CHAPTER OBJECTIVES

This chapter introduces the bipolar junction transistor (BJT) operation and then presents the theory of the bipolar transistor I-V characteristics, current gain, and output conductance. High-level injection and heavy doping induced band narrowing are introduced. SiGe transistor, transit time, and cutoff frequency are explained. Several bipolar transistor models are introduced, i.e., Ebers–Moll model, small-signal model, and charge control model. Each model has its own areas of applications.

The bipolar junction transistor or BJT was invented in 1948 at Bell Telephone Laboratories, New Jersey, USA. It was the first mass produced transistor, ahead of the MOS field-effect transistor (MOSFET) by a decade. After the introduction of metal-oxide-semiconductor (MOS) ICs around 1968, the high-density and low-power advantages of the MOS technology steadily eroded the BJT’s early dominance. BJTs are still preferred in some high-frequency and analog applications because of their high speed, low noise, and high output power advantages such as in some cell phone amplifier circuits. When they are used, a small number of BJTs are integrated into a high-density complementary MOS (CMOS) chip. Integration of BJT and CMOS is known as the BiCMOS technology.

The term bipolar refers to the fact that both electrons and holes are involved in the operation of a BJT. In fact, minority carrier diffusion plays the leading role just as in the PN junction diode. The word junction refers to the fact that PN junctions are critical to the operation of the BJT. BJTs are also simply known as bipolar transistors.

8.1 • INTRODUCTION TO THE BJT •

A BJT is made of a heavily doped emitter (see Fig. 8–1a), a P-type base, and an N-type collector. This device is an NPN BJT. (A PNP BJT would have a P⁺ emitter, N-type base, and P-type collector.) NPN transistors exhibit higher transconductance and
speed than PNP transistors because the electron mobility is larger than the hole mobility. BJTs are almost exclusively of the NPN type since high performance is BJTs’ competitive edge over MOSFETs.

Figure 8–1b shows that when the base–emitter junction is forward biased, electrons are injected into the more lightly doped base. They diffuse across the base to the reverse-biased base–collector junction (edge of the depletion layer) and get swept into the collector. This produces a **collector current**, $I_C$. $I_C$ is independent of $V_{CB}$ as long as $V_{CB}$ is a reverse bias (or a small forward bias, as explained in Section 8.6). Rather, $I_C$ is determined by the rate of electron injection from the emitter into the base, i.e., determined by $V_{BE}$. You may recall from the PN diode theory that the rate of injection is proportional to $e^{qV_{BE}/kT}$. These facts are obvious in Fig. 8–1c.

Figure 8–2a shows that the emitter is often connected to ground. (The emitter and collector are the equivalents of source and drain of a MOSFET. The base is the equivalent of the gate.) Therefore, the $I_C$ curve is usually plotted against $V_{CE}$ as shown in Fig. 8–2b. For $V_{CE}$ higher than about 0.3 V, Fig. 8–2b is identical to Fig. 8–1c but with a shift to the right because $V_{CE} = V_{CB} + V_{BE}$. Below $V_{CE} = 0.3$ V,
8.2 Collector Current

the base–collector junction is strongly forward biased and $I_C$ decreases as explained in Section 8.6. Because of the parasitic IR drops, it is difficult to accurately ascertain the true base–emitter junction voltage. For this reason, the easily measurable base current, $I_B$, is commonly used as the variable parameter in lieu of $V_{BE}$ (as shown in Fig. 8–2c). We will see later that $I_C$ is proportional to $I_B$.

8.2 COLLECTOR CURRENT

The collector current is the output current of a BJT. Applying the electron diffusion equation [Eq. (4.7.7)] to the base region,

$$\frac{d^2n'}{dx^2} = \frac{n'}{L_B^2}$$  (8.2.1)

$$L_B = \sqrt{\tau_B D_B}$$  (8.2.2)

FIGURE 8–3 $x = 0$ is the edge of the BE junction depletion layer. $W_B$ is the width of the base neutral region.
\[ n'(0) = n_{B0}(e^{qV_{BE}/kT} - 1) \]  
\[ n'(W_B) = n_{B0}(e^{qV_{BC}/kT} - 1) = -n_{B0} = 0 \]

where \( n_{B0} = n_i^2/N_B \), and \( N_B \) is the base doping concentration. \( V_{BE} \) is normally a forward bias (positive value) and \( V_{BC} \) is a reverse bias (negative value). The solution of Eq. (8.2.1) is

\[ n'(x) = n_{B0}(e^{qV_{BE}/kT} - 1) \frac{\sinh(W_B - x)}{\sinh(W_B/L_B)} \]  

Equation (8.2.5) is plotted in Fig. 8–4.

Modern BJTs have base widths of about 0.1 \( \mu \)m. This is much smaller than the typical diffusion length of tens of microns (see Example 4–4 in Section 4.8). In the case of \( W_B \ll L_B \), Eq. (8.2.5) reduces to a straight line as shown in Fig. 8–4.

\[ n'(x) = n'(0)(1 - x/W_B) \]
\[ = \frac{n_i^2}{N_B}(e^{qV_{BE}/kT} - 1)(1 - x/W_B) \]  

\( n_i \) is the intrinsic carrier concentration of the base material. The subscript, \( B \), is added to \( n_i \) because the base may be made of a different semiconductor (such as SiGe alloy, which has a smaller band gap and therefore a larger \( n_i \) than the emitter and collector material).

**FIGURE 8–4** When \( W_B \ll L_B \), the excess minority carrier concentration in the base is approximately a linear function of \( x \).
As explained in the PN diode analysis, the minority-carrier current is dominated by the diffusion current. The sign of $I_C$ is defined in Fig. 8–2a and is positive.

$$I_C = A_E q D_B \frac{dn'}{dx} = A_E q D_B \frac{n'(0)}{W_B} \quad (8.2.7)$$

$A_E$ is the area of the BJT, specifically the emitter area. Notice the similarity between Eq. (8.2.7) and the PN diode IV relation [Eq. (4.9.4)]. Both are proportional to $(e^{qV_{BE}/kT} - 1)$ and to $D_n^2/N$. In fact, the only difference is that $dn'/dx$ has produced the $1/W_B$ term in Eq. (8.2.7) due to the linear $n'$ profile. Equation (8.2.7) can be condensed to

$$I_C = I_S(e^{qV_{BE}/kT} - 1) \quad (8.2.8)$$

where $I_S$ is the saturation current. Equation (8.2.7) can be rewritten as

$$I_C = A_E q n_i^2 \frac{D_B}{n_i B} \left(e^{qV_{BE}/kT} - 1\right) \quad (8.2.9)$$

In the special case of Eq. (8.2.7)

$$G_B = \frac{n_i^2 N_B}{d_i B} W_B = \frac{n_i^2}{d_i B} \frac{P}{D_B} W_B \quad (8.2.10)$$

where $P$ is the majority carrier concentration in the base. It can be shown that Eq. (8.2.9) is valid even for nonuniform base and high-level injection condition if $G_B$ is generalized to [1]

$$G_B = \int_0^{W_B} \frac{n_i^2}{d_i B} \frac{P}{D_B} dx \quad (8.2.11)$$

$G_B$ has the unusual dimension of $s/cm^4$ and is known as the **base Gummel number**. In the special case of $n_i B = n_i$, $D_B$ is a constant, and $p(x) = N_B(x)$ (low-level injection),

$$G_B = \frac{1}{D_B} \int_0^{W_B} N_B(x) dx = \frac{1}{D_B} \times \text{base dopant atoms per unit area} \quad (8.2.12)$$

Equation (8.2.12) illustrates that the base Gummel number is basically proportional to the base dopant density per area. The higher the base dopant density is, the lower the $I_C$ will be for a given $V_{BE}$ as given in Eq. (8.2.9).

