**BIPOLAR JUNCTION TRANSISTOR : BJT operations, voltages and currents in npn BJT, early effect**

**Reference: "Microelectronics circuit theory and application", Sedra and Smith**

# Bipolar Junction Transistors (BJTs)

## IN THIS CHAPTER YOU WILL LEARN

1. The physical structure of the bipolar transistor and how it works.
2. How the voltage between two terminals of the transistor controls the current that flows through the third terminal, and the equations that describe these current–voltage characteristics.
3. How to analyze and design circuits that contain bipolar transistors, resistors, and dc sources.
4. How the transistor can be used to make an amplifier.
5. How to obtain linear amplification from the fundamentally nonlinear BJT.
6. The three basic ways for connecting a BJT to be able to construct amplifiers with different properties.
7. Practical circuits for bipolar-transistor amplifiers that can be constructed by using discrete components.

# 6.1 Device Structure and Physical Operation

## 6.1.1 Simplified Structure and Modes of Operation

Figure 6.1 shows a simplified structure for the BJT. A practical transistor structure will be shown later (see also Appendix A, which deals with fabrication technology).

As shown in Fig. 6.1, the BJT consists of three semiconductor regions: the emitter region (*n* type), the base region (*p* type), and the collector region (*n* type). Such a transistor is called an *npn* transistor. Another transistor, a dual of the *npn* as shown in Fig. 6.2, has a *p*-type emitter, an *n*-type base, and a *p*-type collector, and is appropriately called a *pnp* transistor.

A terminal is connected to each of the three semiconductor regions of the transistor, with the terminals labeled **emitter** (E), **base** (B), and **collector** (C).

The transistor consists of two *pn* junctions, the **emitter–base junction** (EBJ) and the **collector–base junction** (CBJ). Depending on the bias condition (forward or reverse) of each of these junctions, different modes of operation of the BJT are obtained, as shown in Table 6.1. The **active mode** is the one used if the transistor is to operate as an amplifier. Switching applications (e.g., logic circuits) utilize both the **cutoff mode** and the **saturation mode**. As the name implies, in the cutoff mode no current flows because both junctions are reverse biased.

## Introduction

In this chapter, we study the other major three-terminal device: the bipolar junction transistor (BJT). The presentation of the material in this chapter parallels but does not rely on that for the MOSFET in Chapter 5; thus, if desired, the BJT can be studied before the MOSFET.

Three-terminal devices are far more useful than two-terminal ones, such as the diodes studied in Chapter 4, because they can be used in a multitude of applications, ranging from signal amplification to the design of digital logic and memory circuits. The basic principle involved is the use of the voltage between two terminals to control the current flowing in the third terminal. In this way, a three-terminal device can be used to realize a controlled source, which as we learned in Chapter 1 is the basis for amplifier design. Also, in the extreme, the control signal can be used to cause the current in the third terminal to change from zero to a large value, thus allowing the device to act as a switch. The switch is the basis for the realization of the logic inverter, the basic element of digital circuits.

The invention of the BJT in 1948 at the Bell Telephone Laboratories ushered in the era of solid-state circuits, which led to electronics changing the way we work, play, and indeed, live. The invention of the BJT also eventually led to the dominance of information technology and the emergence of the knowledge-based economy.

The bipolar transistor enjoyed nearly three decades as the device of choice in the design of both discrete and integrated circuits. Although the MOSFET had been known

