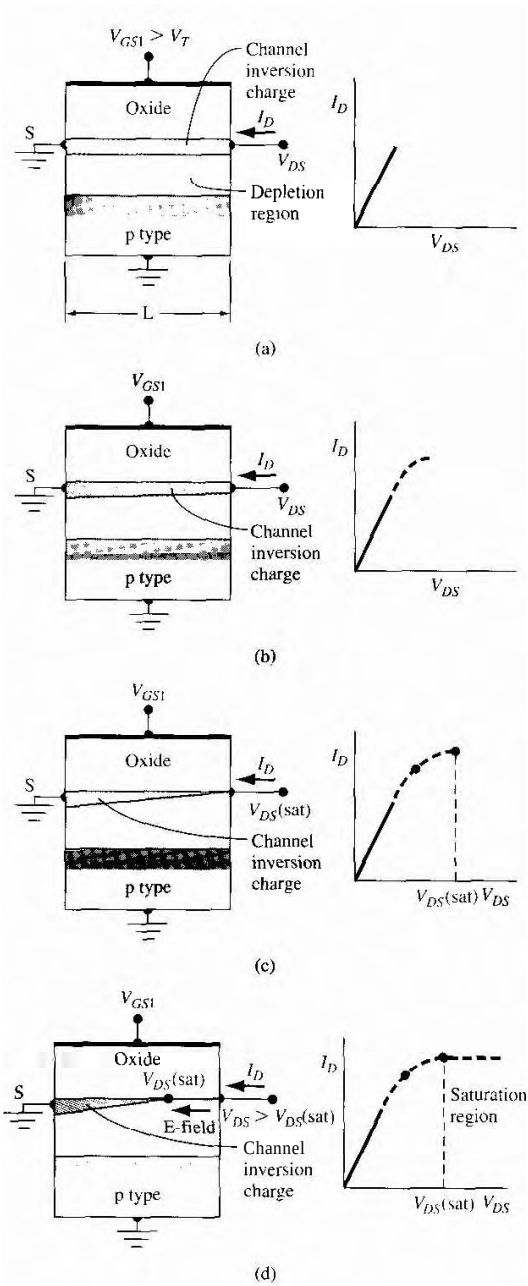


Figure 11.40 | Cross section and I_D versus V_{DS} curve when $V_{GS} < V_T$ for (a) a small V_{DS} value, (b) a larger V_{DS} value, (c) a value of $V_{DS} = V_{DS}$(sat), and (d) a value of $V_{DS} > V_{DS}$(sat).

When V_{DS} becomes larger than the $V_{DS}(\text{sat})$ value, the point in the channel at which the inversion charge is just zero moves toward the source terminal. In this case, electrons enter the channel at the source, travel through the channel toward the drain, and then, at the point where the charge goes to zero, the electrons are injected into the space charge region where they are swept hy the E-field to the drain contact. If we assume that the change in channel length AL is small compared to the original length L, then the drain current will be a constant for $V_{DS} > V_{DS}(\text{sat})$. The region of the I_D versus V_{DS} characteristic is referred to as the *saturation* ***region.*** Figure 11.40d shows this region of operation.

When V_{GS} changes, the I_D versus V_{DS} curve will change. We saw that, if V_{GS} increases, the initial slope of I_D versus V_{DS} increases. We can also note from Equation (11.39b) that the value of $V_{DS}(\text{sat})$ is a function of V_{GS}. We can generate the family of curves for this n-channel enhancement mode MOSFET as shown in Figure 11.41.

Figure 11.42 shows an n-channel depletion mode MOSFET. If the n-channel region is actually an induced electron inversion layer created by the metal–semiconductor work function difference and fixed charge in the oxide, the current–voltage characteristics are exactly the same as we have discussed, except that V_T is a negative quantity. We may also consider the case when the n-channel region is actually an n-type semiconductor region. In this type of device, a negative gate voltage will induce a space charge region under the oxide, reducing the thickness of the n-channel region. The reduced thickness decreases the channel conductance, which reduces the drain current. A positive gate voltage will create an electron accumulation layer, which increases the drain current. One basic requirement for this device is that the channel thickness t_c must be less than the maximum induced space charge width in order to be able to turn the device off. The general I_D versus V_{DS} family of curves for an n-channel depletion mode MOSFET is shown in Figure 11.43.

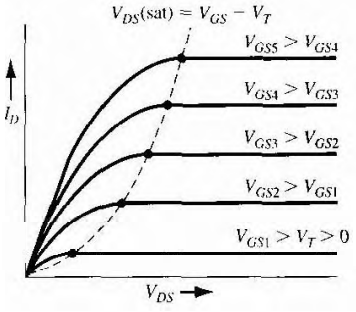


Figure 11.41 | Family of I_D versus V_{DS} curves for an n-channel enhancement-**mods** MOSFET.

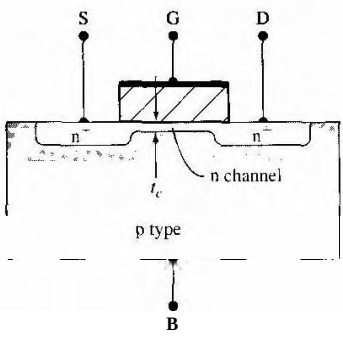


Figure 11.42 | Cross section of an n-channel depletion-mode MOSFET,

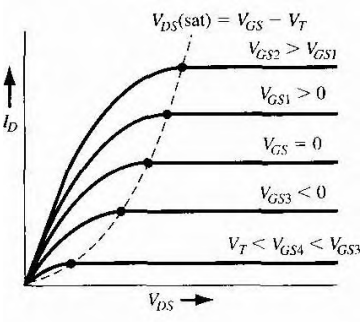


Figure 11.43 | Family of I_D versus V_{DS} curves for an n-channel depletion-mode MOSFET.

In the next section we will derive the ideal current–voltage relation for the n-channel MOSFET. In the nonsaturation region, we will obtain

$$I_D = \frac{W\mu_n C_{ox}}{2L}\left[2(V_{GS} - V_T)V_{DS} - V_{DS}^2\right] \tag{11.40}$$

and, in the saturation region, we will have

$$I_D = \frac{W\mu_n C_{ox}}{2L}(V_{GS} - V_T)^2 \tag{11.41}$$

The operation of a p-channel device is the same as that of the n-channel device. except the charge carrier is the hole and the conventional current direction and voltage polarities are reversed.

*11.3.3 Current–Voltage Relationship—Mathematical Derivation

In the previous section, we qualitatively discussed the current–voltage characteris. tics. In this section, we will derive the mathematical relation between the drain current, the gate-to-source voltage, and the drain-to-source voltage. Figure 11.44 shows the geometry of the device that we will use in this derivation.

In this analysis, we will make the following assumptions:

1. The current in the channel is due to drift rather than diffusion.
2. There is no current through the gate oxide.
3. A gradual channel approximation is used in which $\partial E_y/\partial y \gg \partial E_x/\partial x$. This approximation means that E_x is essentially a constant.
4. Any fixed oxide charge is an equivalent charge density at the oxide-semiconductor interface.
5. The carrier mobility in the channel is constant.

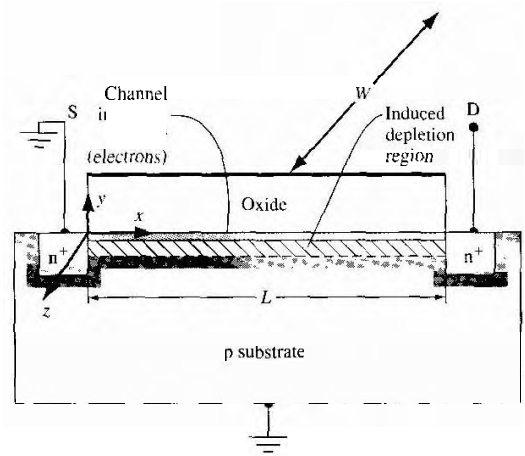


Figure 11.44 | Geometry of a MOSFET for I_D versus V_{DS} derivation.

We start the analysis with Ohm's law, which can be written as

$$J_x = \sigma E_x \tag{11.42}$$

where a is the channel conductivity and E_x is the electric field along the channel created by the drain-to-source voltage. The channel conductivity is given by $\sigma = e\mu_n n(y)$ where μ_n is the electron mobility and $n(y)$ is the electron concentration in the inversion layer.

The total channel current is found by integrating J_x over the cross-sectional area in they- and z-directions. Then

$$I_x = \int_y \int_z J_x \, dy \, dz \tag{11.43}$$

We may write that

$$Q'_n = -\int en(y) \, dy \tag{11.44}$$

where Q'_n is the inversion layer charge per unit area and is a negative quantity for this case.

Equation (11.43) then becomes

$$I_x = -W\mu_n Q'_n E_x \tag{11.45}$$

where Wis the channel width, the result of integrating over z.

Two concepts we will use in the current-voltage derivation are charge neutrality and Gauss's law. Figure 11.45 shows the charge densities through the device for $V_{GS} > V_T$. The charges are all given in terms of charge per unit area. Using the

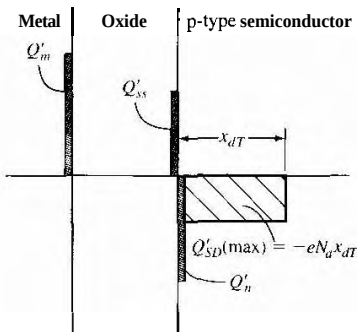


Figure 11.45 | Charge distribution in the n-channel enhancement mode MOSFET for $V_{GS} > V_T$.

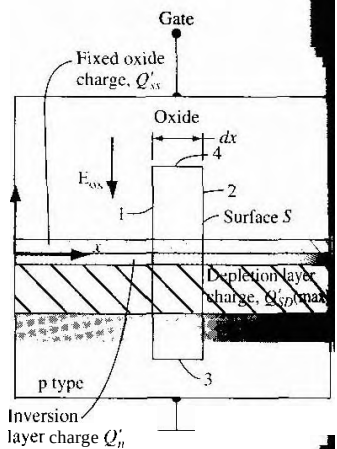


Figure 11.46 | Geometry for applying Gauss's law.

concept of charge neutrality, we can write

$$Q'_m + Q'_{ss} + Q'_n + Q'_{SD}(\max) = 0 \tag{11.46}$$

The inversion layer charge and induced space charge will be negative for this n-channel device.

Gauss's law can be wntten as

$$\oint_S \epsilon E_n \, dS = Q_T \tag{11.47}$$

where the integral is over a closed surface. Q_T is the total charge enclosed by the surface, and E_n, is the outward directed normal component of the electric field crossing the surface S. Gauss's law will he applied to the surface defined in Figure 11.46. Since the surface must he enclosed, we must take into account the two end surfaces in the x-y plane. However, there is no z-component of the electric field so these two end surfaces do not contribute to the integral of Equation (11.47).

Now consider the surfaces labeled 1 and 2 in Figure 11.46. From the gradual channel approximation, we will assume that E_x is essentially a constant along the channel length. This assumption means that E_x into surface 2 is the same as E_x out of surface 1. Since the integral in Equation (11.47) involves the outward component of the E-field, the contributions of surfaces 1 and 2 cancel each other. Surface 3 is in the neutral p-region, so the electric field is zero at this surface.

Surface 4 is the only surface that contributes to Equation (11.47). Taking into account the direction of the electric field in the oxide, Equation (11.47) becomes

$$\oint_S \epsilon E_n \, dS = -\epsilon_{ox} E_{ox} W \, dx = Q_T \tag{11.48}$$

where ϵ_{ox} is the permittivity of the oxide. The total charge enclosed is

$$Q_T = (Q'_{ss} + Q'_n + Q'_{SD}(\max))W\,dx \tag{11.49}$$

Combining Equations (11.48) and (11.49), we have

$$-\epsilon_{ox}E_{ox} = Q'_{ss} + Q'_n + Q'_{SD}(\max) \tag{11.50}$$

We now need an expression for E_{ox}. Figure 11.47a shows the oxide and channel. We will assume that the source is at ground potential. The voltage V_x is the potential in the channel at a point x along the channel length. The potential difference across the oxide at x is a function of V_{GS}, V_x, and the metal-semiconductor work function difference.

The energy-band diagram through the MOS structure at point x is shown in Figure 11.47b. The Fermi level in the p-type semiconductor is E_{Fp} and the Fermi level in the metal is E_{Fm}. We have

$$E_{Fp} - E_{Fm} = e(V_{GS} - V_x) \tag{11.51}$$

Considering the potential barriers, we can write

$$V_{GS} - V_x = (\phi'_m + V_{ox}) - \left(\chi' + \frac{E_g}{2e} - \phi_s + \phi_{fp}\right) \tag{11.52}$$

which can also be written as

$$V_{GS} - V_x = V_{ox} + 2\phi_{fp} + \phi_{ms} \tag{11.53}$$

where ϕ_{ms} is the metal–semiconductor work function difference, and $\phi_s = 2\phi_{fp}$ for the inversion condition.

The electric field in the oxide is

$$E_{ox} = \frac{V_{ox}}{t_{ox}} \tag{11.54}$$

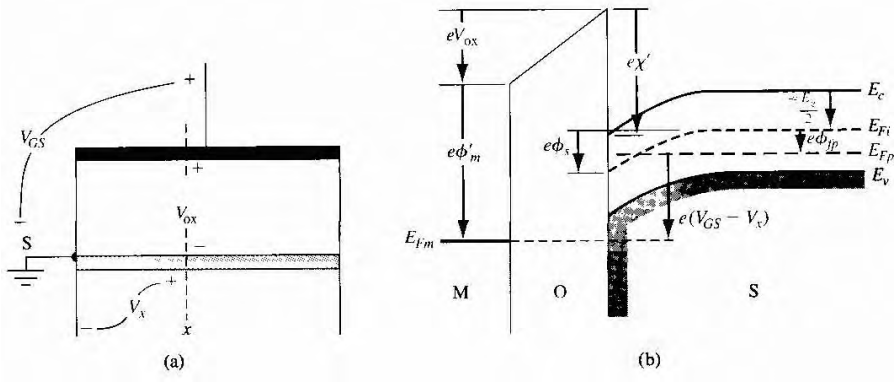


Figure 11.47 | (a) Potentials at a point x along the channel. (b) Energy-band diagram through the MOS structure at the point x.

Combining Equations (11.50), (11.53). and (11.54), we find that

$$-\epsilon_{ox}E_{ox} = -\frac{\epsilon_{ox}}{t_{ox}}[(V_{GS} - V_x) - (\phi_{ms} + 2\phi_{fp})]$$

$$= Q'_{ss} + Q'_n + Q'_{SD}(\max) \tag{11.55}$$

The inversion charge density, Q'_n, from Equation (11.55) can he substituted into Equation (11.45) and we obtain

$$= -W\mu_n C_{ox}\frac{dV_x}{dx}[(V_{GS} \quad V_x) \quad V_T] \tag{11.56}$$

where $E_x = -dV_x/dx$ and V_T is the threshold voltage defined by Equation (11.27).

We can now integrate Equation (11.56) over the length of the channel. We have

$$\int_0^L I_x\,dx = -W\mu_n C_{ox}\int_{V_x(0)}^{V_x(L)}[(V_{GS} - V_T) - V_x]\,dV_x \tag{11.57}$$

We are assuming a constant mobility μ_n. For the n-channel device, the drain current enters the drain terminal and is a constant along the entire channel length. Letting $I_D = -I_x$, Equation (11.57) becomes

$$I_D = \frac{W\mu_n C_{ox}}{2L}\left[2(V_{GS} - V_T)V_{DS} - V_{DS}^2\right] \tag{11.58}$$

Equation (11.58) is valid for $V_{GS} \geq V_T$ and for $0 \leq V_{DS} \leq V_{DS}(\text{sat})$.

Figure 11.48 shows plots of Equation (11.58) as a function of V_{DS} for several values of V_{GS}. We can find the value of V_{DS} at the peak current value from $\partial I_D/\partial V_{DS} = 0$. Then, using Equation (11.58), the peak current occurs when

$$V_{DS} = V_{GS} - V_T \tag{11.59}$$

This value of V_{DS} is just $V_{DS}(\text{sat})$, the point at which saturation occurs. For $V_{DS} > V_{DS}(\text{sat})$, the ideal drain current is a constant and is equal to

$$I_D(\text{sat}) = \frac{W\mu_n C_{ox}}{2L}\left[2(V_{GS} - V_T)V_{DS}(\text{sat}) - V_{DS}^2(\text{sat})\right] \tag{11.60}$$

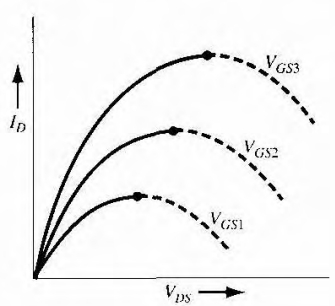


Figure 11.48 | Plots of I_D versus V_{DS} from Equation (11.58).

Using Equation (11.59) for $V_{DS}(\text{sat})$, Equation (11.60) becomes

$$I_D(\text{sat}) = \frac{W\mu_n C_{ox}}{2L}(V_{GS} - V_T)^2 \tag{11.61}$$

Equation (11.58) is the ideal current-voltage relationship of the n-channel MOSFET in the nonsaturation region for $0 \leq V_{DS} \leq V_{DS}(\text{sat})$, and Equation (11.61) is the ideal current-voltage relationship of the n-channel MOSFET in the saturation region for $V_{DS} \geq V_{DS}(\text{sat})$. These *I–V* expressions were explicitly derived for an n-channel enhancement mode device. However, these same equations apply to an n-channel depletion mode MOSFET in which the threshold voltage V_T is a negative quantity.

DESIGN EXAMPLE 11.8

Objective

To design the width of a MOSFET such that a specified current is induced for a given applied bias.

Consider an ideal n-channel MOSFET with parameters $L = 1.25\,\mu\text{m}$, $\mu_n = 650\,\text{cm}^2/\text{V-s}$, $C_{ox} = 6.9 \times 10^{-8}\,\text{F/cm}^2$, and $V_T = 0.65$ V. Design the channel width W such that $I_D(\text{sat}) = 4\,\text{mA}$ for $V_{GS} = 5$ V.

Solution

We have, from Equation (11.61),

$$I_D(\text{sat}) = \frac{W\mu_n C_{ox}}{2L}(V_{GS} - V_T)^2$$

$$4 \times 10^{-3} = \frac{W(650)(6.9 \times 10^{-8})}{2(1.25 \times 10^{-4})} \cdot (5 - 0.65)^2 = 3.39W$$

Then

$$W = 11.8\,\mu\text{m}$$

Comment

The current capability of a MOSFET is directly proportional to the channel width W. The current handling capability can be increased by increasing W.

TEST YOUR UNDERSTANDING

E11.13 The parameters of an n-channel MOSFET are $\mu_n = 650\,\text{cm}^2/\text{V-s}$, $t_{ox} = 200$ Å, $W/L = 50$, and $V_T = 0.40$ V. If the transistor is biased in the saturation region, find the drain current for $V_{GS} = 1$, 2, and 3 V. (Ans. 61 pA, 1.76 mA, 4.51 mA)

E11.14 The n-channel MOSFET in E11.13 is to be redesigned by changing the W/L ratio such that $I_D = 100\,\mu\text{A}$ when the transistor is biased in the saturation region with $V_{GS} = 1.75$ V. (Ans. $W/L = 0.976$)

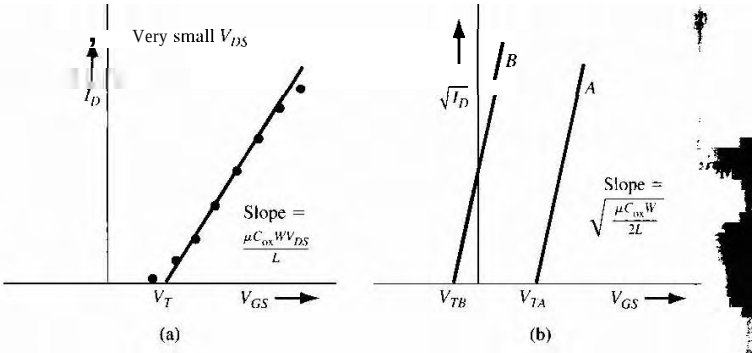


Figure 11.49 | (a) I_D versus V_{GS} (for small V_{DS}) for enhancement mode MOSFET. (b) Ideal $\sqrt{I_D}$ versus V_{GS} in saturation region for enhancement mode (curve A) and depletion mode (curve B) n-channel MOSFETs.