The concept of a Gummel number simplifies the $I_C$ model because it contains all the subtleties of transistor design that affect $I_C$: changing base material through $n_i B(x)$, nonconstant $D_B$, nonuniform base dopant concentration through $p(x) = N_B(x)$, and even the **high-level injection** condition (see Sec. 8.2.1), where $p > N_B$. Although many factors affect $G_B$, $G_B$ can be easily determined from the Gummel plot shown in Fig. 8–5. The (inverse) slope of the
straight line in Fig. 8–5 can be described as 60 mV per decade. The extrapolated intercept of the straight line and \( V_{BE} = 0 \) yields \( I_S \) [Eq. (8.2.8)]. \( G_B \) is equal to \( A \eta' \eta_2 \) divided by the intercept.

### 8.2.1 High-Level Injection Effect

The decrease in the slope of the curve in Fig. 8–5 at high \( I_C \) is called the high-level injection effect. At large \( V_{BE} \), \( n' \) in Eq. (8.2.3) can become larger than the base doping concentration \( N_B \)

\[
n' = p' \gg N_B \quad (8.2.13)
\]

The first part of Eq. (8.2.13) is simply Eq. (2.6.2) or charge neutrality. The condition of Eq. (8.2.13) is called high-level injection. A consequence of Eq. (8.2.13) is that in the base

\[
n = p \quad (8.2.14)
\]

From Eqs. (8.2.14) and (4.9.6)

\[
n = p = n_1 e^{qV_{BE}/2kT} \quad (8.2.15)
\]

Equations (8.2.15) and (8.2.11) yield

\[
G_B \approx n_1 e^{qV_{BE}/2kT} \quad (8.2.16)
\]

Equations (8.2.16) and (8.2.9) yield

\[
I_C = n_1 e^{qV_{BE}/2kT} \quad (8.2.17)
\]

Therefore, at high \( V_{BE} \) or high \( I_C \), \( I_C \approx e^{qV_{BE}/2kT} \) and the (inverse) slope in Fig. 8–5 becomes 120 mV/decade. \( I_{kF} \), the knee current, is the current at which the slope changes. It is a useful parameter in the BJT model for circuit simulation. The IR drop in the parasitic resistance significantly increases \( V_{BE} \) at very high \( I_C \) and further flattens the curve.
8.3 **BASE CURRENT**

Whenever the base–emitter junction is forward biased, some holes are injected from the P-type base into the N+ emitter. These holes are provided by the base current, $I_B$. $I_B$ is an undesirable but inevitable side effect of producing $I_C$ by forward biasing the BE junction. The analysis of $I_B$, the base to emitter injection current, is a perfect parallel of the $I_C$ analysis. Figure 8–6b illustrates the mirror equivalence. At an ideal ohmic contact such as the contact of the emitter, the equilibrium condition holds and $p' = 0$ similar to Eq. (8.2.4). Analogous to Eq. (8.2.9), the base current can be expressed as

$$I_B = A_E \frac{q}{G_E} \left( e^{qV_{BE}/kT} - 1 \right)$$

(8.3.1)

$$G_E = \int_0^{W_E} \frac{n_{iE}^2}{n_{iE}^2 - D_E} dx$$

(8.3.2)

$G_E$ is the **emitter Gummel number**. As an exercise, please verify that in the special case of a uniform emitter, where $n_{iE}, N_E$ (emitter doping concentration) and $D_E$ are not functions of $x$,

$$I_B = A_E q \frac{D_E n_{iE}^2}{W_E N_E} \left( e^{qV_{BE}/kT} - 1 \right)$$

(8.3.3)

![Diagram of electron and hole flow paths in BJT](image)

**FIGURE 8–6** (a) Schematic of electron and hole flow paths in BJT; (b) hole injection into emitter closely parallels electron injection into base.  

1 In older transistors with VERY long bases, $I_B$ also supplies holes at a significant rate for recombination in the base. Recombination is negligible in the narrow base of a typical modern BJT.

2 A good metal–semiconductor ohmic contact (at the end of the emitter) is an excellent source and sink of carriers. Therefore, the excess carrier concentration is assumed to be zero.
Perhaps the most important DC parameter of a BJT is its common-emitter current gain, \( \beta_F \).

Another current ratio, the common-base current gain, is defined by

\[
\alpha_F = \frac{I_C}{I_E} = \frac{I_C}{I_B + I_C} = \frac{I_C/I_B}{1 + I_C/I_B} = \frac{\beta_F}{1 + \beta_F}
\]

\( \alpha_F \) is typically very close to unity, such as 0.99, because \( \beta_F \) is large. From Eq. (8.4.3), it can be shown that

\[
\beta_F = \frac{\alpha_F}{1 - \alpha_F}
\]

\( I_B \) is a load on the input signal source, an undesirable side effect of forward biasing the BE junction. \( I_B \) should be minimized (i.e., \( \beta_F \) should be maximized). Dividing Eq. (8.2.9) by Eq. (8.3.1),

\[
\beta_F = \frac{G_E}{G_B} = \frac{D_E W B N_B n_B^2}{D_W W E N_B n_E^2}
\]

A typical good \( \beta_F \) is 100. \( D \) and \( W \) in Eq. (8.4.5) cannot be changed very much. The most obvious way to achieve a high \( \beta_F \), according to Eq. (8.4.5), is to use a large \( N_E \) and a small \( N_B \). A small \( N_B \), however, would introduce too large a base resistance, which degrades the BJT’s ability to operate at high current and high frequencies. Typically, \( N_B \) is around \( 10^{18} \text{ cm}^{-3} \).

An emitter is said to be efficient if the emitter current is mostly the useful electron current injected into the base with little useless hole current (the base current). The emitter efficiency is defined as

\[
\gamma_E = \frac{I_E - I_B}{I_E} = \frac{I_C}{I_C + I_B} = \frac{1}{1 + G_B/G_E}
\]

**EXAMPLE 8–1  Current Gain**

A BJT has \( I_C = 1 \text{ mA} \) and \( I_B = 10 \mu\text{A} \). What are \( I_E, \beta_F \) and \( \alpha_F \)?

\[
I_E = I_C + I_B = 1 \text{ mA} + 10 \mu\text{A} = 1.01 \text{ mA}
\]

\[
\beta_F = \frac{I_C}{I_B} = \frac{1 \text{ mA}}{10 \mu\text{A}} = 100
\]

\[
\alpha_F = \frac{I_C}{I_E} = \frac{1 \text{ mA}}{1.01 \text{ mA}} = 0.9901
\]
8.4 • Current Gain

8.4.1 Emitter Band Gap Narrowing

To raise $\beta_F$, $N_E$ is typically made larger than $10^{20}$ cm$^{-3}$. Unfortunately, when $N_E$ is very large, $n_i^2$ becomes larger than $n_i$. This is called the heavy doping effect. Recall Eq. (1.8.12)

$$n_i^2 = N_cN_v e^{-E_g/kT} \quad (8.4.7)$$

Heavy doping can modify the Si crystal sufficiently to reduce $E_g$ and cause $n_i^2$ to increase significantly. Therefore, the heavy doping effect is also known as band gap narrowing.

$$n_{iE}^2 = n_i^2 e^{\Delta E_g/kT} \quad (8.4.8)$$

$\Delta E_g$ is the narrowing of the emitter band gap relative to lightly doped Si and is negligible for $N_E < 10^{18}$ cm$^{-3}$, 50 meV at $10^{19}$ cm$^{-3}$, 95 meV cm$^{-3}$ at $10^{20}$ cm$^{-3}$, and 140 meV at $10^{21}$ cm$^{-3}$ [2].