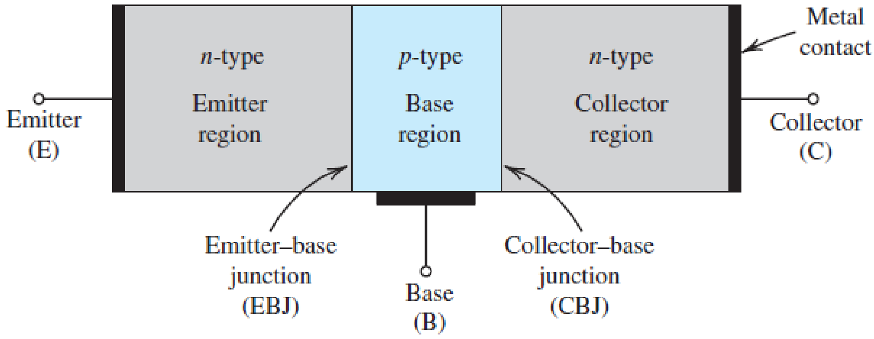



**Figure 6.1** A simplified structure of the *npn* transistor.

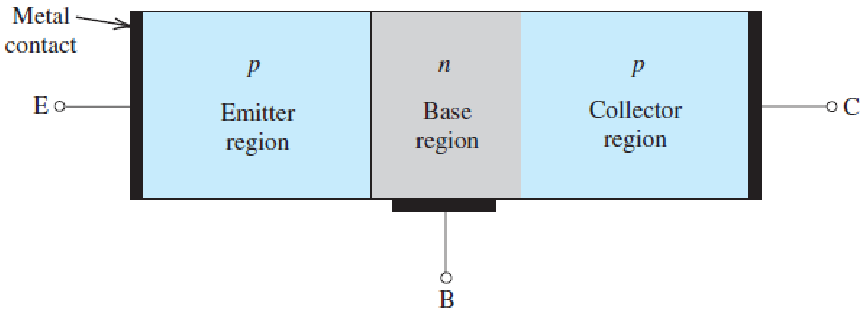



**Figure 6.2** A simplified structure of the *pnp* transistor.

**Table 6.1** BJT Modes of Operation

| Mode | EBJ | CBJ |
|---|---|---|
| Cutoff | Reverse | Reverse |
| Active | Forward | Reverse |
| Saturation | Forward | Forward |

As we will see shortly, charge carriers of both polarities—that is, electrons and holes—participate in the current-conduction process in a bipolar transistor, which is the reason for the name *bipolar*.[1]

## 6.1.2 Operation of the *npn* Transistor in the Active Mode

Of the three modes of operation of the BJT, the active mode is the most important. Therefore, we begin our study of the BJT by considering its physical operation in the active mode.[2] This situation is illustrated in Fig. 6.3 for the *npn* transistor. Two external voltage sources (shown as batteries) are used to establish the required bias conditions for active-mode operation. The voltage

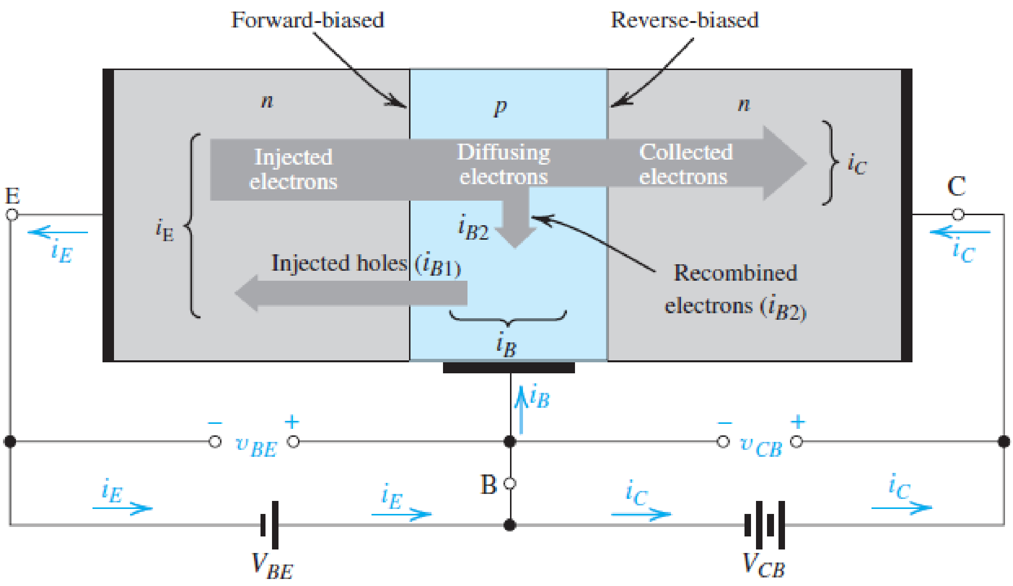



**Figure 6.3** Current flow in an *npn* transistor biased to operate in the active mode. (Reverse current components due to drift of thermally generated minority carriers are not shown.)

$V_{BE}$ causes the *p*-type base to be higher in potential than the *n*-type emitter, thus forward-biasing the emitter–base junction. The collector–base voltage $V_{CB}$ causes the *n*-type collector to be at a higher potential than the *p*-type base, thus reverse-biasing the collector–base junction.

**Current Flow** The forward bias on the emitter–base junction will cause current to flow across this junction. Current will consist of two components: electrons injected from the emitter into the base, and holes injected from the base into the emitter. As will become apparent shortly, it is highly desirable to have the first component (electrons from emitter to base) at a much higher level than the second component (holes from base to emitter). This can be accomplished by fabricating the device with a heavily doped emitter and a lightly doped base; that is, the device is designed to have a high density of electrons in the emitter and a low density of holes in the base.

The current that flows across the emitter–base junction will constitute the emitter current $i_E$, as indicated in Fig. 6.3. The direction of $i_E$ is "out of" the emitter lead, which, following the usual conventions, is in the direction of the positive-charge flow (hole current) and opposite to the direction of the negative-charge flow (electron current), with the emitter current $i_E$ being equal to the sum of these two components. However, since the electron component is much larger than the hole component, the emitter current will be dominated by the electron component.

