CHAPTER 12

The Bipolar Transistor

The 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.

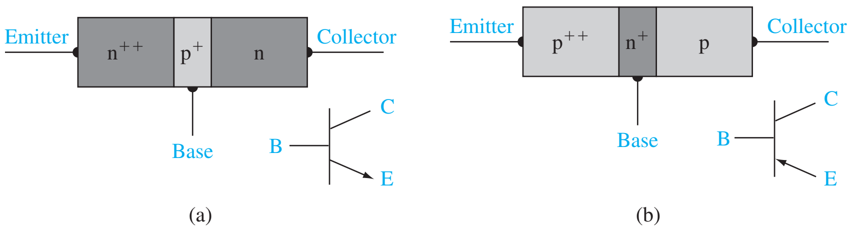

Figure 12.1 | Simplified block diagrams and circuit symbols of (a) npn and (b) pnp bipolar transistors.

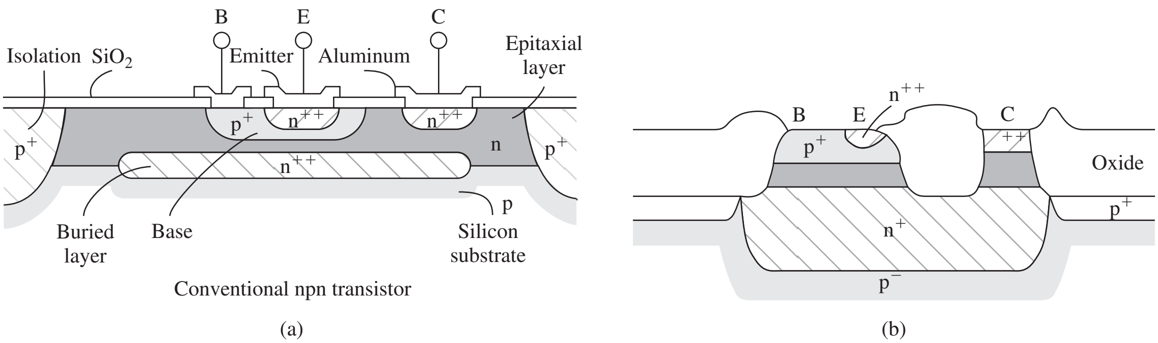

Figure 12.2 | Cross section of (a) a conventional integrated circuit npn bipolar transistor and (b) an oxide-isolated npn bipolar transistor.
(From Muller and Kamins [4].)

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

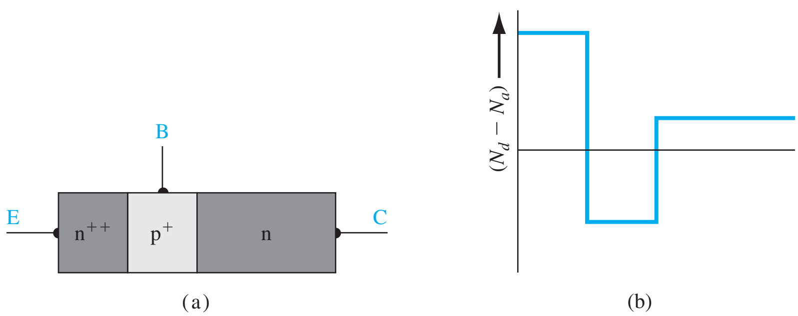

Figure 12.3 | Idealized doping profile of a uniformly doped npn bipolar transistor.

