The Bipolar Transistor

T he transistor is a multijunction 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 the voltage applied across the other two terminals of the device.

The **B**ipolar **J**unction **T**ransistor (BJT) is one of two major types of transistors. The fundamental physics of the BJT is developed in this chapter. The bipolar transistor is used extensively in analog electronic circuits because of its high current gain.

Two complementary configurations of BJTs, the npn and pnp devices, can be fabricated. Electronic circuit design becomes very versatile when the two types of devices are used in the same circuit. ∎

12.0 | PREVIEW

In this chapter, we will:

- Discuss the physical structure of the bipolar transistor, which has three separately doped regions and two pn junctions that are sufficiently close together so interactions occur between the two junctions.
- Discuss the basic principle of operation of the bipolar transistor, including the various possible modes of operation.
- Derive expressions for the minority carrier concentrations through the device for various operating modes.
- Derive expressions for the various current components in the bipolar transistor.
- Define common-base and common-emitter current gains.
- Define the limiting factors and derive expressions for the current gain.
- Discuss several nonideal effects in bipolar transistors, including base width modulation and high-level injection effects.

- ■ Develop the small-signal equivalent circuit of the bipolar transistor. This circuit is used to relate small-signal currents and voltages in analog circuits.
- ■ Define and derive expressions for the frequency limiting factors.
- ■ Present the geometries and characteristics of a few specialized bipolar transistor designs.

12.1 | THE BIPOLAR TRANSISTOR ACTION

The bipolar transistor has three separately doped regions and two pn junctions. Figure 12.1 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.

The block diagrams of Figure 12.1 show the basic structure of the transistor, but in very simplified sketches. Figure 12.2a shows a cross section of a classic npn bipolar transistor fabricated in an integrated circuit configuration, and Figure 12.2b shows the cross section of an npn bipolar transistor fabricated by a more modern technology. One can immediately observe that the actual structure of the bipolar transistor is not nearly as simple as the block diagrams of Figure 12.1 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 12.2a, or they are isolated by large oxide regions as shown in Figure 12.2b.
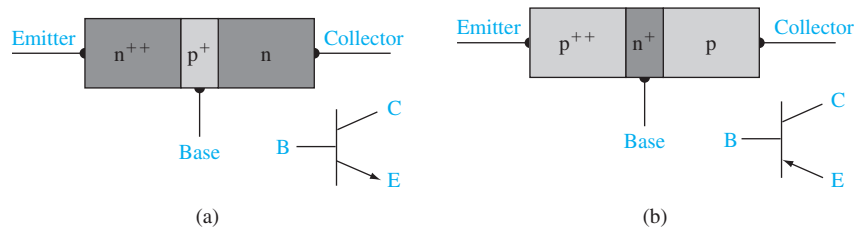

Figure 12.1 | Simplified block diagrams and circuit symbols of (a) npn and (b) pnp bipolar transistors.
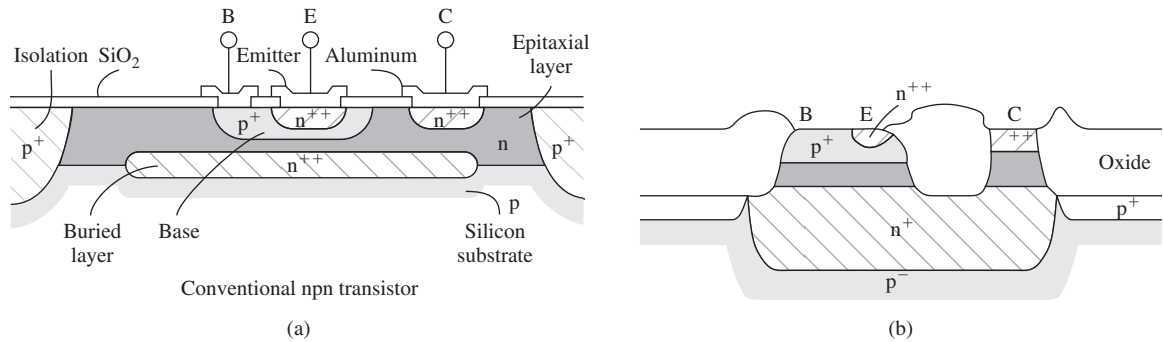
Figure 12.2 | Cross section of (a) a conventional integrated circuit npn bipolar transistor and (b) an oxide-isolated npn bipolar transistor.
(From Muller and Kamins [4].)

An important point to note from the devices shown in Figure 12.2 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 12.1 are highly simplified, but useful, concepts in the development of the basic transistor theory.

12.1.1   The Basic Principle of Operation

The npn and pnp transistors are complementary devices. We develop the bipolar transistor theory using the npn transistor, but the same basic principles and equations also apply to the pnp device. Figure 12.3 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.

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
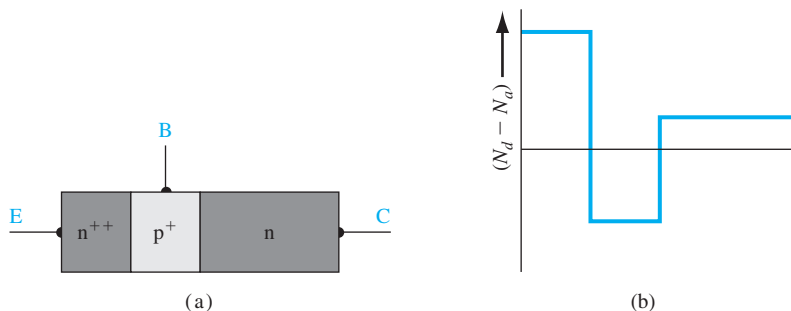

Figure 12.3 | Idealized doping profile of a uniformly doped npn bipolar transistor.
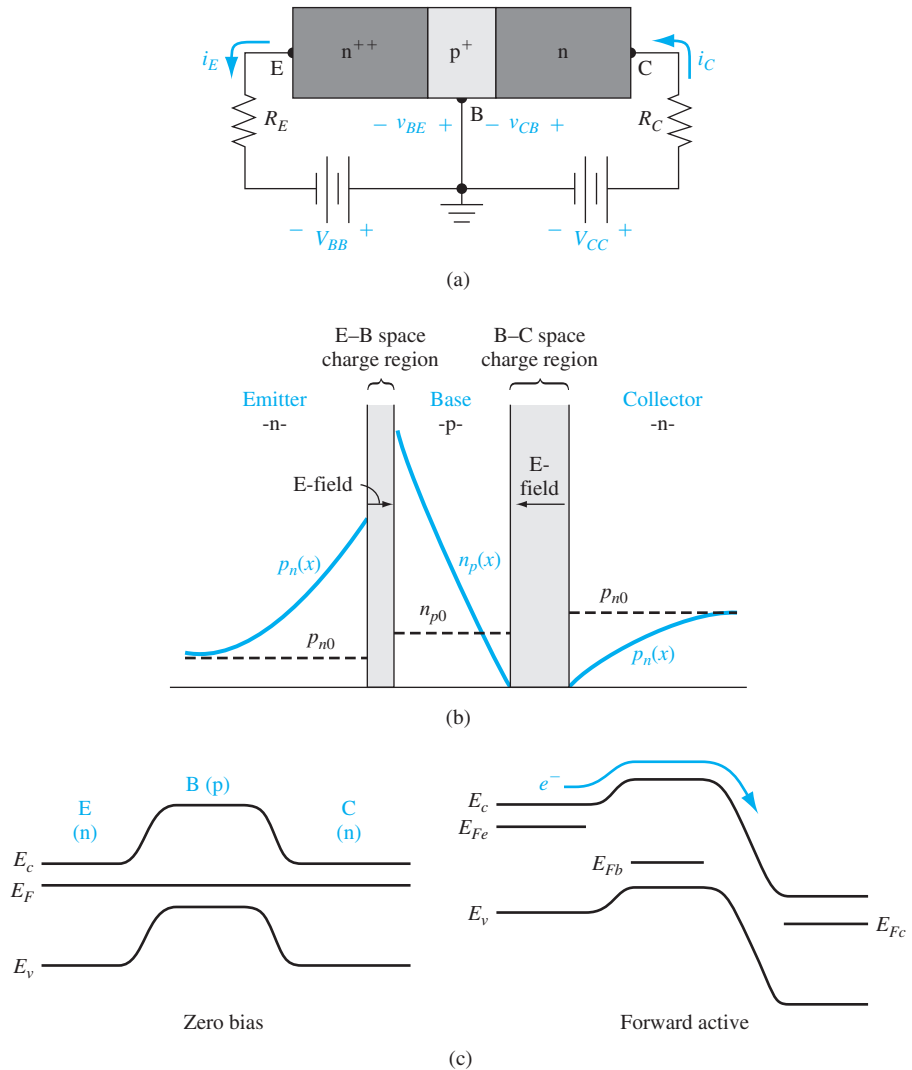
Figure 12.4 | (a) Biasing of an npn bipolar transistor in the forward-active mode, (b) minority carrier distribution in an npn bipolar transistor operating in the forward-active mode, and (c) energy-band diagram of the npn bipolar transistor under zero bias and under a forward-active mode bias.

Figure 12.4a. 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
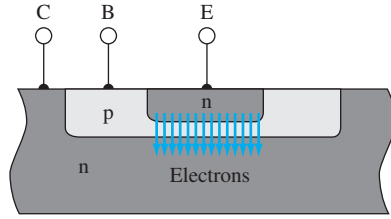
Figure 12.5 | Cross section of an npn bipolar transistor showing the injection and collection of electrons in the forward-active mode.

Figure 12.4b. 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 12.5 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).

12.1.2 Simplified Transistor Current Relation—Qualitative Discussion

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 delve into a more detailed analysis of the physics of the bipolar transistor.

The minority carrier concentrations are again shown in Figure 12.6 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.

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

$$i_C = eD_nA_{BE}\frac{dn(x)}{dx} = eD_nA_{BE}\left[\frac{n_B(0) - 0}{0 - x_B}\right] = \frac{-eD_nA_{BE}}{x_B} \cdot n_{B0}\exp\left(\frac{v_{BE}}{V_t}\right) \quad (12.1)$$

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
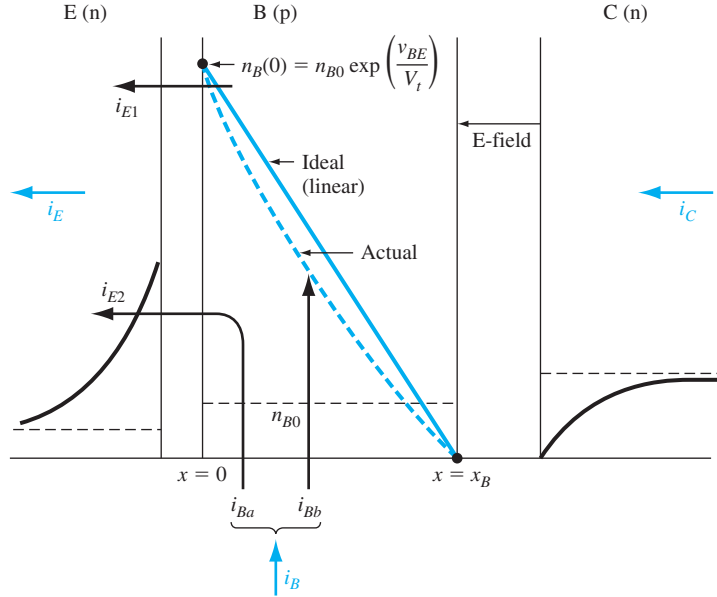
Figure 12.6 | Minority carrier distributions and basic currents in a forward-biased npn bipolar transistor.

diffusion of electrons is in the $+x$ direction so that the conventional current is in the $-x$ direction. Considering magnitudes only, Equation (12.1) can be written as

$$i_C = I_s \exp\left(\frac{v_{BE}}{V_t}\right) \tag{12.2}$$

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.

Emitter Current   One component of emitter current, i_{E1}, shown in Figure 12.6 is due to the flow of electrons injected from the emitter into the base. This current, then, is equal to the collector current given by Equation (12.1).

Since the base–emitter junction is forward biased, majority carrier holes in the base are injected across the B–E junction into the emitter. These injected holes produce a pn junction current i_{E2} as indicated in Figure 12.6. This current is only a B–E junction current so this component of emitter current is not part of the collector current. Since i_{E2} is a forward-biased pn junction current, we can write (considering magnitude only)

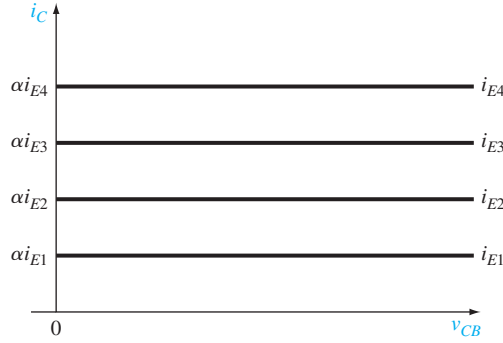$$i_{E2} = I_{S2} \exp\left(\frac{v_{BE}}{V_t}\right) \tag{12.3}$$

Figure 12.7 | Ideal bipolar transistor common-base current–voltage characteristics.

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

$$i_E = i_{E1} + i_{E2} = i_C + i_{E2} = I_{SE}\exp\left(\frac{v_{BE}}{V_t}\right) \tag{12.4}$$

Since all current components in Equation (12.4) are functions of $\exp(v_{BE}/V_t)$, the ratio of collector current to emitter current is a constant. We can write

$$\frac{i_C}{i_E} \equiv \alpha \tag{12.5}$$

where α is called the *common-base current gain*. By considering Equation (12.4), 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 12.4a and Equation (12.4), 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 base–collector voltage as long as the B–C junction is reverse biased. We can sketch the common-base transistor characteristics as shown in Figure 12.7. The bipolar transistor acts like a constant current source.

Base Current As shown in Figure 12.6, the component of emitter current i_{E2} is a B–E junction current so that this current is also a component of base current shown 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 majority carrier holes in the base. However, in reality, there will be some recombination. Since majority carrier holes in the base are disappearing, they must be resupplied by a flow of positive charge into the base terminal. This flow of charge is indicated as a current i_{Bb} in Figure 12.6. 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 $\exp(\nu_{BE}/V_t)$. The total base current is the sum of i_{Ba} and i_{Bb} and is proportional to $\exp(\nu_{BE}/V_t)$.

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

$$\frac{i_C}{i_B} \equiv \beta \qquad (12.6)$$

where β is called the *common-emitter current gain*. Normally, the base 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).

12.1.3   The Modes of Operation

Figure 12.8 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 base. 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.