We can use the *I–V* relations to experimentally determine the mobility an threshold voltage parameters. From Equation (11.58), we can write, for very sma values of V_{DS},

$$I_D = \frac{W\mu_n C_{ox}}{L}(V_{GS} - V_T)V_{DS} \qquad (11.62$$

Figure 11.49a shows a plot of Equation (11.62a) as a function of V_{GS} for constant V_D A straight line is fitted through the points. The deviation from the straight line at lo values of V_{GS} is due to subthreshold conduction and the deviation at higher values o V_{GS} is due to mobility being afunctionof gate voltage. Both of these effects will be con sidered in the next chapter. The extrapolation of the straight line to zero current give the threshold voltage and the slope is proportional to the inversion carrier mobility.

If we take the square root of Equation (11.61), we obtain

$$\sqrt{I_D(\text{sat})} = \sqrt{\frac{W\mu_n C_{ox}}{2L}}(V_{GS} - V_T) \qquad (11.62b$$

Figure 11.49b is a plot of Equation (11.62b). In the ideal case, we can obtain the same information from both curves. However, as we will see in the next chapter. the thresh- oldvoltage may be afunction of V_{DS} in short-channel devices. Since Equation (11.62b) applies to devices biased in the saturation region. the V_t parameter in this equation may differ from the extrapolated value determined in Figure 11.49a. In general, the nonsat- uration current–voltagc characteristics will produce the more reliable data.

EXAMPLE 11.9 | Objective

To determine the inversion carrier mobility from experimental results.

Consider an n-channel MOSFET with $W = 15\ \mu$m, $L = 2\ \mu$m, and $C_{ox} = 6.9 \times 10^{-8}$ F/cm^2. Assume that the drain current in the nonsaturation region for $V_{DS} = 0.10$ V is $I_D = 35\ \mu$A at $V_{GS} = 1.5$ V and $I_D = 75\ \mu$A at $V_{GS} = 2.5$ V.

■ **Solution**

From Equation (11.62a), we can write

$$I_{D2} - I_{D1} = \frac{W\mu_n C_{ox}}{L}(V_{GS2} - V_{GS1})V_{DS}$$

so that

$$75 \times 10^{-6} - 35 \times 10^{-6} = \left(\frac{15}{2}\right)\mu_n(6.9 \times 10^{-8})(2.5 - 1.5)(0.10)$$

which yields

$$\mu_n = 773\ \text{cm}^2/\text{V-s}$$

We can then determine

$$V_T = 0.625\ \text{V}$$

■ **Comment**

The mobility of carriers in the inversion layer is less than that in the bulk semiconductor due to the surface scattering effect. We will discuss this effect in the next chapter.

The current-voltage relationship of a p-channel device can be obtained by the same type of analysis. Figure 11.50 shows a p-channel enhancement mode MOSFET. The voltage polarities and current direction are the reverse of those in the n-channel device. We may note the change in the subscript notation for this device. For the current direction shown in the figure, the I–V relations for the p-channel MOSFET are

$$I_D = \frac{W\mu_p C_{ox}}{2L}\left[2(V_{SG} + V_T)V_{SD} - V_{SD}^2\right] \tag{11.63}$$

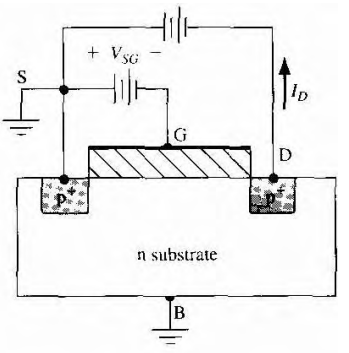


Figure 11.50 | Cross section and bias configuration for a p-channel enhancement-mode MOSFET.

for $0 \leq V_{SD} \leq V_{SD}(\text{sat})$, and

$$I_D(\text{sat}) = \frac{W\mu_p C_{\text{ox}}}{2L}(V_{SG} + V_T)^2 \tag{11}$$

for $V_{SD} \geq V_{SD}(\text{sat})$, where

$$V_{SD}(\text{sat}) = V_{SG} + V_T \tag{11.65}$$

Note the change in the sign in front of V_T and note that the mobility is now the mobility of the holes in the hole inversion layer charge. Keep in mind that V_T is negative for a p-channel enhancement mode MOSFET and positive for a depletion mode p-channel device.

TEST YOUR UNDERSTANDING

E11.15 The parameters of a p-channel MOSFET are $\mu_p = 310\,\text{cm}^2/\text{V-s}$, $t_{\text{ox}} = 220\,\text{Å}$, $W/L = 60$. and $V_T = -0.40$ V. If the transistor is biased in the saturation region, find the drain current for $V_{SG} = 1$, 1.5. and 2 V.
(Ans. 0.526, 1.77 and 3.74 mA)

E11.16 The p-channel MOSFET in E11.15 is to be redesigned by changing the (W/L) ratio such that $I_D = 200\,\mu\text{A}$ when the transistor is biased in the saturation region with $V_{SG} = 1.25$ V. (Ans. $W/L = 11.4$)

One assumption we made in the derivation of the current-voltage relationship was that the charge neutrality condition given by Equation (11.46) was valid over the entire length of the channel. We implicitly assumed that $Q'_{SD}(\text{max})$ was constant along the length of the channel. The space charge width, however, varies between source and drain due to the drain-to-source voltage; it is widest at the drain when $V_{DS} > 0$. A change in the space charge density along the channel length must be balanced by a corresponding change in the inversion layer charge. An increase in the space charge width means that the inversion layer charge is reduced, implying that the drain current and drain-to-source saturation voltage are less than the ideal values. The actual saturation drain current may be as much as 20 percent less than the predicted value due to this bulk charge effect.

11.3.4 Transconductance

The MOSFET transconductance is defined as the change in drain current with respect to the corresponding change in gate voltage, or

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \tag{11.66}$$

The transconductance is sometimes referred to as the transistor gain.

If we consider an n-channel MOSFET operating in the nonsaturation region, thm. using Equation (11.58). we have

$$g_{mL} = \frac{\partial I_D}{\partial V_{GS}} = \frac{W\mu_n C_{ox}}{L} \cdot V_{DS} \tag{11.67}$$

The transconductance increases linearly with V_{DS} hut is independent of V_{GS} in the nonsaturation region.

The I–V characteristics of an n-channel MOSFET in the saturation region were given by Equation (11.61). The transconductance in this region of operation is given by

$$g_{ms} = \frac{\partial I_D(\text{sat})}{\partial V_{GS}} = \frac{W\mu_n C_{ox}}{L}(V_{GS} - V_T) \tag{11.68}$$

In the saturation region, the transconductance is a linear function of V_{GS} and is independent of V_{DS}.

The transconductance is a function of the geometry of the device as well as of carrier mobility and threshold voltage. The transconductance increases as the width of the device increases, and it also increases as the channel length and oxide thickners decrease. In the design of MOSFET circuits, the sire of the transistor. in particular the channel width W. is an important engineering design parameter.

11.3.5 Substrate Bias Effects

In all of our analyses so far, the substrate, or hody, has been connected to the source and held at ground potential. In MOSFET circuits. the source and hody may not he at the same potential. Figure 11.51a shows an n-channel MOSFET and the associated double-subscripted voltage variahles. The source-to-substrate pn junction must always he zero or reverse biased, so V_{SB} must always he greeter than or equal to zero.

If $V_{SB} = 0$, threshold is defined as the condition when $\phi_s = 2\phi_{fp}$ as we discussed previously and as shown in Figure 11.51b. When $V_{SB} > 0$ the surface will still try to invert when $\phi_s = 2\phi_{fp}$. However, these electrons are at a higher potential

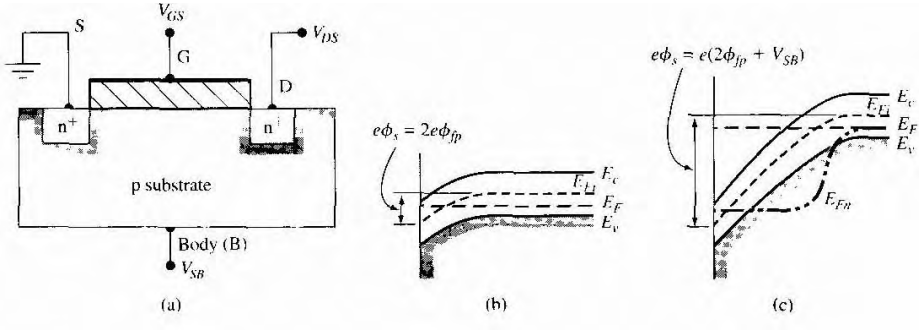


Figure 11.51 | (a) Applied voltages on an n-channel MOSFET (b) Energy-band diagram at inversion point when $V_{SB} = 0$. (c) Energy-band diagram at inversion point when $V_{SB} > 0$ is applied.

energy than are the electrons in the source. The newly created electrons will move laterally and Row out of the source terminal. When $\phi_s = 2\phi_{fp} + V_{SB}$, the surface reaches an equilibrium inversion condition. The energy-band diagram for this condition is shown in Figure 11.51c. The curve represented as E_{Fn} is the Fermi level from the p substrate through the reverse-biased source-substrate junction to the source contact.

The space charge region width under the oxide increases from the original x_{dT} value when a reverse-biased source-substrate junction voltage is applied. With an applied $V_{SB} > 0$, there is more charge associated with this region. Considering the charge neutrality condition through the MOS structure, the positive charge on the top metal gate must increase to compensate for the increased negative space charge in order to reach the threshold inversion point. So when $V_{SB} > 0$, the threshold voltage of the n-channel MOSFET increases.

When $V_{SB} = 0$, we had

$$Q'_{SD}(\text{max}) = -eN_a x_{dT} = -\sqrt{2e\epsilon_s N_a(2\phi_{fp})} \tag{11.69}$$

When $V_{SB} > 0$, the space charge width increases and we now have

$$Q'_{SD} = -eN_a x_d = -\sqrt{2e\epsilon_s N_a(2\phi_{fp} + V_{SB})} \tag{11.70}$$

The change in the space charge density is then

$$\Delta Q'_{SD} = -\sqrt{2e\epsilon_s N_a}\left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}\right] \tag{11.71}$$

To reach the threshold condition, the applied gate voltage must be increased. The change in threshold voltage can be written as

$$\boxed{\Delta V_T = -\frac{\Delta Q'_{SD}}{C_{ox}} = \frac{\sqrt{2e\epsilon_s N_a}}{C_{ox}}\left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}\right]} \tag{11.72}$$

where $\Delta V_T = V_T(V_{SB} > 0) - V_T(V_{SB} = 0)$. We may note that V_{SB} must alwaysbe positive so that, for the n-channel device, ΔV_T is always positive. The threshold voltage of the n-channel MOSFET will increase as a function of the source-substrate junction voltage.

EXAMPLE 11.10

Objective

To calculate the change in the threshold voltage due to an applied source-to-body voltage.

Consider an n-channel silicon MOSFET at T = 300 K. Assume the substrate is doped to $N_a = 3 \times 10^{16}$ cm^{-3} and assume the oxide is silicon dioxide with a thickness of $t_{ox} = 500$ Å. Let $V_{SB} = 1$ V.

■ **Solution**

We can calculate that

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259)\ln\left(\frac{3 \times 10^{16}}{1.5 \times 10^{10}}\right) = 0.376 \text{ V}$$

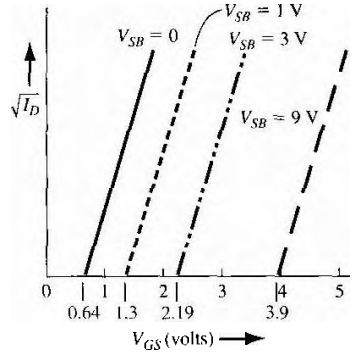


Figure 11.52 | Plots of $\sqrt{I_D}$ versus V_{GS} at several values of V_{SB} for an n-channel MOSFET.

We can also find

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{(3.9)(8.85 \times 10^{-14})}{500 \times 10^{-8}} = 6.9 \times 10^{-8} \text{ F/cm}^2$$

Then from Equation (11.72), we can obtain

$$\Delta V_T = \frac{[2(1.6 \times 10^{-19})(11.7)(8.85 \times 10^{-14})(3 \times 10^{16})]^{1/2}}{6.9 \times 10^{-8}} \times \{[2(0.376) + 1]^{1/2} - [2(0.376)]^{1/2}\}$$

or

$$\Delta V_T = 1.445(1.324 - 0.867) = 0.66 \text{ V}$$

Comment

Figure 11.52 shows plots of $\sqrt{I_D(\text{sat})}$ versus V_{GS} for various values of applied V_{SB}. The original threshold voltage, V_{T0}, is 0.64 V.

If a body or substrate bias is applied to a p-channel device, the threshold voltage is shifted to more negative values. Because the threshold voltage of a p-channel enhancement mode MOSFET is negative, a body voltage will increase the applied negative gate voltage required to create inversion. The same general observation was made for the n-channel MOSFET.

TEST YOUR UNDERSTANDING

E11.17 A silicon MOS device has the following parameters: $N_a = 10^{16}$ cm^{-3} and $t_{ox} = 200$ Å. Calculate (*a*) the body-effect coefficient and (*b*) the change in threshold voltage for (*i*) $V_{SB} = 1$ V and (ii) $V_{SB} = 2$ V.

[Ans. (*a*) $\gamma = 0.333\ V^{1/2}$, (*b*) (*i*) $\Delta V_T = 0.158$ V, (*ii*) $\Delta V_T = 0.269$ V]

E11.18 Repeat exercise E11.17 for a substrate impurity doping concentration of $N_a = 10^{15}\ \text{cm}^{-3}$.

11.4 | FREQUENCY LIMITATIONS

In many applications, the MOSFET is used in a linear amplifier circuit. A small-signal equivalent circuit for the MOSFET is needed in order to mathematically analyze the electronic circuit. The equivalent circuit contains capacitances and resistances that introduce frequency effects. We will initially develop a small-signal equivalent circuit and then discuss the physical factors that limit the frequency response of the MOSFET. A transistor cutoff frequency, which is a figure of merit, will then be defined and an expression derived for this factor.

11.4.1 Small-Signal Equivalent Circuit

The small-signal equivalent circuit of the MOSFET is constructed from the basic MOSFET geometry. A model based on the inherent capacitances and resistances within the transistor structure, along with elements that represent the basic device equations, is shown in Figure 11.53. One simplifying assumption we will make in the equivalent circuit is that the source and substrate are both tied to ground potential.

Two of the capacitances connected to the gate are inherent in the device. These capacitances are C_{gs} and C_{gd}, which represent the interaction between the gate and the channel charge near the source and drain terminals, respectively. The remaining two gate capacitances, C_{gsp} and C_{gdp}, are parasitic or overlap capacitances. In real devices, the gate oxide will overlap the source and drain contacts because of tolerance or fabrication factors. As we will see, the drain overlap capacitance—C_{gdp}, in particular—will lower the frequency response of the device. The parameter C_{ds} is the

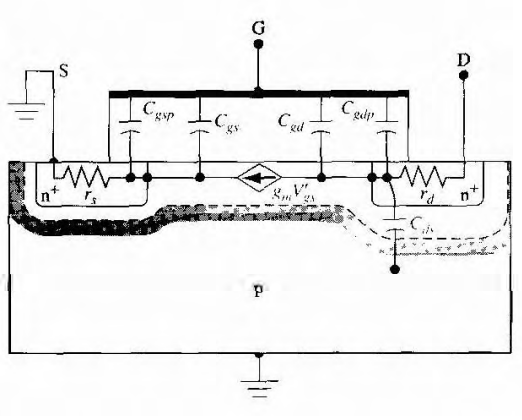


Figure 11.53 | Inherent resistances and capacitances in the n-channel MOSFET structure.

drain-to-substrate pn junction capacitance, and r_s and r_d are the series resistances associated with the source and drain terminals. The small-signal channel current is controlled by the internal gate-to-source voltage through the transconductance.

The small-signal equivalent circuit for the n-channel common-source MOSFET is shown in Figure 11.54. The voltage V'_{gs} is the internal gate-to-source voltage that controls the channel current. The parameters C_{gsT} and C_{gdT} are the total gate-to-source and total gate-to-drain capacitances. One parameter, r_{ds}, shown in Figure 11.54, is not shown in Figure 11.53. This resistance is associated with the slope I_D versus V_{DS}. In the ideal MOSFET biased in the saturation region, I_D is independent of V_{DS} so that r_{ds} would be infinite. In short-channel-length devices, in particular, r_{ds} is finite because of channel length modulation, which we will consider in the next chapter.

A simplified small-signal equivalent circuit valid at low frequency is shown in Figure 11.55. The series resistances, r_s and r_d, have been neglected, so the drain current is essentially only a function of the gate-to-source voltage through the transconductance. The input gate impedance is infinite in this simplified model.

The source resistance r_s can have a significant effect on the transistor characteristics. Figure 11.56 shows a simplified, low-frequency equivalent circuit including r_s but neglecting r_{ds}. The drain current is given by

$$I_d = g_m V'_{gs} \tag{11.73}$$

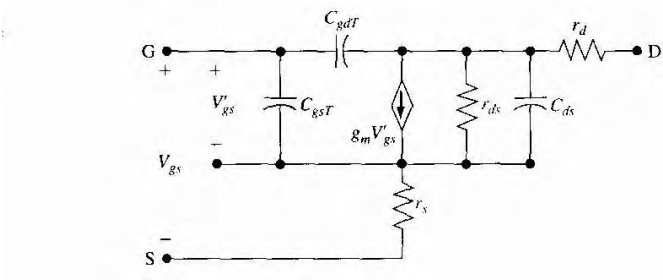


Figure 11.54 | Small-signal equivalent circuit of a common-source n-channel MOSFET.

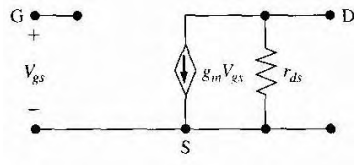


Figure 11.55 | Simplified, low-frequency small-signal equivalent circuit of a common-source n-channel MOSFET.

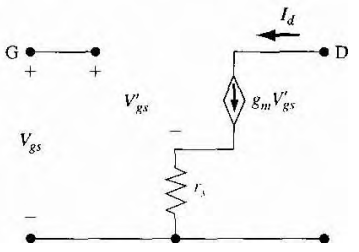


Figure 11.56 | Simplified, low-frequency small-signal equivalent circuit of common-source n-channel MOSFET including source resistance r_s

and the relation between V_{gs} and V'_{gs} can be found from

$$V_{gs} = V'_{gs} + (g_m V'_{gs}) r_s = (1 + g_m r_s) V'_{gs} \tag{11.74}$$

The drain current from Equation (11.73) can now be written as

$$I_d = \left(\frac{g_m}{1 + g_m r_s} \right) V_{gs} = g'_m V_{gs} \tag{11.75}$$

The source resistance reduces the effective transconductance or transistor gain.

The equivalent circuit of the p-channel MOSFET is exactly the same as that of the n-channel except that all voltage polarities and current directions are reversed. The same capacitances and resistances that are in the n-channel model apply to the p-channel model.

11.4.2 Frequency Limitation Factors and Cutoff Frequency

There are two basic frequency limitation factors in the MOSFET. The first factor is the channel transit time. If we assume that carriers are traveling at their saturation drift velocity v_{sat}, then the transit time is $\tau_t = L/v_{sat}$ where L is the channel length. If $v_{sat} = 10^7$ cm/s and L = 1 μm, then $\tau_t = 10$ ps, which translates into a maximum frequency of 100 GHz. This frequency is much larger than the typical maximum frequency response of a MOSFET. The transit time of carriers through the channel is usually not the limiting factor in the frequency responses of MOSFETs.

The second limiting factor is the gate or capacitance charging time. If we neglect r_s, r_d, r_{ds}, and C_{ds}, the resulting equivalent small-signal circuit is shown in Figure 11.57 where R_L is a load resistance.

The input gate impedance in this equivalent circuit is no longer infinite. Summing currents at the input gate node, we have

$$I_i = j\omega C_{gsT} V_{gs} + j\omega C_{gdT} (V_{gs} - V_d) \tag{11.76}$$

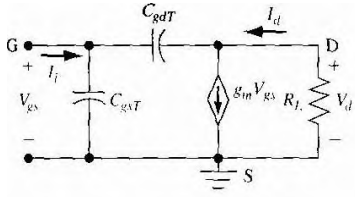


Figure 11.57 | High-frequency small-signal equivalent circuit of common-source n-channel MOSFET.