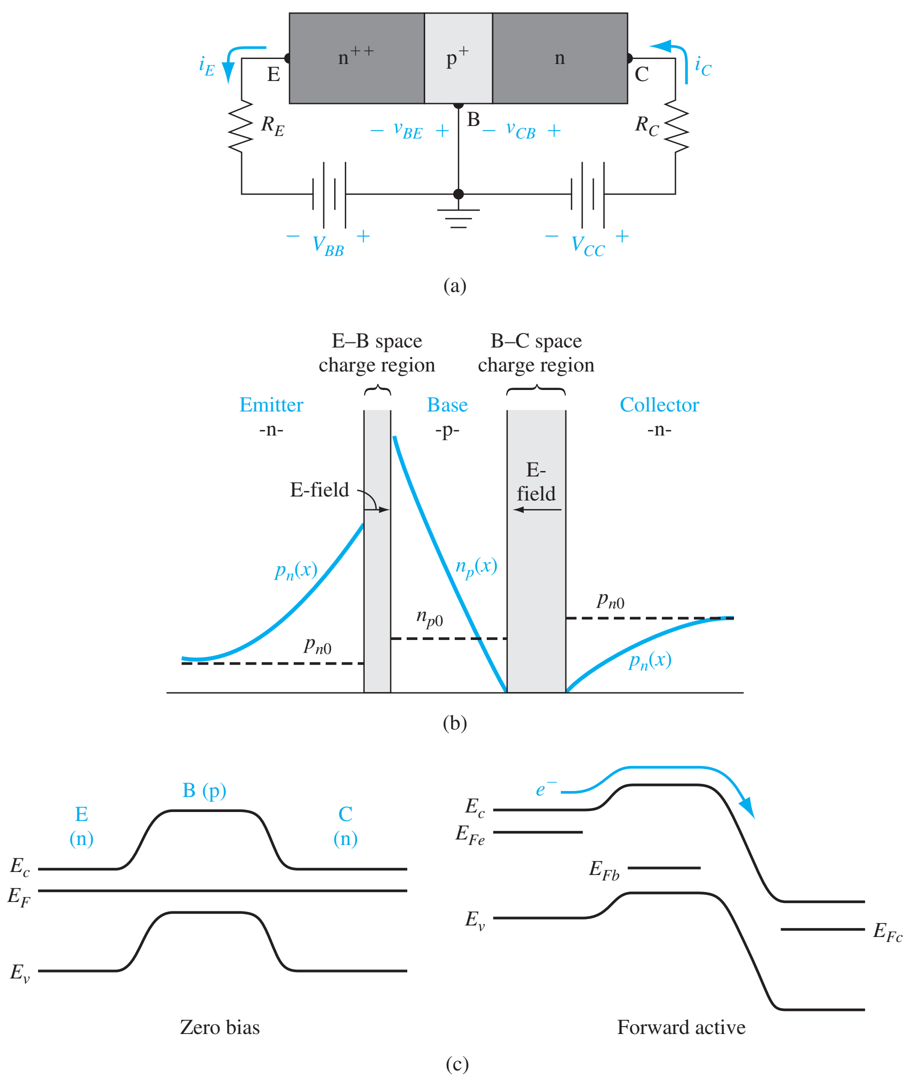

Figure 12.4 | (a) Biasing of an npn bipolar transistor in the forward-active mode, (b) minority carrier distribution in an npn bipolar transistor operating in the forward-active mode, and (c) energy-band diagram of the npn bipolar transistor under zero bias and under a forward-active mode bias.

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

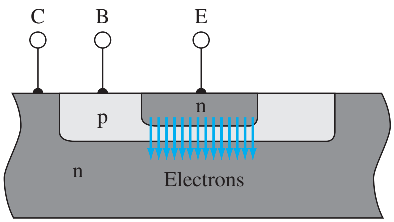

Figure 12.5 | Cross section of an npn bipolar transistor showing the injection and collection of electrons in the forward-active mode.

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

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

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

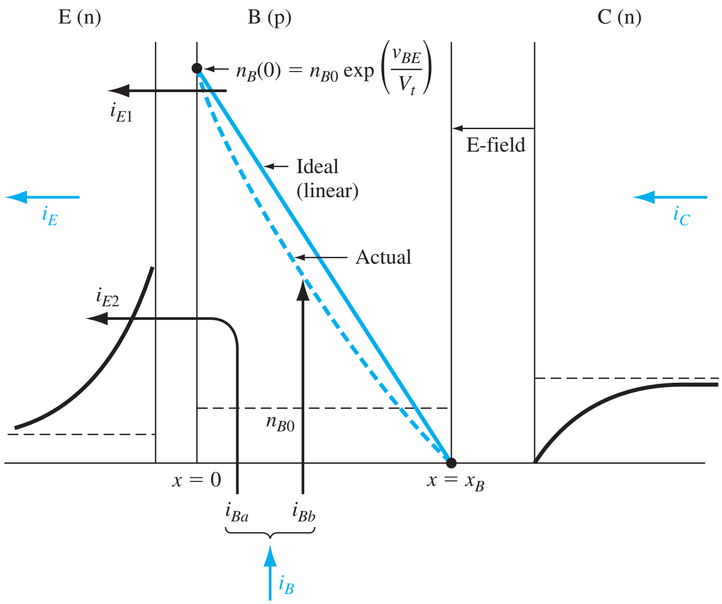

Figure 12.6 | Minority carrier distributions and basic currents in a forward-biased npn bipolar transistor.