8.4.2 Narrow Band-Gap Base and Heterojunction BJT

To further elevate $\beta_F$, we can raise $n_{iB}$ by using a base material that has a smaller band gap than the emitter material. $\text{Si}_{1-\eta}\text{Ge}_\eta$ is an excellent base material candidate for an Si emitter. With $\eta = 0.2$, $E_{gB}$ is reduced by 0.1 eV. In an SiGe BJT, the base is made of high-quality P-type epitaxial SiGe. In practice, $\eta$ is graded such that $\eta = 0$ at the emitter end of the base and 0.2 at the drain end to create a built-in field that improves the speed of the BJT (see Section 8.7.2).

Because the emitter and base junction is made of two different semiconductors, the device is known as a heterojunction bipolar transistor or HBT. HBTs made of InP emitter ($E_g = 1.35$ eV) and InGaAs base ($E_g = 0.68$ eV) and GaAlAs emitter with GaAs base are other examples of well-studied HBTs. The ternary semiconductors are used to achieve lattice constant matching at the heterojunction (see Section 4.13.1).

---

*3 Heavy doping also affects $n_i$ by altering $N_c$ and $N_v$ in a complex manner. It is customary to lump all these effects into an effective narrowing of the band gap.*
8.4.3 Poly-Silicon Emitter

Whether the base material is SiGe or plain Si, a high-performance BJT would have a relatively thick (>100 nm) layer of As doped N⁺ poly-Si film in the emitter (as shown in Fig. 8–7). Arsenic is thermally driven into the “base” by ~20 nm and converts that single-crystalline layer into a part of the N⁺ emitter. This way, $\beta_F$ is larger due to the large $W_E$, mostly made of the N⁺ poly-Si. This is the poly-Silicon emitter technology. The simpler alternative, a deeper implanted or diffused N⁺ emitter without the poly-Si film, is known to produce a higher density of crystal defects in the thin base (causing excessive emitters to collector leakage current or even shorts in a small number of the BJTs).

8.4.4 Gummel Plot and $\beta_F$ Fall-Off at High and Low $I_C$

High-speed circuits operate at high $I_C$, and low-power circuits may operate at low $I_C$. Current gain, $\beta$, drops at both high $I_C$ and at low $I_C$. Let us examine the causes.
We have seen in Fig. 8–5 (Gummel plot) that $I_C$ flattens at high $V_{BE}$ due to the high-level injection effect in the base. That $I_C$ curve is replotted in Fig. 8–8. $I_B$, arising from hole injection into the emitter, does not flatten due to this effect (Fig. 8–8) because the emitter is very heavily doped, and it is practically impossible to inject a higher density of holes than $N_E$.

Over a wide mid-range of $I_C$ in Fig. 8–8, $I_C$ and $I_B$ are parallel, indicating that the ratio of $I_C/I_B$, i.e., $\beta_F$, is a constant. This fact is obvious in Fig. 8–9. Above 1 mA, the slope of $I_C$ in Fig. 8–8 drops due to high-level injection. Consequently, the $I_C/I_B$ ratio or $\beta_F$ decreases rapidly as shown in Fig. 8–9. This fall-off of current gain unfortunately degrades the performance of BJTs at high current where the BJT’s speed is the highest (see Section 8.9). $I_B$ in Fig. 8–8 is the base–emitter junction forward-bias current. As shown in Fig. 4–22, forward-bias current slope decreases at low $V_{BE}$ or very low current due to the space-charge region (SCR) current (see Section 4.9.1). A similar slope change is sketched in Fig. 8–8. As a result, the $I_C/I_B$ ratio or $\beta_F$ decreases at very low $I_C$. The weak $V_{BC}$ dependence of $\beta_F$ in Fig. 8–9 is explained in the next section.

**FIGURE 8–7** Schematic illustration of a poly-Si emitter, a common feature of high-performance BJTs.

**FIGURE 8–8** Gummel plot of $I_C$ and $I_B$ indicates that $\beta_F = I_C/I_B$ decreases at high and low $I_C$. 

\[ I_C \quad I_B \quad \beta_F \quad \text{Excess base current} \]

\[ V_{BE} \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8 \quad 1.0 \quad 1.2 \]

\[ I_C (A) \quad 10^{-12} \quad 10^{-10} \quad 10^{-8} \quad 10^{-6} \quad 10^{-4} \quad 10^{-2} \]
8.5 BASE-WIDTH MODULATION BY COLLECTOR VOLTAGE

Instead of the flat $I_C-V_{CE}$ characteristics shown in Fig. 8–2c, Fig. 8–10a (actual $I_C-V_{CE}$ data) clearly indicates the presence of finite slopes. As in MOSFETs, a large output conductance, $\partial I_C/\partial V_{CE}$, of BJTs is deleterious to the voltage gain of circuits. The cause of the output conductance is base-width modulation, explained in Fig. 8–11. The thick vertical line indicates the location of the base-collector junction. With increasing $V_{CE}$, the base-collector depletion region widens and the neutral base width decreases. This leads to an increase in $I_C$ as shown in Fig. 8–11.

If the $I_C-V_{CE}$ curves are extrapolated as shown in Fig. 8–10b, they intercept the $I_C = 0$ axis at approximately the same point. Figure 8–10b defines the Early voltage, $V_A$. $V_A$ is a parameter that describes the flatness of the $I_C$ curves. Specifically, the output resistance can be expressed as $V_A/I_C$:

$$r_0 = \left(\frac{\partial I_C}{\partial V_{CE}}\right)^{-1} = \frac{V_A}{I_C} \quad \text{(8.5.1)}$$

A large $V_A$ (i.e., a large $r_0$) is desirable for high voltage gains. A typical $V_A$ is 50 V. $V_A$ is sensitive to the transistor design. Qualitatively, we can expect $V_A$ and $r_0$ to increase (i.e., expect the base-width modulation to be a smaller fraction of the base width) if we:

(a) increase the base width
(b) increase the base doping concentration, $N_B$, or
(c) decrease the collector doping concentration, $N_C$.

Clearly, (a) would reduce the sensitivity to any given $\Delta W_B$ (see Fig. 8–11). (b) would reduce the depletion region thickness on the base side because the depletion region penetrates less into the more heavily doped side of a PN junction
8.5  Base-Width Modulation by Collector Voltage

**FIGURE 8–10** BJT output conductance: (a) measured BJT characteristics. $I_B = 4, 8, 12, 16,$ and 20 $\mu$A. (From [3]); (b) schematic drawing illustrates the definition of Early voltage, $V_A$.

**FIGURE 8–11** As $V_C$ increases, the BC depletion layer width increases and $W_B$ decreases causing $dn'/dx$ and $I_C$ to increase. In reality, the depletion layer in the collector is usually much wider than that in the base.
Chapter 8  •  Bipolar Transistor

[see Eq. (4.2.5)]. For the same reason, (c) would tend to move the depletion region into the collector and thus reduce the depletion region thickness on the base side, too. Both (a) and (b) would depress \( \beta_F \) [see Eq. (8.4.5)], (c) is the most acceptable course of action. It also reduces the base–collector junction capacitance, which is a good thing. Therefore, the collector doping is typically ten times lighter than base doping. In Fig. 8–10, the larger slopes at \( V_{CE} > 3V \) are caused by impact ionization (Section 4.5.3). The rise of \( I_C \) due to base-width modulation is known as the Early effect, after its discoverer.

● Early on Early Voltage ●

Anecdote contributed by Dr. James Early, November 10, 1990

“In January, 1952, on my way to a Murray Hill Bell Labs internal meeting, I started to think about how to model the collector current as a function of the collector voltage. Bored during the meeting, I put down the expression for collector current \( I_C = \beta_F I_B \). Differentiating with respect to \( V_C \) while \( I_B \) was held constant gave:

\[
\frac{\partial I_C}{\partial V_C} = \frac{\partial \beta_F}{\partial V_C} I_B
\]

How can \( \beta_F \) change with \( V_C \)? Of course! The collector depletion layer thickens as collector voltage is raised. The base gets thinner and current gain rises. Obvious! And necessarily true.