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

$$V_{CC} = I_C R_C + V_{CB} + V_{BE} = V_R + V_{CE} \qquad (12.7)$$

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
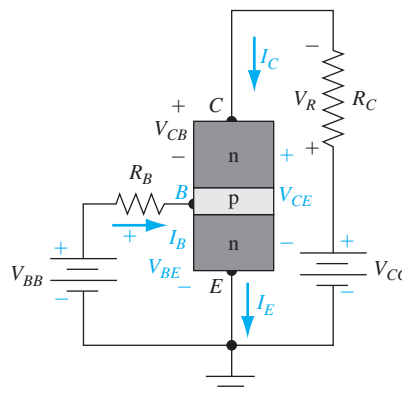

Figure 12.8 | An npn bipolar transistor in a common-emitter circuit configuration.
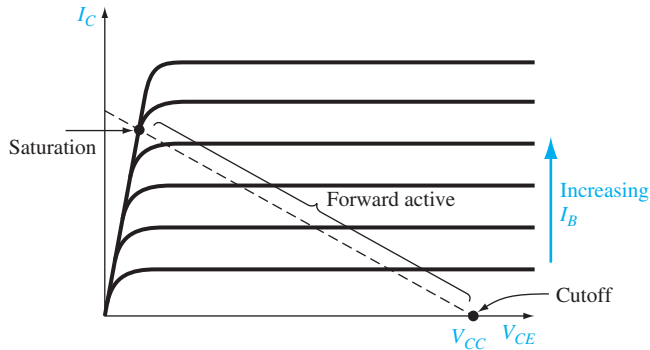
Figure 12.9 | Bipolar transistor common-emitter current–voltage characteristics with load line superimposed.

enough that the combination of V_R and V_{CC} produces 0 V 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*.[1] 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 12.9 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 12.8). 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

$$V_{CE} = V_{CC} - I_C R_C \tag{12.8}$$

Equation (12.8) shows a linear relation between collector current and collector–emitter voltage. This linear relation is called a *load line* and is plotted in Figure 12.9. 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 $I_C = 0$, saturation occurs when there is no longer a change in collector current for a 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 12.8. This fourth mode, known as

---

[1]The concept of "saturation" for the bipolar transistor is not the same as the principle of the "saturation region" for the MOSFET described in Chapter 10. The term "saturation" as applied to the BJT means that the output current and output voltage do not change as the base–emitter voltage changes. The term "saturation region" as applied to the MOSFET means that the output current does not change (ideally) with a change in the drain-to-source voltage.

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 12.10.

12.1.4 Amplification with Bipolar Transistors

Voltages and currents can be amplified by bipolar transistors in conjunction with other elements. We demonstrate this amplification qualitatively in the following discussion. Figure 12.11 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 12.12 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 time-varying collector current induces a time-varying voltage across the R_C resistor which, by Kirchhoff's voltage law, means that a sinusoidal voltage, superimposed on a dc value, exists between the collector and emitter of the bipolar transistor. The sinusoidal voltages in the collector–emitter 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.*

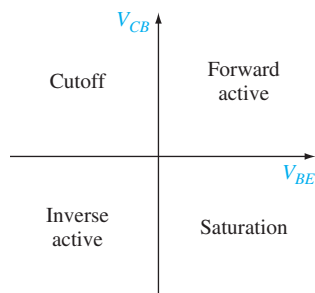**Figure 12.10** | Junction voltage conditions for the four operating modes of a bipolar transistor.
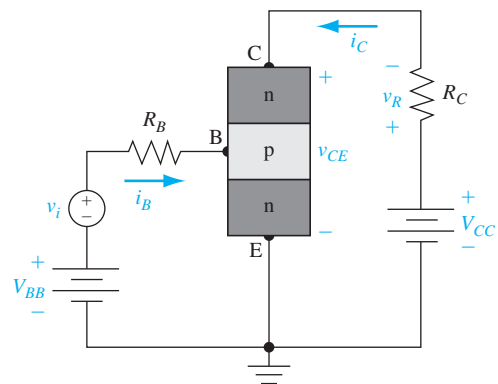

Figure 12.11 | Common-emitter npn bipolar circuit configuration with a time-varying signal voltage v_i included in the base–emitter loop.
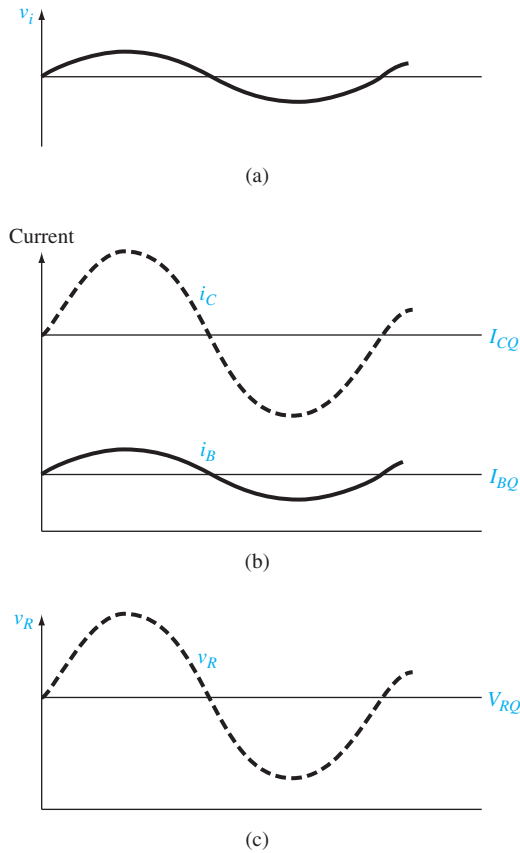
(a)

(b)

(c)

Figure 12.12 | Currents and voltages existing in the circuit shown in Figure 12.11. (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.

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

12.2 | MINORITY CARRIER DISTRIBUTION

We are interested in calculating currents in the bipolar transistor that, 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 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 12.1 summarizes the notation used in the following analysis.

Table 12.1 | 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
τ_{E0}, τ_{B0}, τ_{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
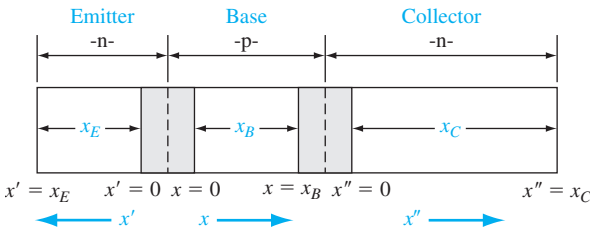
Figure 12.13 | Geometry of the npn bipolar transistor used to calculate the minority carrier distribution.

12.2.1   Forward-Active Mode

Consider a uniformly doped npn bipolar transistor with the geometry shown in Figure 12.13. When we consider the individual emitter, base, and collector regions, we 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
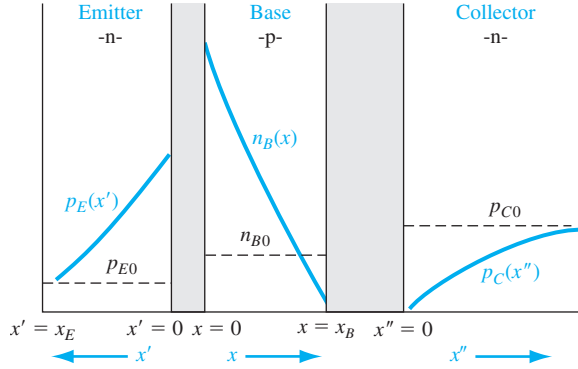
Figure 12.14 | Minority carrier distribution in an npn bipolar transistor operating in the forward-active mode.

in Figure 12.14. As there are two n regions, we have minority carrier holes in both emitter and collector. To distinguish between these two minority carrier hole distributions, we use the notation shown in the figure. Keep in mind that we are 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 assume that the neutral collector length x_C is long compared to the minority carrier diffusion length L_C in the collector, but we 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

$$D_B \frac{\partial^2(\delta n_B(x))}{\partial x^2} - \frac{\delta n_B(x)}{\tau_{B0}} = 0 \qquad (12.9)$$

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

$$\delta n_B(x) = n_B(x) - n_{B0} \qquad (12.10)$$

The general solution to Equation (12.9) can be written as

$$\delta n_B(x) = A \exp\left(\frac{+x}{L_B}\right) + B \exp\left(\frac{-x}{L_B}\right) \qquad (12.11)$$

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 (12.11) must be retained.

The excess minority carrier electron concentrations at the two boundaries become

$$\delta n_B(x = 0) \equiv \delta n_B(0) = A + B \tag{12.12a}$$

and

$$\delta n_B(x = x_B) \equiv \delta n_B(x_B) = A \exp\left(\frac{+x_B}{L_B}\right) + B \exp\left(\frac{-x_B}{L_B}\right) \tag{12.12b}$$

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

$$\delta n_B(0) = n_B(x = 0) - n_{B0} = n_{B0}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \tag{12.13a}$$

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

$$\delta n_B(x_B) = n_B(x = x_B) - n_{B0} = 0 - n_{B0} = -n_{B0} \tag{12.13b}$$

From the boundary conditions given by Equations (12.13a) and (12.13b), the coefficients A and B from Equations (12.12a) and (12.12b) can be determined. The results are

$$A = \frac{-n_{B0} - n_{B0}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right]\exp\left(\frac{-x_B}{L_B}\right)}{2 \sinh\left(\frac{x_B}{L_B}\right)} \tag{12.14a}$$

and

$$B = \frac{n_{B0}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right]\exp\left(\frac{x_B}{L_B}\right) + n_{B0}}{2 \sinh\left(\frac{x_B}{L_B}\right)} \tag{12.14b}$$

Then, substituting Equations (12.14a) and (12.14b) into Equation (12.9), we can write the excess minority carrier electron concentration in the base region as

$$\delta n_B(x) = \frac{n_{B0}\left\{\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right]\sinh\left(\frac{x_B - x}{L_B}\right) - \sinh\left(\frac{x}{L_B}\right)\right\}}{\sinh\left(\frac{x_B}{L_B}\right)} \tag{12.15a}$$

Equation (12.15a) 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 becomes 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 be much less than unity. Figure 12.15 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 (12.15a) is approximately a linear function of* x *through the neutral base region.*
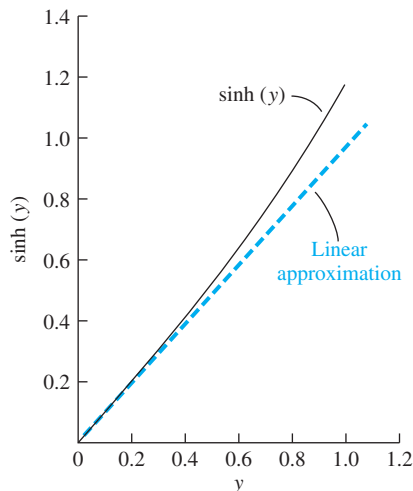
Figure 12.15 | Hyperbolic sine function and its linear approximation.

Using the approximation that sinh $(x) \approx x$ for $x \ll 1$, the excess electron concentration in the base is given by

$$\delta n_B(x) \approx \frac{n_{B0}}{x_B} \left\{ \left[\exp \left(\frac{e V_{BE}}{kT} \right) - 1 \right] (x_B - x) - x \right\} \qquad (12.15b)$$

We use this linear approximation later in some of the example calculations. The difference in the excess carrier concentrations determined from Equations (12.15a) and (12.15b) is demonstrated in the following exercise.

TEST YOUR UNDERSTANDING

TYU 12.1   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 $x_B = 2$ μm and the minority carrier diffusion length in the base is $L_B = 10$ μm. 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 (12.15a)] to that in the ideal case of a linear minority carrier distribution [Equation (12.15b)].

[Ans. (a) $n_B(0) = 3.81 \times 10^{14}$ cm^{-3}; (b) $\delta n_B(x_B/2) \approx n_B(x_B/2) = 1.8947 \times 10^{14}$ cm^{-3}; (c) Ratio $= (1.8947 \times 10^{14}/1.9042 \times 10^{14}) = 0.9950$]

Table 12.2 shows the Taylor expansions of some of the hyperbolic functions that are encountered in this section of the chapter. In most cases, we consider only the linear terms when expanding these functions.

Table 12.2 | Taylor expansions of hyperbolic functions

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

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

$$D_E \frac{\partial^2 [\delta p_E(x')]}{\partial x'^2} - \frac{\delta p_E(x')}{\tau_{E0}} = 0 \tag{12.16}$$

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

$$\delta p_E(x') = p_E(x') - p_{E0} \tag{12.17}$$

The general solution to Equation (12.16) can be written as

$$\delta p_E(x') = C \exp\left(\frac{+x'}{L_E}\right) + D \exp\left(\frac{-x'}{L_E}\right) \tag{12.18}$$

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 (12.18) must be retained.

The excess minority carrier hole concentrations at the two boundaries are

$$\delta p_E(x' = 0) \equiv \delta p_E(0) = C + D \tag{12.19a}$$

and

$$\delta p_E(x' = x_E) \equiv \delta p_E(x_E) = C \exp\left(\frac{x_E}{L_E}\right) + D \exp\left(\frac{-x_E}{L_E}\right) \tag{12.19b}$$

Again, the B–E junction is forward biased, so

$$\delta p_E(0) = p_E(x' = 0) - p_{E0} = p_{E0}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \tag{12.20a}$$

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

$$\delta p_E(x_E) = 0 \tag{12.20b}$$

Solving for C and D using Equations (12.19) and (12.20) yields the excess minority carrier hole concentration in Equation (12.18):

$$\delta p_E(x') = \frac{p_{E0}\left[\exp\left(\dfrac{eV_{BE}}{kT}\right) - 1\right]\sinh\left(\dfrac{x_E - x'}{L_E}\right)}{\sinh\left(\dfrac{x_E}{L_E}\right)} \tag{12.21a}$$

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

$$\delta p_E(x') \approx \frac{p_{E0}}{x_E}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right](x_E - x') \tag{12.21b}$$

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

TEST YOUR UNDERSTANDING

TYU 12.2 Consider a silicon npn bipolar transistor with emitter and base regions uniformly doped at concentrations of $10^{18}\,\text{cm}^{-3}$ and $10^{16}\,\text{cm}^{-3}$, respectively. A forward-bias B–E voltage of $V_{BE} = 0.610$ V is applied. The neutral emitter width is $x_E = 4\ \mu\text{m}$ and the minority carrier diffusion length in the emitter is $L_E = 4\ \mu\text{m}$. Calculate the excess minority carrier concentration in the emitter at (*a*) $x' = 0$ and (*b*) $x' = x_E/2$.