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)

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

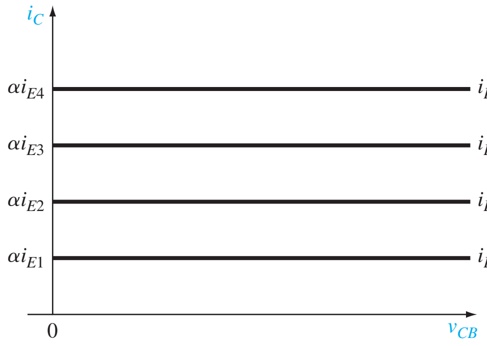

Figure 12.7 | Ideal bipolar transistor common-base current–voltage characteristics.

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 $\exp(v_{BE}/V_t)$. The total base current is the sum of i_{Ba} and i_{Bb} and is proportional to $\exp(v_{BE}/V_t)$.

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

$$\frac{i_C}{i_B} \equiv \beta \tag{12.6}$$

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

$$V_{CC} = I_C R_C + V_{CB} + V_{BE} = V_R + V_{CE} \tag{12.7}$$

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

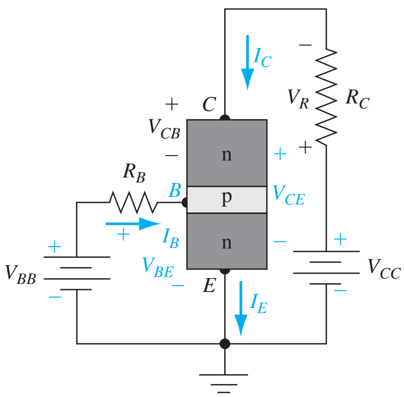

Figure 12.8 | An npn bipolar transistor in a common-emitter circuit configuration.

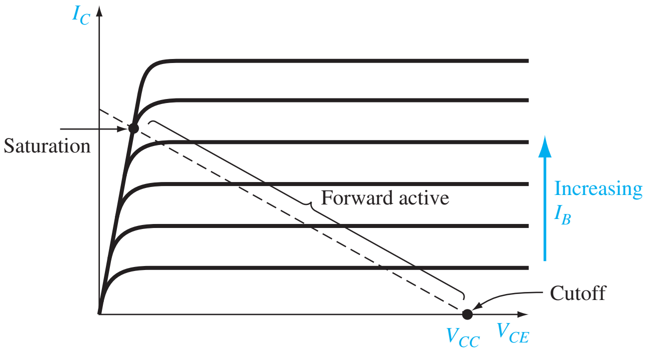

Figure 12.9 | Bipolar transistor common-emitter current–voltage characteristics with load line superimposed.

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.*

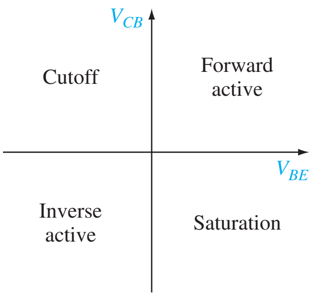

Figure 12.10 | Junction voltage conditions for the four operating modes of a bipolar transistor.

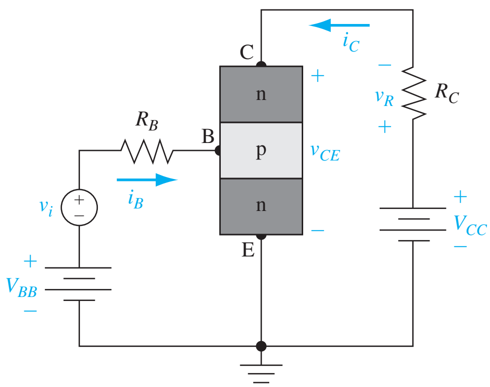

Figure 12.11 | Common-emitter npn bipolar circuit configuration with a time-varying signal voltage v_i included in the base–emitter loop.