Some of the electrons that are diffusing through the base region will combine with holes, which are the majority carriers in the base. However, since the base is usually very thin and lightly doped, the proportion of electrons "lost" through this **recombination process** will be quite small.

**The Collector Current** From the description above we see that most of the diffusing electrons will reach the boundary of the collector–base depletion region. Because the collector is more positive than the base (by $v_{CB}$ volts), these successful electrons will be swept across the CBJ depletion region into the collector. They will thus get "collected" to constitute the collector current $i_C$. Thus $i_C = I_n$, which will yield a negative value for $i_C$, indicating that $i_C$ flows in the negative direction of the $x$ axis (i.e., from right to left). Since we will take this to be the positive direction of $i_C$, we can drop the negative sign in Eq. (6.2). Doing this and substituting for $n_p(0)$ from Eq. (6.1), we can thus express the collector current $i_C$ as

$$i_C = I_S e^{v_{BE}/V_T} \tag{6.3}$$

where the **saturation current $I_S$** is given by

$$I_S = A_E q D_n n_{p0}/W$$

Substituting $n_{p0} = n_i^2/N_A$, where $n_i$ is the intrinsic carrier density and $N_A$ is the doping concentration in the base, we can express $I_S$ as

$$I_S = \frac{A_E q D_n n_i^2}{N_A W} \tag{6.4}$$

An important observation to make here is that the magnitude of $i_C$ is independent of $v_{CB}$. That is, as long as the collector is positive with respect to the base, the electrons that reach the collector side of the base region will be swept into the collector and register as collector current.

The saturation current $I_S$ is inversely proportional to the base width $W$ and is directly proportional to the area of the EBJ. Typically $I_S$ is in the range of $10^{-12}$ A to $10^{-18}$ A (depending on the size of the device). Because $I_S$ is proportional to $n_i^2$, it is a strong function of temperature, approximately doubling for every 5°C rise in temperature.

**The Base Current** The base current $i_B$ is composed of two components. The first component $i_{B1}$ is due to the holes injected from the base region into the emitter region. This current component is proportional to $e^{v_{BE}/V_T}$. The second component of base current, $i_{B2}$, is due to holes that have to be supplied by the external circuit in order to replace the holes lost from the base through the recombination process. Because $i_{B2}$ is proportional to the number of electrons injected into the base, it also will be proportional to $e^{v_{BE}/V_T}$. Thus the total base current, $i_B = i_{B1} + i_{B2}$, will be proportional to $e^{v_{BE}/V_T}$, and can be expressed as a fraction of the collector current $i_C$ as follows:

$$i_B = \frac{i_C}{\beta} \tag{6.5}$$

That is,

$$i_B = \left(\frac{I_S}{\beta}\right)e^{v_{BE}/V_T} \tag{6.6}$$

where $\beta$ is a transistor parameter.

For modern *npn* transistors, $\beta$ is in the range 50 to 200, but it can be as high as 1000 for special devices. For reasons that will become clear later, the parameter $\beta$ is called the **common-emitter current gain**.

The above description indicates that the value of $\beta$ is highly influenced by two factors: the width of the base region, $W$, and the relative dopings of the base region and the emitter region, $N_A/N_D$. To obtain a high $\beta$ (which is highly desirable since $\beta$ represents a gain parameter) the base should be thin ($W$ small) and lightly doped and the emitter heavily doped (making $N_A/N_D$ small). For modern integrated circuit fabrication technologies, $W$ is in the nanometer range.

**The Emitter Current** Since the current that enters a transistor must leave it, it can be seen from Fig. 6.3 that the emitter current $i_E$ is equal to the sum of the collector current $i_C$ and the base current $i_B$; that is,

$$i_E = i_C + i_B \tag{6.7}$$

Use of Eqs. (6.5) and (6.7) gives

$$i_E = \frac{\beta+1}{\beta}i_C \tag{6.8}$$

That is,

$$i_E = \frac{\beta+1}{\beta}I_S e^{v_{BE}/V_T} \tag{6.9}$$

Alternatively, we can express Eq. (6.8) in the form

$$i_C = \alpha i_E \tag{6.10}$$

where the constant $\alpha$ is related to $\beta$ by

$$\alpha = \frac{\beta}{\beta+1} \tag{6.11}$$

Thus the emitter current in Eq. (6.9) can be written

$$i_E = (I_S/\alpha)e^{v_{BE}/V_T} \tag{6.12}$$

Finally, we can use Eq. (6.11) to express $\beta$ in terms of $\alpha$, that is,

$$\beta = \frac{\alpha}{1-\alpha} \tag{6.13}$$

It can be seen from Eq. (6.11) that $\alpha$ is a constant (for a particular transistor) that is less than but very close to unity. For instance, if $\beta = 100$, then $\alpha \simeq 0.99$. Equation (6.13) reveals an important fact: Small changes in $\alpha$ correspond to very large changes in $\beta$. This mathematical observation manifests itself physically, with the result that transistors of the same type may have widely different values of $\beta$. For reasons that will become apparent later, $\alpha$ is called the **common-base current gain**.