Why wasn’t this found sooner? Of those who had thought about it at all before, none was educated in engineering analysis of electron devices, used to setting up new models, and bored at a meeting.”

8.6  •  EBERS–MOLL MODEL  ●

So far, we have avoided examining the part of the \( I-V \) curves in Fig. 8–12 that is close to \( V_{CE} = 0 \). This portion of the \( I-V \) curves is known as the saturation region because the base is saturated with minority carriers injected from both the emitter and the collector. (Unfortunately the MOSFET saturation region is named in exactly the opposite manner.) The rest of the BJT operation region is known as the active region or the linear region because that is where BJT operates in active circuits such as the linear amplifiers.

**FIGURE 8–12** In the saturation region, \( I_C \) drops because the collector–base junction is significantly forward biased.
The Ebers–Moll model is a way to visualize as well as to mathematically describe both the active and the saturation regions of BJT operation. It is also the basis of BJT SPICE models for circuit simulation. The starting point is the idea that $I_C$ is driven by two forces, $V_{BE}$ and $V_{BC}$, as shown in Fig. 8–13. Let us first assume that a $V_{BE}$ is present but $V_{BC} = 0$. Using Eq. (8.2.8),

$$I_C = I_S(e^{qV_{BE}/kT} - 1) \quad (8.6.1)$$

$$I_B = \frac{I_S}{\beta_F}(e^{qV_{BE}/kT} - 1) \quad (8.6.2)$$

Now assume that the roles of the collector and emitter are reversed, i.e., a (possibly forward bias) $V_{BC}$ is present and $V_{BE} = 0$. Electrons would be injected from the collector into base and flow to the emitter. The collector now functions as the emitter and the emitter functions as the collector$^4$

$$I_E = I_S(e^{qV_{BC}/kT} - 1) \quad (8.6.3)$$

$$I_B = \frac{I_S}{\beta_R}(e^{qV_{BC}/kT} - 1) \quad (8.6.4)$$

$$I_C = -I_E - I_B = -I_S\left(1 + \frac{1}{\beta_R}\right)(e^{V_{BC}/kT} - 1) \quad (8.6.5)$$

$\beta_R$ is the reverse current gain. (This is why $\beta_F$ has F as the subscript, $\beta_F$ is the forward current gain.) While $\beta_F$ is usually quite large, $\beta_R$ is small because the doping concentration of the collector, which acts as the “emitter” in the reverse mode, is not high. When both $V_{BE}$ and $V_{BC}$ are present, Eqs. (8.6.1) and (8.6.5) are superimposed as are Eqs. (8.6.2) and (8.6.4).

$$I_C = I_S(e^{qV_{BE}/kT} - 1) - I_S\left(1 + \frac{1}{\beta_R}\right)(e^{qV_{BC}/kT} - 1) \quad (8.6.6)$$

$$I_B = \frac{I_S}{\beta_F}(e^{qV_{BE}/kT} - 1) + \frac{I_S}{\beta_R}(e^{qV_{BC}/kT} - 1) \quad (8.6.7)$$

Equations (8.6.6) and (8.6.7) compromise the Ebers–Moll model as commonly used in SPICE models. These two equations can generate $I_C$ vs. $V_{CE}$ plots with excellent agreement with measured data as shown in Fig. 8–14.

$^4$ When the emitter and collector roles are interchanged, the upper and lower limits of integration in Eq. (8.2.11) are interchanged with no effect on $G_B$ or $I_S$. 
What causes $I_C$ to decrease at low $V_{CE}$? In this region, both the BE and BC junctions are forward biased. (For example: $V_{BE} = 0.8 \, \text{V}$, $V_{BC} = 0.6 \, \text{V}$, thus $V_{CE} = 0.2 \, \text{V}$.) A forward-biased $V_{BC}$ causes the $n'$ at $x = W_B$ to rise in Fig. 8–4. This depresses $d n'/dx$ and therefore $I_C$.

### 8.7 TRANSIT TIME AND CHARGE STORAGE

Static IV characteristics are only one part of the BJT story. Another part is its dynamic behavior or its speed. When the BE junction is forward biased, excess holes are stored in the emitter, the base, and even the depletion layers. We call the sum of all the excess hole charges everywhere $Q_F$. $Q_F$ is the stored excess carrier charge. If $Q_F = 1 \, \text{pC}$ (pico coulomb), there is $+1 \, \text{pC}$ of excess hole charge and $-1 \, \text{pC}$ of excess electron charge stored in the BJT.\(^5\) The ratio of $Q_F$ to $I_C$ is called the forward transit time, $\tau_F$.

$$\tau_F \equiv \frac{Q_F}{I_C} \quad (8.7.1)$$

Equation (8.7.1) states the simple but important fact that $I_C$ and $Q_F$ are related by a constant ratio, $\tau_F$. Some people find it more intuitive to think of $\tau_F$ as the storage time. In general, $Q_F$ and therefore $\tau_F$ are very difficult to predict accurately for a complex device structure. However, $\tau_F$ can be measured experimentally (see Sec. 8.9) and once $\tau_F$ is determined for a given BJT, Eq. (8.7.1) becomes a powerful conceptual and mathematical tool giving $Q_F$ as a function of $I_C$, and vice versa. $\tau_F$ sets a high-frequency limit of BJT operation.

#### 8.7.1 Base Charge Storage and Base Transit Time

To get a sense of how device design affects the transit time, let us analyze the excess hole charge in the base, $Q_{FB}$, from which we will obtain the base transit time, $\tau_{FB}$.

$Q_{FB}$ is $q A_E$ times the area under the line in Fig. 8–15.

\(^5\)This results from Eq. (2.6.2), $n' = p'$. 
Transit Time and Charge Storage

8.7

8.7.2 Drift Transistor

Built-In Base Field

The base transit time can be further reduced by building into the base a drift field that aids the flow of electrons from the emitter to the collector. There are two ways of accomplishing this. The classical method is to use graded base doping, i.e., a large $N_B$ near the EB junction, which gradually decreases toward the CB junction as shown in Fig. 8–16a.

Such a doping gradient is automatically achieved if the base is produced by dopant diffusion. The changing $N_B$ creates a $dE_v/dx$ and a $dE_c/dx$. This means that there is a drift field [Eq. (2.4.2)]. Any electrons injected into the base would drift toward the collector with a base transit time shorter than the diffusion transit time, $W_B^2/2D_B$.

Figure 8–16b shows a more effective technique. In a SiGe BJT, P-type epitaxial Si$_{1-x}$Ge$_x$ is grown over the Si collector with a constant $N_B$ and $\eta$ linearly varying from about 0.2 at the collector end to 0 at the emitter end [4]. A large

$\tau_{FB} = \tau_{FB} W_B^2$
dE_C/dx can be produced by the grading of E_{gB}. These high-speed BJTs are used in high-frequency communication circuits. Drift transistors can have a base transit time several times less than as short as 1 ps.