[Ans. (*a*) $3.808 \times 10^{12}\,\text{cm}^{-3}$; (*b*) $1.689 \times 10^{12}\,\text{cm}^{-3}$]

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

$$D_C \frac{\partial^2[\delta p_C(x'')]}{\partial x''^2} - \frac{\delta p_C(x'')}{\tau_{C0}} = 0 \tag{12.22}$$

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

$$\delta p_C(x'') = p_C(x'') - p_{C0} \tag{12.23}$$

The general solution to Equation (12.22) can be written as

$$\delta p_C(x'') = G \exp\left(\frac{x''}{L_C}\right) + H \exp\left(\frac{-x''}{L_C}\right) \tag{12.24}$$

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

$$\delta p_C(x'' = 0) \equiv \delta p_C(0) = p_C(x'' = 0) - p_{C0} = 0 - p_{C0} = -p_{C0} \tag{12.25}$$

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

$$\delta p_C(x'') = -p_{C0} \exp\left(\frac{-x''}{L_C}\right) \tag{12.26}$$

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

TEST YOUR UNDERSTANDING

TYU 12.3 Consider the collector region of an npn bipolar transistor biased in the forward-active region. At what value of x'', compared to L_C, does the magnitude of the minority carrier concentration reach 95 percent of the thermal-equilibrium value? (Ans. $x''/L_C \approx 3$)

12.2.2   Other Modes of Operation

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

Figure 12.16a 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 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 12.16b 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 12.17a 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
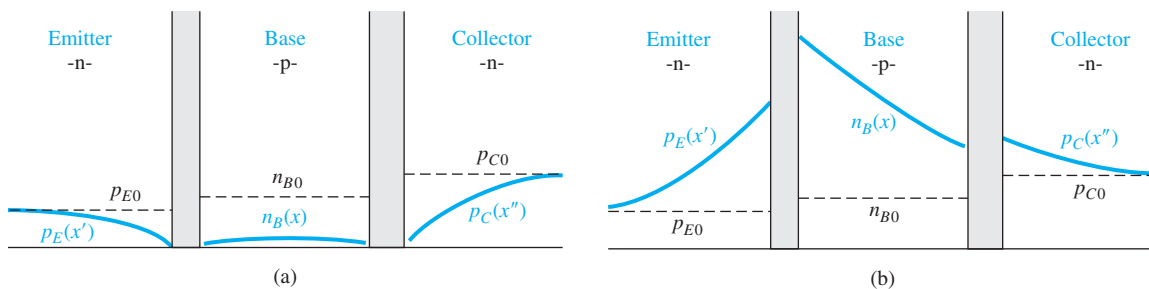

Figure 12.16 | Minority carrier distribution in an npn bipolar transistor operating in (a) cutoff and (b) saturation.
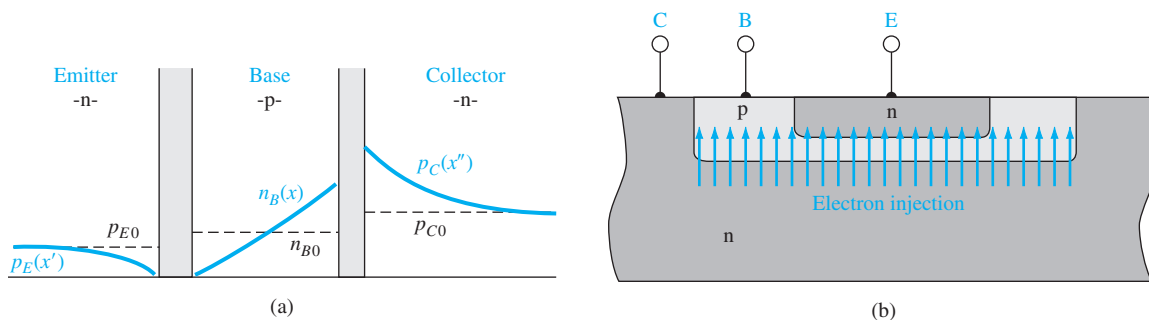

Figure 12.17 | (a) Minority carrier distribution in an npn bipolar transistor operating in the inverse-active mode.
(b) Cross section of an npn bipolar transistor showing the injection and collection of electrons in the inverse-active mode.

direction compared with the forward-active mode, so the emitter and collector currents will change direction. Figure 12.17b 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.

12.3 | TRANSISTOR CURRENTS AND 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.

12.3.1   Current Gain—Contributing Factors

Figure 12.18 shows the various particle flux components in the npn bipolar transistor. We 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 12.14.

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 recombination 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 carrier holes from the base that are injected back into the emitter result in a hole flux denoted by J_{pE}^{+}. Some electrons and holes
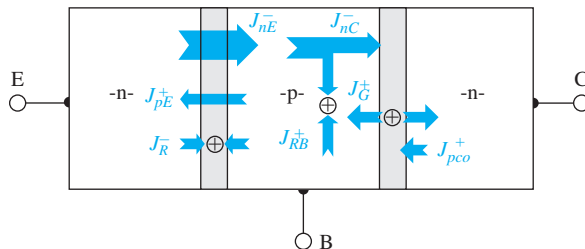

Figure 12.18 | Particle current density or flux components in an npn bipolar transistor operating in the forward-active mode.
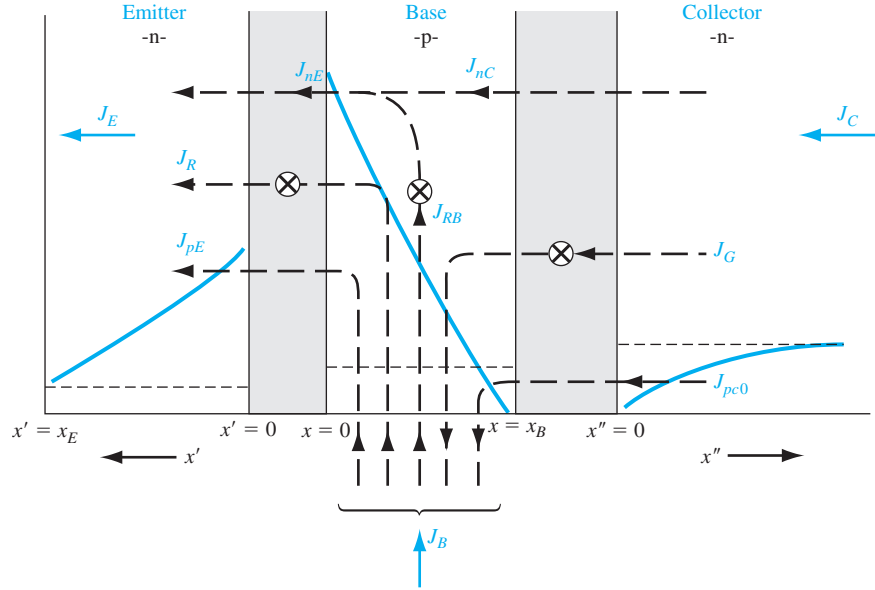
Figure 12.19 | Current density components in an npn bipolar transistor operating in the forward-active mode.

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 12.19 along with the minority carrier distributions for the forward-active mode. The curves are the same as in Figure 12.14. 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 recombination 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 $x' = 0$.

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

$$\alpha_0 = \frac{I_C}{I_E} \tag{12.27}$$

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

$$\alpha_0 = \frac{J_C}{J_E} = \frac{J_{nC} + J_G + J_{pc0}}{J_{nE} + J_R + J_{pE}} \tag{12.28}$$

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

$$\alpha = \frac{\partial J_C}{\partial J_E} = \frac{J_{nC}}{J_{nE} + J_R + J_{pE}} \tag{12.29}$$

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

We can rewrite Equation (12.29) in the form

$$\alpha = \left(\frac{J_{nE}}{J_{nE} + J_{pE}}\right)\left(\frac{J_{nC}}{J_{nE}}\right)\left(\frac{J_{nE} + J_{pE}}{J_{nE} + J_R + J_{pE}}\right) \tag{12.30a}$$

or

$$\alpha = \gamma \alpha_T \delta \tag{12.30b}$$

The factors in Equation (12.30b) are defined as:

$$\gamma = \left(\frac{J_{nE}}{J_{nE} + J_{pE}}\right) \quad \equiv \text{emitter injection efficiency factor} \tag{12.31a}$$

$$\alpha_T = \left(\frac{J_{nC}}{J_{nE}}\right) \quad \equiv \text{base transport factor} \tag{12.31b}$$

$$\delta = \frac{J_{nE} + J_{pE}}{J_{nE} + J_R + J_{pE}} \equiv \text{recombination factor} \tag{12.31c}$$

We would like to have the change in collector current be exactly the same as the change in emitter current or, ideally, to have $\alpha = 1$. However, a consideration of Equation (12.29) shows that α will always be less than unity. The goal is to make α as close to unity as possible. To achieve this goal, we must make each term in Equation (12.30b) as close to unity as possible, since each factor is less than unity.

The *emitter injection efficiency factor* γ takes into account the minority carrier hole diffusion current in the emitter. This current is part of the emitter current, but does not contribute to the transistor action in that J_{pE} is not part of the collector current. The *base transport factor* α_T takes into account any recombination of excess minority carrier electrons in the base. Ideally, we want no recombination in the base. The *recombination 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 not contribute to collector current.

12.3.2 Derivation of Transistor Current Components and Current Gain Factors

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

Emitter Injection Efficiency Factor   Consider, initially, the emitter injection efficiency factor. We have from Equation (12.31a)

$$\gamma = \left(\frac{J_{nE}}{J_{nE} + J_{pE}}\right) = \frac{1}{\left(1 + \dfrac{J_{pE}}{J_{nE}}\right)} \tag{12.32}$$

We derived the minority carrier distribution functions for the forward-active mode in Section 12.2.1. Noting that J_{nE}, as defined in Figure 12.19, is in the negative x direction, we can write the current densities as

$$J_{pE} = -eD_E \frac{d[\delta p_E(x')]}{dx'}\bigg|_{x'=0} \tag{12.33a}$$

and

$$J_{nE} = (-)eD_B \frac{d[\delta n_B(x)]}{dx}\bigg|_{x=0} \tag{12.33b}$$

where $\delta p_E(x')$ and $\delta n_B(x)$ are given by Equations (12.21) and (12.15), respectively.

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

$$J_{pE} = \frac{eD_E p_{E0}}{L_E} \left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \cdot \frac{1}{\tanh(x_E/L_E)} \tag{12.34a}$$

and

$$J_{nE} = \frac{eD_B n_{B0}}{L_B} \left\{\frac{1}{\sinh(x_B/L_B)} + \frac{[\exp(eV_{BE}/kT) - 1]}{\tanh(x_B/L_B)}\right\} \tag{12.34b}$$

Positive J_{pE} and J_{nE} values imply that the currents are in the directions shown in Figure 12.19. 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 (12.32), then becomes

$$\gamma = \cfrac{1}{1 + \cfrac{p_{E0} D_E L_B}{n_{B0} D_B L_E} \cdot \cfrac{\tanh{(x_B/L_B)}}{\tanh{(x_E/L_E)}}} \qquad (12.35a)$$

If we assume that all the parameters in Equation (12.35a) 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

$$\gamma \approx \cfrac{1}{1 + \cfrac{N_B}{N_E} \cdot \cfrac{D_E}{D_B} \cdot \cfrac{x_B}{x_E}} \qquad (12.35b)$$

Objective: Calculate the emitter injection efficiency.

Assume the following transistor parameters: $N_B = 10^{15}\ \text{cm}^{-3}$, $N_E = 10^{17}\ \text{cm}^{-3}$, $D_E = 10\ \text{cm}^2/\text{s}$, $D_B = 20\ \text{cm}^2/\text{s}$, $x_B = 0.80\ \mu\text{m}$, and $x_E = 0.60\ \mu\text{m}$.

■ **Solution**
From Equation (12.35b), we find

$$\gamma \cong \cfrac{1}{1 + \left(\cfrac{N_B}{N_E}\right)\left(\cfrac{D_E}{D_B}\right)\left(\cfrac{x_B}{x_E}\right)} = \cfrac{1}{1 + \left(\cfrac{10^{15}}{10^{17}}\right)\left(\cfrac{10}{20}\right)\left(\cfrac{0.80}{0.60}\right)} = 0.9934$$

■ **Comment**
This simple example shows a typical magnitude of the emitter injection efficiency.

■ **EXERCISE PROBLEM**
Ex 12.1 Repeat Example 12.1 if the base and emitter doping concentrations are
$N_B = 5 \times 10^{15}\ \text{cm}^{-3}$ and $N_E = 10^{18}\ \text{cm}^{-3}$, respectively.
(Ans. $\gamma = 0.9967$)

Base Transport Factor The next term to consider is the base transport factor, given by Equation (12.31b) as $\alpha_T = J_{nC}/J_{nE}$. From the definitions of the current directions shown in Figure 12.19, we can write

$$J_{nC} = (-)eD_B \frac{d[\delta n_B(x)]}{dx}\Bigg|_{x=x_B} \qquad (12.36a)$$

and

$$J_{nE} = (-)eD_B \frac{d[\delta n_B(x)]}{dx}\bigg|_{x=0} \tag{12.36b}$$

Using the expression for $\delta n_B(x)$ given in Equation (12.15), we find that

$$J_{nC} = \frac{eD_B n_{B0}}{L_B} \left\{ \frac{[\exp(eV_{BE}/kT) - 1]}{\sinh(x_B/L_B)} + \frac{1}{\tanh(x_B/L_B)} \right\} \tag{12.37}$$

The expression for J_{nE} is given in Equation (12.34a).

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 (12.37) and (12.34b) into Equation (12.31b), we have

$$\alpha_T = \frac{J_{nC}}{J_{nE}} \approx \frac{\exp(eV_{BE}/kT) + \cosh(x_B/L_B)}{1 + \exp(eV_{BE}/kT)\cosh(x_B/L_B)} \tag{12.38}$$

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

$$\boxed{\alpha_T \approx \frac{1}{\cosh(x_B/L_B)}} \tag{12.39a}$$

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

$$\boxed{\alpha_T \approx \frac{1}{\cosh(x_B/L_B)} \approx \frac{1}{1 + \frac{1}{2}(x_B/L_B)^2} \approx 1 - \frac{1}{2}(x_B/L_B)^2} \tag{12.39b}$$

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.

EXAMPLE 12.2

Objective: Calculate the base transport factor.
   Assume transistor parameters of $x_B = 0.80~\mu m$ and $L_B = 10.0~\mu m$.

■ **Solution**
From Equation (12.39a), we find

$$\alpha_T \cong \frac{1}{\cosh\left(\frac{x_B}{L_B}\right)} = \frac{1}{\cosh\left(\frac{0.80}{10.0}\right)} = 0.9968$$

■ **Comment**
This simple example shows a typical magnitude of the base transport factor.

■ **EXERCISE PROBLEM**
Ex 12.2   Repeat Example 12.2 for $x_B = 1.2~\mu m$ and $L_B = 10.0~\mu m$. (Ans. $\alpha_T = 0.9928$)

Recombination Factor   The recombination factor is given by Equation (12.31c). We can write

$$\delta = \frac{J_{nE} + J_{pE}}{J_{nE} + J_R + J_{pE}} \approx \frac{J_{nE}}{J_{nE} + J_R} = \frac{1}{1 + J_R/J_{nE}} \tag{12.40}$$

We have assumed in Equation (12.40) 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

$$J_R = \frac{ex_{BE}n_i}{2\tau_0} \exp\left(\frac{eV_{BE}}{2kT}\right) = J_{r0} \exp\left(\frac{eV_{BE}}{2kT}\right) \tag{12.41}$$

where x_{BE} is the B–E space charge width.
   The current J_{nE} from Equation (12.34b) can be approximated as

$$J_{nE} = J_{s0} \exp\left(\frac{eV_{BE}}{kT}\right) \tag{12.42}$$

where

$$J_{s0} = \frac{eD_B n_{B0}}{L_B \tanh (x_B/L_B)} \tag{12.43}$$

The recombination factor, from Equation (12.40), can then be written as

$$\delta = \frac{1}{1 + \dfrac{J_{r0}}{J_{s0}} \exp\left(\dfrac{-eV_{BE}}{2kT}\right)} \tag{12.44}$$

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.