### 6.1.3 Structure of Actual Transistors

Figure 6.7 shows a more realistic (but still simplified) cross section of an *npn* BJT. Note that the collector virtually surrounds the emitter region, thus making it difficult for the electrons injected into the thin base to escape being collected. In this way, the resulting $\alpha$ is close to unity and $\beta$ is large. Also, observe that the device is *not* symmetrical, and thus the emitter and collector cannot be interchanged.

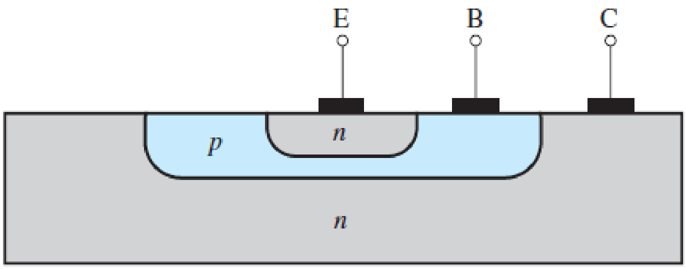



**Figure 6.7** Cross-section of an *npn* BJT.

### 6.1.4 Operation in the Saturation Mode[5]

As mentioned above, for the BJT to operate in the active mode, the CBJ must be reverse biased. Thus far, we have stated this condition for the *npn* transistor as $v_{CB} \geq 0$. However, we know that a *pn* junction does not effectively become forward biased until the forward voltage across it exceeds approximately 0.4 V. It follows that one can maintain active-mode operation of an *npn* transistor for negative $v_{CB}$ down to approximately –0.4 V. This is illustrated in Fig. 6.8, which is a sketch of $i_C$ versus $v_{CB}$ for an *npn* transistor operated with a constant emitter current $I_E$. As expected, $i_C$ is independent of $v_{CB}$ in the active mode, a situation that extends for $v_{CB}$ going negative to approximately –0.4 V. Below this value of $v_{CB}$, the CBJ begins to conduct sufficiently that the transistor leaves the active mode and enters the saturation mode of operation, where $i_C$ decreases.

To see why $i_C$ decreases in saturation, we can construct a model for the saturated *npn* transistor as follows.

$$i_C = I_S e^{v_{BE}/V_T} - I_{SC} e^{v_{BC}/V_T} \tag{6.14}$$

The second term will play an increasing role as $v_{BC}$ exceeds 0.4 V or so, causing $i_C$ to decrease and eventually reach zero.

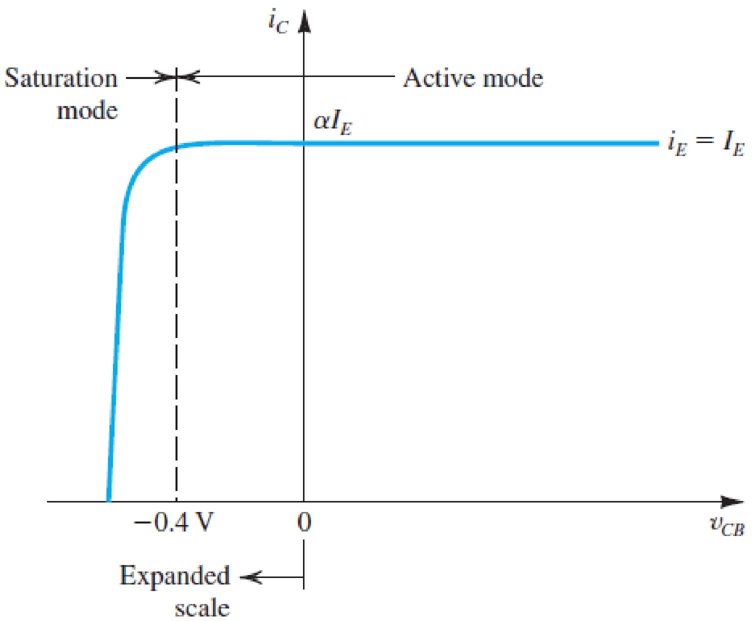



**Figure 6.8** The $i_C$–$v_{CB}$ characteristic of an *npn* transistor fed with a constant emitter current $I_E$. The transistor enters the saturation mode of operation for $v_{CB} < -0.4$ V, and the collector current diminishes.