8.7.3 Emitter-to-Collector Transit Time and Kirk Effect

The total forward transit time, \( \tau_F \), is also known as the emitter-to-collector transit time. \( \tau_{FB} \) is only one portion of \( \tau_F \). The base transit time typically contributes about half of \( \tau_F \). To reduce the transit (or storage) time in the emitter and collector, the emitter and the depletion layers must be kept thin. \( \tau_F \) can be measured, and an example of \( \tau_F \) is shown in Fig. 8–17. \( \tau_F \) starts to increase at a current density where the electron density corresponding to the dopant density in the collector \( (n = N_C) \) is insufficient to support the collector current even if the dopant-induced electrons move at the saturation velocity (see Section 6.8). This intriguing condition of too few dopant atoms and too much current leads to a reversal of the sign of the charge density in the “depletion region.”

\[
I_C = A_E q n v_{sat} \tag{8.7.4}
\]

\[
\rho = q N_C - q n
\]

\[
= q N_C - \frac{I_C}{A_E v_{sat}} \tag{8.7.5}
\]

\[
\frac{d^2 \xi}{dx} = \rho / \varepsilon_s \tag{4.1.5}
\]

\( \xi \) This section may be omitted in an accelerated course.
When $I_C$ is small, $\rho = qN_C$ as expected from the PN junction analysis (see Section 4.3), and the electric field in the depletion layer is shown in Fig. 8–18a. The shaded area is the potential across the junction, $V_{CB} + \phi_B$. The N$^+$ collector is always present to reduce the series resistance (see Fig. 8–22). No depletion layer is

**FIGURE 8–17** Transit time vs. $I_C/A_E$. From top to bottom: $V_{CE} = 0.5, 0.8, 1.5, \text{ and } 3 \text{ V}$. The rise at high $I_C$ is due to base widening (Kirk effect). (From [3].)

**FIGURE 8–18** Electric field $\mathbf{E}(x)$, location of the depletion layer, and base width at (a) low $I_C$ such as 0.1 mA/µm$^2$ in Fig. 8–17; (b) larger $I_C$; (c) even larger $I_C$ (such as 1 mA/µm$^2$) and base widening is evident; and (d) very large $I_C$ with severe base widening.
Chapter 8  •  Bipolar Transistor

shown in the base for simplicity because the base is much more heavily doped than
the collector. As $I_C$ increases, $\rho$ decreases [Eq. (8.7.5)] and $d\zeta/dx$ decreases as
shown in Fig. 8–18b. The electric field drops to zero in the very heavily doped N^+ collector as expected. Note that the shaded area under the $\zeta(x)$ line is basically
equal to the shaded area in the Fig. 8–18a because $V_{CB}$ is kept constant. In Fig.
8–18c, $I_C$ is even larger such that $\rho$ in Eq. (8.7.5) and therefore $d\zeta/dx$ has changed
sign. The size of the shaded areas again remains unchanged. In this case, the high-
field region has moved to the right-hand side of the N collector. As a result, the
base is effectively widened. In Fig. 8–18d, $I_C$ is yet larger and the base become yet
wider. Because of the base widening, $\tau_F$ increases as a consequence [see Eq. (8.7.3)].
This is called the Kirk effect. Base widening can be reduced by increasing $N_C$ and
$V_{CE}$. The Kirk effect limits the peak BJT operating speed (see Fig. 8–21).

8.8  •  SMALL-SIGNAL MODEL  •

Figure 8–19 is an equivalent circuit for the behavior of a BJT in response to a small
input signal, e.g., a 10 mV sinusoidal signal, superimposed on the DC bias. BJTs are
often operated in this manner in analog circuits.

If $V_{BE}$ is not close to zero, the “1” in Eq. (8.2.8) is negligible; in that case

$$I_C = I_S e^{qV_{BE}/kT}$$  

(8.8.1)

When a signal $v_{BE}$ is applied to the BE junction, a collector current $g_m v_{BE}$ is
produced. $g_m$, the transconductance, is

$$g_m = \frac{dI_C}{dV_{BE}} = \frac{d}{dV_{BE}}(I_S e^{qV_{BE}/kT})$$

$$= \frac{d}{dV_{BE}} I_S e^{qV_{BE}/kT} = I_C/kT$$

(8.8.2)

$$g_m = I_C/kT$$

(8.8.3)

At room temperature, for example, $g_m = I_C/26$ mV. The transconductance is
determined by the collector bias current, $I_C$.

The input node, the base, appears to the input drive circuit as a parallel RC
circuit as shown in Fig. 8–19.

$$\frac{1}{\tau_\pi} = \frac{dI_B}{dV_{BE}} = \frac{1}{\beta_F dV_{BE}} \frac{dI_C}{dV_{BE}} = g_m$$

(8.8.4)

$$\tau_\pi = \beta_F/g_m$$

(8.8.5)

$Q_F$ in Eq. (8.7.1) is the excess carrier charge stored in the BJT. If $Q_F = 1$ pC,
there is +1 pC of excess holes and −1 pC of excess electrons in the BJT. All the
excess hole charge, $Q_F$, is supplied by the base current, $I_B$. Therefore, the base
presents this capacitance to the input drive circuit:

$$C_\pi = \frac{dQ_F}{dV_{BE}} = \frac{d}{dV_{BE}} \tau_F I_C = \tau_F g_m$$

(8.8.6)
The capacitance in Eq. (8.8.6) may be called the charge-storage capacitance, better known as the diffusion capacitance. In addition, there is one charge component that is not proportional to $I_C$ and therefore cannot be included in $Q_F$ [see Eq. (8.7.1)]. That is the junction depletion-layer charge. Therefore, a complete model of $C_\pi$ should include the BE junction depletion-layer capacitance, $C_{dBE}$

$$C_\pi = \tau_F g_m + C_{dBE} \quad (8.8.7)$$

**EXAMPLE 8–4  Small-Signal Model Parameters**

The BJT represented in Figs. 8–9 and 8–17 is biased at $I_C = 1$ mA and $V_{CE} = 3$ V. $T = 300$ K and $A_E = 5.6 \mu m^2$. Find (a) $g_m$, (b) $r_\pi$, and (c) $C_\pi$.

**SOLUTION:**

a. $g_m = I_C/kT/q = 1 \text{ mA} / 26 \text{ mV} = 39 \text{ mA/V} = 39 \text{ mS (milli siemens)}$

b. From Fig. 8–9, $\beta_F = 90$ at $I_C = 1$ mA and $V_{CB} = 2$ V. ($V_{CB} = V_{CE} + V_{EB} = 3 \text{ V} + 1 \text{ V} = 2 \text{ V}$.)

$$r_\pi = \beta_F / g_m = \frac{90}{39 \text{ mS}} = 2.3 \text{ k}\Omega$$

c. From Fig. 8–17, at $J_C = I_C/A_E = 1 \text{ mA} / 5.6 \mu \text{m}^2 = 0.18 \text{ mA/}\mu \text{m}^2$ and $V_{CE} = 3$ V, we find $\tau_F = 5 \text{ ps}$.

$$C_\pi = \tau_F g_m = 5 \times 10^{-12} \times 0.039 = 2.0 \times 10^{-13} \text{ F} = 200 \text{ fF (femtofarad)}.$$
parameters are difficult to predict from theory with the accuracy required for commercial circuit design. Therefore, the parameters are routinely determined through comprehensive measurement of the BJT AC and DC characteristics.

8.9 CUTOFF FREQUENCY

Consider a special case of Fig. 8–20a. The load is a short circuit. The signal source is a current source, $i_b$, at a frequency $f$. At what frequency does the AC current gain $\beta (\equiv i_c/i_b)$ fall to unity?

$$v_{bc} = \frac{i_b}{\text{input admittance}} = \frac{i_b}{1/r_\pi + j\omega C_\pi} \quad (8.9.1)$$

$$i_c = g_m v_{bc} \quad (8.9.2)$$

Using Eqs. (8.9.1), (8.9.2), (8.8.7), and (8.8.3)

$$\beta(\omega) = \left| \frac{i_c}{i_b} \right| = \left| \frac{g_m}{1/r_\pi + j\omega C_\pi} \right| = \frac{1}{|1/g_m r_\pi + j\omega \tau_c + j\omega C_{dBE}/g_m|}$$

$$= \frac{1}{|1/\beta_F + j\omega \tau_c + j\omega C_{dBE} kT/qI_c|} \quad (8.9.3)$$
At $\omega = 0$, i.e., DC, Eq. (8.9.3) reduces to $\beta_F$ as expected. As $\omega$ increases, $\beta$ drops. By carefully analyzing the $\beta(\omega)$ data, one can determine $\tau_F$. If $\beta_F >> 1$ so that $1/\beta$ is negligible, Eq. (8.9.3) shows that $\beta(\omega) \propto 1/\omega$ and $\beta = 1$ at

$$f_T = \frac{1}{2\pi(\tau_F + C_{dBE}kT/qI_C)}$$  \hspace{1cm} (8.9.4)