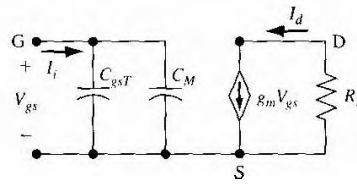


Figure 11.58 | Small-signal equivalent circuit including Miller capacitance.

where I_i is the input current. Likewise, summing currents at the output drain node, we have

$$\frac{V_d}{R_L} + g_m V_{gs} + j\omega C_{gdT}(V_d - V_{gs}) = 0 \tag{11.77}$$

Combining Equations (11.76) and (11.77) to eliminate the voltage variable V_d, we can determine the input current as

$$I_i = j\omega\left[C_{gsT} + C_{gdT}\left(\frac{1 + g_m R_L}{1 + j\omega R_L C_{gdT}}\right)\right]V_{gs} \tag{11.78}$$

Normally, $\omega R_L C_{gdT}$ is much less than unity: therefore we may neglect the $(j\omega R_L C_{gdT})$ term in the denominator. Equation (11.78) then simplifies to

$$I_i = j\omega[C_{gsT} + C_{gdT}(1 + g_m R_L)]V_{gs} \tag{11.79}$$

Figure 11.58 shows theequivalent circuit with theequivalent input impedance described by Equation (11.79). The parameter C_M is the Miller capacitance and is given by

$$C_M = C_{gdT}(1 + g_m R_L) \tag{11.80}$$

The serious effect of the drain overlap capacitance now becomes apparent. When the transistor is operating in the saturation region, C_{gd} essentially becomes zero, but C_{gdp} is a constant. This parasitic capacitance is multiplied by the gain of the transistor and can become a significant factor in the input impedance.

The cutoff frequency f_T is defined to he the frequency at which the magnitude of the current gain of the device is unity, or when the magnitude of the input current I_i is equal to the ideal load current I_d. From Figure 11.58, we can see that

$$I_i = j\omega(C_{gsT} + C_M)V_{gs} \tag{11.81}$$

and the ideal load current is

$$I_d = g_m V_{gs} \tag{11.82}$$

The magnitude of the current gain is then

$$\left|\frac{I_d}{I_i}\right| = \frac{g_m}{2\pi f(C_{gsT} + C_M)} \tag{11.83}$$

Setting the magnitude of the current gain equal to unity at the cutoff frequency, we find

$$f_T = \frac{g_m}{2\pi(C_{gsT} + C_M)} = \frac{g_m}{2\pi C_G} \tag{11.84}$$

where C_G is the equivalent input gate capacitance.

In the ideal MOSFET, the overlap or parasitic capacitances, C_{gsp} and C_{gdp}, are zero. Also, when the transistor is biased in the saturation region, C_{gd} approaches zero and C_g, is approximately C_{ox}WL. The transconductance of the ideal MOSFET biased in the saturation region and assuming a constant mobility was given by Equation (11.68) as

$$g_{ms} = \frac{W\mu_n C_{ox}}{L}(V_{GS} - V_T)$$

Then, for this ideal case, the cutoff frequency is

$$f_T = \frac{W\mu_n C_{ox}(V_{GS} - V_T)}{2\pi(C_{ox}WL)} = \frac{\mu_n(V_{GS} - V_T)}{2\pi L^2} \tag{11.85}$$

EXAMPLE 11.11 Objective

To calculate the cutoff frequency of an ideal MOSFET with a constant mobility.

Assume that the electron mobility in an n-channel device is $\mu_n = 400\ \text{cm}^2/\text{V-s}$ and that the channel length is $L = 4\ \mu\text{m}$. Also assume that $V_T = 1$ V and let $V_{GS} = 3$ V.

■ Solution

From Equation (11.85), the cutoff frequency is

$$f_T = \frac{\mu_n(V_{GS} - V_T)}{2\pi L^2} = \frac{400(3-1)}{2\pi(4 \times 10^{-4})^2} = 796\ \text{MHz}$$

Comment

In an actual MOSFET, the effect of the parasitic capacitance will substantially reduce the cutoff frequency from that calculated in this example.

TEST YOUR UNDERSTANDING

E11.19 An n-channel MOSFET has the following parameters: $\mu_n = 400\ \text{cm}^2/\text{V-s}$, $t_{ox} = 200$ Å, $W/L = 20$, and $V_T = 0.4$ V. The transistor is biased at $V_{GS} = 2.5$ V in the saturation region and is connected to an effective load of $R_L = 100\ \text{k}\Omega$. Calculate the ratio of Miller capacitance C_M to gate-to-drain capacitance C_{gdT}. (Ans. 267)

E11.20 An n-channel MOSFET has the same parameters as described in E11.19. The channel length is $L = 0.5\ \mu\text{m}$. Determine the cutoff frequency. (Ans. 796 MHz)

*11.5 | THE CMOS TECHNOLOGY

The primary objective of this text is topresent the basic physics of semiconductormaterials and devices without considering in detail the various fabrication processes; this important subject is left to other texts. However, there is one MOS technology that is usedextensively, for which the basic fabrication techniques must be considered in order tounderstand essential characteristics of these devices andcircuits. The one MOS technology we will consider briefly is the complementary MOS, or CMOS, process.

We have considered the physics of both n-channel and p-channel enhancement mode MOSFETs. Both devices are used in a CMOS inverter, which is the basis of CMOS digital logic circuits. The dc power dissipation in a digital circuit can he reduced to very low levels by using a complementary p-channel and n-channel pair.

It is necessary to form electrically isolated p- and n-substrate regions in an integrated circuit to accommodate the n- and p-channel transistors. The p-well process has been a commonly used technique for CMOS circuits. The process starts with a fairly low doped n-type silicon substrate in which the p-channel MOSFET will be fabricated. A diffused p-region, called a p well, is formed in which the n-channel MOSFET will be fabricated. In most cases. the p-type substrate doping level must be larger than the n-type substrate doping level to obtain the desired threshold voltages. The larger p doping can easily compensate the initial n doping to form the p well. A simplifiedcross section of the p-well CMOS structure is shown in Figure 11.59a. The

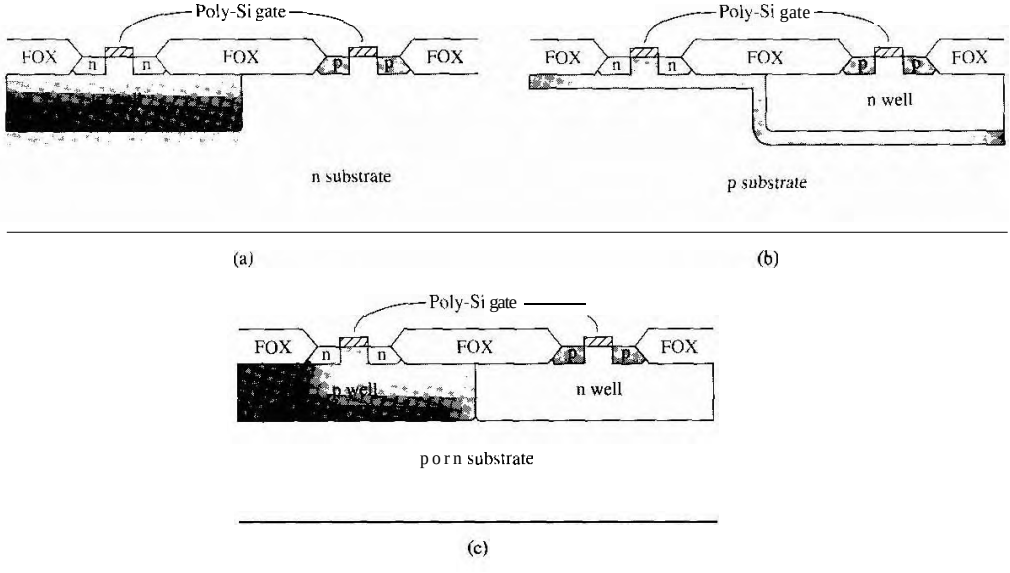


Figure 11.59 | CMOS structures: (a) p well, (b) n well, and (c) twin **well**
(From Yang [211])

notation FOX stands for field oxide, which is a relatively thick oxide separating the devices. The field oxide prevents either the n or p substrate from becoming inverted and helps maintain isolation between the two devices. In practice, additional processing steps must he included; for example, providing connections so that the p well and n suhstrate can be electrically connected to the appropriate voltages. The n substrate must always be at a higher potential than the p well; therefore, this pn junction will always be reverse biased.

With ion implantation now being extensively used for threshold voltage control, both the n-well CMOS process and twin-well CMOS process can be used. The n-well CMOS process, shown in Figure 11.59b, starts with an optimized p-type suhstrate that is used to form the n-channel MOSFETs. (The n-channel MOSFETs, in general, have superior characteristics, so this starting point should yield excellent n-channel devices.) The n well is then added, in which the p-channel devices are fabricated. The n-well doping can be controlled by ion implantation.

The twin-well CMOS process, shown in Figure 11.59c, allows both the p-well an n-well regions to be optimally doped to control the threshold voltage and transconductance of each transistor. The twin-well process allows a higher packing density because of self-aligned channel stops.

One major problem in CMOS circuits has been ***latch-up.*** Latch-up refers to a high-current, low-voltage condition that may occur in a four-layer pnpn structure. Figure 11.60a shows the circuit of a CMOS inverter and Figure 11.60b shows a simplified integrated circuit layout of the inverter circuit. In the CMOS layout, the p^+-source to n-substrate to p-well to n+-source forms such a four-layer structure.

The equivalent circuit of this four-layer structure is shown in Figure 11.61. The silicon controlled rectifier action involves the interaction of the parasitic pnp and npn transistors. The npn transistor corresponds to the vertical n^+ source top well ton substrate structure and the pnp transistor corresponds to the lateral p-well to n-substrate to p^+-source structure. Under normal CMOS operation, both parasitic bipolar transistors are cut off. However, under certain conditions, avalanche breakdown may occur in the p-well to n-substrate junction, driving both bipolar transistors into saturation. This high-current, low-voltage condition—latch-up—can sustain itself by

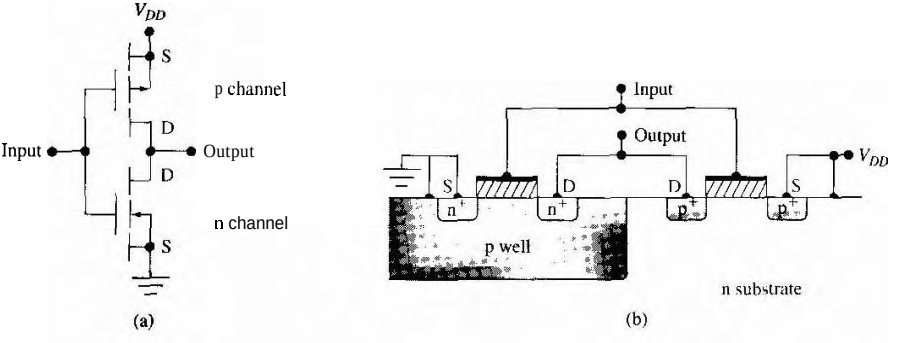


Figure 11.60 | (a) CMOS inverter circuit. (b) Simplified integrated circuit cross section of CMOS inverter.

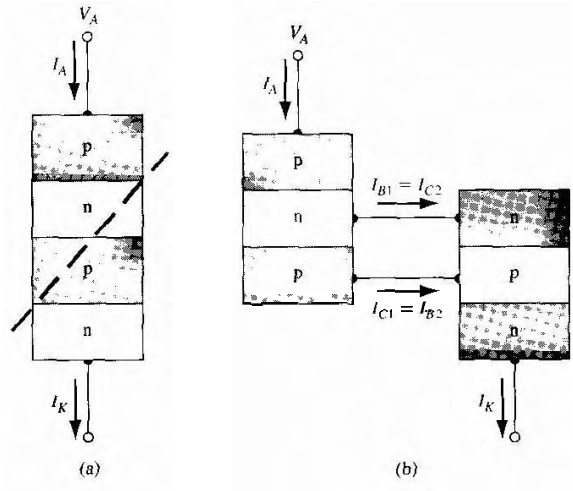


Figure 11.61 | (a) The splitting of the basic pnpn structure. (b) The two-transistor equivalent circuit of the four-layered pnpn device.

positive feedback. The condition can prevent the CMOS circuit from operating and can also cause permanent damage and burn-out of the circuit.

Latch-up can be prevented if the product $\beta_n \beta_p$ is less than unity at all times, where β_n and β_p are the common-emitter current gains of the npn and pnp parasitic bipolar transistors, respectively. One method of preventing latch-up is to "kill" the minority carrier lifetime. Minority carrier lifetime degradation can be accomplished by gold doping or neutron irradiation. either of which introduces deep traps within the semiconductor. The deep traps increase the excess minority carrier recombination rate and reduce current gain. A second method of preventing latch-up is by using proper circuit layout techniques. If the two bipolar transistors can he effectively decoupled, then latch-up can be minimized or prevented. The two parasitic bipolar transistors can also be decoupled by using a different fabrication technology. The silicon-on-insulator technology, for example, allows the n-channel and the p-channel MOSFETs to be isolated from each other by an insulator. This isolation decouples the parasitic bipolar transistors.

11.6 | SUMMARY

- The fundamental physics and characteristics of the metalkoxide–semiconductor field-effect transistor (MOSFET) have been considered in this chapter
- The heart of the MOSFET is the MOS capacitor. The energy hands in the semiconductor adjacent to the oxide–semiconductor interface bend, depending upon the voltage applied across the MOS capacitor. The position of the conduction and valence bands relative to the Fermi level at the surface is a function of the MOS capacitor voltage.
- The semiconductor surface at the oxide–semiconductor interface can be inverted from p type to n type by applying a positive gate voltage, or from n type to p type by applying

a negative gate voltage. Thus, an inversion layer of mobile charge can be created adjacent to the oxide. The basic MOS field-effect action is the modulation of the inversion charge density, or channel conductance. by the gatc voltage.

- The C-V characteristics of the MOS capacitor were considered. The amount of equivalent oxide trapped charge and the density of interface states, for example. can be determined from the C-V measurements.
- Two basic types of MOSFETs are the n channel, in which current is due to the flow of electrons in the inversion layrr. and the p channel, in which current is due to the flow of holes in the inversion layer Each of these devices can be either enhancement mode, in which the device is normally "off" and is turned on by applying a gate voltage. or depletion mode. in which the device is normally "on" and is turned off hy applying a gate voltage.
- The flat-band voltage is the gate voltage that must be applied to achieve the flat-band condition, in which the conduction and valence bands in the semiconductor do not bend and there is no space charge region in the semiconductor. The Hat-hand voltage is a function of the metal–oxide barrier height. the semiconductor–oxide barrier height, and the amount of fixed trapped oxide charge.
- The threshold voltage is the applied gate voltage required to reach the threshold inversion point, which is the condition at which the inversion charge density is equal in magnitude to the semiconductor doping concentration. The threshold voltage is a function of the flat-band voltage. semiconductor doping concentration, and oxide thickness.
- The current in a MOSFET is due lo the flow of carriers in the inversion layer between the source and drain terminals. The inversion layer charge density and channel conductance are controlled by the gate voltage, which means that the channel current is also controlled by the gate voltage.
- When the transistor is biased in the nonsaturation region ($V_{DS} < V_{DS}(\text{sat})$), the inversion charge extends completely across the channel from the source to the drain terminals. The drain current is a function of bath the gate-to-source and drain-to-source voltages. When the transistor is biased in the saturation region ($V_{DS} > V_{DS}(\text{sat})$), the inversion charge density is pinched off near the drain terminal. and the ideal drain current is only a function of the gate-to-source voltage.
- The MOSFET is actually a four-terminal device. with the substrate or body being the fourth terminal. As the magnitude of the reverse-bias source-to-substrate voltage increases, the magnitude of the threshold voltage increases. The substrate bias effect may become important in integrated circuits in which the source and substrate are not electrically tied together.
- A small-signal equivalent circuit. including capacitances, of the MOSFET was developed. The various physical factors in the MOSFET that affect the frequency limitations were considered. In particular, the drain overlap capacitance may be a limiting factor in the frequency response of the MOSFET because of the Miller effect. The cutoff frequency, a figure of merit for the frequency response of the device. is inversely proportional to channel length; thus. a reduction in channel length results in an increased frequency capability of the MOSFET.
- The CMOS technology, in which both n-channel and p-channel devices are fabricated in the same semiconductor chip, was briefly considered. Electrically isolated p- and n-substrate regions arc required to accommodate the two types of transistors. Various orocesses are used to fabricate this structure. One potential problem encountered in the CMOS structure is latch-up—the high-current, low-voltage condition that may occur in a four-layer pnpn structure.

GLOSSARY OF IMPORTANT TERMS

accumulation layer chargc The induced charge directly under an oxide that is in excess of the thermal-equilibrium majority carrier concentration.

bulk charge effect The deviation in drain current from the ideal due to the space charge width variation along the channel length caused by a drain-to-source voltage.

channel conductance The ratio uf drain current to drain-to-source voltage in the limit as $V_{DS} \to 0$.

channel conductance modulation The process whereby the channel conductance varies with gate-to-source voltage.

CMOS Complementary MOS; the technology that uses both p- and n-channel devices in an electronic circuit fabricated in a single semiconductor chip.

cutoff frequency The signal frequency at which the input ac gate current is equal to the output ac drain current.

depletion mode MOSFET The type of MOSFET in which a gate voltage must be applied to turn the device off.

enhancement mode MOSFET The type of MOSFET in which a gate voltage must he applied to turn the device on.

equivalent fined oxide charge The effective fixed charge in the oxide, Q'_{ss}, directly adjacent to the oxide–semiconductor interface.

flat-band voltage The gate voltage that must be applied to create the flat-band condition in which there is no space charge region in the semiconductor under the oxide.

gate capacitance charging time The time during which the input gate capacitance is being charged or discharged because of a step change in the gate signal.

interface states The allowed electronic energy states within the bandgap energy at the onide–semiconductor interface.

inversion layer charge The induced charge directly under the oxide, which is the opposite type compared with the semiconductor doping.

inrersion layer mobility The mohility of carriers in the inversion layer.

latch-up The high-current, low-voltage condition that may occur in a four-layer pnpn structure such as in CMOS.

maximum induced space charge width The width of the induced space charge region under the oxide at the threshold inversion condition.

metal–semiconductor work function difference The parameter ϕ_{ms}, a function of the difference between the metal work function and semiconductor electron affinity.

moderate inversion The condition in which the induced space charge width is changing slightly when the gate voltage is at or near the threshold voltage and the inversion charge density is of the same magnitude as the semiconductor doping concentration.

oxide capacitance The ratio of oxide permittivity to oxide thickness, which is the capacitance per unit area, C_{ox}

saturation The condition in which the inversion charge density is zero at the drain and the drain current is no longer a function of the drain-to-source voltage.

strong inversion The condition in which the inversion charge density is larger than the magnitude of the semiconductor doping concentration.

threshold inversion point The condition in which the inversion charge density is equal in magnitude to the semiconductor doping concentration.

threshold voltage The gate voltage that must be applied to achieve the threshold inversion point.

transconductance The ratio of an incremental change in drain current to the corresponding incremental change in gate voltage.

weak inversion The condition in which the inversion charge density is less than the magnitude of the semiconductor doping concentration.

CHECKPOINT

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

- Sketch the energy band diagrams in the semiconductor of the MOS capacitor under various bias conditions.
- Describe the process by which an inversion layer of charge is created in an MOS capacitor.
- Discuss the reason the space charge width reaches a maximum value once the inversion layer is formed.
- Discuss what is meant by the metal–semiconductor work function difference, and discuss why this value is different between aluminum. n^+ polysilicon, and p^+ polysilicon gates.
- Describe what is meant by flat-band voltage.
- Define threshold voltage.
- Sketch the C-Vcharacteristics of an MOS capacitor with p-type and n-type semiconductor substrates under high-frequency and low-frequency conditions.
- Discuss the effects of fixed trapped oxide charge and interface states on the C-V characteristics.
- Sketch the cross-sections of n-channel and p-channel MOSFET structures.
- Explain the basic operation of the MOSFET.
- Discuss the *I–V* characteristics of the MOSFET when biased in the nonsaturation and saturation regions.
- Describe the substrate bias effects on the threshold voltage.
- Sketch the small-signal equivalent circuit. including capacitances, of the MOSFET, and explain the physical origin of each capacitance.
- Discuss the condition that defines the cutoff frequency of a MOSFET.
- Sketch the cross section of a CMOS structure.
- Discuss what is meant by latch-up in a CMOS structure.