## 6.1.5 The *pnp* Transistor

The *pnp* transistor operates in a manner similar to that of the *npn* device described above. Figure 6.10 shows a *pnp* transistor biased to operate in the active mode. Here the voltage $V_{EB}$ causes the *p*-type emitter to be higher in potential than the *n*-type base, thus forward-biasing the emitter–base junction. The collector–base junction is reverse biased by the voltage $V_{BC}$, which keeps the *p*-type collector lower in potential than the *n*-type base.

Unlike the *npn* transistor, current in the *pnp* device is mainly conducted by holes injected from the emitter into the base as a result of the forward-bias voltage $V_{EB}$. Since the component of emitter current contributed by electrons injected from base to emitter is kept small by using a lightly doped base, most of the emitter current will be due to holes. The electrons injected from base to emitter give rise to the first component of base current, $i_{B1}$. Also, a number of the holes injected into the base will recombine with the majority carriers in the base (electrons) and will thus be lost. The disappearing base electrons will have to be replaced from the external circuit, giving rise to the second component of base current, $i_{B2}$. The holes that succeed in reaching the boundary of the depletion region of the collector–base junction will be attracted by the negative voltage on the collector. Thus these holes will be swept across the depletion region into the collector and appear as collector current.

It can easily be seen from the above description that the current–voltage relationship of the *pnp* transistor will be identical to that of the *npn* transistor except that $v_{BE}$ has to be replaced by $v_{EB}$. Also, the large-signal, active-mode operation of the *pnp* transistor can be modeled by any of four equivalent circuits similar to those for the *npn* transistor in Fig. 6.5. Two of these four circuits are shown in Fig. 6.11. Finally, we note that the *pnp* transistor can operate in the saturation mode in a manner analogous to that described for the *npn* device.

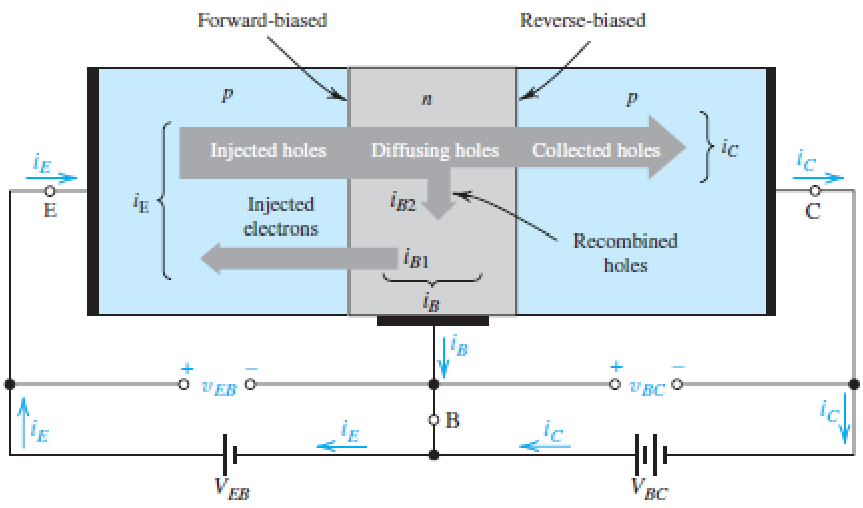



**Figure 6.10** Current flow in a *pnp* transistor biased to operate in the active mode.

# 6.2 Current–Voltage Characteristics

## 6.2.1 Circuit Symbols and Conventions

The physical structure used thus far to explain transistor operation is rather cumbersome to employ in drawing the schematic of a multitransistor circuit. Fortunately, a very descriptive and convenient circuit symbol exists for the BJT. Figure 6.12(a) shows the symbol for the *npn* transistor; the *pnp* symbol is given in Fig. 6.12(b). In both symbols the emitter is distinguished by an arrowhead. This distinction is important because, as we have seen in the last section, practical BJTs are not symmetric devices.

The polarity of the device—*npn* or *pnp*—is indicated by the direction of the arrowhead on the emitter. This arrowhead points in the direction of normal current flow in the emitter, which is also the forward direction of the base–emitter junction. Since we have adopted a drawing convention by which currents flow from top to bottom, we will always draw *pnp* transistors in the manner shown in Fig. 6.12(b) (i.e., with their emitters on top).