Using a more complete small-signal model similar to Fig. 8–20b, it can be shown that

$$f_T = \frac{1}{2\pi[\tau_F + (C_{dBE} + C_{dBC})kT/(qI_C) + C_{dBC}(r_e + r_c)]}$$  \hspace{1cm} (8.9.5)

$f_T$ is the cutoff frequency and is commonly used to compare the speed of transistors. Equations (8.9.4) and (8.9.5) predict that $f_T$ rises with increasing $I_C$ due to increasing $g_m$, in agreement with the measured $f_T$ shown in Fig. 8–21. At very high $I_C$, $\tau_F$ increases due to base widening (Kirk Effect, Fig. 8–17), and therefore, $f_T$ falls. BJTs are often biased near the $I_C$ where $f_T$ peaks in order to obtain the best high-frequency performance.

$f_T$ is the frequency of unity power gain. The frequency of unity power gain, called the maximum oscillation frequency, can be shown to be [5]

$$f_{\text{max}} = \left(\frac{f_T}{8\pi r_b C_{dBC}}\right)^{1/2}$$  \hspace{1cm} (8.9.6)

It is therefore important to reduce the base resistance, $r_b$.

![FIGURE 8–21 Cutoff frequency of a SiGe bipolar transistor. A compact BJT model matches the measured $f_T$ well. (From [6]. © 1997 IEEE.)](image)
Chapter 8  •  Bipolar Transistor

8.10  Charge Control Model

The small-signal model is ideal for analyzing circuit response to small sinusoidal signals. What if the input signal is large? For example, what \( I_C(t) \) is produced by a step-function \( I_B \) switching from zero to 20 \( \mu A \) or by any \( I_B(t) \)? The response can be

---

7 This section may be omitted. Charge control model is used for analysis of digital switching operations.
conveniently analyzed with the **charge control model**, a simple extension of the charge storage concept (Eq. (8.7.1)).

\[ I_C = \frac{Q_F}{\tau_F} \]  
\[ (8.10.1) \]

Assume that Eq. (8.10.1) holds true even if \( Q_F \) varies with time

\[ I_C(t) = \frac{Q_F(t)}{\tau_F} \]  
\[ (8.10.2) \]

\( I_C(t) \) becomes known if we can solve for \( Q_F(t) \). (\( \tau_F \) has to be characterized beforehand for the BJT being used.) Equation (8.10.2) suggests the concept that \( I_C \) is controlled by \( Q_F \), hence the name of the charge control model. From Eq. (8.10.1), at DC condition,

\[ I_B = I_C \beta_F = \frac{Q_F}{\tau_F \beta_F} \]  
\[ (8.10.3) \]

Equation (8.10.3) has a straightforward physical meaning: *In order to sustain a constant excess hole charge in the transistor, holes must be supplied to the transistor through \( I_B \) to replenish the holes that are lost to recombination. Therefore, DC \( I_B \) is proportional to \( Q_F \). When holes are supplied by \( I_B \) at the rate of \( \frac{Q_F}{\tau_F \beta_F} \), the rate of hole supply is exactly equal to the rate of hole loss to recombination and \( Q_F \) remains at a constant value.* What if \( I_B \) is larger than \( \frac{Q_F}{\tau_F \beta_F} \)? In that case, holes flow into the BJT at a higher rate than the rate of hole loss—and the stored hole charge (\( Q_F \)) increases with time.

\[ \frac{dQ_F}{dt} = I_B(t) - \frac{Q_F}{\tau_F \beta_F} \]  
\[ (8.10.4) \]

Equations (8.10.4) and (8.10.2), together constitute the basic charge control model. For any given \( I_B(t) \), Eq. (8.10.4) can be solved for \( Q_F(t) \) analytically or by numerical integration. Once \( Q_F(t) \) is found, \( I_C(t) \) becomes known from Eq. (8.10.2). We may interpret Eq. (8.10.4) with the analogy of filling a very leaky bucket from a faucet shown in Fig. 8–23. \( Q_F \) is the amount of water in the bucket, and \( \frac{Q_F}{\tau_F \beta_F} \) is the rate of \( Q_F \) changing with time.

![FIGURE 8–23 Water analogy of the charge control model. Excess hole charge (\( Q_F \)) rises (or falls) at the rate of supply (\( I_B \)) minus loss (\( \propto Q_F \)).](image)
of water leakage. $I_B$ is the rate of water flowing from the faucet into the bucket. If the faucet is turned fully open, the water level rises in the bucket; if it is turned down, the water level falls.

**EXAMPLE 8-5 Finding $I_C(t)$ for a Step $I_B(t)$**

**QUESTION:** $\tau_F$ and $\beta_F$ of a BJT are given. $I_B(t)$ is a step function rising from zero to $I_{B0}$ at $t = 0$ as shown in Fig. 8–24. Find $I_C(t)$.

**SOLUTION:**

At $t \geq 0$, $I_B(t) = I_{B0}$ and the solution of Eq. (8.10.4)

$$\frac{dQ_F}{dt} = I_B(t) - \frac{Q_F}{\tau_F \beta_F}$$

is

$$Q_F(t) = \frac{\tau_F \beta_F I_{B0}}{1 - e^{-t/\tau_F \beta_F}}$$

(8.10.5)

Please verify that Eq. (8.10.5) is the correct solution by substituting it into Eq. (8.10.4). Also verify that the initial condition $Q_F(0) = 0$ is satisfied by Eq. (8.10.5). $I_C(t)$ follows Eq. (8.10.2).

$$I_C(t) = \frac{Q_F(t)}{\tau_F} = \frac{\beta_F I_{B0}}{1 - e^{-t/\tau_F \beta_F}}$$

(8.10.6)

$I_C(t)$ is plotted in Fig. 8–24. At $t \to \infty$, $I_C = \beta_F I_{B0}$ as expected. $I_C(t)$ can be determined for any given $I_B(t)$ by numerically solving Eq. (8.10.4).

![FIGURE 8–24](image)

**FIGURE 8–24** From the given step-function $I_B(t)$, charge control analysis can predict $I_C(t)$.

What we have studied in this section is a basic version of the charge control model. For a more exact analysis, one would introduce the junction depletion-layer capacitances into Eq. (8.10.4). Diverting part of $I_B$ to charge the junction capacitances would produce an additional delay in $I_C(t)$.

**8.11 MODEL FOR LARGE-SIGNAL CIRCUIT SIMULATION**

The BJT model used in circuit simulators such as SPICE can accurately represent the DC and dynamic currents of the transistor in response to $V_{BE}(t)$ and $V_{CE}(t)$. A typical circuit simulation model or compact model is made of the Ebers–Moll
model (with $V_{BE}$ and $V_{BC}$ as the two driving forces for $I_C$ and $I_B$) plus additional enhancements for high-level injection, voltage-dependent capacitances that accurately represent the charge storage in the transistor, and parasitic resistances as shown in Fig. 8–25. This BJT model is known as the **Gummel–Poon model**.