EXAMPLE 12.3

Objective: Calculate the recombination factor.
   Assume the following transistor parameters: $x_{BE} = 0.10 \ \mu m$, $\tau_o = 10^{-7} \ s$, $N_B = 5 \times 10^{15} \ cm^{-3}$, $D_B = 20 \ cm^2/s$, $L_B = 10 \ \mu m$, and $x_B = 0.80 \ \mu m$. Assume $V_{BE} = 0.50 \ V$.

■ **Solution**
From Equation (12.41), we find

$$J_{r0} = \frac{ex_{BE}n_i}{2\tau_o} = \frac{(1.6 \times 10^{-19})(0.10 \times 10^{-4})(1.5 \times 10^{10})}{2(10^{-7})} = 1.2 \times 10^{-7} \ A/cm^2$$

and from Equation (12.43), we find

$$J_{s0} = \frac{eD_B n_{B0}}{L_B \tanh (x_B/L_B)} = \frac{eD_B (n_i^2/N_B)}{L_B \tanh (x_B/L_B)}$$

$$= \frac{(1.6 \times 10^{-19})(20)[(1.5 \times 10^{10})^2/5 \times 10^{15}]}{(10 \times 10^{-4})\tanh (0.80/10.0)} = 1.804 \times 10^{-9} \ A/cm^2$$

Then from Equation (12.44), the recombination factor is found as

$$\delta = \frac{1}{1 + \frac{J_{r0}}{J_{s0}} \cdot \exp\left(\frac{-V_{BE}}{2V_t}\right)} = \frac{1}{1 + \left(\frac{1.2 \times 10^{-7}}{1.804 \times 10^{-9}}\right) \cdot \exp\left(\frac{-0.50}{2(0.0259)}\right)}$$

$$= 0.99574$$

■ **Comment**

This simple example shows a typical magnitude of the recombination factor.

■ **EXERCISE PROBLEM**

Ex 12.3   Repeat Example 12.3 for $V_{BE} = 0.65$ V. (Ans. $\delta = 0.99976$)

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 12.20a shows the B–E junction of an npn transistor near the semiconductor surface. We assume that the B–E junction is forward biased. Figure 12.20b 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 12.20c shows the excess minority carrier electron concentration along the cross section C-C′ from the surface. We have 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 12.20d shows the injection of electrons from the emitter into the base and the diffusion of
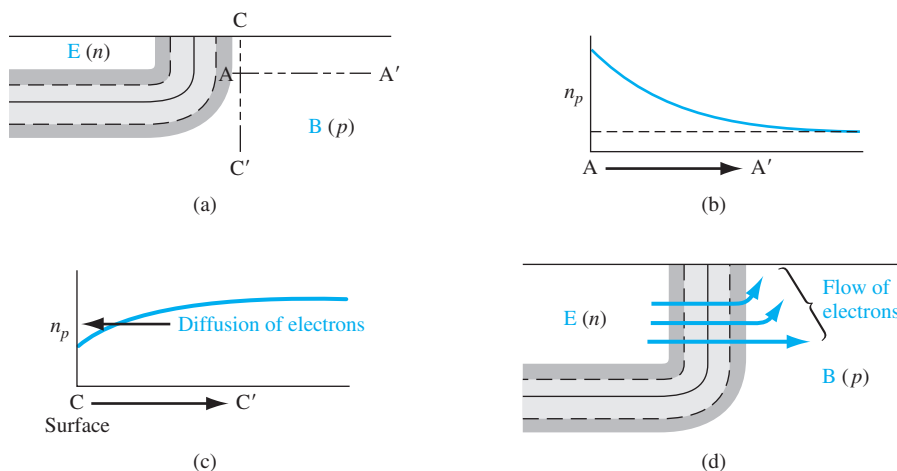

Figure 12.20 | The surface at the E–B junction showing the diffusion of carriers toward the surface.

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 (12.41).

12.3.3 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 are obtained except that the electron concentrations become hole concentrations and vice versa. The current directions and voltage polarities also change.

We have been considering the common-base current gain, defined in Equation (12.27) as $\alpha_0 = I_C/I_E$. The common-emitter current gain is defined as $\beta_0 = I_C/I_B$. From Figure 12.8 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 both 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 12.3 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.

12.3.4 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.

Table 12.3 | Summary of limiting factors

Emitter injection efficiency

$$\gamma \approx \frac{1}{1 + \dfrac{N_B}{N_E} \cdot \dfrac{D_E}{D_B} \cdot \dfrac{x_B}{x_E}} \qquad (x_B \ll L_B), (x_E \ll L_E)$$

Base transport factor

$$\alpha_T \approx \frac{1}{1 + \dfrac{1}{2}\left(\dfrac{x_B}{L_B}\right)^2} \qquad (x_B \ll L_B)$$

Recombination factor

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

Common-base current gain

$$\alpha = \gamma\alpha_T\delta \approx \frac{1}{1 + \dfrac{N_B}{N_E} \cdot \dfrac{D_E}{D_B} \cdot \dfrac{x_B}{x_E} + \dfrac{1}{2}\left(\dfrac{x_B}{L_B}\right)^2 + \dfrac{J_{r0}}{J_{s0}} \exp\left(\dfrac{-eV_{BE}}{2kT}\right)}$$

Common-emitter current gain

$$\beta = \frac{\alpha}{1 - \alpha} \approx \frac{1}{\dfrac{N_B}{N_E} \cdot \dfrac{D_E}{D_B} \cdot \dfrac{x_B}{x_E} + \dfrac{1}{2}\left(\dfrac{x_B}{L_B}\right)^2 + \dfrac{J_{r0}}{J_{s0}}\exp\left(\dfrac{-eV_{BE}}{2kT}\right)}$$

**DESIGN
EXAMPLE 12.4**

Objective: Design the ratio of emitter doping to base doping in order to achieve an emitter injection efficiency factor of $\gamma = 0.9967$.

Consider an npn bipolar transistor. Assume, for simplicity, that $D_E = D_B$, $L_E = L_B$, and $x_E = x_B$.

■ Solution

Equation (12.35b) 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

$$\frac{N_B}{N_E} = 0.00331 \qquad \text{or} \qquad \frac{N_E}{N_B} = 302$$

■ Comment

The emitter doping concentration must be much larger than the base doping concentration to achieve a high emitter injection efficiency.

■ EXERCISE PROBLEM

Ex 12.4 Assume that transistor parameters are the same as described in Example 12.4. In addition, let $N_E = 6 \times 10^{18}$ cm^{-3}. Determine the base doping concentration such that the emitter injection efficiency is $\gamma = 0.9950$. (Ans. $N_B = 3.02 \times 10^{16}$ cm^{-3}.)

Objective: Design the base width required to achieve a base transport factor of $\alpha_T = 0.9967$.

Consider a pnp bipolar transistor. Assume that $D_B = 10$ cm²/s and $\tau_{B0} = 10^{-7}$s.

<div style="text-align: right">**DESIGN**
EXAMPLE 12.5</div>

■ 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

$$x_B = 0.814 \times 10^{-4} \text{ cm} = 0.814 \ \mu\text{m}$$

■ Comment

If the base width is less than approximately 0.8 μm, 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.

■ EXERCISE PROBLEM

Ex 12.5 Assume that transistor parameters are the same as described in Example 12.5. Determine the minimum base width x_B such that the base transport factor is $\alpha_T = 0.9980$. (Ans. $x_B = 0.633 \ \mu\text{m}$)

Objective: Determine the forward-biased B–E voltage required to achieve a recombination factor equal to $\delta = 0.9967$.

Consider an npn bipolar transistor at $T = 300$ K. Assume that $J_{r0} = 10^{-8}$ A/cm² and that $J_{s0} = 10^{-11}$ A/cm².

<div style="text-align: right">**DESIGN**
EXAMPLE 12.6</div>

■ Solution

The recombination factor, from Equation (12.44), is

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

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

$$\exp\left(\frac{+eV_{BE}}{2kT}\right) = \frac{0.9967 \times 10^3}{1 - 0.9967} = 3.02 \times 10^5$$

Then

$$V_{BE} = 2(0.0259) \ln (3.02 \times 10^5) = 0.654 \text{ V}$$

■ **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 δ will fall below the desired 0.9967 value.

■ **EXERCISE PROBLEM**

Ex 12.6   If $J_{r0} = 10^{-8}$ A/cm^2 and $J_{s0} = 10^{-11}$ A/cm^2, determine the value of V_{BE} such that $\delta = 0.9950$. (Ans. $V_{Be} = 0.6320$ V.)

EXAMPLE 12.7

Objective: 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:

$$
\begin{array}{ll}
D_E = 10 \text{ cm}^2\text{/s} & x_B = 0.70 \ \mu\text{m} \\
D_B = 25 \text{ cm}^2\text{/s} & x_E = 0.50 \ \mu\text{m} \\
\tau_{E0} = 1 \times 10^{-7} \text{ s} & N_E = 1 \times 10^{18} \text{ cm}^{-3} \\
\tau_{B0} = 5 \times 10^{-7} \text{ s} & N_B = 1 \times 10^{16} \text{ cm}^{-3} \\
J_{r0} = 5 \times 10^{-8} \text{ A/cm}^2 & V_{BE} = 0.65 \text{ V}
\end{array}
$$

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 (12.35a), is

$$\gamma = \frac{1}{1 + \frac{(2.25 \times 10^2)(10)(3.54 \times 10^{-3})}{(2.25 \times 10^4)(25)(10^{-3})} \cdot \frac{\tanh (0.0198)}{\tanh (0.050)}} = 0.9944$$

The base transport factor, from Equation (12.39a), is

$$\alpha_T = \frac{1}{\cosh\left(\frac{0.70 \times 10^{-4}}{3.54 \times 10^{-3}}\right)} = 0.9998$$

The recombination factor, from Equation (12.44), is

$$\delta = \frac{1}{1 + \frac{5 \times 10^{-8}}{J_{s0}} \exp\left(\frac{-0.65}{2(0.0259)}\right)}$$

where

$$J_{s0} = \frac{eD_B n_{B0}}{L_B \tanh\left(\frac{x_B}{L_B}\right)} = \frac{(1.6 \times 10^{-19})(25)(2.25 \times 10^4)}{3.54 \times 10^{-3} \tanh(1.977 \times 10^{-2})} = 1.29 \times 10^{-9} \text{ A/cm}^2$$

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

$$\beta = \frac{\alpha}{1 - a} = \frac{0.99406}{1 - 0.99406} = 167$$

■ Comment
In this example, the emitter injection efficiency is the limiting factor in the current gain.

■ EXERCISE PROBLEM
Ex 12.7 Assume that $\gamma = \alpha_T = 0.9980$, $J_{r0} = 5 \times 10^{-9}$ A/cm², and $J_{s0} = 2 \times 10^{-11}$ A/cm². Determine the common-emitter current gain β for (a) $V_{BE} = 0.550$ V and (b) $V_{BE} = 0.650$ V. [Ans. (a) $\beta = 98.5$; (b) $\beta = 204$]

TEST YOUR UNDERSTANDING

NOTE: In the following Test Your Understanding questions, assume a silicon npn bipolar transistor at $T = 300$ K has the following minority carrier parameters: $D_E = 8$ cm²/s, $D_B = 20$ cm²/s, $D_C = 12$ cm²/s, $\tau_{E0} = 10^{-8}$ s, $\tau_{B0} = 10^{-7}$ s, and $\tau_{C0} = 10^{-6}$ s.

TYU 12.4 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 $x_E = 2x_B = 2$ μm.
(Ans. $N_B = 1.80 \times 10^{17}$ cm⁻³)

TYU 12.5 Assume that $\alpha_T = \delta = 0.9967$, $x_B = x_E = 1$ μm, $N_B = 5 \times 10^{16}$ cm⁻³, and $N_E = 5 \times 10^{18}$ cm⁻³. Determine the common-emitter current gain β.
(Ans. $\beta = 92.4$)

TYU 12.6 Assume that $\gamma = \delta = 0.9967$ and $x_B = 0.80$ μm. Determine the common-emitter current gain β.
(Ans. $\beta = 121$)

12.4 | 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 that are not in breakdown. If any of these ideal conditions is not present, then the transistor properties will deviate from the ideal characteristics we have derived.

12.4.1   Base Width Modulation

We have implicitly assumed that the neutral base width x_B is 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-biased 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 12.21. 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 12.22. 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-V range.
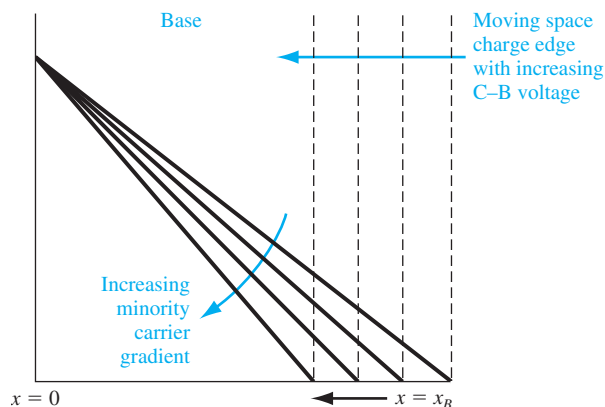

Figure 12.21 | The change in the base width and the change in the minority carrier gradient as the B–C space charge width changes.
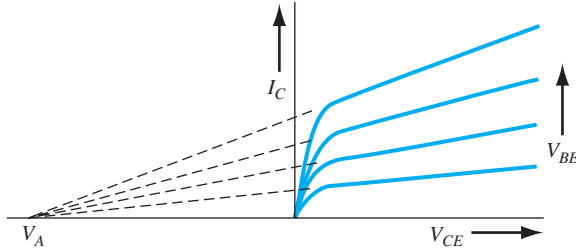
Figure 12.22 | The collector current versus collector–emitter voltage showing the Early effect and Early voltage.

From Figure 12.22, we can write that

$$\frac{dI_C}{dV_{CE}} \equiv g_o = \frac{I_C}{V_{CE} + V_A} = \frac{1}{r_o} \qquad (12.45a)$$

where V_A and V_{CE} are defined as positive quantities, g_o is defined as the output conductance, and r_o is defined as the output resistance. Equation (12.45a) can be rewritten in the form

$$I_C = g_o (V_{CE} + V_A) = \frac{1}{r_o}(V_{CE} + V_A) \qquad (12.45b)$$

showing that the collector current is now an explicit function of the collector–emitter voltage or the collector–base voltage.

Objective: Calculate the change in collector current with a change in neutral base width, and estimate the Early voltage.

EXAMPLE 12.8

Consider a uniformly doped silicon npn bipolar transistor with the following parameters: $N_B = 5 \times 10^{16}$ cm^{-3}, $N_C = 2 \times 10^{15}$ cm^{-3}, $x_{B0} = 0.70$ μm, and $D_B = 25$ cm^2/s. Assume that $x_{B0} \ll L_B$ and that $V_{BE} = 0.60$ V. The collector–base voltage is in the range $2 \le V_{CB} \le 10$ V.