Figure 6.13 shows *npn* and *pnp* transistors biased to operate in the active mode. It should be mentioned in passing that the biasing arrangement shown, utilizing two dc voltage sources,

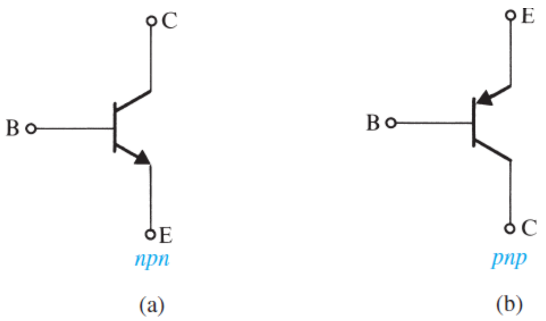



**Figure 6.12** Circuit symbols for BJTs.

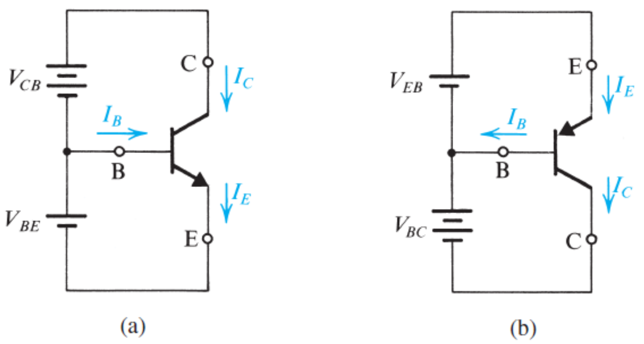



**Figure 6.13** Voltage polarities and current flow in transistors biased in the active mode.

is not a usual one and is used here merely to illustrate operation. Practical biasing schemes will be presented in Section 6.7. Figure 6.13 also indicates the reference and actual directions of current flow throughout the transistor. Our convention will be to take the reference direction to coincide with the normal direction of current flow. Hence, normally, we should not encounter a negative value for $i_E$, $i_B$, or $i_C$.

The convenience of the circuit-drawing convention that we have adopted should be obvious from Fig. 6.13. Note that currents flow from top to bottom and that voltages are higher at the top and lower at the bottom. The arrowhead on the emitter also implies the polarity of the emitter–base voltage that should be applied in order to forward bias the emitter–base junction. Just a glance at the circuit symbol of the *pnp* transistor, for example, indicates that we should make the emitter higher in voltage than the base (by $v_{EB}$) in order to cause current to flow into the emitter (downward). Note that the symbol $v_{EB}$ means the voltage by which the emitter (E) is higher than the base (B). Thus for a *pnp* transistor operating in the active mode $v_{EB}$ is positive, while in an *npn* transistor $v_{BE}$ is positive.

an *npn* transistor whose EBJ is forward biased will operate in the active mode *as long as the collector voltage does not fall below that of the base by more than approximately 0.4 V.* Otherwise, the transistor leaves the active mode and enters the saturation region of operation.[6]

**Table 6.2** Summary of the BJT Current–Voltage Relationships in the Active Mode

$$i_C = I_S e^{v_{BE}/V_T}$$

$$i_B = \frac{i_C}{\beta} = \left(\frac{I_S}{\beta}\right) e^{v_{BE}/V_T}$$

$$i_E = \frac{i_C}{\alpha} = \left(\frac{I_S}{\alpha}\right) e^{v_{BE}/V_T}$$

*Note:* For the *pnp* transistor, replace $v_{BE}$ with $v_{EB}$.

$$i_C = \alpha i_E \qquad i_B = (1-\alpha) i_E = \frac{i_E}{\beta + 1}$$

$$i_C = \beta i_B \qquad i_E = (\beta + 1) i_B$$

$$\beta = \frac{\alpha}{1-\alpha} \qquad \alpha = \frac{\beta}{\beta + 1}$$

$$V_T = \text{thermal voltage} = \frac{kT}{q} \simeq 25 \text{ mV at room temperature}$$

**The Collector–Base Reverse Current ($I_{CBO}$)** In our discussion of current flow in transistors we ignored the small reverse currents carried by thermally generated minority carriers. Although such currents can be safely neglected in modern transistors, the reverse current across the collector–base junction deserves some mention. This current, denoted $I_{CBO}$, is the reverse current flowing from collector to base with the emitter open-circuited (hence the subscript $O$). This current is usually in the nanoampere range, a value that is many times higher than its theoretically predicted value. As with the diode reverse current, $I_{CBO}$ contains a substantial leakage component, and its value is dependent on $v_{CB}$. $I_{CBO}$ depends strongly on temperature, approximately doubling for every 10°C rise.[7]