REVIEW QUESTIONS

1. Sketch the energy band diagrams in an MOS capacitor with an n-type substrate in accumulation, depletion, and inversion modes.
2. Describe what is meant by an inversion layer of charge. Describe haw an inversion layer of charge can be formed in an MOS capacitor with a p-type substrate.
3. Why does the space-charge region in the semicunductor of an MOS capacitor reach a maximum width once the inversion layer is formed?
4. Define electron affinity in the semiconductor of an MOS capacitor
5. Sketch the energy band diagram through an MOS structure with a p-type substrate and an n^+ polysilicon gate under zero bias.

6. Define the Rat-band voltage.
7. Define the threshold voltage.
8. Sketch the *C-V* characteristics of an **MOS** capacitor with an n-type substrate under the low-frequency condition. How do the characteristics change for the high-frequency condition?
9. indicate the approximate capacitance at flat-band on the C-Vcharacteristic of an MOS capacitor with a p-type substrate under the high-frequency condition.
10. What is the effect on the C-Vcharacteristics **of** an **MOS** capacitor with a p-type substrate if the amount of positive trapped oxide charge increases?
11. Qualitatively sketch the inversion charge density in the channel region when the transistor is biased in the nonsaturation region. Repeat for the case when the transistor is biased in the saturation region.
12. Define $V_{DS}(\text{sat})$.
13. Define enhancement mode and depletion mode for both n-channel and p-channel devices.
14. Sketch the charge distribution through an MOS capacitor with a p-type substrate when biased in the inversion mode. Write the charge neutrality equation.
15. Discuss why the threshold voltage changes when a reverse-biased source-to-substrate voltage is applied to a **MOSFET.**

PROBLEMS

(*Note:* In the following problems, assume the semiconductor and oxide in the **MOS** system are silicon and silicon dioxide, respectively. and assume the temperature is $T = 300$ K unless otherwise stated. Use Figure 11.15 to determine the metal–semiconductor work function difference.)

Section 11.1 The Two-Terminal MOS Structure

11.1 The dc charge distributions of four ideal **MOS** capacitors are shown in Figure 11.62. For each case: (*a*) Is the semiconductor n- or p-type? (*b*) Is the device biased in the accumulation. depletion. or inversion mode'?(*c*) Draw the energy-band diagram in the semiconductor region.

11.2 (*a*)Calculate the maximum space charge width x_{dT} and the maximum space charge density $|Q'_{SD}(\text{max})|$ in p-type silicon. gallium arsenide, and germanium semiconductors of an MOS structure. Let T = 300 K and assume $N_a = 10^{16}$ c m '. (*b*) Repeat part (*a*) if T = 200 K.

11.3 (*a*) Consider n-type silicon in an **MOS** structure. Let $T = 300$ K. Determine the semiconductor doping so that $|Q'_{SD}(\text{max})| = 7.5 \times 10^{-9}$ C/cm^2. (*b*) Determine the surface potential that results in the maximum space charge width.

11.4 Determine the metal–semiconductor work function difference ϕ_{ms} in an **MOS** structure with p-type silicon for the case when the gate is (*a*) aluminum. (*b*) n^+ polysilicon, and (*c*) p^+ polysilicon. Let $N_a = 6 \times 10^{15}$ cm^{-3}

11.5 Consider an **MOS** structure with n-type silicon. A metal–semiconductor work function difference of $\phi_{ms} = -0.35$ V is required. Determine the silicon doping required to meet this specification when the gats is (*a*) n^+ polysilicon, (*h*) p^+ polysilicon, and (*c*) aluminum. If a particular gate cannot meet this requirement, explain why.

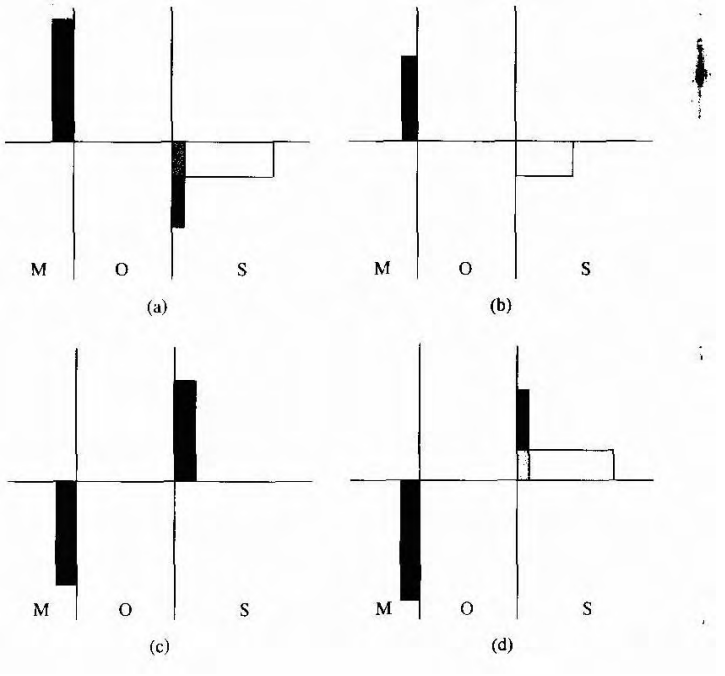


Figure 11.62 | Figure for Problem 11.1

11.6 Consider an n^+ polysilicon–silicon dioxide-n-type silicon MOS capacitor. Let $N_d = 10^{15}$ cm^{-3}. Calculate the flat-band voltage for (a) $t_{ox} = 500$ Å when Q'_{ss} is (*i*) 10^{10} cm^{-2}, (*ii*) 10^{11} cm^{-2}, and (*iii*) 5×10^{11} cm^{-2}. (*b*) Repeat pan ***(a)*** when $t_{ox} = 250$ Å.

11.7 Consider an aluminum gate–silicon dioxide–p-type silicon MOS structure with $t_{ox} = 450$ Å. The silicon doping is $N_a = 2 \times 10^{16}$ cm^{-3} and the flat-band voltage is $V_{FB} = -1.0$ V Determine the fixed oxide charge Q'_{ss}.

11.8 An MOS transistor is fabricated on a p-type silicon substrate with $N_a = 2 \times 10^{15}$ cm^{-3}. The oxide thickness is $t_{ox} = 450$ Å and the equivalent fixed oxide charge is $Q'_{ss} = 2 \times 10^{11}$ cm^{-2}. Calculate the threshold voltage for *(a)* an aluminum gate, (*b*) an n^+ polysilicon gate, and (*c*) a p^+ polysilicon gate.

11.9 Repeat Problem 11.8 for an n-type silicon substrate with $N_d = 10^{15}$ cm^{-3}.

11.10 A 400 Å oxide is grown on p-type silicon with $N_a = 5 \times 10^{15}$ cm^{-3}. The flat-band voltage is –0.9 V. Calculate the surface potential at the threshold inversion point as well as the threshold voltage assuming negligible oxide charge. Also find the maximum space charge width for this device.

***11.11** An MOS transistor with an aluminum gate is fabricated on a p-type silicon substrate. The oxide thickness is $t_{ox} = 750$ A, and the equivalent fixed oxide charge is $Q'_{ss} = 10^{11}$ cm^{-2}. The measured threshold voltage is $V_T = +0.80$ V. Determine the p-type doping concentration.

*11.12 Kepeat Problem 11.11 for an n-type silicon substrate if the measured threshold voltage is $V_T = -1.50$ V. Determine the n-type doping concentmtion.

11.13 An Al–silicon dioxide–silicon MOS capacitor has an oxide thickness of 450 Å and a doping of $N_a = 10^{15}$ cm^{-3}. The oxide charge density is $Q'_{ss} = 3 \times 10^{11}$ cm^{-2}. Calculate (a) the flat-band voltage and (*b*) the threshold voltage. Sketch the electric field through the structure at the onset of inversion.

11.14 An n-channel depletion mode MOSFET with an n^+ polysilicon gate is shown in Figure 11.42. The n-channel doping is $N_d = 10^{15}$ cm^{-3}, and the oxide thickness is $t_{ox} = 500$ A. The equivalent fixed oxide charge is $Q'_{ss} = 10^{10}$ cm^{-2}. The n-channel thickness t_c is equal to the maximum induced space charge width. (Disregard the space charge region at the n-channel–p-substratejunction.) (a) Determine the channel thickness t_c, and (*b*) calculate the threshold voltage.

11.15 Consider an MOS capacitor with an n^+ polyrilicon gate and n-type silicon substrate. Assume $N_a = 10^{16}$ cm^{-3} and let $E_F - E_i = 0.2$ eV in the n^+ polysilicon. Assume the oxide has a thickness of $t_{ox} = 300$ Å. Also assume that χ' (polysilicon) = χ' (single-crystal silicon). (o) Sketch the energy-band diagrams (*i*) for $V_G = 0$ and (ii) at flat band. (h) Calculate the metal–semiconductar work function difference. (c) Calculate the threshold voltage for the ideal case of zero fixed oxide charge and zero interface states.

11.16 The threshold voltage of an n-channel MOSFET is given by Equation (11.27). Plot V_T versus temperature over the range $200 \leq T \leq 450$ K. Consider both an aluminum gate and an n^+ polysilicon gate. Assume the work functions are independent of temperature and use device parameters similar to those in Example 11.4.



11.17 Plot the threshold voltage of an n-channel MOSFET versus p-type substrate doping concentration similar to Figure 11.20. Consider hoth n^+ and p^+ polysilicon gales. Use reasonable device parameters.



11.18 Plot the threshold voltage of a p-channel MOSFET versus n-type substrate doping concentration similar to Figure 11.21. Consider both n^+ and p^+ polysilicon gates. Use reasonable device parameters.



11.19 Consider an NMOS device with the parameters given in Problem 11.10. Plot V_t versus t_{ox} over the range $20 \leq t_{ox} \leq 500$ Å.



Section 11.2 Capacitance-Voltage Characteristics

11.20 An ideal MOS capacitor with an aluminum gate has a silicon dioxide thickness of $t_{ox} = 400$ Å on a p-type silicon substrate doped with an acceptor concentration of $N_a = 10^{16}$ cm^{-3}. Determine the capacitances C_{ox}, C'_{FB}, C'_{min}, and C'(inv) at (a) $f = 1$ Hz and (*b*) $f = 1$ MHz. (*c*) Determine V_{FB} and V_T. Sketch C'/C_{ox} versus V_G for parts (*a*) and (*b*).

11.21 Repeat Problem 11.20 for an n-type silicon substrate duped with a donor concentration of $N_d = 5 \times 10^{14}$ cm^{-3}.

*11.22 Using superposition, show that the shift in the Rat-band voltage due to a fixed charge distribution $\rho(x)$ in the oxide is given by

$$\Delta V_{FB} = -\frac{1}{C_{ox}} \int_0^{t_{ox}} \frac{x\rho(x)}{t_{ox}} dx$$

*11.23 Using the results of Problem 11.22, calculate the shift in the flat-band voltage for the following oxide charge distributions: (*a*) $Q'_{ss} = 5 \times 10^{11}$ cm^{-2} is entirely

located at the oxide-semiconductor interface. Let $t_{ox} = 750$ Å. (*b*) $Q'_{ss} = 5 \times 10^{11}$ cm^{-2} is uniformly distributed throughout the oxide. which has a thickness of $t_{ox} = 750$ Å. (*c*) $Q'_{ss} = 5 \times 10^{11}$ cm^{-2} forms a triangular distribution with the peak at $x = t_{ox} = 750$ Å (the oxide–semiconductor interface) and which goes to zero at $x = 0$ (the metal–oxide interface).

11.24 **An** ideal MOS capacitor is fabricated by using intrinsic silicon and an n^+ polysilicon gate. (a) Sketch the energy-band diagram through the MOS structure under flat-band conditions. (b) Sketch the low-frequency *C-V* characteristics from negative to positive gate voltage.

11.25 Consider an MOS capacitor with a p-type substrate. Assume that donor-type interface traps exist only at midgap (i.e., at E_{Fi}). Sketch the high-frequency C-V curve from accumulation to inversion. Compare this sketch to the ideal C-Vplot.

11.26 Consider an SOS capacitor as shown in Figure 11.63. Assume the SiO_2 is ideal (no trapped charge) and has a thickness of $t_{ox} = 500$ Å. The doping concentrations are $N_d = 10^{16}$ cm^{-3} and $N_a = 10^{16}$ cm^{-3}, (*a*) Sketch the energy band diagram through the device for (*i*) flat-band, (ii) $V_G = +3$ V, and (*iii*) $V_G = -3$ V. (*b*) Calculate the flat-band voltage. (*c*) Estimate the voltage across the oxide for (*i*) $V_G = +3$ *V* and (*ii*) $V_G = -3$ *V*. (d) Sketch the high-frequency *C-V* characteristic curve.

11.27 The high-frequency *C-V* characteristic curve of an MOS capacitor is shown in Figure 11.64. The area of the device is 2×10^{-3} cm'. The metal–semiconductor work function difference is $\phi_{ms} = -0.50$ V, the oxide is SiO_2, the semiconductor is silicon, and the semiconductor doping concentration is 2×10^{16} cm^{-3}. (a) Is the

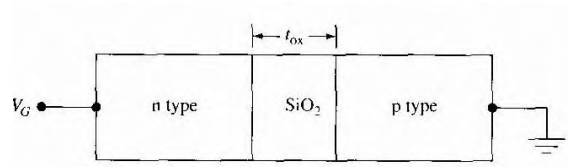


Figure 11.63 | Figure for Problem 11.26.

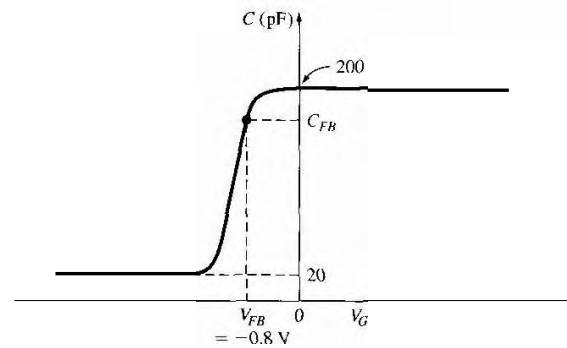


Figure **11.64** | Figure for Problem 112 7

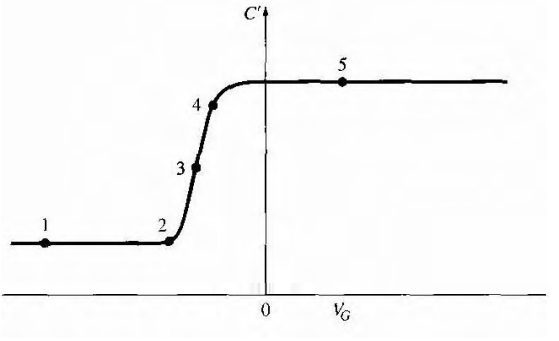


Figure 11.65 | Figure for Problem 11.28

semiconductor n or p type'?(*b*) What is the oxide thickness? ***(r)*** What is the equivalent trapped oxide charge density? (*d*) Determine the flat-hand capacitance.

11.28 Consider the high-frequency C-Vplot shown in Figure 11.65. (a) Indicate which points correspond to flat-band, inversion, accumulation, threshold, and depletion mode. (*b*) Sketch the energy band diagram in the semiconductor for each condition

Section 11.3 The Basic MOSFET Operation

11.29 An expression that includes the inversion charge density was given by Equation (11.55). Consider the definition of threshold voltage and show that the inversion charge density goes to zero at the drain terminal at saturation. (Hint: Let $V_x = V_{DS} = V_{DS}(\text{sat})$.)

11.30 An ideal n-channel MOSFET has the following parameters:

$$\text{W} = 30\ \mu\text{m} \qquad \mu_n = 450\ \text{cm}^2/\text{V-s}$$
$$L = 2\ \mu\text{m} \qquad t_{ox} = 350\ \text{A}$$
$$V_T = +0.80\ \text{V}$$

(*a*) Plot I_D versus V_{DS} for $0 \leq V_{DS} \leq 5$ V and for $V_{GS} = 0, 1, 2, 3.4$, and 5 V. Indicate on each curve the $V_{DS}(\text{sat})$ point. (*b*) Plot $\sqrt{I_D(\text{sat})}$ versus V_{GS} for $0 \leq V_{GS} \leq 5\text{V}$. **(c)** Plot I_D versus V_{GS} for $V_{DS} = 0.1$ V and for $0 \leq V_{GS} \leq 5\text{V}$.

11.31 An ideal p-channel MOSFET has the following parameters:

$$\text{W} = 15\ \mu\text{m} \qquad \mu_p = 300\ \text{cm}^2/\text{V-s}$$
$$L = 1.5\ \mu\text{m} \qquad t_{ox} = 350\ \text{Å}$$
$$V_T = -0.80\ \text{V}$$

(a) Plot I_D versus V_{SD} for $0 \leq V_{SD} \leq 5$ V and for $V_{SG} = 0, 1, 2, 3, 4$, and 5 V. Indicate on each curve the $V_{SD}(\text{sat})$ point. (*b*) Plot I_D versus V_{SG} for $V_{SD} = 0.1$ V and for $0 \leq V_{SG} \leq 5$ V.

11.32 Consider an n-channel MOSFET with the same parameters as given in Problem 11.30 except that $V_T = -2.0$ V. (a) Plot I_D versus V_{DS} for $0 \leq V_{DS} \leq 5$ V and for $V_{GS} = -2, -1, 0, +1$, and $+2$ V. (*b*) Plot $\sqrt{I_D(\text{sat})}$ versus V_{GS} for $-2\ 5\ V_{GS} \leq +3$ V.

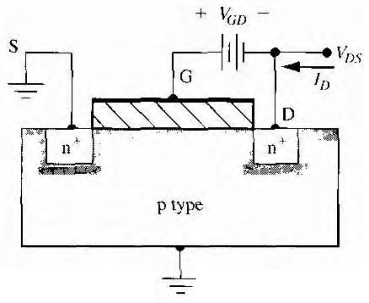


Figure **11.66** | Figure for Problem 11.33

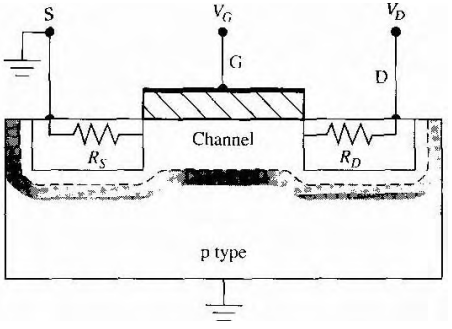


Figure **11.67** | Figure for Problem 11.34

11.33 Consider an n-channel enhancement mode MOSFET biased as shown in Figure 11.66. Sketch the current–voltage characteristics, I_D versus V_{DS}, for (*a*) $V_{GD} = 0$, (*b*) $V_{GD} = V_T/2$, and (*c*) $V_{GD} = 2V_T$.

11.34 Figure 11.67 shows the cross section of an NMOS device that includes source and drain resistances. These resistances take into account the bulk n^+ semiconductor resistance and the ohmic contact resistance. The current–voltage relations can be generated by replacing V_{GS} by $V_G - I_D R_S$ and V_{DS} by $V_D - I_D(R_S + R_D)$ in the ideal equations. Assume transistor parameters of $V_T = 1$ V and $K_n = 1$ mA/V^2 (a) Plot the following curves on the same graph: I_D versus V_D for $V_G = 2$ V and $V_G = 3$ V over the range $0 \le V_D \le 5$ V for (*i*) $R_S = R_D = 0$ and (*ii*) $R_S = R_D = 1$ kΩ. (b) Plot the following curves on the same graph: $\sqrt{I_D}$ versus V_G for $V_D = 0.1$ V and $V_D = 5$ V over the range $0 \le I_D \le 1$ mA for (*i*) $R_S = R_D = 0$ and (*ii*) $R_S = R_D = 1$ kΩ.

11.35 An n-channel MOSFET has the same parameters as given in Problem 11.30. The gate terminal is connected to the drain terminal. Plot I_D versus V_{DS} for $0 \le V_{DS} \le 5$ V. Determine the range of V_{DS} over which the transistor is biased in the nonsaturation and saturation regions.

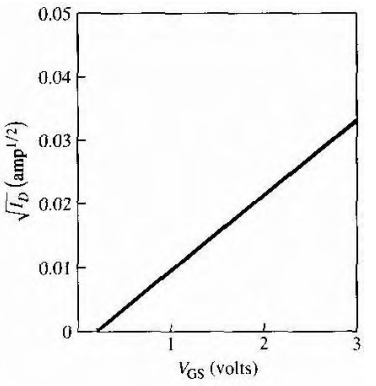


Figure 11.68 | Figure for Problem 11.37.