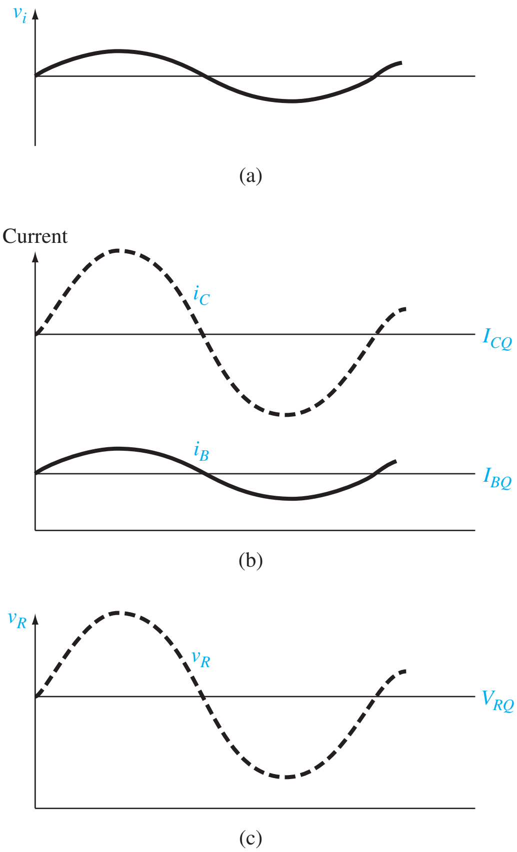

Figure 12.12 | Currents and voltages existing in the circuit shown in Figure 12.11. (a) Input sinusoidal signal voltage. (b) Sinusoidal base and collector currents superimposed on the quiescent dc values. (c) Sinusoidal voltage across the R_C resistor superimposed on the quiescent dc value.

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

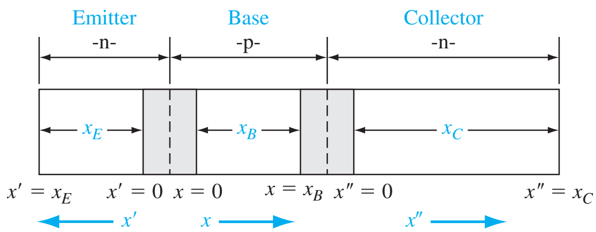

Figure 12.13 | Geometry of the npn bipolar transistor used to calculate the minority carrier distribution.

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

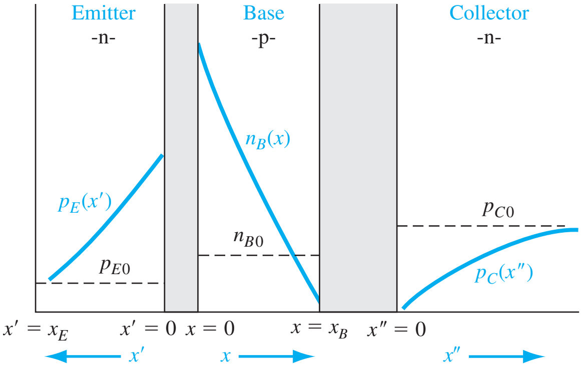

Figure 12.14 | Minority carrier distribution in an npn bipolar transistor operating in the forward-active mode.

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

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

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

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

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

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

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

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

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.*

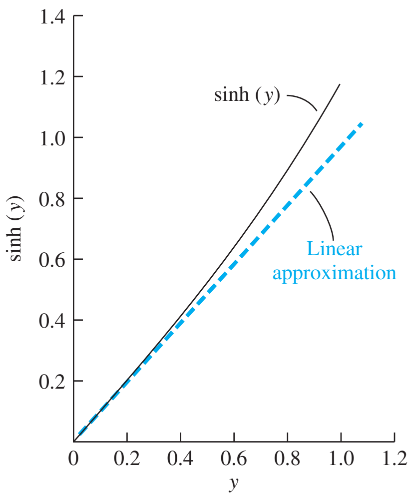

Figure 12.15 | Hyperbolic sine function and its linear approximation.

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

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

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 $x_B = 2\ \mu$m and the minority carrier diffusion length in the base is $L_B = 10\ \mu$m. Calculate the excess minority carrier concentration in the base at (*a*) $x = 0$ and (*b*) $x = x_B/2$. (*c*) Determine the ratio of the actual minority carrier concentration at $x = x_B/2$ [Equation (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)	$x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$
cosh (x)	$1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$
tanh (x)	$x - \frac{x^3}{3} + \frac{2x^5}{15} + \cdots$

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):

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

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

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

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 $x_E = 4\ \mu$m and the minority carrier diffusion length in the emitter is $L_E = 4\ \mu$m. Calculate the excess minority carrier concentration in the emitter at (*a*) $x' = 0$ and (*b*) $x' = x_E/2$.
[Ans. (*a*) 3.808×10^{12} cm^{-3}; (*b*) 1.689×10^{12} cm^{-3}]

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

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

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

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

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

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

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

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

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

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

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

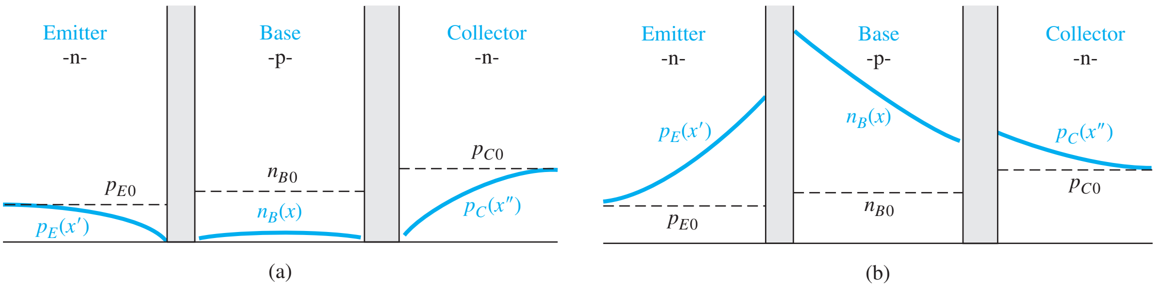

Figure 12.16 | Minority carrier distribution in an npn bipolar transistor operating in (a) cutoff and (b) saturation.