## 6.2.2 Graphical Representation of Transistor Characteristics

It is sometimes useful to describe the transistor $i$–$v$ characteristics graphically. Figure 6.15 shows the $i_C$–$v_{BE}$ characteristic, which is the exponential relationship

$$i_C = I_S e^{v_{BE}/V_T}$$

which is identical to the diode $i$–$v$ relationship. The $i_E$–$v_{BE}$ and $i_B$–$v_{BE}$ characteristics are also exponential but with different scale currents: $I_S/\alpha$ for $i_E$, and $I_S/\beta$ for $i_B$. Since the constant of the exponential characteristic, $1/V_T$, is quite high ($\simeq 40$), the curve rises very sharply. For $v_{BE}$ smaller than about 0.5 V, the current is negligibly small.[8] Also, over most of the normal current range $v_{BE}$ lies in the range of 0.6 V to 0.8 V. In performing rapid first-order dc calculations, we normally will assume that $V_{BE} \simeq 0.7$ V, which is similar to the approach used in the analysis of diode circuits (Chapter 4). For a *pnp* transistor, the $i_C$–$v_{EB}$ characteristic will look identical to that of Fig. 6.15 with $v_{BE}$ replaced with $v_{EB}$.

As in silicon diodes, the voltage across the emitter–base junction decreases by about 2 mV for each rise of 1°C in temperature, provided the junction is operating at a constant current. Figure 6.16 illustrates this temperature dependence by depicting $i_C$–$v_{BE}$ curves for an *npn* transistor at three different temperatures.

**BIPOLAR JUNCTION TRANSISTOR : BJT operations, voltages and currents in npn BJT, early effect**

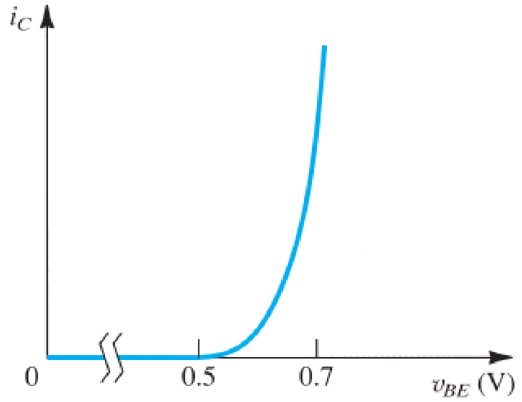



**Figure 6.15** The $i_C$–$v_{BE}$ characteristic for an *npn* transistor.

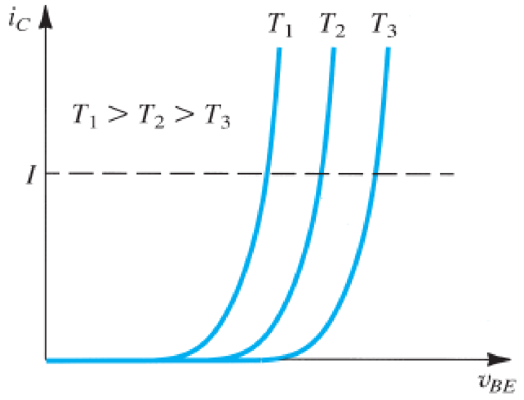
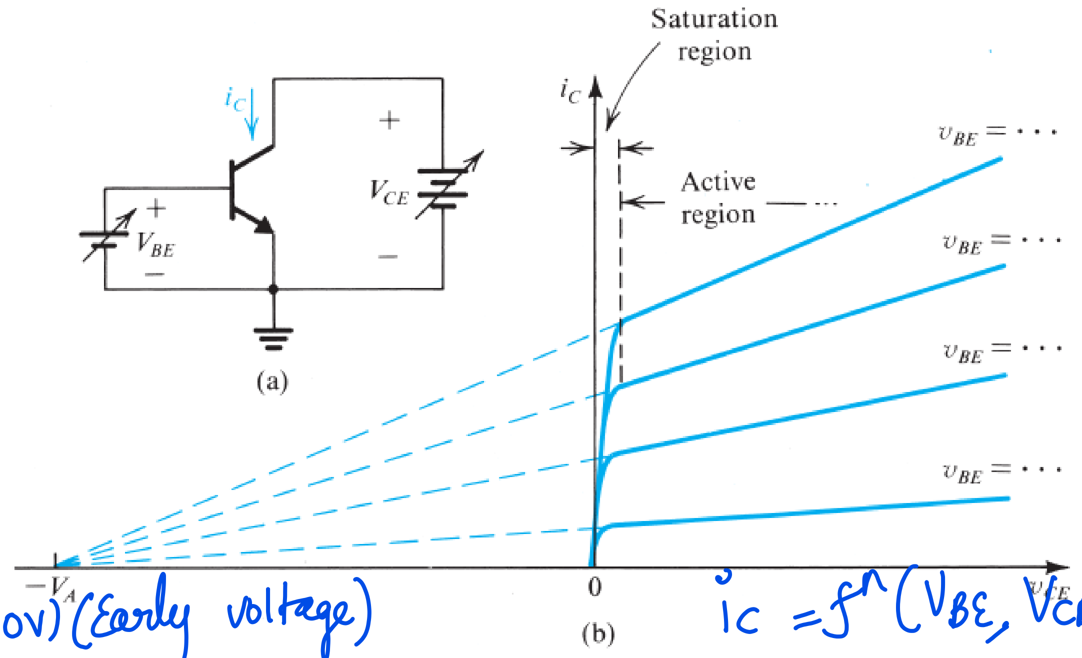



**Figure 6.16** Effect of temperature on the $i_C$–$v_{BE}$ characteristic. At a constant emitter current (broken line), $v_{BE}$ changes by $-2$ mV/°C.