The two diodes represent the two $I_B$ terms due to $V_{BE}$ and $V_{BC}$ similar to Eq. (8.6.7). The capacitor labeled $Q_F$ is voltage dependent such that the charge stored in it is equal to the $Q_F$ described in Section 8.7. $Q_R$ is the counterpart of $Q_F$ produced by a forward bias at the BC junction. Inclusion of $Q_R$ makes the dynamic response of the model accurate even when $V_{BC}$ is sometimes forward biased. $C_{BE}$ and $C_{BC}$ are the junction depletion-layer capacitances. $C_{CS}$ is the collector-to-substrate capacitance (see Fig. 8–22).

$$I_C = I_S'(e^{qV_{BE}/kT} - e^{qV_{BC}/kT})(1 + \frac{V_{CB}}{V_A}) - \frac{I_S}{\beta_R}(e^{qV_{BC}/kT} - 1) \quad (8.11.1)$$

The similarity between Eqs. (8.11.1) and (8.6.6) is obvious. The $1 + V_{CB}/V_A$ factor is added to represent the Early effect—$I_C$ increasing with increasing $V_{CB}$. $I_S'$ differs from $I_S$ in that $I_S'$ decreases at high $V_{BE}$ due to the high-level injection effect in accordance with Eq. (8.2.11) and as shown in Fig. 8–5.

$$I_B = \frac{I_S}{\beta_F}(e^{qV_{BE}/kT} - 1) + \frac{I_S}{\beta_R}(e^{qV_{BC}/kT} - 1) + I_{SE}(e^{qV_{BE}/n_EkT} - 1) \quad (8.11.2)$$

Equation (8.11.2) is identical to Eq. (8.6.7) except for the additional third term, which represents the excess base junction current shown in Fig. 8–8. $I_{SE}$ and $n_E$ parameters are determined from the measured BJT data as are all of the several dozens of model parameters. The continuous curves in Figs. 8–9, 8–10a, and 8–17 are all examples of compact models. The excellent agreement between the models and the discrete data points in the same figures are necessary conditions for the
circuit simulation results to be accurate. The other necessary condition is that the capacitance in Fig. 8–24 be modeled accurately.

8.12 CHAPTER SUMMARY

The base–emitter junction is usually forward biased while the base–collector junction is reverse biased (as shown in Fig. 8–1b). \( V_{BE} \) determines the rate of electron injection from the emitter into the base, and thus uniquely determines the collector current, \( I_C \), regardless of the reverse bias, \( V_{CB} \)

\[
I_C = A_E \frac{q n^2}{G_B} (e^{qV_{BE}/kT} - 1) \quad (8.2.9)
\]

\[
G_B \equiv \int_0^{W_B} \frac{n_1^2}{n_{iB}} \frac{P}{D_B} \, dx \quad (8.2.11)
\]

\( G_B \) is the base Gummel number, which represents all the subtleties of BJT design that affect \( I_C \): base material, nonuniform base doping, nonuniform material composition, and the high-level injection effect.

An undesirable but unavoidable side effect of the application of \( V_{BE} \) is a hole current flowing from the base, mostly into the emitter. This base (input) current, \( I_B \), is related to \( I_C \) by the common-emitter current gain, \( \beta_F \).

\[
\beta_F = \frac{I_C}{I_B} = \frac{G_E}{G_B} \quad (8.4.1), (8.4.5)
\]

where \( G_E \) is the emitter Gummel number. The common-base current gain is

\[
\alpha_F \equiv \frac{I_C}{I_E} = \frac{\beta_F}{1 + \beta_F} \quad (8.4.3)
\]

The Gummel plot, Fig. 8–8, indicates that \( \beta_F \) falls off in the high \( I_C \) region due to high-level injection in the base and also in the low \( I_C \) region due to excess base current.

Base-width modulation by \( V_{CB} \) results in a significant slope of the \( I_C-V_{CE} \) curve in the active region. This is the Early effect. The slope, called the output conductance, limits the voltage gain that can be produced with a BJT. The Early effect can be suppressed with a lightly doped collector. A heavily doped subcollector (see Fig. 8–22) is routinely used to reduce the collector resistance.

Due to the forward bias, \( V_{BE} \), a BJT stores a certain amount of excess hole charge, which is equal but of opposite sign to the excess electron charge. Its magnitude is called the excess carrier charge, \( Q_F \): \( Q_F \) is linearly proportional to \( I_C \).

\[
Q_F \equiv I_C \tau_F \quad (8.7.1)
\]

\( \tau_F \) is the forward transit time. If there were no excess carriers stored outside the base

\[
\tau_F = \tau_{FB} = \frac{W_B^2}{2D_B} \quad (8.7.3)
\]
Problems

$\tau_{FB}$ is the base transit time. In general, $\tau_F > \tau_{FB}$ because excess carrier storage in the emitter and in the depletion layer are also significant. All these regions should be made small in order to minimize $\tau_F$. Besides minimizing the base width, $W_B$, $\tau_{FB}$ may be reduced by building a drift field into the base with graded base doping (or better, with graded Ge content in a SiGe base). $\tau_{FB}$ is significantly increased at large $I_C$ due to base widening, also known as the Kirk effect.

For computer simulation of circuits, the Gummel–Poon model, shown in Fig. 8–25, is widely used. Both the DC and the dynamic (charge storage) currents are well modeled. The Early effect and high-level injection effect are included. Simpler models consisting of $R$, $C$, and current source are used for hand analysis of circuits. The small-signal models (Figs. 8–19 and 8–20b) employ parameters such as transconductance

$$g_m = \frac{dI_C}{dV_{BE}} = \frac{I_C}{q} \frac{kT}{q}$$

(8.8.2)

input capacitance

$$C_\pi = \frac{dQ_F}{dV_{BE}} = \tau_F g_m$$

(8.8.6)

and input resistance.

$$r_\pi = \frac{dV_{BE}}{dI_B} = \frac{\beta F g_m}{\tau_F}$$

(8.8.5)

The BJT’s unity-gain cutoff frequency (at which $\beta$ falls to unity) is $f_T$. In order to raise device speed, device density, or current gain, a modern high-performance BJT usually employs (see Fig. 8–22) poly-Si emitter, self-aligned poly-Si base contacts, graded Si-Ge base, shallow oxide trench, and deep trench isolation. High-performance BJTs excel over MOSFETs in circuits requiring the highest device $g_m$ and speed.

**PROBLEMS**

**Energy Band Diagram of BJT**

8.1 A Silicon PNP BJT with $N_{AE} = 5 \times 10^{18}$ cm$^{-3}$, $N_{DB} = 10^{17}$ cm$^{-3}$, $N_{AC} = 10^{15}$ cm$^{-3}$, and $W_B = 3 \mu$m is at equilibrium at room temperature.

(a) Sketch the energy band diagram for the device, properly positioning the Fermi level in the three regions.

(b) Sketch (i) the electrostatic potential, setting $V = 0$ in the emitter region, (ii) the electric field, and (iii) the charge density as a function of position inside the BJT.

(c) Calculate the net built-in potential between the collector and the emitter.

(d) Determine the quasi-neutral region width of the base. Bias voltages of $V_{EB} = 0.6$ V and $V_{CB} = -2$ V are now applied to the BJT.

(e) Sketch the energy band diagram for the device, properly positioning the Fermi level in the three regions.

(f) On the sketches completed in part (b), sketch the electrostatic potential, electric field, and charge density as a function of position inside the biased BJT.
Chapter 8  ●  Bipolar Transistor

IV Characteristics and Current Gain

8.2 Derive Eq. (8.4.4) from the definitions of $\beta_F$ (Eq. 8.4.1) and $\alpha_F$ (Eq. 8.4.2).

8.3 Consider a conventional NPN BJT with uniform doping. The base–emitter junction is forward biased, and the base–collector junction is reverse biased.

(a) Qualitatively sketch the energy band diagram.

(b) Sketch the minority carrier concentrations in the base, emitter, and collector regions.

(c) List all the causes contributing to the base and collector currents. You may neglect thermal recombination–generation currents in the depletion regions.