■ **Solution**

Assuming $x_{B0} \ll L_B$, the excess minority carrier electron concentration in the base can be approximated by Equation (12.15b), which is

$$\delta n_B (x) \cong \frac{n_{B0}}{x_B}\left\{\left[\exp\left(\frac{V_{BE}}{V_t}\right) - 1\right](x_B - x) - x\right\}$$

The collector current is

$$|J_C| = eD_B\frac{d[\delta n_B(x)]}{dx} \cong \frac{eD_B n_{B0}}{x_B}\exp\left(\frac{V_{BE}}{V_t}\right)$$

The value of n_{B0} is found as

$$n_{B0} = \frac{n_i^2}{N_B} = \frac{(1.5 \times 10^{10})^2}{5 \times 10^{16}} = 4.5 \times 10^3 \text{ cm}^{-3}$$

For $V_{CB} = 2$ V, we find (see the following Exercise Problem Ex 12.8)

$$x_B = x_{B0} - x_{dB} = 0.70 - 0.0518 = 0.6482 \ \mu\text{m}$$

and

$$|J_C| = \frac{(1.6 \times 10^{-19})(25)(4.5 \times 10^3)}{0.6482 \times 10^{-4}} \exp\left(\frac{0.60}{0.0259}\right) = 3.195 \text{ A/cm}^2$$

For $V_{CB} = 10$ V, we find (see the following Exercise Problem Ex 12.8)

$$x_B = 0.70 - 0.103 = 0.597 \ \mu m$$

and

$$|J_C| = \frac{(1.6 \times 10^{-19})(25)(4.5 \times 10^3)}{0.597 \times 10^{-4}} \exp\left(\frac{0.60}{0.0259}\right) = 3.469 \text{ A/cm}^2$$

We now can find, from Equation (12.45a)

$$\frac{dJ_C}{dV_{CE}} = \frac{\Delta J_C}{\Delta V_{CB}} = \frac{J_C}{V_{CE} + V_A} = \frac{J_C}{V_{BE} + V_{CB} + V_A}$$

or

$$\frac{3.469 - 3.195}{8} = \frac{3.195}{0.60 + 2 + V_A}$$

The Early voltage is then determined to be

$$V_A = 90.7 \text{ V}$$

■ 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.

■ EXERCISE PROBLEM

Ex 12.8   Consider a silicon npn bipolar transistor with parameters described in Example 12.8. Determine the neutral base width for a C–B voltage of (a) $V_{CB} = 2$ V and (b) $V_{CB} = 10$ V. Neglect the B–E space charge width.

[Ans. (a) $x_B = 0.6482 \ \mu m$; (b) $x_B = 0.597 \ \mu m$]

The previous example and exercise problem demonstrate, 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.

12.4.2   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 12.23 because of the excess holes.
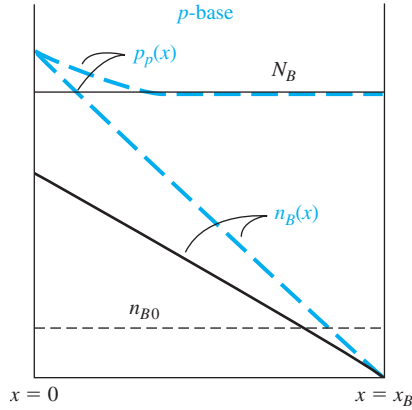
Figure 12.23 | Minority and majority carrier concentrations in the base under low and high injection (solid line: low injection; dashed line: high injection).
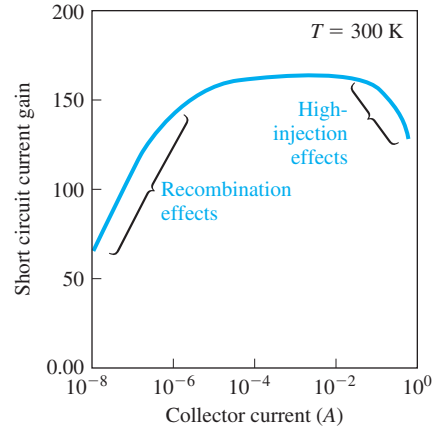

Figure 12.24 | Common-emitter current gain versus collector current. *(From Shur [14].)*

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 12.24 shows a typical common-emitter current gain versus collector current 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 now consider the second high-injection effect. At low injection, the majority carrier hole concentration at $x = 0$ for the npn transistor is

$$p_p(0) = p_{p0} = N_a \tag{12.46a}$$

and the minority carrier electron concentration is

$$n_p(0) = n_{p0} \exp\left(\frac{eV_{BE}}{kT}\right) \tag{12.46b}$$

The pn product is

$$p_p(0)n_p(0) = p_{p0}n_{p0} \exp\left(\frac{eV_{BE}}{kT}\right) \tag{12.46c}$$

At high injection, Equation (12.46c) 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

$$n_p(0) \approx n_{p0} \exp\left(\frac{eV_{BE}}{2kT}\right) \tag{12.47}$$
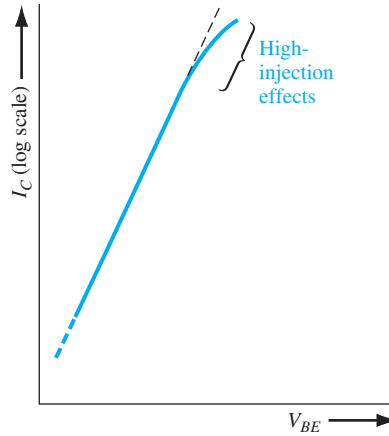
Figure 12.25 | Collector current versus base–emitter voltage showing high-injection effects.

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 12.25. The high-injection effect is very similar to the effect of a series resistance in a pn junction diode.

12.4.3   Emitter Bandgap Narrowing

Another phenomenon affecting the emitter injection efficiency is bandgap narrowing. We have implied from our previous discussion that the emitter injection efficiency factor will continue to increase and approach unity as the ratio of emitter doping to base doping continues 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 band widens, becomes skewed, and moves up toward the conduction band, eventually merging with it. At this point, the effective bandgap energy has decreased. Figure 12.26 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

$$n_i^2 = N_c N_v \exp\left(\frac{-E_g}{kT}\right) \tag{12.48}$$

In a heavily doped emitter, the intrinsic carrier concentration can be written as

$$n_{iE}^2 = N_c N_v \exp\left[\frac{-(E_{g0} - \Delta E_g)}{kT}\right] = n_i^2 \exp\left(\frac{\Delta E_g}{kT}\right) \tag{12.49}$$
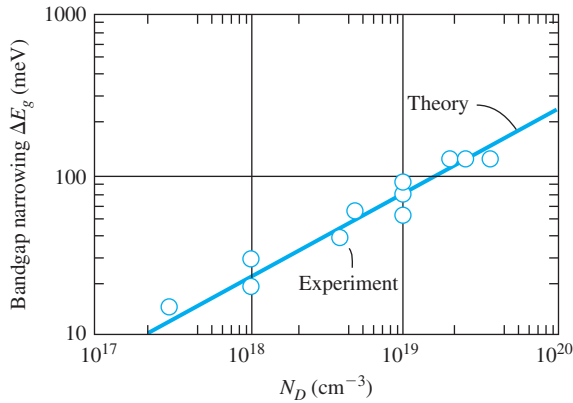
Figure 12.26 | Bandgap narrowing factor versus donor impurity concentration in silicon.
(From Sze [19].)

where E_{g0} is the bandgap energy at a low doping concentration and ΔE_g is the bandgap narrowing factor.

The emitter injection efficiency factor is given by Equation (12.35) as

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

The term p'_{E0} is the thermal-equilibrium minority carrier concentration in the emitter, taking into account bandgap narrowing, and can be written as

$$p'_{E0} = \frac{n_{iE}^2}{N_E} = \frac{n_i^2}{N_E}\exp\left(\frac{\Delta E_g}{kT}\right) \tag{12.50}$$

As the emitter doping concentration increases, ΔE_g increases; thus, p'_{E0} does not continue to decrease with increasing emitter doping N_E. If p'_{E0} starts to increase because of the bandgap narrowing, the emitter injection efficiency begins to fall off instead of continuing to increase with increased emitter doping.

Objective: Determine the increase in p_{E0} in the emitter due to bandgap narrowing.

EXAMPLE 12.9

Consider a silicon emitter at $T = 300$ K. Assume the emitter doping increases from 10^{18} to 10^{19} cm^{-3}. Determine the new value of p'_{E0} and determine the ratio p'_{E0}/p_{E0}.

■ **Solution**
For emitter doping concentrations of $N_E = 10^{18}$ 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

$$p_{E0} = \frac{n_i^2}{N_E} = \frac{(1.5 \times 10^{10})^2}{10^{19}} = 2.25 \times 10^1 \text{ cm}^{-3}$$

Taking into account the bandgap narrowing effect shown in Figure 12.26, we obtain, respectively, for $N_E = 10^{18}$ and 10^{19} cm^{-3},

$$p'_{E0} = \frac{n_i^2}{N_E} \exp\left(\frac{\Delta E_g}{kT}\right) = \frac{(1.5 \times 10^{10})^2}{10^{18}} \exp\left(\frac{0.020}{0.0259}\right) = 4.87 \times 10^2 \text{ cm}^{-3}$$

and

$$p'_{E0} = \frac{(1.5 \times 10^{10})^2}{10^{19}} \exp\left(\frac{0.080}{0.0259}\right) = 4.94 \times 10^2 \text{ cm}^{-3}$$

Taking the ratio of p'_{E0}/p_{E0} for $N_E = 10^{18}$ cm^{-3}, we find

$$\frac{p'_{E0}}{p_{E0}} = \exp\left(\frac{\Delta E_g}{kT}\right) = \exp\left(\frac{0.020}{0.0259}\right) = 2.16$$

and for $N_E = 10^{19}$ cm^{-3}, we find

$$\frac{p'_{E0}}{p_{E0}} = \exp\left(\frac{0.080}{0.0259}\right) = 21.95$$

■ Comment
If the emitter doping concentration increases from 10^{18} to 10^{19} cm^{-3}, the thermal-equilibrium minority carrier concentration actually increases rather than decreasing by a factor of 10 as would be expected.

■ EXERCISE PROBLEM
Ex 12.9 Determine the thermal-equilibrium minority carrier concentration for an emitter doping concentration of $N_E = 10^{20}$ cm^{-3} taking into account bandgap narrowing.

(Ans. $p_{E0} = 2.25$ cm^{-3}, $p'_{E0} = 1.618 \times 10^4$ cm^{-3})

As the emitter doping increases, the bandgap narrowing factor, ΔE_g, 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 shown in Figure 12.24. A very high emitter doping may result in a smaller current gain than we anticipate because of the bandgap narrowing effect.

12.4.4 Current Crowding

It is tempting to neglect 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 12.27 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
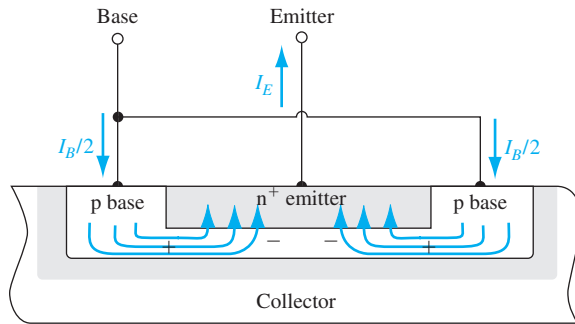
Figure 12.27 | Cross section of an npn bipolar transistor showing the base current distribution and the lateral potential drop in the base region.
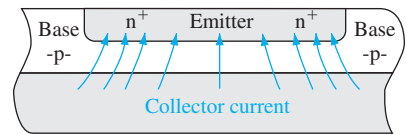

Figure 12.28 | Cross section of an npn bipolar transistor showing the emitter current crowding effect.

current crowding effect is schematically shown in Figure 12.28. The larger current density near the emitter edge may cause localized heating effects 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 be 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 12.29 shows the basic geometry. In effect, many narrow emitters are connected in parallel to achieve the required emitter area.
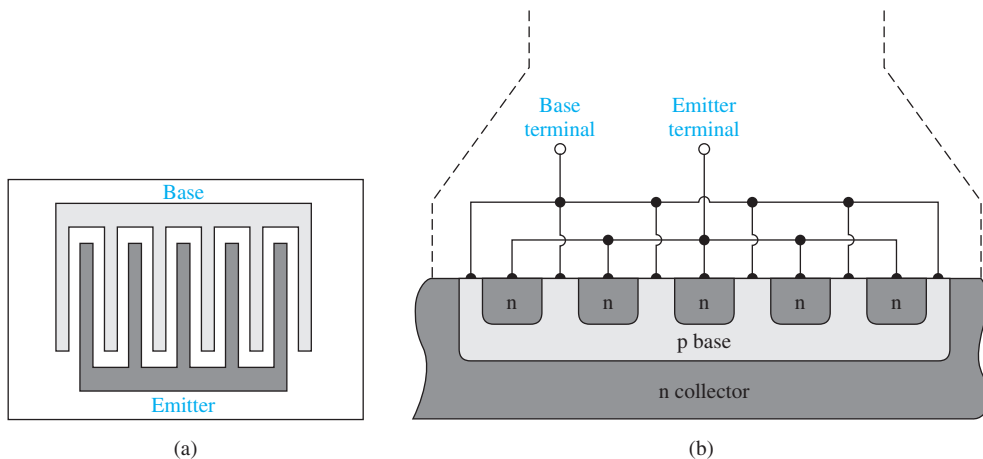

Figure 12.29 | (a) Top view and (b) cross section of an interdigitated npn bipolar transistor structure.
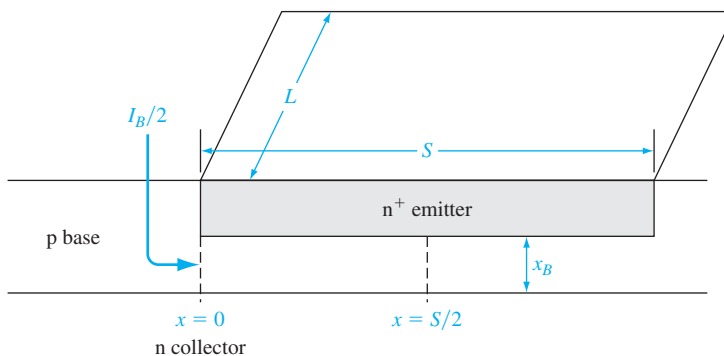
Figure 12.30 | Figure for TYU 12.7.

TEST YOUR UNDERSTANDING

TYU 12.7 Consider the geometry shown in Figure 12.30. The base doping concentra-
tion is $N_B = 10^{16}$ cm^{-3}, the neutral base width is $x_B = 0.80$ μm, the emitter
width is $S = 10$ μm, and the emitter length is $L = 10$ μm. (*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
$I_B/2 = 5$ μA, 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*) 48.83 mV; (*c*) 6.59]

*12.4.5 Nonuniform Base Doping

In the analysis of the bipolar transistor, we have assumed uniformly doped regions.
However, uniform doping rarely occurs. Figure 12.31 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.

We determined in Chapter 5 that a graded impurity concentration leads to an in-
duced electric field. For the p-type base region in thermal equilibrium, we can write

$$J_p = e\mu_p N_a E - eD_p \frac{dN_a}{dx} = 0 \tag{12.51}$$

Then

$$E = +\left(\frac{kT}{e}\right) \frac{1}{N_a} \frac{dN_a}{dx} \tag{12.52}$$

According to the example of Figure 12.31, dN_a/dx is negative; hence, the induced
electric field is in the negative x direction.