## 6.2.3 Dependence of $i_C$ on the Collector Voltage—The Early Effect

When operated in the active region, practical BJTs show some dependence of the collector current on the collector voltage, with the result that, unlike the graph shown in Fig. 6.8, their $i_C$–$v_{CB}$ characteristics are not perfectly horizontal straight lines. To see this dependence more clearly, consider the conceptual circuit shown in Fig. 6.17(a). The transistor is connected in




(-100V) (Early voltage)

$i_C = f^n(V_{BE}, V_{CE})$

**Figure 6.17** (a) Conceptual circuit for measuring the $i_C$–$v_{CE}$ characteristics of the BJT. (b) The $i_C$–$v_{CE}$ characteristics of a practical BJT.

the **common-emitter configuration**; that is, here the emitter serves as a common terminal between the input and output ports. The voltage $V_{BE}$ can be set to any desired value by adjusting the dc source connected between base and emitter. At each value of $V_{BE}$, the corresponding $i_C$–$v_{CE}$ characteristic curve can be measured point by point by varying the dc source connected between collector and emitter and measuring the corresponding collector current. The result is the family of $i_C$–$v_{CE}$ characteristic curves shown in Fig. 6.17(b) and known as **common-emitter characteristics**.

At low values of $v_{CE}$ (lower than about 0.3 V), as the collector voltage goes below that of the base by more than 0.4 V, the collector–base junction becomes forward biased and the transistor leaves the active mode and enters the saturation mode. Shortly, we shall look at the details of the $i_C$–$v_{CE}$ curves in the saturation region. At this time, however, we wish to examine the characteristic curves in the active region in detail. We observe that the characteristic curves, though still straight lines, have finite slope. In fact, when extrapolated, the characteristic lines meet at a point on the negative $v_{CE}$ axis, at $v_{CE} = -V_A$. The voltage $V_A$, a positive number, is a parameter for the particular BJT, with typical values in the range of 10 V to 100 V. It is called the **Early voltage**, after J. M. Early, the engineering scientist who first studied this phenomenon.

At a given value of $v_{BE}$, increasing $v_{CE}$ increases the reverse-bias voltage on the collector–base junction, and thus increases the width of the depletion region of this junction (refer to Fig. 6.3). This in turn results in a decrease in the **effective base width** $W$. Recalling that $I_S$ is inversely proportional to $W$ (Eq. 6.4), we see that $I_S$ will increase and that $i_C$ increases proportionally. This is the Early effect. For obvious reasons, it is also known as the **base-width moduation effect**.[9]

The linear dependence of $i_C$ on $v_{CE}$ can be explicitly accounted for by assuming that $I_S$ remains constant and including the factor $(1 + v_{CE}/V_A)$ in the equation for $i_C$ as follows:

$$i_C = I_S e^{v_{BE}/V_T}\left(1 + \frac{v_{CE}}{V_A}\right) \tag{6.18}$$

The nonzero slope of the $i_C$–$v_{CE}$ straight lines indicates that the **output resistance** looking into the collector is not infinite. Rather, it is finite and defined by

$$r_o \equiv \left[\left.\frac{\partial i_C}{\partial v_{CE}}\right|_{v_{BE} = \text{constant}}\right]^{-1} \tag{6.19}$$

Using Eq. (6.18) we can show that

$$r_o = \frac{V_A + V_{CE}}{I_C} \tag{6.20}$$

where $I_C$ and $V_{CE}$ are the coordinates of the point at which the BJT is operating on the particular $i_C$–$v_{CE}$ curve (i.e., the curve obtained for $v_{BE}$ equal to constant value $V_{BE}$ at which Eq. (6.19) is evaluated). Alternatively, we can write

$$r_o = \frac{V_A}{I'_C} \tag{6.21}$$

where $I'_C$ is the value of the collector current with the Early effect neglected; that is,

$$I'_C = I_S e^{V_{BE}/V_T} \tag{6.22}$$

The finite output resistance $r_o$ can have a significant effect on the gain of transistor amplifiers.