11.36 The channel conduclance for a p-channel MOSFET is defined as

$$g_d = \left.\frac{\partial I_D}{\partial V_{SD}}\right|_{V_{SD}\to 0}$$

Plot the channel conductance for the p-channel MOSFET in Problem 11.31 for $0 \le V_{SG} \le 5$ V.

1137 The experimental characteristics of an ideal n-channel MOSFET biased in the saturation region are shown in Figure 11.68. If $W/L = 10$ and $t_{ox} = 425$ Å, determine V_T and μ_n.

11.38 One curve of an n-channel MOSFET is characterized by the following parameters: $I_D(\text{sat}) = 2 \times 10^{-4}$ A, $V_{DS}(\text{sat}) = 4$ V, and $V_T = 0.8$ V.

(a) What is the gate voltage?
(*b*) What is the value of the conduction parameter?
(*c*) If $V_G = 2$ V and $V_{DS} = 2$ V, determine I_D.
(*d*) If $V_G = 3$ V and $V_{DS} = 1$ V, determine I_D
(e) For each of the conditions given in (c) and (d), sketch the inversion charge density and depletion region through the channel.

11.39 (*a*) An ideal n-channel MOSFET has an inversion carrier mobility $\mu_n = 525$ cm^2/V-s, a threshold voltage $V_T = +0.75$ V, and an oxide thickness $t_{ox} = 400$ Å. When biased in the saturation region, the required rated current is $I_D(\text{sat}) = 6$ mA when $V_{GS} = 5$ V. Determine the required W/L ratio. (*b*) Ap-channel MOSFET has the same requirements when $V_{GS} = 5$ V and has the same parameters as part (*a*) except $\mu_p = 300$ cm^2/V-s and $V_T = -0.75$ V. Determine the W/L ratio.

11.40 Consider the transistor described in Problem 11.30. ***(a)*** Calculate g_{mL} for $V_{DS} = 0.5$ V. (*b*) Calculate g_{ms} for $V_{GS} = 4$ V.

11.41 Consider the transistor described in Problem 11.31. (a) Calculate g_{mL} for $V_{SD} = 0.5$ V. (*b*) Calculate g_{ms} for $V_{SG} = 4$ V.

11.42 An n-channel MOSFET has the following parameters:

$$t_{ox} = 400 \text{ Å} \qquad N_a = 5 \times 10^{16} \text{ cm}^{-3}$$
$$V_{FB} = -0.5 \text{ V} \qquad L = 2 \ \mu\text{m}$$
$$W = 10 \ \mu\text{m} \qquad \mu_n = 450 \text{ cm}^2/\text{V-s}$$

Plot $\sqrt{I_D}$ versus V_{GS} over the range $0 \le I_D \le 1$ mA when the transistor is biased in the saturation region for the following values of source-to-body voltage: $V_{SB} = 0, 1, 2$ and 4 V.

11.43 Consider a p-channel MOSFET with $t_{ox} = 600$ Å and $N_d = 5 \times 10^{15} \text{ cm}^{-3}$. Determine the body-to-source voltage, V_{BS}, such that the shift in threshold voltage, ΔV_T, from the $V_{BS} = 0$ curve is $\Delta V_T = -1.5$ V.

11.44 An NMOS device has the following parameters: n^+ poly gate, $t_{ox} = 400$ A. $N_a = 10^{15} \text{ cm}^{-3}$, and $Q'_{ss} = 5 \times 10^{10} \text{ cm}^{-2}$. (*a*) Determine V_T. (*b*) Is it possible to apply a V_{SB} voltage such that $V_T = 0$? If so, what is the value of V_{SB}?

11.45 Investigate the threshold voltage shift due to substrate bias. The threshold shift is given by Equation (11.72). Plat ΔV_I versus V_{SB} over the range $0 \le V_{SB} \le 5$ V for several values of N_a and t_{ox}. Determine the conditions for which ΔV_I is limited to a maximum value of 0.7 V over the range of V_{SB}

Section 11.4 Frequency Limitations

11.46 Consider an ideal n-channel MOSFET with a width-to-length ratio of $(W/L) = 10$, an electron mobility of $\mu_n = 400 \text{ cm}^2/\text{V-s}$, an oxide thickness of $t_{ox} = 475$ Å, and a threshold voltage of $V_T = +0.65$ V. (a) Determine the maximum value of source resistance so that the saturation transconductance g_{ms} is reduced by no more than 20 percent from its ideal value when $V_{GS} = 5$ V. (*b*) Using the value of r, calculated in part (*a*), how much is g_{ms} reduced from its ideal value when $V_{GS} = 3$ V!

11.47 An n-channel MOSFET has the following parameters:

$$\mu_n = 400 \text{ cm}^2/\text{V-s} \qquad t_{ox} = 500 \text{ A}$$
$$L = 2 \ \mu\text{m} \qquad W = 20 \ \mu\text{m}$$
$$V_I = +0.75 \text{ V}$$

Assume the transistor is biased in the saturation region at $V_{GS} = 4$ V. (*a*) Calculate the ideal cutoff frequency. (b) Assume that the gate oxide overlaps both the source and drain contacts by 0.75 μm. If a load resistance of $R_L = 10 \text{ k}\Omega$ is connected to the output. calculate the cutoff frequency.

11.48 Repeat Problem 11.47 for the case when the electrons are traveling at a saturation velocity of $v_{sat} = 4 \times 10^6$ cm/s.

Summary and Review

***11.49** Design an ideal silicon n-channel MOSFET with a polysilicon gate to have a threshold voltage of $V_T = 0.65$ V Assume an oxide thickness of $t_{ox} = 300$ Å. a channel length of $L = 1.25 \ \mu$m, and a nominal value of $Q'_{ss} = 1.5 \times 10^{11} \text{ cm}^{-2}$. It is desired to have a drain current of $I_D = 50 \ \mu$A at $V_{GS} = 2.5$ V and $V_{DS} = 0.1$ V. Determine the substrate doping concentration, channel width. and type of gale required.

***11.50** Design an ideal silicon n-channel depletion mode MOSFET with a polysilicon gate to have a threshold voltage of $V_T = -0.65$ V. Assume an oxide thickness of $t_{ox} = 300$ Å, a channel length of L = 1.25 μm, and a nominal value of $Q'_{ss} = 1.5 \times 10^{11}$ cm^{-2}. It is desired to have a drain current of $I_D(\text{sat}) = 50\ \mu$A at $V_{GS} = 0$. Determine the type of gate, substrate doping concentration, and channel width required.

***11.51** Consider the CMOS inverter circuit shown in Figure 11.60a. Ideal n- and p-channel devices are to be designed with channel lengths of L = 2.5 μm and oxide thicknesses of $t_{ox} = 450$ Å. Assume the inversion channel mobilities are one-half the bulk values. The threshold voltages of the n and p-channel transistors are to be +0.5 V and −0.5 V, respectively. The drain current is to be $I_D = 0.256$ mA when the input voltage to the inverter is 1.5 V and 1.5 V with $V_{DD} = 5$ V. The gate material is to be the same in each device. Determine the type of gate, substrate doping concentrations, and channel widths.

***11.52** A complementary pair of ideal n-channel and p-channel MOSFETs are to be designed to produce the same *I–V* characteristics when they are equivalently biased. The devices are to have the same oxide thickness of 250 Å and the same channel length of L = 2 μm. Assume the SiO_2 layer is ideal. The n-channel device is to have a channel width of W = 20 μm. Assume constant inversion layer mobilities of $\mu_n = 600$ cm^2/V-s and $\mu_p = 220$ cm^2/V-s. (*a*) Determine p-type and n-type substrate doping concentrations. (*b*) What are the threshold voltages? (*c*) What is the width of the p-channel device?

READING LIST

1. Dimitrijev, S. *Understanding Semiconductor Devices.* New York: Oxford University Press. 2000.
2. Kano, K. *Semiconductor Devices.* Upper Saddle River, NJ: Prentice Hall, 1998.
3. Muller, R. S., and T. I. Kamins. *Device Electronics for Integrated Circuits.* 2nd ed. New York: Wiley, 1986.
4. Ng, K. K. *Complete Guide to Semiconductor Devices.* New York: McGraw-Hill, 1995.
5. Nicollian, E. H., and J. R. Brews. *MOS Physics and Technology.* New York: Wiley, 1982.
6. Ong, D. G. *Modern MOS Technology: Processes, Devices, and Design.* New York: McGraw-Hill, 1984.
7. Pirret. R. F. *Semiconductor Device Fundamentals.* Reading. MA: Addison-Wesley. 1996.
8. Roulston. D. J. *An Introduction to the Physics of Semiconductor Devices.* New York: Oxford University Press, 1999.
9. Schroder. D. K. *Advanced MOS Devices, Modular Series on Solid State Devices.* Reading. MA: Addison-Wesley. 1987.
10. Shur, M. *Introduction to Electronic Devices.* New York: John Wiley & Sons. Inc., 1996.
11. Shur. M. Physics *of Semiconductor Devices.* Englewood Cliffs. NJ: Prentice Hall, 1990.
12. Singh. J. *Semiconductor Devices: An Introduction.* New York: McGraw-Hill, 1994.
13. Singh, J. *Semiconductor Devices: Basic Principles.* New York: Wiley, 2001
14. Streetman, B. G.. and S. Banerjee. *Solid State Electronic Devices.* 5th ed. Upper Saddle River. NJ: Prentice Hall. 2000.

15. Sze, S. M. ***High-speed Semiconductor Devices.*** New York: Wiley. 1990.

16. Sze, S. M. ***Physics of Semiconductor Devices.*** 2nd ed. New York: Wiley, 1981.

*17. Taur, Y., and T. H. Ning. ***Fundamentals of Modem VLSI Devices.*** New York: Cambridge University Press, 1998.

*18. Tsividis, Y. ***Operation and Modeling of the MOS Transistor.*** 2nd ed. ***Burr*** Ridge, IL.: McGraw-Hill, 1999.

19. Werner, W. M. "The Work Function Difference of the MOS System with Aluminum Field Plates and Polycrystalline Silicon Field Plates." *Solid State Electronics* 17 (1974), pp. 769–75.

20. Yamaguchi, T., S. Monmoto, G. H. Kawamoto, and J. C. DeLacy. "Process and Device Performance of 1 μm-Channel n-Well CMOS Technology." ***IEEE Transactions on Electron Devices*** ED-31 (February 1984), pp. 205–14.

21. Yang, E. S. ***Microelectronic Devices.*** New York: McGraw-Hill, 1988.

APPENDIX



Selected List of Symbols

This list does not include some symbols that are defined and used specifically in only one section. Some symbols have more than one meaning; however, the context in which the symbol is used should make the meaning unambiguous. The usual unit associated with each symbol is given.

	Unit cell dimension (Å), potential well width, acceleration, gradient of impurity concentration, channel thickness of a one-sided JFET (cm)
a_0	Bohr radius (Å)
c	Speed of light (cm/s)
d	Distance (cm)
e	Electronic charge (magnitude) (C), Napierian base
f	Frequency (Hz)
$f_F(E)$	Fermi–Dirac probability function
f_T	Cutoff frequency (Hz)
g	Generation rate ($cm^{-3}\,s^{-1}$)
g'	Generation rate of excess carriers ($cm^{-3}\,s^{-1}$)
$g(E)$	Density of states function ($cm^{-3}\,eV^{-1}$)
g_c, g_v	Density of states function in the conduction band and valence band ($cm^{-3}\,eV^{-1}$)
g_d	Channel conductance (S), small-signal diffusion conductance (S)
g_m	Transconductance (A/V)
g_n, g_p	Generation rate for electrons and holes ($cm^{-3}\,s^{-1}$)
h	Planck's constant (J-s), induced space charge width in a JFET (cm)
$\hbar$	Modified Planck's constant ($h/2\pi$)

h_f	Small-signal common emitter current gain
j	Imaginary constant, $\sqrt{-1}$
k	Boltzmann's constant (J/K), wavenumber (cm^{-1})
k_n	Conduction parameter (A/V^2)
m	Mass (kg)
m_0	Rest mass of the electron (kg)
m^*	Effective mass (kg)
m_n^*, m_p^*	Effective mass of an electron and hole (kg)
n	Integer
n, l, m, s	Quantum numbers
n, p	Electron and hole concentration (cm^{-3})
$\bar{n}$	Index of refraction
n', p'	Constants related to the trap energy (cm^{-3})
n_{B0}, p_{E0}, p_{C0}	Thermal-equilibrium minority carrier electron concentration in the base and minority carrier hole concentration in the emitter and collector (cm^{-3})
n_d	Density of electrons in the donor energy level (cm^{-3})
n_i	Intrinsic concentration of electrons (cm^{-3})
n_0, p_0	Thermal-equilibrium concentration of electrons and holes (cm^{-3})
n_p, p_n	Minority carrier electron and minority carrier hole concentration (cm^{-3})
n_{p0}, p_{n0}	Thermal-equilibrium minority carrier electron and minority carrier hole concentration (cm^{-3})
n_s	Density of a two-dimensional electron gas (cm^{-2})
p	Momentum
p_a	Density of boles in the acceptor energy level (cm^{-3})
p_i	Intrinsic hole concentration ($= n_i$)(cm^{-3})
q	Charge (C)
r, θ, ϕ	Spherical coordinates
r_d, r_π	Small-signal diffusion resistance (Ω)
r_{ds}	Small-signal drain-to-source resistance (Ω)
s	Surface recombination velocity (cm/s)
t	Time (s)
t_d	Delay time (s)
t_{ox}	Gate oxide thickness (cm or A)
t_s	Storage time (s)
$u(x)$	Periodic wave function
v	Velocity (cm/s)
v_d	Carrier drift velocity (cm/s)

v_{ds}, v_s, v_{sat}	Carrier saturation drift velocity (cm/s)
x, y, z	Cartesian coordinates
x	Mole fraction in compound semiconductors
x_B, x_E, x_C	Neutral base, emitter, and collector region widths (cm)
x_d	Induced space charge width (cm)
x_{dT}	Maximum space charge width (cm)
x_n, x_p	Depletion width from the metallurgical junction into n-type and p-type semiconductor regions (cm)
A	Area (cm^2)
A^*	Effective Richardson constant ($A/K^2/cm^2$)
B	Magnetic flux density (Wb/m^2)
B, E, C	Base, emitter, and collector
BV_{CBO}	Breakdown voltage of collector-base junction with emitter open (volt)
BV_{CEO}	Breakdown voltage of collector-emitter with base open (volt)
C	Capacitance (F)
C'	Capacitance per unit area (F/cm^2)
C_d, C_π	Diffusion capacitance (F)
C_{FB}	Flat-band capacitance (F)
C_{gs}, C_{gd}, C_{ds}	Gate-source, gate-drain, and drain-source capacitance (F)
C_j'	Junction capacitance per unit area (F/cm^2)
C_M	Miller capacitance (F)
C_n, C_p	Constants related to capture rate of electrons and holes
C_{ox}	Gate oxide capacitance per unit area (F/cm^2)
C_μ	Reverse-biased B-C junction capacitance (F)
D, S, G	Drain, source, and gate of an FET
D'	Ambipolar diffusion coefficient (cm^2/s)
D_B, D_E, D_C	Base, emitter, and collector minority carrier diffusion coefficients (cm^2/s)
D_{it}	Density of intertace states ($\#/eV\text{-}cm^3$)
D_n, D_p	Minority carrier electron and minority carrier hole diffusion coefficient (cm^2/s)
E	Energy (joule or eV)
E_a	Acceptor energy level (eV)
E_c, E_v	Energy at the bottom edge of the conduction band and top edge of the valence band (eV)
$AE_{,}, AE,$	Difference in conduction band energies and valence band energies at a heterojunction (eV)
	Donor energy level (eV)

E_F	Fermi energy (eV)
E_{Fi}	Intrinsic Fermi energy (eV)
E_{Fn}, E_{Fp}	Quasi-Fermi energy levels for electrons and holes (eV)
E_g	Bandgap energy (eV)
ΔE_g	Bandgap narrowing factor (eV), difference in bandgap energies at a heterojunction (eV)
E_t	Trap energy level (eV)
F	Force (N)
F_n^-, F_p^+	Electron and hole particle flux ($cm^{-2}\,s^{-1}$)
$F_{1/2}(\eta)$	Fermi–Dirac integral function
G	Generation rate of electron-hole pairs ($cm^{-3}\,s^{-1}$)
G_L	Excess carrier generation rate ($cm^{-3}\,s^{-1}$)
G_{n0}, G_{p0}	Thermal equilibrium generation rate for electrons and holes ($cm^{-3}\,s^{-1}$)
G_{01}	Conductance (S)
I	Current (A)
I_A	Anode current (A)
I_B, I_E, I_C	Base, emitter, and collector current (A)
I_{CBO}	Reverse-bias collector-base junction current with emitter open (A)
I_{CEO}	Reverse-bias collector–emitter current with base open (A)
I_D	Diode current (A), drain current (A)
I_D(sat)	Saturation drain current (A)
I_L	Photocurrent (A)
I_{P1}	Pinchoff current (A)
I_S	Ideal reverse-bias saturation current (A)
I_{SC}	Short-circuit current (A)
I_ν	Photon intensity (energy/cm^2/s)
J	Electric current density (A/cm^2)
J_{gen}	Generation current density (A/cm^2)
J_L	Photocurrent density (A/cm^2)
J_n, J_p	Electron and hole electric current density (A/cm^2)
J_n^-, J_p^+	Electron and hole particle current density ($cm^{-2}\,s^{-1}$)
J_{rec}	Recombination current density (A/cm^2)
J_{r0}	Zero-bias recombination current density (A/cm^2)
J_R	Reverse-bias current density (A/cm^2)
J_S	Ideal reverse-bias saturation current density (A/cm^2)
J_{sT}	Ideal reverse saturation current density in a Schottky diode (A/cm^2)
L	Length (cm), inductance (H), channel length (cm)

ΔL	Channel length modulation factor (cm)
L_B, L_E, L_C	Minority carrier diffusion length in the base, emitter, and collector (cm)
L_D	Debye length (cm)
L_n, L_p	Minority carrier electron and hole diffusion length (cm)
M, M_n	Multiplication constant
N	Number density (cm^{-3})
N_a	Density of acceptor impurity atoms (cm^{-3})
N_B, N_E, N_C	Base, emitter, and collector doping concentrations (cm^{-3})
N_c, N_v	Effective density of states function in the conduction band and valence band (cm^{-3})
N_d	Density of donor impurity atoms (cm^{-3})
N_{it}	Interface state density (cm^{-2})
N_t	Trap density (cm^{-3})
P	Power (watt)
$P(r)$	Probability density function
Q	Charge (C)
Q'	Charge per unit area (C/cm^2)
Q_B	Gate controlled bulk charge (C)
Q'_n	Inversion channel charge density per unit area (C/cm^2)
Q'_{sig}	Signal charge density per unit area (C/cm^2)
$Q'_{SD}(max)$	Maximum space charge density per unit area (C/cm^2)
Q'_{SS}	Equivalent trapped oxide charge per unit area (C/cm^2)
R	Reflection coefficient, recombination rate ($cm^{-3}\,s^{-1}$), resistance (52)
$R(r)$	Radial wave function
R_c	Specific contact resistance (Ω-cm^2)
R_{cn}, R_{cp}	Capture rate for electrons and holes ($cm^{-3}\,s^{-1}$)
R_{en}, R_{ep}	Emission rate for electrons and holes ($cm^{-3}\,s^{-1}$)
R_n, R_p	Recombination rate for electrons and holes ($cm^{-3}\,s^{-1}$)
R_{n0}, R_{p0}	Thermal equilibrium recombination rate of electrons and holes ($cm^{-3}\,s^{-1}$)
	Temperature (K), kinetic energy (J or eV), transmission coefficient
V	Potential (volt), potential energy (J or eV)
V_a	Applied forward-bias voltage (volt)
V_A	Early voltage (volt), anode voltage (volt)
V_{bi}	Built-in potential barrier (volt)
V_B	Breakdown voltage (volt)
V_{BD}	Breakdown voltage at the drain (volt)