5×10^{19} N_d n-type emitter

5×10^{17} N_a p-type base N_d n-type collector
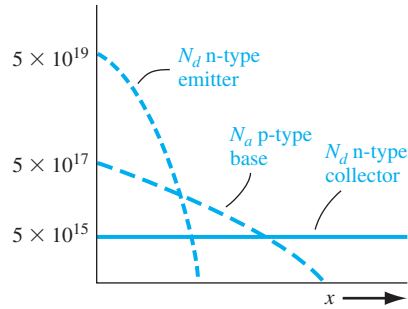
5×10^{15}

x

Figure 12.31 | Impurity concentration profiles of a double-diffused npn bipolar transistor.

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 be 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.

12.4.6 Breakdown Voltage

There are two breakdown mechanisms to consider in a bipolar transistor. The first is called punch-through. As the reverse-biased 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 12.32a shows the energy-band diagram of an npn bipolar transistor in thermal equilibrium, and Figure 12.32b shows the energy-band diagram for two values of reverse-biased 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-biased 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 barrier 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.
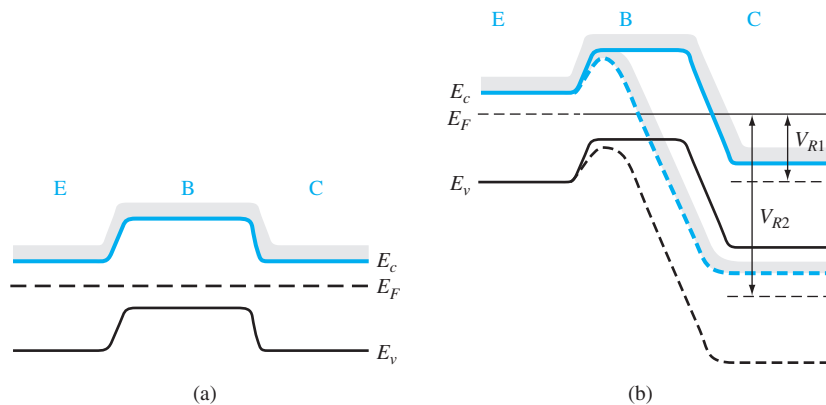
Figure 12.32 | Energy-band diagram of an npn bipolar transistor (a) in thermal equilibrium, and (b) with a reverse-biased B–C voltage before punch-through, V_{R1}, and after punch-through, V_{R2}.
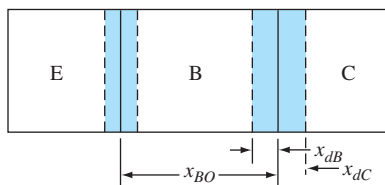

Figure 12.33 | Geometry of a bipolar transistor to calculate the punch-through voltage.

Figure 12.33 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 x_{BO} 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 narrow space charge width of a zero-biased or forward-biased B–E junction, then punch-through, assuming the abrupt junction approximation, occurs when $x_{dB} = x_{BO}$. We can write that

$$x_{dB} = x_{BO} = \left\{ \frac{2\epsilon_s(V_{bi} + V_{pt})}{e} \cdot \frac{N_C}{N_B} \cdot \frac{1}{N_C + N_B} \right\}^{1/2} \tag{12.53}$$

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

$$V_{pt} = \frac{e x_{BO}^2}{2\epsilon_s} \cdot \frac{N_B(N_C + N_B)}{N_C} \tag{12.54}$$

Objective: Design the collector doping concentration 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 $N_B = 10^{16}$ cm^{-3}. The punch-through voltage is to be $V_{pt} = 25$ V.

■ **Solution**

The maximum collector doping concentration can be determined from Equation (12.54) as

$$25 = \frac{(1.6 \times 10^{-19})(0.5 \times 10^{-4})^2(10^{16})(N_C + 10^{16})}{2(11.7)(8.85 \times 10^{-14})N_C}$$

or

$$12.94 = 1 + \frac{10^{16}}{N_C}$$

which yields

$$N_C = 8.38 \times 10^{14} \text{ cm}^{-3}$$

Using this n-type doping concentration for the collector, we can determine the minimum width of the collector region such that the depletion region extending into the collector will not reach the substrate and cause breakdown in the collector region. We have, using the results of Chapter 7,

$$x_{dC} = x_C = 5.97 \ \mu\text{m}$$

■ **Comment**

From Figure 7.15, the expected avalanche breakdown voltage for this junction is greater than 300 V. Obviously punch-through will occur before the normal breakdown voltage in this case. For a larger punch-through voltage, a larger metallurgical base 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.

■ **EXERCISE PROBLEM**

Ex 12.10   The metallurgical base width of a silicon npn bipolar transistor is $x_{B0} = 0.80 \ \mu$m. The base and collector doping concentrations are $N_B = 5 \times 10^{16}$ cm^{-3} and $N_C = 2 \times 10^{15}$ cm^{-3}, respectively. (a) Determine the punch-through voltage. (b) What is the expected avalanche breakdown voltage?

[Ans. (a) $V_{pt} = 643$ V; (b) $BV = 180$ V]

The second breakdown mechanism to consider is avalanche breakdown, but taking into account the gain of the transistor.[2] Figure 12.34a is an npn transistor with a reverse-biased voltage applied to the B–C junction and with the emitter left open. The current I_{CBO} is the reverse-biased junction current. Figure 12.34b shows the transistor with an applied C–E voltage and with the base terminal left open. This bias

---

[2]The doping concentrations in the base and collector of the transistor are small enough that Zener breakdown is not a factor to be considered.
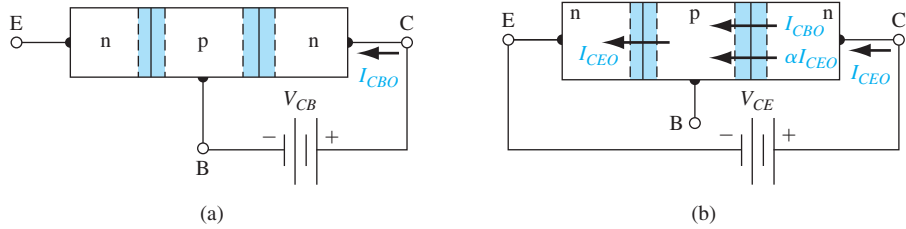
Figure 12.34 | (a) Open-emitter configuration with saturation current I_{CBO}: (b) Open-base configuration with saturation current I_{CEO}.

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 12.34b is the normal reverse-biased B–C junction current. Part 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 flow of holes into the 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

$$I_{CEO} = \alpha I_{CEO} + I_{CBO} \tag{12.55a}$$

or

$$I_{CEO} = \frac{I_{CBO}}{1 - \alpha} \approx \beta I_{CBO} \tag{12.55b}$$

where β is the common-emitter current gain. 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 12.34a, the current 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

$$M = \frac{1}{1 - (V_{CB}/BV_{CBO})^n} \tag{12.56}$$

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 12.34b, the currents in the B–C junction at breakdown are multiplied, so that

$$I_{CEO} = M(\alpha I_{CEO} + I_{CBO}) \tag{12.57}$$

Solving for I_{CEO}, we obtain

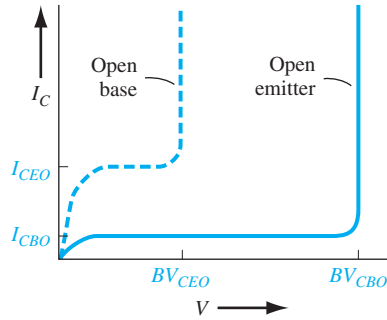$$I_{CEO} = \frac{M I_{CBO}}{1 - \alpha M} \tag{12.58}$$

Figure 12.35 | Relative breakdown voltages and saturation currents of the open-base and open-emitter configurations.

The condition for breakdown corresponds to

$$\alpha M = 1 \tag{12.59}$$

Using Equation (12.56) and assuming that $V_{CB} \approx V_{CE}$, Equation (12.59) becomes

$$\frac{\alpha}{1 - (BV_{CEO}/BV_{CBO})^n} = 1 \tag{12.60}$$

where BV_{CEO} is the C–E voltage at breakdown in the open-base configuration. Solving for BV_{CEO}, we find

$$BV_{CEO} = BV_{CBO}\sqrt[n]{1 - \alpha} \tag{12.61}$$

where, again, α is the common-base current gain. The common-emitter and common-base current gains are related by

$$\beta = \frac{\alpha}{1 - \alpha} \tag{12.62a}$$

Normally $\alpha \approx 1$, so that

$$1 - \alpha \approx \frac{1}{\beta} \tag{12.62b}$$

Then Equation (12.61) can be written as

$$BV_{CEO} = \frac{BV_{CBO}}{\sqrt[n]{\beta}} \tag{12.63}$$

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 12.35.

Objective: 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 V.

■ Solution

From Equation (12.63), 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 7.15, 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 have determined previously, an increase in breakdown voltage can be achieved by decreasing the collector doping concentration.

■ EXERCISE PROBLEM

Ex 12.11   A uniformly doped silicon bipolar transistor has base and collector doping concentrations of $N_B = 7 \times 10^{16}$ cm^{-3} and $N_C = 3 \times 10^{15}$ cm^{-3}, respectively. The common-emitter current gain is $\beta = 125$. Assuming an empirical constant of $n = 3$, determine (a) BV_{CBO} and (b) BV_{CEO}.

[Ans. (a) $BV_{CBO} \cong 125$ V; (b) $BV_{CEO} = 25$ V]

TEST YOUR UNDERSTANDING

TYU 12.8   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. $\Delta I_C = 30 \ \mu$A)

TYU 12.9   (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) $109 \leq \beta \leq 121$]

TYU 12.10   The base impurity doping concentration is $N_B = 3 \times 10^{16}$ cm^{-3} and the metallurgical base width is $x_B = 0.70 \ \mu$m. The minimum required punch-through breakdown voltage is specified to be $V_{pt} = 70$ V. What is the maximum allowed collector doping concentration? (Ans. $N_C = 5.81 \times 10^{15}$ cm^{-3})

12.5 | 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.

A detailed study of all possible models is beyond the scope of this chapter. 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—defined by 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.

*12.5.1 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 12.36 shows the current directions and voltage polarities used in the Ebers–Moll model. The currents are defined as all entering the terminals so that

$$I_E + I_B + I_C = 0 \tag{12.64}$$

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

$$I_C = \alpha_F I_F - I_R \tag{12.65a}$$

where α_F is the common-base current gain in the forward-active mode. In this mode, Equation (12.65a) becomes

$$I_C = \alpha_F I_F + I_{CS} \tag{12.65b}$$

where the current I_{CS} is the reverse-biased B–C junction current. The current I_F is given by

$$I_F = I_{ES}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \tag{12.66}$$
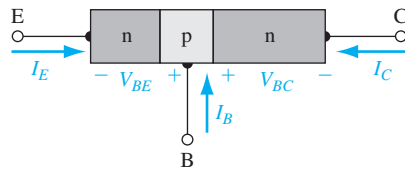

Figure 12.36 | Current direction and voltage polarity definitions for the Ebers–Moll model.

If the B–C junction becomes forward biased, such as in saturation, then we can write the current I_R as

$$I_R = I_{CS}\left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right] \tag{12.67}$$

Using Equations (12.66) and (12.67), the collector current from Equation (12.65a) can be written as

$$I_C = \alpha_F I_{ES}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] - I_{CS}\left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right] \tag{12.68}$$

We can also write the emitter current as

$$I_E = \alpha_R I_R - I_F \tag{12.69}$$

or

$$I_E = \alpha_R I_{CS}\left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right] - I_{ES}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \tag{12.70}$$

The current I_{ES} is the reverse-biased B–E junction current and α_R is the common-base current gain for the inverse-active mode. Equations (12.68) and (12.70) are the classic Ebers–Moll equations.

Figure 12.37 shows the equivalent circuit corresponding to Equations (12.68) and (12.70). The current sources in the equivalent circuit represent current components that depend on voltages across other junctions. 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

$$\alpha_F I_{ES} = \alpha_R I_{CS} \tag{12.71}$$

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.
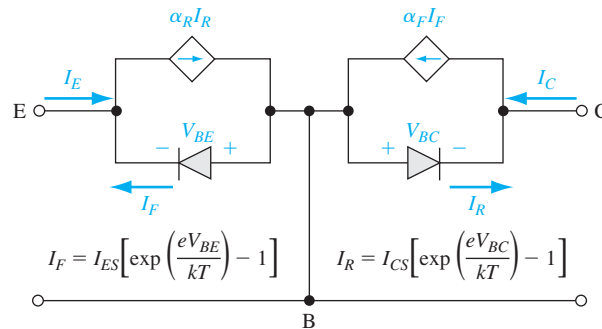

Figure 12.37 | Basic Ebers–Moll equivalent circuit.

Normally in electronic circuit applications, the collector–emitter voltage at saturation is of interest. We can define the C–E saturation voltage as

$$V_{CE}(\text{sat}) = V_{BE} - V_{BC} \tag{12.72}$$

We 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 (12.64) and (12.70), we have

$$-(I_B + I_C) = \alpha_R I_{CS}\left[\exp\left(\frac{eV_{BC}}{kT}\right) - 1\right] - I_{ES}\left[\exp\left(\frac{eV_{BE}}{kT}\right) - 1\right] \tag{12.73}$$

If we solve for $[\exp(eV_{BC}/kT) - 1]$ from Equation (12.73), and substitute the resulting expression into Equation (12.68), we can then find V_{BE} as

$$V_{BE} = V_t \ln\left[\frac{I_C(1 - \alpha_R) + I_B + I_{ES}(1 - \alpha_F\alpha_R)}{I_{ES}(1 - \alpha_F\alpha_R)}\right] \tag{12.74}$$

where V_t is the thermal voltage. Similarly, if we solve for $[\exp(eV_{BE}/kT) - 1]$ from Equation (12.68), and substitute this expression into Equation (12.73), we can find V_{BC} as

$$V_{BC} = V_t \ln\left[\frac{\alpha_F I_B - (1 - \alpha_F)I_C + I_{CS}(1 - \alpha_F\alpha_R)}{I_{CS}(1 - \alpha_F\alpha_R)}\right] \tag{12.75}$$

We may neglect the I_{ES} and I_{CS} terms in the numerators of Equations (12.74) and (12.75). Solving for $V_{CE}(\text{sat})$, we have

$$V_{CE}(\text{sat}) = V_{BE} - V_{CB} = V_t \ln\left[\frac{I_C(1 - \alpha_R) + I_B}{\alpha_F I_B - (1 - \alpha_F)I_C} \cdot \frac{I_{CS}}{I_{ES}}\right] \tag{12.76}$$

The ratio of I_{CS} to I_{ES} can be written in terms of α_F and α_R from Equation (12.71). We can finally write

$$V_{CE}(\text{sat}) = V_t \ln\left[\frac{I_C(1 - \alpha_R) + I_B}{\alpha_F I_B - (1 - \alpha_F)I_C} \cdot \frac{\alpha_F}{\alpha_R}\right] \tag{12.77}$$

Objective: Calculate the collector–emitter saturation voltage of a bipolar transistor at $T = 300$ K.

EXAMPLE 12.12

Assume that $\alpha_F = 0.99$, $\alpha_R = 0.20$, $I_C = 1$ mA, and $I_B = 50$ μA.