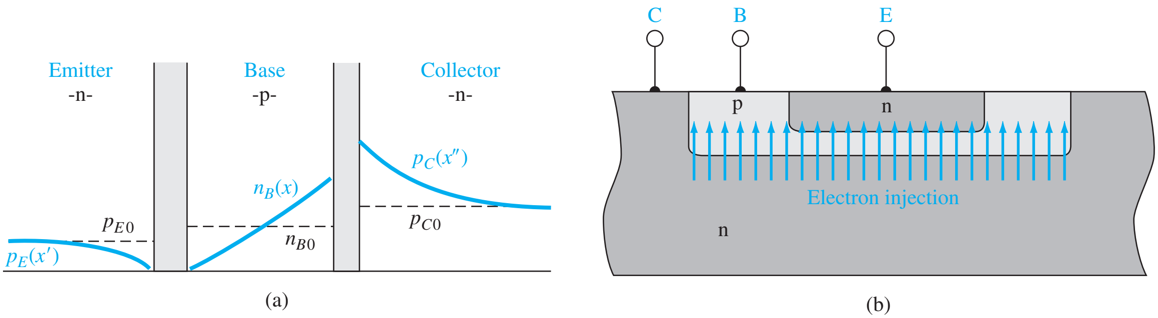

Figure 12.17 | (a) Minority carrier distribution in an npn bipolar transistor operating in the inverse-active mode. (b) Cross section of an npn bipolar transistor showing the injection and collection of electrons in the inverse-active mode.

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

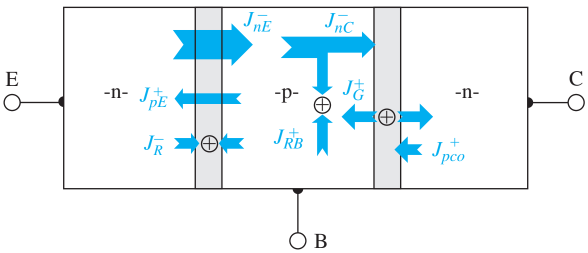

Figure 12.18 | Particle current density or flux components in an npn bipolar transistor operating in the forward-active mode.

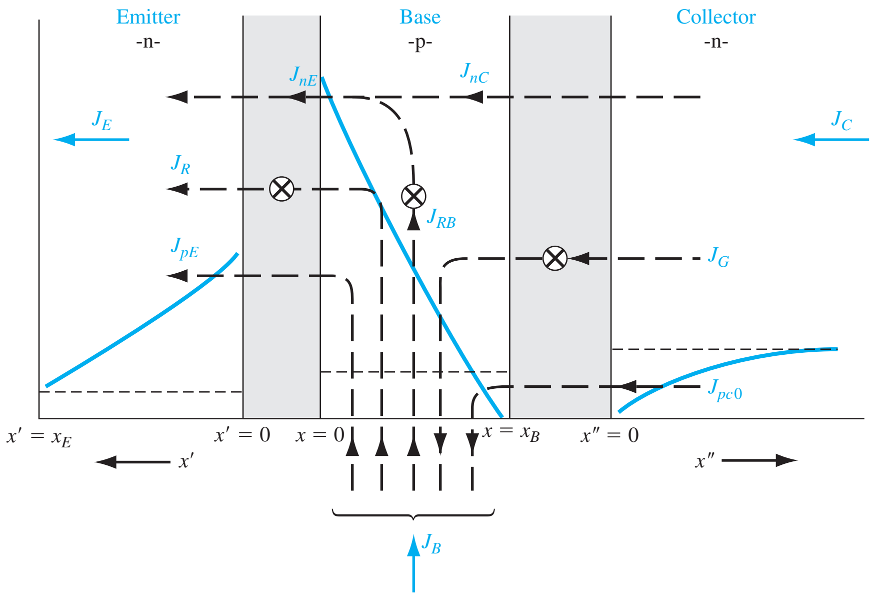

Figure 12.19 | Current density components in an npn bipolar transistor operating in the forward-active mode.

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.