V_{BE}, V_{CB}, V_{CE}	Base-emitter, collector-base, and collector-emitter voltage (volt)
V_{DS}, V_{GS}	Drain-source and gate-source voltage (volt)
V_{DS}(sat)	Drain-source saturation voltage (volt)
V_{FB}	Flat-band voltage (volt)
V_G	Gate voltage (volt)
V_H	Hall voltage (volt)
V_{oc}	Open-circuit voltage (volt)
V_{ox}	Potential difference across an oxide (volt)
V_{p0}	Pinchoff voltage (volt)
V_{pt}	Punch-through voltage (volt)
V_R	Applied reverse-bias voltage (volt)
V_{SB}	Source-body voltage (volt)
V_t	Thermal voltage (kT/e)
V_T	Threshold voltage (volt)
ΔV_T	Threshold voltage shift (volt)
W	Total space charge width (cm), channel width (cm)
W_B	Metallurgical base width (cm)
Y	Admittance
α	Photon absorption coefficient (cm^{-1}), ac common base current gain
α_n, α_p	Electron and hole ionization rates (cm^{-1})
α_0	dc common base current gain
α_T	Base transport factor
β	Common-emitter current gain
γ	Emitter injection efficiency factor
δ	Recombination factor
δn, δp	Excess electron and hole concentration (cm^{-3})
δn_p, δp_n	Excess minority carrier electron and excess minority carrier hole concentration (cm^{-3})
ϵ	Permittivity (F/cm^2)
ϵ_0	Permittivity of free space (F/cm^2)
ϵ_{ox}	Permittivity of an oxide (F/cm^2)
ϵ_r	Relative permittivity or dielectric constant
ϵ_s	Permittivity of a semiconductor (F/cm^2)
λ	Wavelength (cm or μm)
μ	Permeability (H/cm)
μ'	Ambipolar mobility (cm^2/V-s)
μ_n, μ_p	Electron and hole mobility (cm^2/V-s)

	Permeability of free space (H/cm)
	Frequency (Hz)
	Resistivity (R-cm), volume charge density (C/cm^3)
σ	Conductivity ($\Omega^{-1}\,cm^{-1}$)
$\Delta\sigma$	Photoconductivity ($\Omega^{-1}\,cm^{-1}$)
σ_i	Intrinsic conductivity ($\Omega^{-1}\,cm^{-1}$)
σ_n, σ_p	Conductivity of n-type and p-type semiconductor ($\Omega^{-1}\,cm^{-1}$)
τ	Lifetime (s)
τ_n, τ_p	Electron and hole lifetime (s)
τ_{n0}, τ_{p0}	Excess minority carrier electron and hole lifetime (s)
τ_0	Lifetime in space charge region (s)
ϕ	Potential (volt)
$\phi(t)$	Time-dependent wave function
$\Delta\phi$	Schottky barrier lowering potential (volt)
ϕ_{Bn}	Schottky barrier height (volt)
ϕ_{B0}	Ideal Schottky barrier height (volt)
ϕ_{fn}, ϕ_{fp}	Potential difference (magnitude) between E_{Fi} and E_F in n-type and p-type semiconductor (volt)
ϕ_{Fn}, ϕ_{Fp}	Potential difference (with sign) between E_{Fi} and E_F in n-type and p-type semiconductor (volt)
ϕ_m	Metal work function (volt)
ϕ'_m	Modified metal work function (volt)
ϕ_{ms}	Metal-semiconductor work function difference (volt)
ϕ_n, ϕ_p	Potential difference (magnitude) between E_c and E_F in n-type and between E_v and E_F in p-type semiconductor (volt)
ϕ_s	Semiconductor work function (volt). surface potential (volt)
χ	Electron affinity (volt)
χ'	Modified electron affinity (volt)
$\psi(x)$	Time-independent wave function
ω	Radian frequency (s^{-1})
Γ	Reflection coefficient
E	Electric field (V/cm)
E_H	Hall electric field (V/cm)
E_{crit}	Critical electric field at breakdown (V/cm)
$\Theta(\theta)$	Angular wave function
Φ	Photon flux ($cm^{-2}\,s^{-1}$)
$\Phi(\phi)$	Angular wave function
$\Psi(x, t)$	Total wave function

APPENDIX B

System of Units, Conversion Factors, and General Constants

Table B.1 | International system of units*

Quantity	Unit	Symbol	Dimension
Length	meter	m	
Mass	kilogram	kg	
Time	second	s or sec	
Temperature	kelvin	K	
Current	ampere	A	
Frequency	hertz	Hz	1/s
Force	newton	N	kg-m/s^2
Pressure	pascal	Pa	N/m^2
Energy	joule	J	N-m
Power	watt	W	J/s
Electric charge	coulomb	C	A-s
Potential	volt	V	J/C
Conductance	siemens	S	AN
Resistance	ohm	Ω	V/A
Capacitance	farad	F	C N
Magnetic flux	weber	Wb	V-s
Magnetic flux density	tesla	T	Wb/m^2
Inductance	henry	H	Wb/A

*The cm is the common unit of length and the electron-volt is the common unit of energy (*see* Appendix F) used in the study of semiconductors. However. the joule and in some cases the meter should be used in moat formulas.

Table B.2 | Conversion factors

	Prefixes		
1 Å (angstrom) = 10^{-8} cm = 10^{-10} m	10^{-15}	femto-	= f
1 μm (micron) = 10^{-4} cm	10^{-12}	pico-	= p
1 mil = 10^{-3} in. = 25.4 μm	10^{-9}	nano-	= n
2.54 cm = 1 in.	10^{-6}	micro-	= μ
1 eV = 1.6 x 10^{-19} J	10^{-3}	milli-	= m
1 J = 10^{7} erg	10^{+3}	kilo-	= k
	10^{+6}	mega-	= M
	10^{+9}	giga-	= G
	10^{+12}	tera	= T

Table B.3| Physical constants

Avogadro's number	$N_A = 6.02 \times 10^{+23}$ atoms per gram molecular weight
Boltzmann's constant	$k = 1.38 \times 10^{-23}$ J/K $= 8.62 \times 10^{-5}$ eV/K
Electronic charge (magnitude)	$e = 1.60 \times 10^{-19}$ C
Free electron rest mass	$m_0 = 9.11 \times 10^{-31}$ kg
Permeability of free space	$\mu_0 = 4\pi \times 10^{-7}$ Wm
Permittivity of free space	$\epsilon_0 = 8.85 \times 10^{-14}$ F/cm $= 8.85 \times 10^{-12}$ F/m
Planck's constant	$h = 6.625 \times 10^{-34}$ J-s $= 4.135 \times 10^{-15}$ eV-s $\frac{h}{2\pi} = \hbar = 1.054 \times 10^{-34}$ J-s
Proton rest mass	$M = 1.67 \times 10^{-27}$ kg
Speed of light in vacuum	$c = 2.998 \times 10^{10}$ cm/s
Thermal voltage (T = 300 K)	$V_t = \frac{kT}{e} = 0.0259$ volt $kT = 0.0259$ eV

Table B.4 | Silicon, gallium arsenide, and germanium properties ($T = 300$ K)

Property	Si	GaAs	Ge
Atoms (cm^{-3})	5.0×10^{22}	4.42×10^{22}	4.42×10^{22}
Atomic weight	28.09	144.63	72.60
Crystal structure	Diamond	Zincblende	Diamond
Density (g/cm^{-3})	2.33	5.32	5.33
Lattice constant (Å)	5.43	5.65	5.65
Melting point ('C)	1415	1238	937
Dielectric constant	11.7	13.1	16.0
Bandgap energy (eV)	1.12	1.42	0.66
Electron affinity. χ (volts)	4.01	4.07	4.13
Effective density of states in conduction band, N_c (cm^{-3})	2.8×10^{19}	4.7×10^{17}	1.04×10^{19}
Effective density of states in valence band. N, (cm^{-3})	1.04×10^{19}	7.0×10^{18}	6.0×10^{18}
Intrinsic carrier concentration (cm^{-3})	1.5×10^{10}	1.8×10^{6}	2.4×10^{13}
Mobility (cm^2/V-s)			
Electron, μ_n	1350	8500	3900
Hole, μ_p	480	400	1900
Effective mass $\left(\frac{m^*}{m_0}\right)$			
Electrons	$m_l^* = 0.98$	0.067	1.64
	$m_t^* = 0.19$		0.082
Holes	$m_{lh}^* = 0.16$	0.082	0.044
	$m_{hh}^* = 0.49$	0.45	0.28
Effective mass (density of states)			
Electrons $\left(\frac{m_n^*}{m_0}\right)$	1.08	0.067	0.55
Holes $\left(\frac{m_p^*}{m_0}\right)$	0.56	0.48	**0.37**

Table B.5 | Other semiconductor parameters

Material	E_g(eV)	a(Å)	ϵ_r	χ	n
Aluminum arsenide	2.16	5.66	12.0	3.5	2.97
Gallium phosphide	2.26	5.45	10	4.3	3.37
Aluminum phosphide	2.43	5.46	9.8		3.0
Indium phosphide	1.35	5.87	12.1	4.35	3.37

Table R.6 | Properties of SiO_2 and Si_3N_4 ($T = 300\,K$)

Property	SiO_2	Si_3N_4
Crystal structure	[Amorphous for most integrated circuit applications]	
Atomic or molecular density (cm^{-3})	2.2×10^{22}	1.48×10^{22}
Density ($g\text{-}cm^{-3}$)	2.2	**3.4**
Energy gap	≈ 9 eV	4.7 eV
Dielectric constant	3.9	7.5
Melting point (°C)	≈1700	≈1900

APPENDIX C

The Periodic Table

Period	Group I a	Group I b	Group II a	Group II b	Group III a	Group III b	Group IV a	Group IV b	Group V a	Group V b	Group VI a	Group VI b	Group VII a	Group VII b	Group VIII a			Group VIII b
I	1 H																	2 He
	1.0079																	*4.003*
II	3 Li		4 Be		5 B		6 C		7 N		8 O		9 F					10 Ne
	6.94		9.02		10.82		12.01		14.01		16.00		19.00					20.18
III	*11 Na*		*12 Mg*		*13 Al*		*14 Si*		*15 P*		*16 S*		*17 Cl*					*18 Ar*
	22.99		24.32		26.97		28.06		30.98		32.06		35.45					39.94
	19 K		20 Ca		21 Sc		22 Ti		23 V		24 Cr		25 Mn		26 Fe	27 Co	28 Ni	
IV	39.09		40.08		44.96		47.90		50.95		52.01		54.93		55.85	58.94	58.69	
		29 Cu		30 Zn		31 Ga		32 Ge		33 As		34 Se		35 Br				36 Kr
		63.54		65.38		69.72		72.60		74.91		78.96		79.91				83.7
	37 Rb		38 Sr		39 Y		40 Zr		41 Nb		42 Mo		43 Tc		44 Ru	45 Rh	46 Pd	
V	85.48		87.63		88.92		91.22		92.91		95.95		99		101.7	102.91	106.4	
		47 Ag		48 Cd		49 In		50 Sn		51 Sb		52 Te		53 I				54 Xe
		107.88		112.41		114.76		118.70		121.76		127.61		126.92				131.3
	55 Cs		56 Ba		57–71		72 Hf		73 Ta		74 W		75 Re		76 Os	77 Ir	28 Pt	
VI	132.91		137.36		*Rare earths*		178.6		180.88		183.92		186.31		190.2	193.1	**195.2**	
		79 Au		80 Hg		81 Tl		82 Pb		83 Bi		84 Po		85 At				86 Rn
		197.2		200.61		204.39		207.21		209.00		210		211				222
VII	87 Fr		88 Ra		89 Ac		90 Th		91 Pa		92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md
	223		226.05		227		232.12		231		238.07	237	239	241	242	246	249	254	256	256

Rare earths

VI	57 La	58 Ce	59 R	60 Nd	61 Pm	62 Sm	63 Eu	M Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
57-71	138.92	140.13	140.92	144.27	147	150.43	152.0	156.9	159.2	162.46	164.90	167.2	169.4	173.04	174.99

The numbers in front of the symbols of the elements denote the atomic numbers; the numbers underneath are the atomic weights.

A P P E N D I X

The Error Function

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$$

$$\operatorname{erf}(0) = 0 \qquad \operatorname{erf}(\infty) = 1$$

$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$$

z	erf(z)	z	erf(z)
0.00	0.00000	1.00	0.84270
0.05	0.05637	1.05	0.86244
0.10	0.11246	1.10	0.88021
0.15	0.16800	1.15	0.89612
0.20	0.22270	1.20	0.91031
0.25	0.27633	1.25	0.92290
0.30	0.32863	1.30	0.93401
0.35	0.37938	1.35	0.94376
0.40	0.42839	1.40	0.95229
0.45	0.47548	1.45	0.95970
0.50	0.52050	1.50	0.96611
0.55	0.56332	1.55	0.97162
0.60	0.60386	1.60	0.97635
0.65	0.64203	1.65	0.98038
0.70	0.67780	1.70	0.98379
0.75	0.71116	1.75	0.98667
0.80	0.74210	1.80	0.98909
0.85	0.77067	1.85	0.99111
0.90	0.79691	1.90	0.99279
0.95	0.82089	1.95	0.99418
1.00	0.84270	2.00	0.99532

APPENDIX

E

"Derivation" of Schrodinger's Wave Equation

Schrodinger's wave equation was stated in Equation *(2.6)*.The time-independent form of Schrodinger's wave equation was then developed and given by Equation *(2.13)*.The time-independent Schrodinger's wave equation can also be developed from the classical wave equation. We may think of this development more in terms of a justification of the Schrodinger's time-independent wave equation rather than a strict derivation.

The time-independent classical wave equation, in terms of voltage, is given as

$$\frac{\partial^2 V(x)}{\partial x^2} + \left(\frac{\omega^2}{v_p^2}\right) V(x) = 0 \tag{E.1}$$

where ω is the radian frequency and v_p is the phase velocity.

If we make a change of variable and let $\psi(x) = V(x)$, then we have

$$\frac{\partial^2 \psi(x)}{\partial x^2} + \left(\frac{\omega^2}{v_p^2}\right) \psi(x) = 0 \tag{E.2}$$

We can write that

$$\frac{\omega^2}{v_p^2} = \left(\frac{2\pi v}{v_p}\right)^2 = \left(\frac{2\pi}{\lambda}\right)^2 \tag{E.3}$$

where v and A are the wave frequency and wavelength, respectively.

From the wave-particle duality principle, we can relate the wavelength and momentum as

$$\lambda = \frac{h}{p} \tag{E.4}$$

Then

$$\left(\frac{2\pi}{\lambda}\right)^2 = \left(\frac{2\pi}{h} \cdot p\right)^2 \tag{E.5}$$

and since $\hbar = \frac{h}{2\pi}$, we can write

$$\left(\frac{2\pi}{\lambda}\right)^2 = \left(\frac{p}{\hbar}\right)^2 = \frac{2m}{\hbar^2}\left(\frac{p^2}{2m}\right) \tag{E.6}$$

Now

$$\frac{p^2}{2m} = T = E - V \tag{E.7}$$

where T, E, and V are the kinetic energy, total energy, and potential energy terms. respectively.

We can then write

$$\frac{\omega^2}{v_p^2} = \left(\frac{2\pi}{\lambda}\right)^2 = \frac{2m}{\hbar^2}\left(\frac{p^2}{2m}\right) = \frac{2m}{\hbar^2}(E - V) \tag{E.8}$$

Substituting Equation (E.8) into Equation (E.2), we have

$$\frac{\partial^2\psi(x)}{\partial x^2} + \frac{2m}{\hbar^2}(E - V)\psi(x) = 0 \tag{E.9}$$

which is the one-dimensional, time-independent Schrodinger's wave equation

APPENDIX F

Unit of Energy—The Electron-Volt

The electron-volt (eV) is a unit of energy that is used constantly in the study of semiconductor physics and devices. This short discussion may help in "petting a feel" for the electron-volt.

Consider a parallel plate capacitor with an applied voltage as shown in Figure F.1. Assume that an electron is released at $x = 0$ at time $t = 0$. We may write

$$F = m_0 a = m_0 \frac{d^2x}{dt^2} = eE \tag{F.1}$$

where e is the magnitude of the electronic charge and E is the magnitude of the electric field as shown. Upon integrating, the velocity and distance versus time are given by

$$v = \frac{eEt}{m_0} \tag{F.2}$$

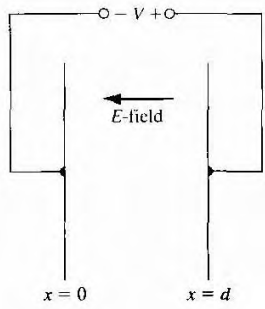


Figure F.1 | Parallel plate capacitor.

and

$$x = \frac{eEt^2}{2m_0} \tag{F.3}$$

where we have assumed that $v = 0$ at $t = 0$.

Assume that at $t = t_0$ the electron reaches the positive plate of the capacitor so that x = d. Then

$$d = \frac{eEt_0^2}{2m_0} \tag{F.4a}$$

or

$$t_0 = \sqrt{\frac{2m_0 d}{eE}} \tag{F.4b}$$

The velocity of the electron when it reaches the positive plate of the capacitor is

$$v(t_0) = \frac{eEt_0}{m_0} = \sqrt{\frac{2eEd}{m_0}} \tag{F.5}$$

The kinetic energy of the electron at this time is

$$T = \frac{1}{2}m_0 v(t_0)^2 = \frac{1}{2}m_0\left(\frac{2eEd}{m_0}\right) = eEd \tag{F.6}$$

The electric field is

$$E = \frac{V}{d} \tag{F.7}$$

so that the energy is

$$T = e \cdot V \tag{F.8}$$

If an electron is accelerated through a potential of 1 volt, then the energy is

$$T = e \cdot V = (1.6 \times 10^{-19})(1) = 1.6 \times 10^{-19} \text{ joule} \tag{F.9}$$

The electron-volt (eV) unit of energy is defined as

$$\text{Electron-volt} = \frac{\text{joule}}{e} \tag{F.10}$$

Then, the electron that is accelerated through a potential of 1 volt will have an energy of

$$T = 1.6 \times 10^{-19}\,J = \frac{1.6 \times 10^{-19}}{1.6 \times 10^{-19}}(\text{eV}) \tag{F.11}$$

or 1 eV.

We may note that the magnitude of the potential (1 volt) and the magnitude of the electron energy (1 eV) are the same. However, it is important to keep in mind that the unit associated with each number is different.

APPENDIX J

ANSWERS TO SELECTED PROBLEMS

Chapter **1**

1.1 (*a*)4 atoms. (h) 2 atoms, (*c*) 8 atoms

1.3 (*a*) 52.4 percent, (h) 74 percent. (c) 68 percent, (*d*) 34 percent

1.5 (*a*) 2.36 Å. (*b*) 5 x 10^{22} atoms/cm^3

1.7 (b) a = 2.8 Å, ***(c)*** 2.28 x 10^{22} cm^{-3} for both Na and Cl, (*d*) 2.21 gm/cm^3

1.9 (*a*) 3.31 x 10^{14} atoms/cm^2; Same for A atoms and B atoms. ***(b)*** Same as ***(a)***.(*c*) Same material.