■ **Solution**

Substituting the parameters into Equation (12.77), 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–emitter saturation voltages. Because of the log function, $V_{CE}(\text{sat})$ is not a strong function of I_C or I_B.

■ **EXERCISE PROBLEM**

Ex 12.12   Repeat Example 12.12 for transistor parameters of $\alpha_F = 0.992$, $\alpha_R = 0.05$, $I_C = 0.5$ mA, and $I_B = 50$ μA. (Ans. V_{CE} (sat) $= 0.141$ V)

12.5.2   Gummel–Poon Model

The Gummel–Poon model of the BJT considers more physics of the transistor than the Ebers–Moll model. This model can be used if, for example, there is a nonuniform doping concentration in the base.

The electron current density in the base of an npn transistor can be written as

$$J_n = e\mu_n n(x)E + eD_n \frac{dn(x)}{dx} \tag{12.78}$$

An electric field will occur in the base if nonuniform doping exists in the base. This is discussed in Section 12.4.5. The electric field, from Equation (12.52), can be written in the form

$$E = \frac{kT}{e} \cdot \frac{1}{p(x)} \cdot \frac{dp(x)}{dx} \tag{12.79}$$

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 12.31, 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 (12.79) into Equation (12.78), we obtain

$$J_n = e\mu_n n(x) \cdot \frac{kT}{e} \cdot \frac{1}{p(x)} \cdot \frac{dp(x)}{dx} + eD_n \frac{dn(x)}{dx} \tag{12.80}$$

Using Einstein's relation, we can write Equation (12.80) in the form

$$J_n = \frac{eD_n}{p(x)} \left[n(x) \frac{dp(x)}{dx} + p(x) \frac{dn(x)}{dx} \right] = \frac{eD_n}{p(x)} \cdot \frac{d(pn)}{dx} \tag{12.81}$$

Equation (12.81) can be written in the form

$$\frac{J_n p(x)}{eD_n} = \frac{d(pn)}{dx} \tag{12.82}$$

Integrating Equation (12.82) through the base region while assuming that the electron current density is essentially a constant and the diffusion coefficient is a constant, we find

$$\frac{J_n}{eD_n} \int_0^{x_B} p(x)\,dx = \int_0^{x_B} \frac{dp(x)}{dx}\,dx = p(x_B)n(x_B) - p(0)n(0) \tag{12.83}$$

Assuming that 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 (12.83) can be written as

$$J_n = \frac{-eD_n n_i^2 \exp(V_{BE}/V_t)}{\displaystyle\int_0^{x_B} p(x)\,dx} \tag{12.84}$$

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

$$J_p = \frac{-eD_p n_i^2 \exp(V_{BE}/V_t)}{\int_0^{x_E} n(x')\,dx'} \tag{12.85}$$

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 for nonuniformly doped transistors.

The Gummel–Poon model can also take into account nonideal effects, such as the Early effect and high-level injection. As the B–C voltage changes, the neutral base width changes so that the base Gummel number Q_B changes. The change in Q_B with B–C voltage then makes the electron current density given by Equation (12.84) a function of the B–C voltage. This is the base width modulation effect or Early effect as discussed previously in Section 12.4.1.

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 (12.84), that the electron current density will also change. High-level injection has also been previously discussed in Section 12.4.2.

The Gummel–Poon model can then be used to describe the basic operation of the transistor as well as to describe nonideal effects.

12.5.3 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 12.38a shows an npn bipolar transistor in a common-emitter configuration with the small-signal terminal voltages and currents. Figure 12.38b 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, base, and emitter regions.

We can begin constructing the equivalent circuit of the transistor by considering the various terminals individually. Figure 12.39a 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
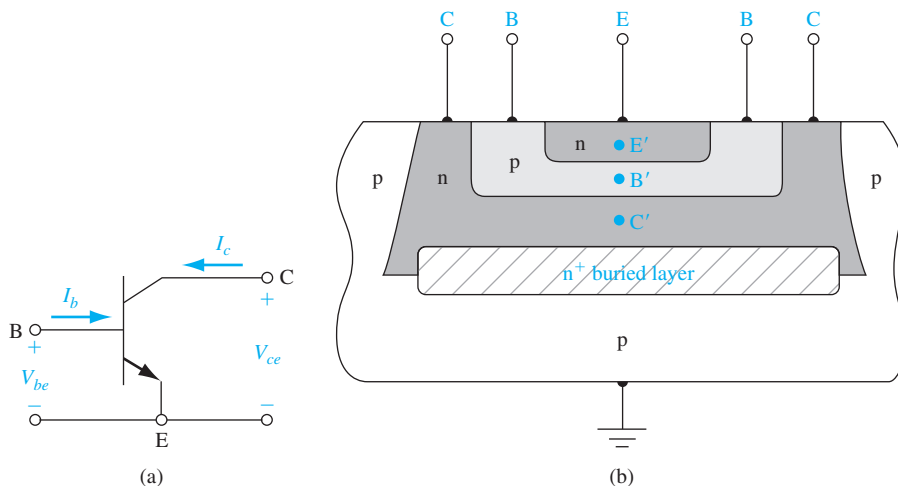
Figure 12.38 | (a) Common-emitter npn bipolar transistor with small-signal current and voltages. (b) Cross section of an npn bipolar transistor for the hybrid-pi model.
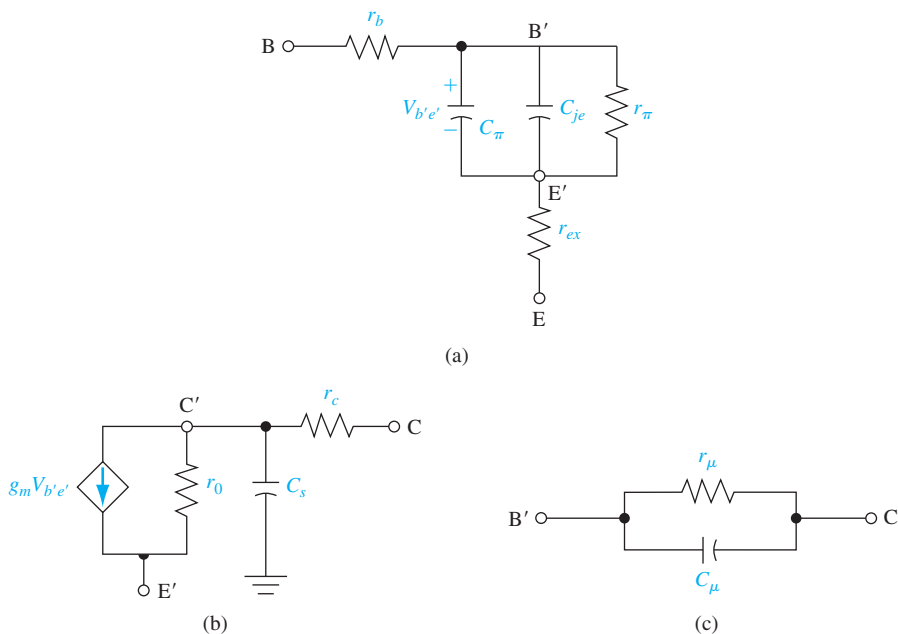

Figure 12.39 | Components of the hybrid-pi equivalent circuit between (a) the base and emitter, (b) the collector and emitter, and (c) the base and collector.

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 (8.105), and the diffusion resistance r_π is the same as the diffusion resistance r_d given by Equation (8.68). The values of both parameters are functions of the junction
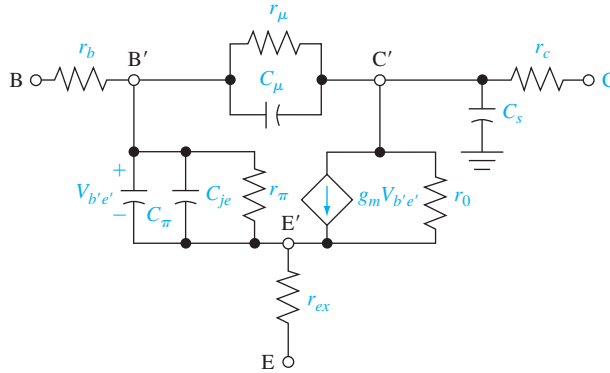
Figure 12.40 | Hybrid-pi equivalent circuit.

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 internal emitter region. This resistance is usually very small and may be on the order of 1 to 2 Ω.

Figure 12.39b 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 12.39c shows the equivalent circuit of the reverse-biased B′−C′ junction. The C_μ 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 that 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_μ between the C′ and E′ terminals at the output. However, the effect on the output characteristics can usually be ignored.

Figure 12.40 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.

Objective: Determine, to a first approximation, the frequency at which the small-signal current gain decreases to $1/\sqrt{2}$ of its low-frequency value.

Consider the simplified hybrid-pi circuit shown in Figure 12.41. We are ignoring C_μ, C_s, r_μ, C_{je}, r_0, and the series resistances. We must emphasize that this is a first-order calculation and that C_μ normally cannot be neglected.
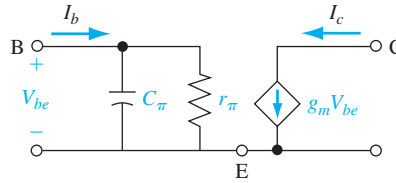
EXAMPLE 12.13

Figure 12.41 | Simplified hybrid-pi equivalent circuit.

■ 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 is then

$$|A_i| = \left| \frac{I_c}{I_b} \right| = \frac{h_{fe0}}{\sqrt{1 + (\omega r_\pi C_\pi)^2}} = \frac{h_{fe0}}{\sqrt{1 + (2\pi f r_\pi C_\pi)^2}}$$

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.

■ EXERCISE PROBLEM

Ex 12.13 Using the results of Example 12.13, determine the maximum value of C_π such that the frequency at which $|A_i| = h_{fe0}/\sqrt{2}$ is $f = 35$ MHz.

(Ans. $C_\pi = 1.75$ pF)

12.6 | FREQUENCY LIMITATIONS

The hybrid-pi equivalent circuit, developed in the last section, introduces frequency effects through the capacitor–resistor circuits. We 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.

12.6.1 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

$$\tau_{ec} = \tau_e + \tau_b + \tau_d + \tau_c \tag{12.86}$$

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 is given in Figure 12.39a. The capacitance C_{je} is the junction capacitance. If we ignore the series resistance, then the emitter–base junction capacitance charging time is

$$\boxed{\tau_e = r_e' \, (C_{je} + C_p)} \tag{12.87}$$

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

$$r_e' = \frac{kT}{e} \cdot \frac{1}{I_E} \tag{12.88}$$

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 be written as

$$J_n = -en_B(x)v(x) \tag{12.89}$$

where $v(x)$ is an average velocity. We can write

$$v(x) = dx/dt \quad \text{or} \quad dt = dx/v(x) \tag{12.90}$$

The transit time can then be found by integrating, or

$$\tau_b = \int_0^{x_B} dt = \int_0^{x_B} \frac{dx}{v(x)} = \int_0^{x_B} \frac{en_B(x)\, dx}{(-J_n)} \tag{12.91}$$

The electron concentration in the base is approximately linear (see Equation (12.15b)) so we can write

$$n_B(x) \cong n_{B0}\left[\exp\left(\frac{eV_{BE}}{kT}\right)\right]\left(1 - \frac{x}{x_B}\right) \tag{12.92}$$

and the electron current density is given by

$$J_n = eD_n \frac{dn_B(x)}{dx} \tag{12.93}$$

The base transit time is then found by combining Equations (12.92) and (12.93) with Equation (12.91). We find that

$$\boxed{\tau_b = \frac{x_B^2}{2D_n}} \tag{12.94}$$

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

$$\boxed{\tau_d = \frac{x_{dc}}{v_s}} \tag{12.95}$$

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

$$\boxed{\tau_c = r_c(C_\mu + C_s)} \tag{12.96}$$

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 are given in the next section as part of the cutoff frequency discussion.

12.6.2 Transistor Cutoff Frequency

The current gain as a function of frequency is developed in Example 12.13 so that we can also write the common-base current gain as

$$\alpha = \frac{\alpha_0}{1 + j\dfrac{f}{f_\alpha}} \tag{12.97}$$

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

$$f_\alpha = \frac{1}{2\pi\tau_{ec}} \tag{12.98}$$

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

$$\beta = \frac{\alpha}{1 - \alpha} \tag{12.99}$$

We may replace α in Equation (12.99) with the expression given by Equation (12.97). When the frequency f is of the same order of magnitude as f_α, then

$$|\beta| = \left| \frac{\alpha}{1 - \alpha} \right| \approx \frac{f_\alpha}{f} \tag{12.100}$$

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

$$\boxed{f_T = \frac{1}{2\pi\tau_{ec}}} \tag{12.101}$$

From the analysis in Example 12.13, we may also write the common-emitter current gain as

$$\beta = \frac{\beta_0}{1 + j(f/f_\beta)} \tag{12.102}$$

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 (12.99) and (12.97), we can write

$$\beta = \frac{\alpha}{1 - \alpha} = \frac{\dfrac{\alpha_0}{1 + j(f/f_T)}}{1 - \dfrac{\alpha_0}{1 + j(f/f_T)}} = \frac{\alpha_0}{1 - \alpha_0 + j(f/f_T)} \tag{12.103}$$

or

$$\beta = \frac{\alpha_0}{(1 - \alpha_0)\left[1 + j\dfrac{f}{(1 - \alpha_0)f_T} \right]} \approx \frac{\beta_0}{1 + j\dfrac{\beta_0 f}{f_T}} \tag{12.104}$$

where

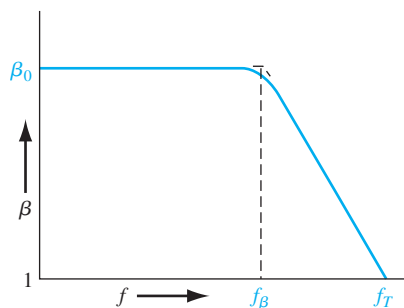$$\beta_0 = \frac{\alpha_0}{1 - \alpha_0} \approx \frac{1}{1 - \alpha_0}$$

Figure 12.42 | Bode plot of common-emitter current gain versus frequency.

Comparing Equations (12.104) and (12.102), the beta cutoff frequency is related to the cutoff frequency by

$$f_\beta \cong \frac{f_T}{\beta_0} \qquad (12.105)$$

Figure 12.42 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 12.14

Objective: Calculate the emitter-to-collector transit time and the cutoff frequency of a bipolar transistor, with the following parameters.

Consider a silicon npn transistor at $T = 300$ K. Assume the following parameters:

$$I_E = 1 \text{ mA} \qquad C_{je} = 1 \text{ pF}$$
$$x_B = 0.5 \ \mu\text{m} \qquad D_n = 25 \text{ cm}^2/\text{s}$$
$$x_{dc} = 2.4 \ \mu\text{m} \qquad r_c = 20 \ \Omega$$
$$C_\mu = 0.1 \text{ pF} \qquad C_s = 0.1 \text{ pF}$$

■ **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

$$\tau_d = \frac{x_{dc}}{v_s} = \frac{2.4 \times 10^{-4}}{10^7} = 24 \text{ ps}$$

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.