8.4 Neglect all the depletion region widths. The emitter, base, and collector of an NPN transistor have doping concentrations $10^{19}$, $10^{17}$, and $10^{15}$ cm$^{-3}$ respectively. $W_E = 0.8 \mu$m, $W_B = 0.5 \mu$m, and $W_C = 2.2 \mu$m as shown in Fig. 8–26. Assume exp(q$V_{BE}/kT$) = $10^{10}$ and the base–collector junction is reverse biased. Assume that the device dimensions are much smaller than the carrier diffusion lengths throughout.

(a) Find and plot the electron current density, $J_n(x)$, and hole current density, $J_p(x)$, in each region ($J_p$ in the base is rather meaningless since it is three-dimensional in reality).

(b) What are $\gamma_E$ and $\beta_F$ (assume $L_B = 10 \mu$m)?

8.5 For the following questions, answer in one or two sentences.

(a) Why should the emitter be doped more heavily than the base?

(b) “The base width is small” is often stated in device analysis. What is it being compared with?

(c) If the base width, $W_B$, were made smaller, explain how it would affect the base width modulation.

(d) Why does $\beta_F$ increase with increasing $I_C$ at small values of collector current?

(e) Explain why $\beta_F$ falls off at large values of collector current.

(f) For a PNP device, indicate the voltage polarity (+ or –) for the following:

<table>
<thead>
<tr>
<th>Region of operation</th>
<th>$V_{EB}$</th>
<th>$V_{CB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Schottky Emitter and Collector

8.6 The emitter of a high-$\beta_F$ BJT should be heavily doped.

(a) Is it desirable to replace the emitter in BJT with a metal?

(b) Considering a metal on N–Si junction. The minority-carrier injection ratio is the number per second of minority carriers injected into the semiconductor divided by the majority carrier injected per second from the semiconductor into the metal when
the device is forward biased. The ratio is \( \frac{I_{\text{diff}}}{I_{\text{te}}} \), where \( I_{\text{diff}} \) and \( I_{\text{te}} \) are respectively the hole diffusion current flowing into the semiconductor and the thermionic emission current of electrons flowing into the metal. Estimate the minority carrier injection ratio in an Si Schottky diode where \( K = 140 \text{ A/cm}^2 \), \( \Phi_B = 0.72 \text{ eV} \), \( N_d = 10^{16} \text{ cm}^{-3} \), \( \tau_p = 10^{-6} \text{ s} \) and \( T = 300 \text{ K} \). \( I_{\text{diff}} \) in the given diode is the same as the hole diffusion current into the N side of a P⁺-N step junction diode with the same \( N_d \) & \( \tau_p \).

(c) If the collector in BJT is replaced with a metal, would it still function as a BJT? (Hint: Compare the energy diagrams of the two cases.)

Gummel Plot

8.7 Consider an NPN transistor with \( W_E = 0.5 \mu\text{m}, W_B = 0.2 \mu\text{m}, W_C = 2 \mu\text{m}, D_B = 10 \text{ cm}^2/\text{s} \).

(a) Find the peak \( \beta_F \) from Fig. 8–27.

(b) Estimate the base doping concentration \( N_B \).

(c) Find the \( V_{BE} \) at which the peak minority carrier concentration in the base is about \( N_B = 10^{17} \text{ cm}^{-3} \).

(d) Find the base transit time.

Ebers–Moll Model

8.8 Consider the excess minority-carrier distribution of a PNP BJT as shown in Fig. 8–28. (The depletion regions at junctions are not shown.) Assume all generation–recombination current components are negligible and each region is uniformly doped. Constant \( D_n = 30 \text{ cm}^2/\text{s} \) and \( D_p = 10 \text{ cm}^2/\text{s} \) are given. This device has a cross-section area of \( 10^{-5} \text{ cm}^2 \) and \( N_E = 10^{18} \text{ cm}^{-3} \).
Chapter 8  * Bipolar Transistor

(a) Find \( N_C \), i.e., the dopant concentration in the collector.
(b) In what region of the IV characteristics is this BJT operating? Explain your answer. (Hints: Are the BE and BE junctions forward or reverse biased?)
(c) Calculate the total stored excess carrier charge in the base (in Coulombs).
(d) Find the emitter current \( I_E \).
(e) Calculate \( \beta_F \), i.e., the common-emitter current gain when the BJT is operated in the nonsaturation region (i.e., \( V_{EB} > 0.7 \) V and \( V_{CE} > 0.3 \) V. Neglect base-width modulation).

8.9 An NPN BJT is biased so that its operating point lies at the boundary between active mode and saturation mode.

(a) Considering the Ebers–Moll of an NPN transistor, draw the simplified equivalent circuit for the transistor at the given operating point.
(b) Employing the simplified equivalent circuit of part (a), or working directly with Ebers–Moll equations, obtain an expression for \( V_{EC} \) at the specified operating point. Your answer should be in terms of \( I_B \) and the Ebers–Moll parameters.

8.10 An NPN BJT with a Si\(_{0.8}\)Ge\(_{0.2}\) base has an \( E_{gB} \), which is 0.1 eV smaller than an Si-base NPN BJT.

(a) At a given \( V_{BE} \), how do \( I_B \) and \( I_C \) change when a SiGe base is used in place of an Si base? If there is a change, indicate whether the currents are larger or smaller.
(b) To reduce the base transit time and increase \( \beta \), the percentage of Ge in an Si\(_{1-x}\)Ge\(_x\) base is commonly graded in order to create a drift field for electrons across the base. Assume that \( E_g \) is linearly graded and that \( x = 0 \) at the emitter–base junction and \( x = 0.2 \) \( \mu \)m at the base–collector junction. What is \( \beta_{(SiGe)}/\beta_{(Si)} \)? (Hint: \( n_{iB}^* = n_{iSi}^* \exp\left[\left(\frac{\Delta E_{g,Si0.8Ge0.2}}{kT}\right)x/W_B\right] \) where \( W_B \) is the base width.)

8.11 An NPN transistor is fabricated such that the collector has a uniform doping of \( 5 \times 10^{15} \) cm\(^{-3}\). The emitter and base doping profiles are given by \( N_{dE} = 10^{20}e^{-x/0.106} \) cm\(^{-3}\). And \( N_{aB} = 4 \times 10^{18}e^{-x/0.19} \) cm\(^{-3}\), where \( x \) is in micrometers.

(a) Find the intercept of \( N_{dE} \) and \( N_{aB} \) and the intercept of \( N_{aB} \) and \( N_c \). What is the difference between the two intercepts? What is the base width ignoring the depletion widths, known as the *metallurgical base width*?
(b) Find base and emitter Gummel numbers. Ignore the depletion widths for simplicity.
(c) Find the emitter injection efficiency \( \gamma_E \).
(d) Now considering only the \( N_{aB} \) doping in the base (ignore the other doping), sketch the energy band diagram of the base and calculate the built-in electric field, defined as \( E_{bi} = (1/q)(dE_c/dx) \), where \( E_c \) is the conduction band level.

8.12 Derive an expression for the “base width” in Fig. 8–18c or Fig. 8–18d as a function of \( I_C \), \( V_{CB} \), and N-collector width, \( W_C \). Assume all common BJT parameters are known.

8.13 Solve the problem for the step-function \( I_B \) in Example 8–5 in Section 8.10 on your own without copying the provided solution.
8.14  A step change in base current occurs as shown in Fig. 8–29. Assuming forward active operation, estimate the collector current $i_C(t)$ for all time by application of the charge control model and reasonable approximations. Depletion region capacitance can be neglected. The following parameters are given: $\alpha_F = 0.9901$, $\tau_F = 10 \text{ ps}$, $i_{B1} = 100 \, \mu\text{A}$, and $i_{B2} = 10 \, \mu\text{A}$.

![Figure 8-29](image)

8.15  After studying Section 8.9, derive expressions for $\beta(\omega)$ and $f_T$.

REFERENCES


GENERAL REFERENCES