1.13 (*a*) 5.63 Å, (*h*)3.98 Å. (*c*) 3.25 Å

1.15 (*a*) 6.78 x 10^{14} cm^{-2}, (b) 9.59 x 10^{14} cm^{-2}, (*c*) 7.83 x 10^{14} cm^{-2}

1.17 2 x 10^{23} cm^{-3}

1.19 (*a*) 4 x 10^{-5} percent. (*b*) 2 x 10^{-6} percent

1.21 d = 7.94 x 10^{-6} cm or d/a_0 = 146

Chapter **2**

2.5 $\lambda = 0.254\ \mu$m (gold). $\lambda = 0.654\ \mu$m (cesium)

2.7 $E_{av} = 0.01727$ eV,
$P_{av} = 7.1 \times 10^{-26}$ kg-m/s, $\lambda = 93.3$ A

2.9 (*a*) $E = 1.14 \times 10^{-3}$ eV,
$p = 1.82 \times 10^{-26}$ kg-m/s, A = 364 Å
(b) $p = 5.3 \times 10^{-26}$ kg-m/s,
$v = 5.82 \times 10^{6}$ cm/s, $E = 9.64 \times 10^{-3}$ eV

2.11 ***(a)*** Ap = 1.054 x 10^{-28} kg-m/s
(*b*) ***AE*** = 0.198 eV

2.13 (*a*) ***Ap*** = 8.78 x 10^{-26} kg-m/s
(*b*) ***AE*** = 4.82 x 10^{-7} eV

2.15 (*a*) Ap = 1.054 x 10^{-24} kg-m/s
(*b*) ***Ar*** = 6.6 x 10^{-16} s

2.17 $|A|^2 = 1$, or $A = +1, -1, +j, -j$

2.19 (*a*) P = 0.393. (hi P = 0.239, (*c*) P = 0.865

2.21 $\Psi(x, t) = A \exp[-j(kx + \omega t)]$ where
k = 6.27 x 10^{8} m^{-1} and
$\omega = 2.28 \times 10^{13}$ rad/s

2.23 (*a*) $E_1 = 0.261$ eV, $E_2 = 1.04$ eV,
(*b*) A = 1.59 μm

2.25 $E_1 = 2.06 \times 10^{6}$ eV (neutron).
$E_1 = 3.76 \times 10^{9}$ eV (electron)

2.29 (*b*) *(i)* A E = 3.85 x 10^{-3} eV,
(ii) ***AE*** = 2.46 x 10^{-17} eV

2.31 **(a) P** = 0.118 percent.
(*b*) P = 1.9 x 10^{-10} percent

2.33 ***(a)*** **T** = 0.138, (*b*) T = 1.27 x 10^{-5}

2.38 $E_1 = -13.58$ eV, $E_2 = -3.395$ eV,
$E_3 = -1.51$ eV, $E_4 = -0.849$ eV

Chapter **3**

3.9 ***(a)*** ***AE*** = 0.488 eV, (b) AE = 1.87 eV,
(c) ***AE*** = 3.83 eV, (*d*) ***AE*** = 6.27 eV

3.11 (*a*) ***AE*** = 0.638 eV, (*b*) ΔE = 2.36 eV,
(c) ***AE*** = 4.73 eV, (d) ***AE*** = 7.39 eV

3.13 m* (A)< m^* (B)

3.15 A, B: velocity = −x; C, D: velocity = +x;
B. C: positive mass; A, D: negative mass

3.17 A: $m/m_0 = 0.476$; B: $m/m_0 = 0.0953$

3.23 g = 3.28 x 10^{17} cm^{-3}

3.25 (*a*) At E_c, $g_c = 0$;
0.05 eV, $g_c = 1.71 \times 10^{21}$ cm^{-3} eV^{-1};
0.10 eV, $g_c = 2.41 \times 10^{21}$ cm^{-3} eV^{-1};

0.15 eV, $g_c = 2.96 \times 10^{21}$ cm^{-3} eV^{-1};
0.20 eV, $g_c = 3.41 \times 10^{21}$ cm^{-3} eV^{-1},
(*b*) At E_v, $g_v = 0$,
−0.05 eV, $g_v = 0.637 \times 10^{21}$ cm^{-3} eV^{-1};
−0.10 eV, $g_v = 0.901 \times 10^{21}$ cm^{-3} eV^{-1};
−0.15 eV, $g_v = 1.10 \times 10^{21}$ cm^{-3} eV^{-1};
−00.20 eV, $g_v = 1.27 \times 10^{21}$ cm^{-3} eV^{-1}

3.29 (*a*) $f(E) = 0.269$, (*b*) $1 - f(E) = 0.269$

3.31 (*a*) $1 - f(E) = 0.269$,
(*b*) $1 - f(E) = 6.69 \times 10^{-3}$,
(*c*) $1 - f(E) = 4.54 \times 10^{-5}$

3.37 (*a*) $f(E) = 6.43 \times 10^{-3}$ percent,
(b) $f(E) = 4.53$ percent, (*c*) $T \approx 756$ K

3.39 (*a*) Far $E = E_1$, $f(E) = 9.3 \times 10^{-6}$;
For $E = E_2$, $1 - f(E) = 1.78 \times 10^{-14}$,
(b) For $E = E_1$, $f(E) = 8.45 \times 10^{-13}$;
For $E = E_2$, $1 - f(E) = 1.Y6 \times 10^{-7}$

3.43 T ≈ 461 K

Chapter 4

4.1 (*a*) $n_i = 7.68 \times 10^4$ cm^{-3}; 2.38×10^{12} cm^{-3}; 9.74×10^{14} cm^{-3},
(*b*) $n_i = 2.16 \times 10^{10}$ cm^{-3}; 8.60×10^{14} cm^{-3}; 3.82×10^{16} cm^{-3},
(*c*) $n_i = 1.38$ cm^{-3}; 3.28×10^9 cm^{-3}; 5.72×10^{12} cm^{-3}

4.5 (*a*) $E = E_c + \frac{kT}{2}$ (*b*) $E = E_v - \frac{kT}{2}$

4.8 $\frac{n_i(A)}{n_i(B)} = 47.5$

4.10 $E_{Fi} - E_{\text{midgap}} = -0.0128$ eV (Si)
$E_{Fi} - E_{\text{midgap}} = -0.0077$ eV (Ge)
E $- E_{\text{midgap}} = +0.038$ eV (GaAs)

4.12 $E_{Fi} - E_{\text{midgap}} = -8.51$ meV, −17.0 meV, −25.5 meV

4.17 $r_1 = 104$ A, E = 0.0053 eV

4.19 $p_0 = 2.13 \times 10^{15}$ cm^{-3}, $n_0 = 2.27 \times 10^4$ cm^{-3}

4.21 $E_c - E_F = 0.88$ eV, $n_0 = 4.9 \times 10^4$ cm^{-3}

4.23 (a) $p_0 = 1.33 \times 10^{12}$ cm^{-3},
(*b*) $E_{Fi} - E_F = 0.207$ eV,
(c) For (a) $n_0 = 2.44$ cm^{-3};
For (*b*) $n_0 = 8.09 \times 10^6$ cm^{-3}

4.25 $E_c - E_F = -0.034$ eV

4.27 (*a*) $n_0 = 2.45 \times 10^{11}$ cm^{-3},
$p_0 = 9.12 \times 10^{16}$ cm^{-3},
(b) $n_0 = 1.87 \times 10^{16}$ cm^{-3},
$p_0 = 9.20 \times 10^{16}$ cm^{-3}

4.29 (*a*) $p_0 = 2.95 \times 10^{11}$ cm^{-3}.
$n_0 = 1.95 \times 10^{13}$ cm^{-3}, (h) $n_0 = 5 \times 10^{15}$ cm^{-3},
$p_0 = 1.15 \times 10^{11}$ cm^{-3}

4.31 (*a*) $n_0 = 2 \times 10^{15}$ cm^{-3}, $p_0 = 1.125 \times 10^5$ cm^{-3},
(b) $p_0 = 10^{16}$ cm^{-3}, $n_0 = 2.25 \times 10^4$ cm^{-3},
(*c*) $n_0 = p_0 = n_i = 1.5 \times 10^{10}$ cm^{-3},
(*d*) $p_0 = 1.0 \times 10^{14}$ cm^{-3},
$n_0 = 5.66 \times 10^{10}$ cm^{-3},
(*e*) $n_0 = 1.49 \times 10^{14}$ cm^{-3},
$p_0 = 4.89 \times 10^{13}$ cm^{-3}

4.33 (*a*) p type, (h) Si: $p_0 = 1.5 \times 10^{13}$ cm^{-3},
$n_0 = 1.5 \times 10^7$ cm^{-3},
Ge: $p_0 = 3.26 \times 10^{13}$ cm^{-3},
$n_0 = 1.77 \times 10^{13}$ cm^{-3}, GaAs:
$p_0 = 1.5 \times 10^{13}$ cm^{-3}, $n_0 = 0.216$ cm^{-3}

4.35 $n_0 = 1.125 \times 10^{15}$ cm^{-3}, n-type

4.41 (*a*) n type; $n_0 = 10^{16}$ cm^{-3},
$p_0 = 2.25 \times 10^4$ cm^{-3}, (*b*) p-type;
$p_0 = 2.8 \times 10^{16}$ cm^{-3}, $n_0 = 8.04 \times 10^3$ cm^{-3}

4.43 $T = 200$ K $\Rightarrow E_{Fi} - E_F = 0.1855$ eV
T = 400 K $\Rightarrow E_{Fi} - E_F = 0.01898$ eV
T = 600 K $\Rightarrow E_{Fi} - E_F = 0.000674$ eV

4.45 $T \approx 762$ K

4.49 $N_d = 1.2 \times 10^{16}$ cm^{-3}

4.51 (a) $E_F - E_{Fi} = 0.2877$ eV,
(b) $E_{Fi} - E_F = 0.2877$ eV. (*c*) For (*a*) $n_0 = 10^{15}$ cm^{-3}, For (*h*) $n_0 = 2.25 \times 10^5$ cm^{-3}

4.53 (*a*) $E_F - E_{Fi} = 0.3056$ eV,
(*b*) $E_{Fi} - E_F = 0.3473$ eV,
(*c*) $E_F = E_{Fi}$. (*d*) $E_{Fi} - E_t = 0.1291$ eV,
(*e*) $E_t - E_{Fi} = 0.0024$ eV

4.55 p type, $E_{Fi} - E_t = 0.3294$ eV

Chapter 5

5.1 (*a*) $n_0 = 10^{16}$ cm^{-3}, $p_0 = 3.24 \times 10^{-4}$ cm^{-3}
(b) $\mu_n \approx 7500$ cm^2/V-s so
$J = 120$ A/cm^2, (*c*) (*i*) $p_0 = 10^{16}$ cm^{-3},
$n_0 = 3.24 \times 10^{-4}$ cm^{-3};
(ii) $\mu_p \approx 310$ cm^2/V-s
so $J = 4.96$ A/cm^2

5.3 (a) $I = 0.44$ mA, (*b*) I = 4.4 mA,
(*c*) For (*a*) $v_d = 5.5 \times 10^7$ cm/s,
For (b) $v_d = 5.5 \times 10^5$ cm/s

5.5 (*a*) $\mu_n = 3333$ cm^{2}N-s.
(*h*) $v_d = 2.4 \times 10^7$ cm/s

5.7 (*a*) $\sigma_i = 4.39 \times 10^{-6}$ (Ω-cm)$^{-1}$,
(b) $\sigma_i = 1.03 \times 10^{-6}$ (Ω-cm)$^{-1}$

5.9 $n_i(300\text{ K}) = 3.91 \times 10^9\text{ cm}^{-3}$, $E_g = 1.122\text{ eV}$; $n_i(500\text{ K}) = 2.27 \times 10^{13}\text{ cm}^{-3}$, $\sigma(500\text{ K}) = 5.81 \times 10^{-3}\ (\Omega\text{-cm})^{-1}$

5.11 (a) $N_d = 9.26 \times 10^{14}\text{ cm}^{-3}$, (b) $\rho(200\text{ K}) \approx 2.7\ \Omega\text{-cm}$, $p \approx 9.64\ \Omega\text{-cm}$

5.13 (a) $T = 5.6 \times 10^{-8}$ eV, (b) $T = 5.6 \times 10^{-4}$ eV

5.17 $\mu = 316\text{ cm}^2/\text{V-s}$

5.19 $\mu = 167\text{ cm}^2/\text{V-s}$

5.23 $I = 18\text{ mA}$

5.25 $J = 16\text{ A/cm}^2$

5.27 $J_n = 3.41 \exp\left(\dfrac{-x}{22.5}\right)\text{ A/cm}^2$

5.29 (a) $J_{h,diff} = 1.6 \exp\left(\dfrac{-x}{L}\right)\text{ A/cm}^2$,

(b) $J_{e,drf} = 4.8 - 1.6 \exp\left(\dfrac{-x}{L}\right)\text{ A/cm}^2$

(c) $E = \left[3 - 1 \cdot \exp\left(\dfrac{-x}{L}\right)\right]\text{ V/cm}$

5.31 (a) $n = n_i \exp\left[\dfrac{0.4 - 2.5 \times 10^2 x}{kT}\right]$,

(b) $J_n = -5.79 \times 10^{-4} \exp\left[\dfrac{0.4 - 2.5 \times 10^2 x}{0.0259}\right]$,

(i) $J_n(x = 0) = -2.95 \times 10^3\text{ A/cm}^2$,

(ii) $J_n(x = 5\ \mu\text{m}) = -23.7\text{ A/cm}^2$

5.33 (a) $E = \alpha\left(\dfrac{kT}{e}\right)$, (b) $V = -\left(\dfrac{kT}{e}\right)$

5.35 $N_d(x) = A \exp(-\alpha x)$ where $\alpha = 3.86 \times 10^4\text{ cm}^{-1}$

5.39 (a) $V_H = 2.19\text{ mV}$, (b) $E_H = 0.219\text{ V/cm}$

5.41 (a) p type, (b) $p = 8.08 \times 10^{15}\text{ cm}^{-3}$, (b) $\mu_p = 387\text{ cm}^2/\text{V-s}$

5.43 (a) n type, (b) $n = 8.68 \times 10^{14}\text{ cm}^{-3}$, (c) $\mu_n = 8182\text{ cm}^2/\text{V-s}$, (d) $\rho = 0.88\ \Omega\text{-cm}$

Chapter 6

6.1 $R' = 5 \times 10^{19}\text{ cm}^{-3}\text{ s}^{-1}$

6.3 (a) $\tau_{n0} = 8.89 \times 10^{-6}$ s, (b) $G = 1.125 \times 10^9\text{ cm}^{-3}\text{ s}^{-1}$, (c) $G = R = 1.125 \times 10^9\text{ cm}^{-3}\text{ s}^{-1}$

6.7 $\dfrac{\partial F_p^+}{\partial x} = -2 \times 10^{19}\text{ cm}^{-3}\text{ s}^{-1}$

6.9 $D' = 58.4\text{ cm}^2/\text{s}$, $\mu' = -868\text{ cm}^2/\text{V-s}$, $\tau_{n0} = 54\ \mu\text{s}$, $\tau_{p0} = 24\ \mu\text{s}$

6.11 $a = 8 + 0.114(1 - e^{-t/\tau_{p0}})$, $\tau_{p0} = 10^{-7}$ s

6.13 $I = (54 + 2.20e^{-t/\tau_{p0}})\text{ mA}$, $\tau_{p0} = 3 \times 10^{-7}$ s

6.15 (a) $R'_p/R_{p0} = 4.44 \times 10^9$, (b) $\tau_{p0} = 2.25 \times 10^{-7}$ s

6.17 (a) For $0 < t < 2 \times 10^{-6}$, $\delta n = 10^{14}(1 - e^{-t/\tau_{n0}})$ where $\tau_{n0} = 10^{-6}$ s; For $t > T = 2 \times 10^{-6}$, $\delta n = 0.865 \times 10^{14} \exp\left[\dfrac{-(t - \quad)}{\tau_{n0}}\right]$

6.19 (a) $n_{p0} = 2.25 \times 10^6\text{ cm}^{-3}$, (b) $\delta n(0) = -n_{p0} = -2.25 \times 10^6\text{ cm}^{-3}$, (c) $\delta n = -n_{p0}e^{-x/L_n}$

6.25 For $-L < x < +L$, $\delta p = \dfrac{G'_0}{2D_n}(5L^2 - x^2)$;

For $L < x < 3L$, $\delta p = \dfrac{G'_0 L}{D_p}(3L - x)$;

For $-3L < x < -L$, $\delta p = \dfrac{G'_0 L}{D_p}(3L + x)$

6.29 $E_{Fn} - E_{Fi} = 0.3498$ eV, $E_{Fi} - E_{Fp} = 0.2877$ eV

6.31 $\delta n = \delta p = 5 \times 10^{15}\text{ cm}^{-3}$, (a) $E_{Fn} - E_F = 0.0025$ eV, (b) $E_{Fi} - E_{Fp} = 0.5632$ eV

6.33 (a) $\delta p = 5 \times 10^{12}\text{ cm}^{-3}$, (b) $E_{Fn} - E_{Fi} = 0.1505$ eV

6.37 (a) For n-type, $\dfrac{R}{\delta n} = \dfrac{1}{\tau_{p0}} = 10^{+7}\text{ s}^{-1}$,

(b) For intrinsic, $\dfrac{R}{\delta n} = \dfrac{1}{\tau_{p0} + \tau_{n0}} = 1.67 \times 10^6\text{ s}^{-1}$

(c) For p-type, $\dfrac{R}{\delta n} = \dfrac{1}{\tau_{n0}} = 2 \times 10^6\text{ s}^{-1}$

6.39 (a) $\delta n = \dfrac{\delta n_0 \sinh[(W - x)/L_n]}{\sinh[W/L_n]}$, $\delta n_0 = 10^{15}\text{ cm}^{-3}$ and $L_n = 35.4\ \mu\text{m}$.

(b) $\delta n = 10^{15}\left(1 - \dfrac{x}{W}\right)$

6.41 For $-W < x < 0$, $\delta n = \dfrac{G'_0}{2D_n}(-x^2 - 2Wx + 2W^2)$;

For $0 < x < W$, $\delta n = \dfrac{G'_0 W}{D_n}(W - x)$

Chapter 7

7.1 (a) For $N_d = 10^{15}\text{ cm}^{-3}$; (i) $V_{bi} = 0.575$ V, (ii) 0.635 V, (iii) 0.695 V, (iv) 0.754 V, (b) For $N_d = 10^{18}\text{ cm}^{-3}$; (i) 0.754 V, (ii) 0.814 V, (iii) 0.874 V, (iv) 0.933 V

.5 (*a*)n side. $E_F - E_{Fi} = 0.3294$ eV;
p side. $E_{Fi} - E_F = 0.4070$ eV
(*b*) $V_{bi} = 0.3294 + 0.4070 = 0.7364$ V
(*c*) $V_{bi} = 0.7363$ V
(*d*) $x_n = 0.426\ \mu$m, $x_p = 0.0213\ \mu$m,
$|E_{max}| = 3.29 \times 10^4$ Vlcm

7.7 (*b*)(n region), $n_0 = N_d = 8.43 \times 10^{15}$ cm^{-3},
(p region), $p_0 = N_a = 9.97 \times 10^{15}$ cm^{-3},
(*c*) $V_{bi} = 0.690$ V

7.9 (*a*) $V_{bi} = 0.635$ V, (*b*) $x_n = 0.864\ \mu$m,
$x_p = 0.0864\ \mu$m,
(*d*) $|E_{max}| = 1.34 \times 10^4$ Vlcm

7.11 (*a*) $V_{bi} = 0.8556$ V, (*b*) $T = 302.4$ K

7.13 (*a*) $V_{bi} = 0.456$ V, (*b*) $x_n = 2.43 \times 10^{-7}$ cm,
(*c*) $x_p = 2.43 \times 10^{-3}$ cm,
(*d*) $|E_{max}| = 3.75 \times 10^3$ Vlcm

7.17 (*a*) $V_{bi} = 0.856$ V,
(*b*) W = 0.301 x 10^{-4} cm,
(*c*) $E_{max} = 3.89 \times 10^5$ Vlcm,
(*d*) C = 3.44 pF

7.19 (*a*)Neglecting change in V_{bi}, 41.4 percent increase; (*b*) 17.95 mV increase

7.21 (*a*) $V_R = 73$ V, (*b*) $V_R = 7.18$ V,
(*c*) $V_R = 0.570$ V

7.23 $V_{R2} = 18.6$ V

7.25 $N_d = 3.24 \times 10^{17}$ cm^{-3}

7.27 (*a*) $V_{bi} = 0.557$ V, (*b*) $x_n = 5.32 \times 10^{-6}$ cm,
$x_n = 2.66 \times 10^{-4}$ cm, (*c*) $V_R = 70.3$ V

7.29 (*a*) (*i*) $C = 1.14$ pF, (*ii*) C = 0.521 pF,
(*iii*) $C = 0.389$ pF; (*b*) (*i*) $C = 3.69$ pF,
(*ii*) $C = 1.74$ pF, (*iii*) $C = 1.31$ pF

7.33 (*a*)$E(x = 0) = 7.73 \times 10^4$ V/cm,
(*c*) $V_R = 23.2$ V

7.35 $a = 1.1 \times 10^{20}$ cm^{-4}

Chapter 8

8.1 (*a*) 60 mV, (*b*) 120 mV

8.5 (*a*) $$\frac{J_n}{J_n + J_p} = \frac{1}{1 + (2.04)(N_a/N_d)}$$
(*b*) $$\frac{J_n}{J_n + J_n} = \frac{(\sigma_n/\sigma_p)}{(\sigma_n/\sigma_p) + 4.90}$$

8.7 $N_a/N_d = 0.083$

8.9 $I_S = 2.91 \times 10^{-9}$ A, (*a*) $I = 6.55\ \mu$A,
(*b*) $I = -2.91$ nA

8.11 (*a*) $I_{p0} = 4.02 \times 10^{-14}$ A.
(*b*) $I_{n0} = 6.74 \times 10^{-15}$ A,
(*c*) $p_{n0} = 3.42 \times 10^{10}$ cm^{-3},
(*d*) $I_n = 3.43 \times 10^{-9}$ A

8.13 **(a)** $V_a = 0.253$ V, (*b*) $V_a = 0.635$ V

8.15 $E_{g2} = 0.769$ eV

8.18 $T_{max} \approx 519$ K

8.20 For 300 K, $V_D = 0.60$ V; For 310 K,
$V_D = 0.5827$ V; For 320 K. $V_D = 0.5653$ V

8.23 For 10 kHz, $Z = 25.9 - j0.0814$;
For 100 kHz, $Z = 25.9 - j0.814$;
For 1 MHz, $Z = 23.6 - i7.41$:
For 10 MHz, $Z = 2.38 - j7.49$

8.25 $\tau_{p0} = 1.3 \times 10^{-7}$ s; $C_d = 2.5 \times 10^{-9}$ F

8.27 ***(a)*** R = 72.3 Ω, $I = 1.38$ mA

8.29 $V_a = 0.443$ V

8.31 $J_S = 8.57 \times 10^{-18}$ A/cm^2,
$J_{gen} = 1.93 \times 10^{-9}$ A/cm^2

8.33 (*a*)For $V_a = 0.3$ V, $I = 7.96 \times 10^{-11}$ A;
For $V_a = 0.5$ V, $I = 3.36 \times 10^{-9}$ A

8.39 $V_D = 0.548$ V

8.41 $N_d \approx 3 \times 10^{15}$ cm^{-3}, A = 1.99 x 10^{-4} cm^2

8.43 $V_R = 19.9$ V

8.45 $V_R = 5.54$ V

8.47 $V_B \approx 15$ V

8.49 $I_R/I_F = 1.11$. $t_2/\tau_{p0} \approx 0.65$

8.51 $W = 61.9$ Å

Chapter 9

9.1 (*c*) $\phi_n = 0.206$ V, $\phi_{B0} = 0.27$ V,
$V_{bi} = 0.064$ V, $|E_{max}| = 1.41 \times 10^4$ V/cm,
(*d*) $\phi_{Bn} = 0.55$ V, $|E_{max}| = 3.26 \times 10^4$ V/cm

9.3 (u) $\phi_{B0} = 1.03$ V, (*b*) $\phi_n = 0.058$ V,
(*c*) $V_{bi} = 0.972$ V, (*dl* $x_d = 0.416\ \mu$m,
(*e*) $|E_{max}| = 2.87 \times 10^5$ Vlcm

9.5 (a) C = 4.75 pF, (*b*) C = 15 pF

9.7 (*a*) $V_{bi} = 0.334$ V, $x_d = 0.211\ \mu$m,
$|E_{max}| = 3.26 \times 10^4$ Vlcm.
(*h*) **A@**= 20 mV, $x_m = 0.307 \times 10^{-6}$ cm,
(*c*) $|E_{max}| = 1.16 \times 10^5$ Vlcm,
A@= 37.8 mV, $x_m = 0.163 \times 10^{-6}$ cm

9.9 (*a*) $V_{bi} = 0.812$ V, $x_d = 0.153\ \mu$m,
$|E_{max}| = 1.06 \times 10^5$ V/cm,
(*b*) $V_R = 7.47$ V

9.11 (a) $\phi_{B0} = 1.13$ V, (*b*) $\phi_{Bn} = 0.858$ V,
(*c*) $\phi_{B0} = 0.43$ V, $\phi_{Bn} = 0.733$ V

9.13 (*a*) @$_n$= 0.206 V,
(*b*) $V_{bi} = 0.684$ V.