■ **EXERCISE PROBLEM**

Ex 12.14 Consider a bipolar transistor with the same parameters as described in Example 12.14 except that $I_E = 50 \ \mu\text{A}$, $C_{je} = 0.40$ pF, and $C_\mu = 0.05$ pF. Determine the emitter-to-collector transit time, the cutoff frequency, and the beta cutoff frequency.

(Ans. $\tau_{ec} = 282.2$ ps, $f_T = 564$ MHz, $f_\beta = 5.64$ MHz)

12.7 | 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.

12.7.1 Switching Characteristics

Consider an npn transistor in the circuit shown in Figure 12.43a switching from cutoff to saturation, and then switching back from saturation to cutoff. We 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 12.43b. We assume that V_{BB0} is sufficiently positive to eventually drive the transistor into saturation. For $0 \le t \le t_1$, 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
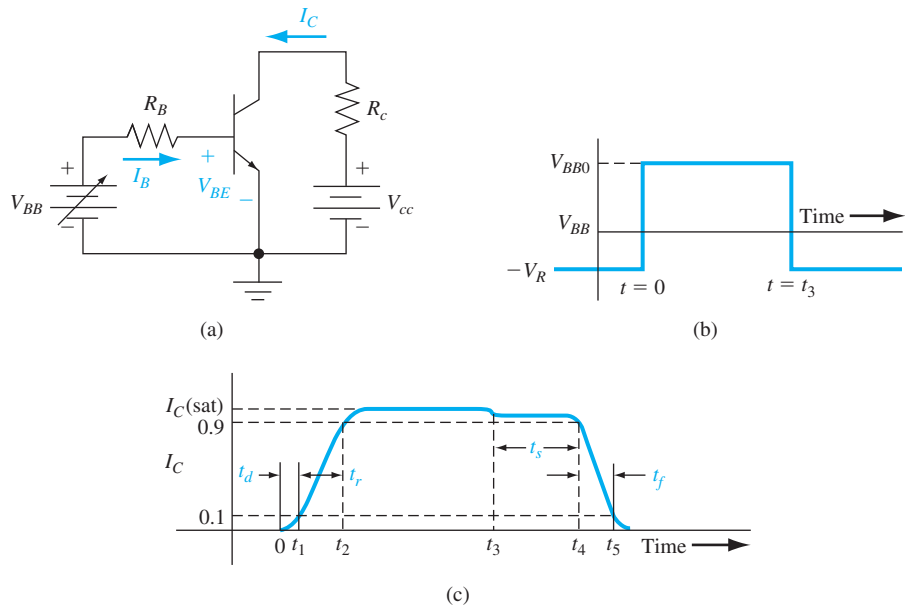
Figure 12.43 | (a) Circuit used for transistor switching. (b) Input base drive for transistor switching. (c) Collector current versus time during transistor switching.

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.

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 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 12.44 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
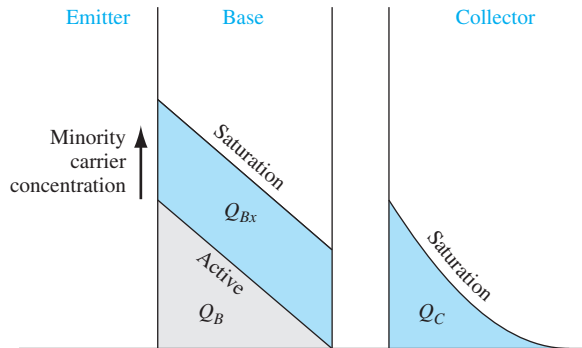
Figure 12.44 | Charge storage in the base and collector at saturation and in the active mode.

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 0 V 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 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 are presented here.

12.7.2 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 12.45a. The circuit symbol for the Schottky-clamped transistor is shown in Figure 12.45b. 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 Chapter 9 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 is shunted through the Schottky diode and away
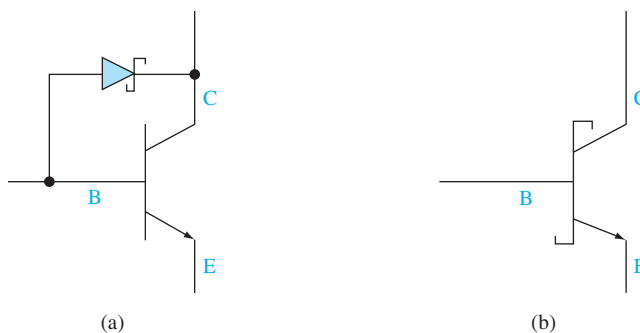
Figure 12.45 | (a) The Schottky-clamped transistor. (b) Circuit symbol of the Schottky-clamped transistor.

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 to 0.3 V, for example, the excess minority carrier concentration is reduced by over three 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.

*12.8 | 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.

12.8.1   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 12.19. 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 Table 12.3.

Figure 12.46 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.
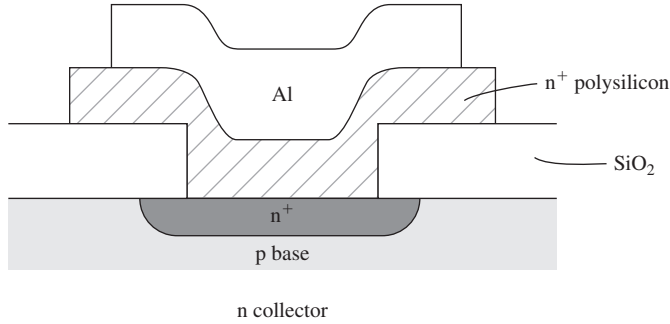
Figure 12.46 | Simplified cross section of an npn polysilicon emitter BJT.

If 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

$$eD_{E(\text{poly})} \frac{d(\delta p_{E(\text{poly})})}{dx} = eD_{E(\text{n}^+)} \frac{d(\delta p_{E(\text{n}^+)})}{dx} \tag{12.106a}$$

or

$$\frac{d(\delta p_{E(\text{n}^+)})}{dx} = \frac{D_{E(\text{poly})}}{D_{E(\text{n}^+)}} \cdot \frac{d(\delta p_{E(\text{poly})})}{dx} \tag{12.106b}$$

Since $D_{E(\text{poly})} < D_{E(\text{n}^+)}$, 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 12.47 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.
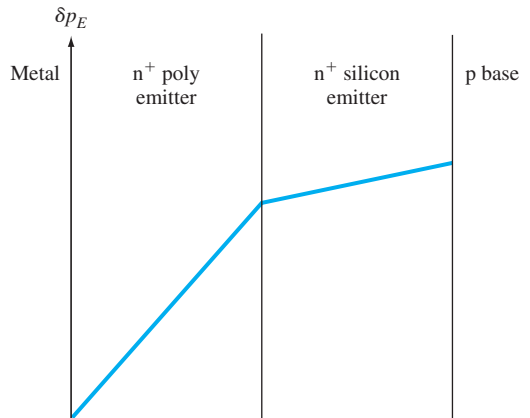

Figure 12.47 | Excess minority carrier hole concentrations in n^+ polysilicon and n^+ silicon emitter.

12.8.2   Silicon–Germanium Base Transistor

The bandgap energy of germanium (Ge) (~0.67 eV) is significantly smaller than the bandgap energy of silicon (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 12.48a 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 12.48a, are shown in Figure 12.48b. 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.
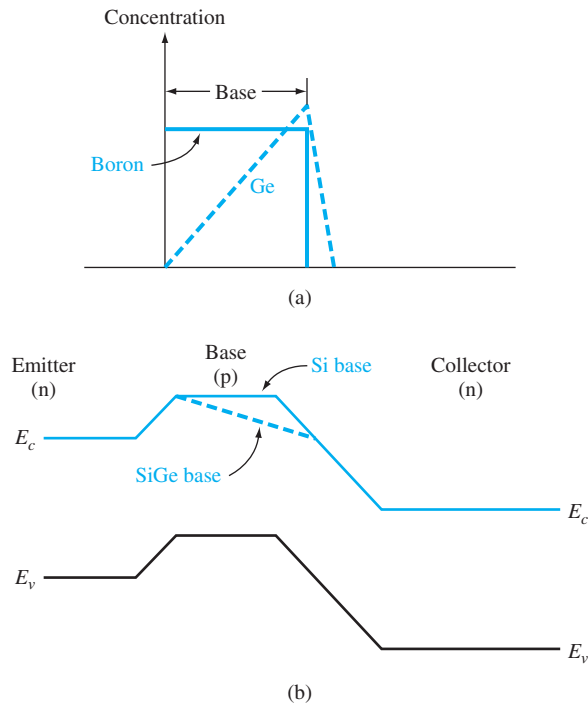

Figure 12.48 | (a) Assumed boron and germanium concentrations in the base of the SiGe-base transistor. (b) Energy-band diagram of the Si- and SiGe-base transistors.
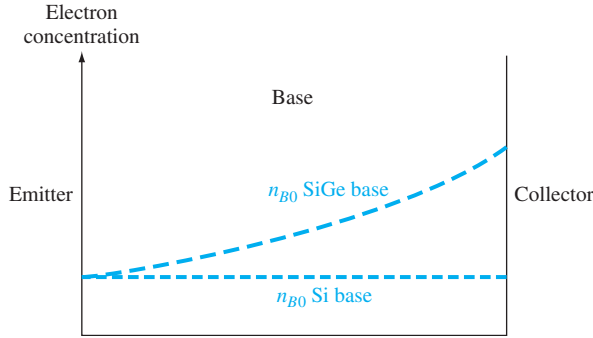
Figure 12.49 | Thermal-equilibrium minority carrier electron concentration through the base of the Si- and SiGe-base transistors.

Collector Current and Current Gain Effects   Figure 12.49 shows the thermal-equilibrium minority carrier electron concentration through the base region of the SiGe and Si transistors. This concentration is given by

$$n_{B0} = \frac{n_i^2}{N_B} \qquad (12.107)$$

where N_B is assumed to be constant. The intrinsic concentration, however, is a function of the bandgap energy. We may write

$$\frac{n_i^2(\text{SiGe})}{n_i^2(\text{Si})} = \exp\left(\frac{\Delta E_g}{kT}\right) \qquad (12.108)$$

where n_i (SiGe) is the intrinsic carrier concentration in the SiGe material, n_i(Si) is the intrinsic carrier concentration in the Si material, and ΔE_g 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 is found from Equation (12.36a), in which the derivative is evaluated at the base–collector junction. This means that the value of n_{B0} in the collector current expression in Equation (12.37) is the value at the base–collector junction. Since this value is larger for the SiGe-base transistor (Figure 12.49), 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 4.

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 (12.87), is directly proportional to the emitter diffusion resistance r'_e. This parameter is inversely proportional to the emitter current, as seen in Equation (12.88). 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.

12.8.3 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 12.50a shows a discrete aluminum gallium arsenide (AlGaAs)/gallium arsenide (GaAs) heterojunction bipolar transistor, and Figure 12.50b shows the energy-band diagram of the 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 number of minority carrier holes injected into the emitter is reduced by a factor of

$$\exp\left(\frac{\Delta E_g}{kT}\right)$$

in changing from a narrow- to wide-bandgap emitter. If $\Delta E_g = 0.30$ eV, for example, n_i^2 would be reduced by approximately 10^5 at $T = 300$ K. The drastic reduction in n_i^2 for the wide-bandgap emitter means 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 bandgap narrowing effect.

The heterojunction GaAs bipolar transistor has the potential of being a very high-frequency device. A lower emitter doping in the wide-bandgap emitter leads to
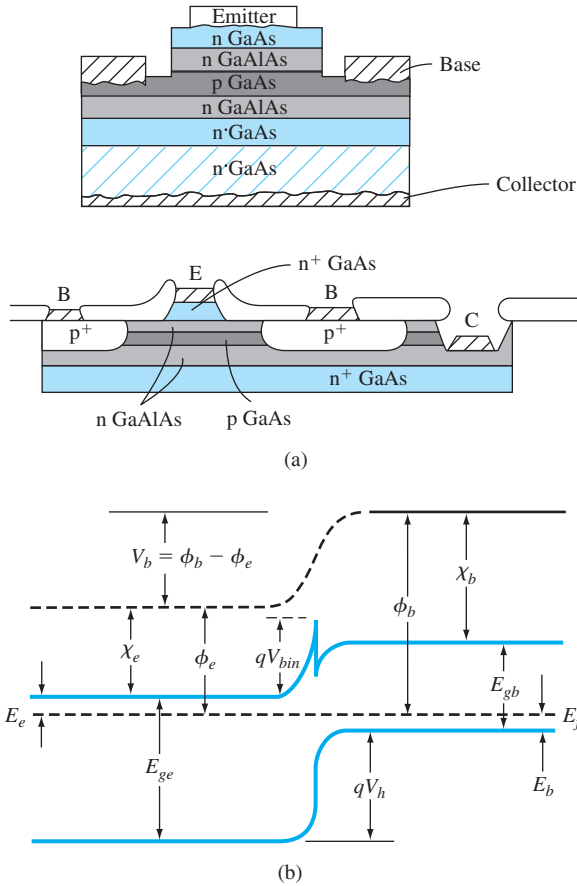
Figure 12.50 | (a) Cross section of AlGaAs/GaAs hetero-junction bipolar transistor showing a discrete and integrated structure. (b) Energy-band diagram of the n-AlGaAs emitter and p-GaAs base junction.
(From Tiwari et al. [20].)

a smaller junction capacitance, increasing the speed of the device. Also, for the GaAs npn device, the minority carriers in the base are electrons with a high mobility. The electron mobility in GaAs is approximately five times that in silicon; thus, the base transit time in the GaAs base is very short. Experimental AlGaAs/GaAs heterojunction transistors with base 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.

12.9 | 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 be interacting junctions.

■ In the forward-active mode, the B–E junction is forward biased and the B–C junction is reverse biased. Majority carriers from the emitter are injected into the base where they become minority carriers. These minority carriers diffuse across the base into the B–C space charge region where they are swept into the collector.

■ When a transistor is biased in the forward-active mode of operation, the current at one terminal of the transistor (collector current) is controlled by the voltage applied across the other two terminals of the transistor (base–emitter voltage). This is the basic transistor action.

■ The minority carrier concentrations are determined in each region of the transistor. The principal currents in the device are determined by 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 are considered:

1. Base width modulation, or Early effect—the change 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 are 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 becomes equal to unity. The frequency response 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-frequency value; also equal to the cutoff frequency.

bandgap narrowing   The reduction in the forbidden energy bandgap with high emitter doping concentration.

base 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 change 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.

common-emitter current gain   The ratio of collector current to base current.

current crowding   The nonuniform current density across the emitter junction area created 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-biased voltages are applied to both transistor junctions, resulting in zero transistor currents.

cutoff frequency   The frequency at which the magnitude of the common-emitter current gain is unity.

early effect   Another term for 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 first 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 frequency 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. What is the relation between the common-emitter and common-base current gains?
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. Why does the base doping concentration affect the emitter injection efficiency?
6. Describe base width modulation. Sketch an $I–V$ curve that shows the base width modulation 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 of 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 12.13. Assume $T=300$ K unless otherwise stated.)

Section 12.1   The Bipolar Transistor Action

12.1   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 (*b*) for the transistor biased in the forward-active region.