(c) $J_{ST} = 1.3 \times 10^{-8}$ A/cm^2,
(d) $V_a = 0.488$ V

9.15 *(a)* $V_a = 0.603$ V, *(b)* $\Delta V_a = 18$ mV

9.17 $V_{bi} = 0.474$ V, *(a)* $I_{R1} = 1.52 \times 10^{-8}$ A, *(b)* $I_{R2} = 1.86 \times 10^{-8}$ A

9.19 For Schottky diode, $V_a = 0.467$ V; for pn diode. $V_a = 0.705$ V

9.21 (a) For Schottky diode, I $\approx 0.5 \times 10^{-3}$ A; for pn diode, I $= 1.02 \times 10^{-8}$ A; *(b)* for Schottky diode. $V_a = 0.239$ V; for pn diode, $V_a = 0.519$ V

9.25 *(b)* $N_d = 1.24 \times 10^{16}$ cm^{-3}, *(c)* 0.20 V

9.27 *(b)* $\phi_{B0} =$ @,= 0.138 V

Chapter 10

10.3 *(a)* $I_S = 3.2 \times 10^{-14}$ A; *(b)* *(i)* $i_C = 7.75\ \mu$A, (ii) $i_C = 0.368$ mA, (iii) $i_C = 17.5$ mA

10.5 *(a)* $\beta = 85$, a $= 0.9884$. $i_E = 516\ \mu$A; *(b)* $\beta = 53$.a $= 0.9815$, $i_E = 2.70$ mA

10.7 *(b)* $I_C = 4.7$ mA

10.9 (a) $p_{E0} = 4.5 \times 10^{}$ cm^{-3}, $n_{B0} = 2.25 \times 10^4$ cm^{-3}, $p_{C0} = 2.25 \times 10^5$ cm^{-3}, *(b)* $n_B(0) = 6.80 \times 10^{14}$ cm^{-3}, $p_E(0) = 1.36 \times 10^{13}$ cm^{-3}

10.11 Assume $\exp(V_{BE}/\text{V},) \gg \cosh(x_B/L_B)$, *(a)* 0.9950, *(b)* 0.648, *(c)* 9.08×10^{-5}

10.15 *(c)* For $x_B \ll L_B$, $J(x_B)/J(0) = 1$; For $x_B = L_B = 10\ \mu$m, $J(x_B)/J(0) = 0.648$

10.17 (a) $V_{CB} = 0.70$ V, *(b)* $V_{EC}(\text{sat}) = 0.05$ V, *(c)* 3.41×10^{11} holes/cm^2, *(d)* 8.82×10^{13} electrons/cm^2

10.19 $V_{CB} = 0.48$ V

10.21 *(o)* $I_C = 17.4\ \mu$A, *(b)* a $= 0.9067$. $I_C = 1.36$ mA, *(c)* $I_C = 19.4\ \mu$A

10.23 *(a)* (i) $\dfrac{\gamma(B)}{\gamma(A)} \approx 1 - \dfrac{Nno}{N_C} \cdot \dfrac{D_E}{D_B} \cdot \dfrac{x_B}{x_E}$,

(ii) $\dfrac{\gamma(C)}{\gamma(A)} = 1$: *(b)* (i) $\dfrac{\alpha_T(B)}{\alpha_T(A)} = 1$,

(ii) $\dfrac{\alpha_T(C)}{\alpha_T(A)} \approx 1 + \dfrac{1}{2} \cdot \dfrac{x_{B0}}{L_B}$; *(c)* neglect changes in space charge width,

(i) $\dfrac{\delta(B)}{\delta(A)} \approx 1 - \dfrac{J_{r0} \exp\left(\dfrac{-V_{BE}}{V_t}\right)}{eD_B n_{B0}}$,

(ii) $\dfrac{\delta(C)}{\delta(A)} \approx 1 + \dfrac{J_{r0} \exp\left(\dfrac{-V_{BE}}{V_t}\right)}{\dfrac{eD_B n_{B0}}{x_B}}$;

(d) Device C

10.25 *(b)* $I_C = 1.19$ mA, $I_E = 0.829$ mA

10.27 *(a)* $\delta = \dfrac{1}{1 + 15.38 \exp\left(\dfrac{-V_{BE}}{0.0518}\right)}$,

(b) $\beta = \dfrac{\text{S}}{1 - \delta}$, *(c)* for $V_{BE} < 0.4$ V, recombination factor will be the limiting factor in current gain.

10.29 *(a)* $x_B = 0.742\ \mu$m, *(b)* S $= 0.9999994$

10.35 (a) $V_A = 47.8$ V, (b) V, $= 33.4$ V, (c) $V_A = 19.0$ V

10.39 (a) $R = 893\ \Omega$, *(b)* $V = 8.93$ mV, *(c)* 70.8 percent

10.41 *(a)* $\text{E} = -\left(\dfrac{a}{x_B}\right)\left(\dfrac{kT}{e}\right)$,

(c) Total solution is

$$n = \frac{J_n}{e\mu_n \text{E}} + n_H(0)\exp(-Ax)$$

where $\text{A} = \dfrac{\text{E}}{V_t}$ and

$$n_H(0) = \frac{n_i^2}{N_B(0)} \exp\left(\frac{V_{BE}}{V_t}\right) - \frac{J_n}{e\mu_n \text{E}}$$

10.43 $BV_{CBO} = 221$ V, $N_C = 1.5 \times 10^{15}$ cm^{-3}, $x_C = 6.75\mu m$

10.45 *(a)* $V_{pt} = 295$ V: however, junction breakdown for these doping concentrations is $V_B \approx 70$ V, so punchthrough will not be reached.

10.47 *(a)* $I_B = 0.105$ mA. *(b)* $I_B = 11.9\ \mu$A, *(c)* I" $= 10.14\ \mu$A

10.53 $f_T = 509$ MHz

Chapter 11

11.1 *(a)* p type, inversion; *(b)* p type, depletion; *(c)* p type. accumulation: *(d)* n type, inversion

11.3 *(a)* By trial and error, $N_d = 3.27 \times 10^{14}$ cm^{-3}, *(b)* $\phi_s = -0.518$ V

11.5 (a) $N_d = 4.98 \times 10^{13}$ cm^{-3}, *(b)* cannot use p$^+$ poly gate, (c) $N_d = 3.43 \times 10^{14}$ cm^{-3}

11.7 $Q'_{ss}/e = 1.2 \times 10^{10}$ cm^{-2}

11.9 $V_{TP} = -1.44 + \phi_{ms}$, *(a)* $V_{TP} = -1.76$ V, *(b)* $V_{TP} = -1.71$ V, *(c)* $V_{TP} = -0.592$ V

11.11 By trial and error, $N_n = 1.71 \times 10^{16}$ cm^{-3}
11.13 (*a*) $V_{FB} = -1.52$ V, (*b*) $V_T = -0.764$ V
11.15 (*b*) $\phi_{ms} = -1.11$ V, (*c*) $V_{TN} = +0.0012$ V
11.21 (*a*) $C_{ox} = 8.63 \times 10^{-8}$ F/cm^2,
$C'_{FB} = 3.42 \times 10^{-8}$ F/cm^2,
$C'_{min} = 0.797 \times 10^{-8}$ F/cm^2,
$C'(\text{inv}) = C_{ox}$
(*b*) Same as (*a*) except $C'(\text{inv}) = C'_{min}$
(*c*) $V_{TP} = -0.989$ V
11.23 (a) $\Delta V_{FB} = -1.74$ V,
(*b*) $\Delta V_{FB} = -0.869$ V,
(*c*) $\Delta V_{FB} = -1.16$ V
11.27 (*a*) n type, (b) $t_{ox} = 345$ Å,
(*c*) $Q'_{ss} = 1.875 \times 10^{11}$ cm^{-2},
(*d*) $C_{FB} = 15$h pF
11.31 $V_{SG} = 1$ V, $I_D(\text{sat}) = 0.00592$ mA;
$V_{SG} = 3$ V, $I_D(\text{sat}) = 0.716$ mA;
$V_{SG} = 5$ V, $I_D(\text{sat}) = 2.61$ mA
11.37 $V_T \approx 0.2$ V, $\mu_n = 342$ cm^2/V-s
11.39 (a) $W/L = 14.7$, (*b*) $W/L = 25.7$
11.41 (*a*) $g_{mL} = 0.148$ mS, (*b*) $g_{ms} = 0.917$ mS
11.43 $V_{BS} = 7.92$ V
11.47 (*a*) $f_T = 5.17$ GHz, (b) $f_T = 1.0$ GHz

Chapter 12

12.1 $I_D = 10^{-15} \exp\left(\dfrac{V_{GS}}{(2.1)V_t}\right)$, $I_T = (10^6)I_D$,
$P = I_T \cdot V_{DD}$; for $V_{GS} = 0.5$ V,
$I_D = 9.83$ pA, $I_T = 9.83\ \mu$A,
$P = 49.2\ \mu$W; for $V_{GS} = 0.7$ V,
$I_D = 0.388$ nA, $I_T = 0.388$ mA,
$P = 1.94$ mW; for $V_{GS} = 0.9$ V,
$I_D = 15.4$ nA, $I_T = 15.4$ mA, $P = 77$ mW
12.3 (*a*) $\Delta V_{DS} = 1$ V, $\Delta L = 0.0451\ \mu$m;
$\Delta V_{DS} = 3$ V, $\Delta L = 0.122\ \mu$m;
$\Delta V_{DS} = 5$ V, $\Delta L = 0.188\ \mu$m;
(*b*) $L = 1.88\ \mu$m
12.7 (*a*) Assume $V_{DS}(\text{sat}) = 1$ V then
$L = 3\ \mu\text{m} \Rightarrow E_{sat} = 3.33 \times 10^3$ V/cm
$L = 1\ \mu\text{m} \Rightarrow E_{sat} = 10^4$ V/cm
$L = 0.5\ \mu\text{m} \Rightarrow E_{sat} = 2 \times 10^4$ V/cm
(*b*) Assume $\mu_n = 500$ cm^2/V-s,
$v = \mu_n E_{sat}$,
$L = 3\ \mu\text{m} \Rightarrow v = 1.67 \times 10^6$ cm/s
$L = 1\ \mu\text{m} \Rightarrow v = 5 \times 10^6$ cm/s
$L \leq 0.5\ \mu\text{m} \Rightarrow v \approx 10^7$ cm/s
12.13 (*a*) Both bias conditions. $I_D \approx kI_D$,
(*b*) $P \approx k^2P$
12.15 (a) (*i*) $I_D = 1.764$ mA;
(*ii*) $I_D = 0.807$ mA;
(b) (i) $P = 8.82$ mW, (ii) $P = 2.42$ mW;
(*c*) current: 0.457; power: 0.274
12.17 $L = 1.59\ \mu$m
12.23 $\Delta V_T = +0.118$ V
12.27 (a) $V_{BD} = 15$ V, (*b*) $V_G = 5$ V
12.31 $L = 0.844\ \mu$m
12.33 (*a*) $V_T = -0.478$ V, (b) implant acceptors.
$D_I = 4.25 \times 10^{11}$ cm^{-2}
12.35 (a) $V_T = -0.624$ V, (b) implant acceptors,
$D_I = 4.37 \times 10^{11}$ cm^{-2}, (*c*) $V_T = 1.24$ V
12.37 (*a*) $V_T = -1.53$ V; enhancement PMOS,
(b) implant acceptors,
$D_I = 4.13 \times 10^{12}$ cm^{-2}
12.39 $\Delta V_t = -2.09$ V

Chapter 13

13.3 (*a*) $V_P = 4.91$ V, (*b*) for $V_{GS} = 1$ V,
(*i*) $a - h = 0.215\ \mu$m,
(*ii*) $a - h = 0.0653\ \mu$m,
(iii) $a - h = -0.045\ \mu$m (zero depletion width)
13.5 (a) $V_{P0} = 15.5$ V, (*b*) $V_{GS} = -4.66$ V
13.7 (a) $V_{P0} = 1.863$ V, $V_P = 0.511$ V;
(*b*) (*i*) $a - h = 4.45 \times 10^{-6}$ cm,
(ii) $a - h = 1.70 \times 10^{-5}$ cm
13.9 (*a*) For $V_{DS} = 0$. $V_{GS} = -1.125$ V:
(*b*) For $V_{DS} = 1$ V. $V_{GS} = -0.125$ V
13.11 $V_{GS} = 0$, $g_d = 0.523 \times 10^{-3}$; $V_{GS} = -0.53$ V,
$g_d = 0.236 \times 10^{-3}$; $V_{GS} = -1.06$ V, $g_d = 0$
13.13 $g_{ms}(\text{max}) = 1.31$ mS/mm
13.15 (*a*) $V_{P0} = 2.59$ V, $V_T = -17.8$ V,
(*b*) depletion mode
13.17 $V_{DS} = 0$, $a - h = 0.716\ \mu$m;
$V_{DS} = 2$ V, $a - h = 0.545\ \mu$m;
$V_{DS} = 5$ V, $a - h = 0.410\ \mu$m
13.19 $N_d \approx 5.45 \times 10^{15}$ cm^{-3}
13.21 (*a*) $V_{bi} = 0.612$ V, $V_{P0} = 2.47$ V,
$V_T = -1.86$ V, $V_{DS}(\text{sat}) = 0.858$ V,
(b) add donors, $N_d = 1.64 \times 10^{16}$ cm^{-3};

$V_{bi} = 0.628$ V, $V_t = -3.87$ V,
$V_{DS}(\text{sat}) = 2.87$ V

13.23 (*a*) W = 2h.4 μm; (*b*) for $V_{GS} = 0.4$ V
$I_{D1} = 78.8$ μA; for $V_{GS} = 0.65$ V.
$I_{D1}(\text{sat}) = 0.56$ *m*A

13.29 (a) With velocity saturation.
$I_{D1}(\text{sat}) = 4.86$ mA; without velocity saturation, $I_{D1}(\text{sat}) = 18.2$ mA

13.31 (*a*) $t_d = 20$ ps, (b) $t_d = 20$ ps

13.33 (*a*) $g_{ms} = 2.82$ mS, (b) $r_s = 88.7$ Ω,
(*c*) L = 0.67 μm

13.35 (*a*) $f_t = 755$ GHz, (b) $f_T = 15.9$ GHz

13.37 (*a*) $g_m/\text{W} = 502$ mS/mm,
(*b*) $I_{D1}(\text{sat})/\text{W} = 537$ mA/mm

Chapter 14

14.1 (*a*) $\lambda = 1.24/E$
(i) E = 0.66 ⇒ A = 1.88 μm
(*ii*) $E = 1.12 \Rightarrow$ A = 1.11 μm
(*iii*) $E = 1.42 \Rightarrow$ A = 0.873 μm
(*b*) (*i*) A = 570 nm ⇒ $E = 2.18$ eV
(*ii*) $\lambda = 700$ nm ⇒ $E = 1.77$ eV

14.3 $\delta n = g'\tau = 1.44 \times 10^{13}$ cm^{-3}

14.5 (n) $x = 1.98$ μm, (b) $x = 0.41$ μm

14.11 $I_L = 500$ mA, $V_s = 0.577$ V,
$I_s = 478.3$ mA, $P_m = 276$ mW

14.13 For $h\nu = 1.7$ eV, $x = 2.3$ μm;
For hv = 2.0 eV, $x = 0.23$ μm

14.15 (*a*) $\delta p = \delta n = 10^{13}$ cm^{-3},
(*b*) $\Delta\sigma = 1.32 \times 10^{-2}$ (Ω-cm)$^{-1}$,
(*c*) $I_L = 0.66$ mA, (*d*) $\Gamma_{ph} = 4.13$

14.17 (*a*) $J_{L1} = 9.92$ mA/cm^2,
(*b*) $J_L = 0.528$ A/cm^2

14.19 W = 1 μm ⇒ $J_L = 4.15$ mA,
W = 10 μm ⇒ $J_L = 15.2$ mA,
W = 100 μm ⇒ $J_L = 16$ mA,

14.21 $0.625 \leq \lambda \leq 0.871$ μm

14.23 (*a*) 8.83 percent, (*b*) 5.95 percent

Chapter 15

15.1 $I_1 = 5.33$ A

15.7 $V_{CC} = 25$ V

15.9 (*a*) $I_1 = 1.84$ A, $I_2 = 1.66$ A, $I_3 = 1.51$ A:
$P_1 = 6.09$ W, $P_2 = 5.48$ W, $P_3 = 4.98$ W
(b) $I_1 = 2.16$ A, $I_2 = 1.08$ A, $I_3 = 1.77$ A;
$P_1 = 8.38$ W, $P_2 = 4.19$ W, $P_3 = 6.85$ W

15.11 (*b*)
(i) $V_{GS} = 5$ V, $I_D = 0.25$ A. $V_{DS} = 37.5$ V,
$P = 9.38$ W
(ii) $V_{GS} = 6$ V, $I_D = 1.0$ A, $V_{DS} = 30$ V,
$P = 30$ W
(iii) $V_{GS} = 7$ V, $I_D = 2.25$ A. $V_{DS} = 17.5$ V,
$P = 39.4$ W
(*iv*) $V_{GS} = 8$ V, $I_D = 4.0$ A, $V_{DS} = 2.92$ V,
$P = 11.7$ W
(*v*) $V_{GS} = 9$ V, $I_D = 6.25$ A, $V_{DS} = 1.88$ V,
$P = 11.7$ W

15.13 $P_D = \dfrac{125}{2.5 + \theta_{\text{case-amb}}}$

15.15 $\theta_{\text{case-amb}} = 4°\text{C/W}$