CHAPTER 1

Models for Integrated-Circuit Active Devices

1.1 Introduction

The analysis and design of integrated circuits depend heavily on the utilization of suitable models for integrated-circuit components. This is true in hand analysis, where fairly simple models are generally used, and in computer analysis, where more complex models are encountered. Since any analysis is only as accurate as the model used, it is essential that the circuit designer have a thorough understanding of the origin of the models commonly utilized and the degree of approximation involved in each.

This chapter deals with the derivation of large-signal and small-signal models for integrated-circuit devices. The treatment begins with a consideration of the properties of *pn* junctions, which are basic parts of most integrated-circuit elements. Since this book is primarily concerned with circuit analysis and design, no attempt has been made to produce a comprehensive treatment of semiconductor physics. The emphasis is on summarizing the basic aspects of semiconductor-device behavior and indicating how these can be modeled by equivalent circuits.

1.2 Depletion Region of a *pn* Junction

The properties of reverse-biased *pn* junctions have an important influence on the characteristics of many integrated-circuit components. For example, reverse-biased *pn* junctions exist between many integrated-circuit elements and the underlying substrate, and these junctions all contribute voltage-dependent parasitic capacitances. In addition, a number of important characteristics of active devices, such as breakdown voltage and output resistance, depend directly on the properties of the depletion region of a reverse-biased *pn* junction. Finally, the basic operation of the junction field-effect transistor is controlled by the width of the depletion region of a *pn* junction. Because of its importance and application to many different problems, an analysis of the depletion region of a reverse-biased *pn* junction is considered below. The properties of forward-biased *pn* junctions are treated in Section 1.3 when bipolar-transistor operation is described.

Consider a *pn* junction under reverse bias as shown in Fig. 1.1. Assume *constant doping densities* of N_D atoms/cm^3 in the *n*-type material and N_A atoms/cm^3 in the *p*-type material. (The characteristics of junctions with nonconstant doping densities will be described later.) Due to the difference in carrier concentrations in the *p*-type and *n*-type regions, there exists a region at the junction where the mobile holes and electrons have been removed, leaving the fixed acceptor and donor ions. Each acceptor atom carries a negative charge and each donor atom carries a positive charge, so that the region near the junction is one of significant space charge and resulting high electric field. This is called

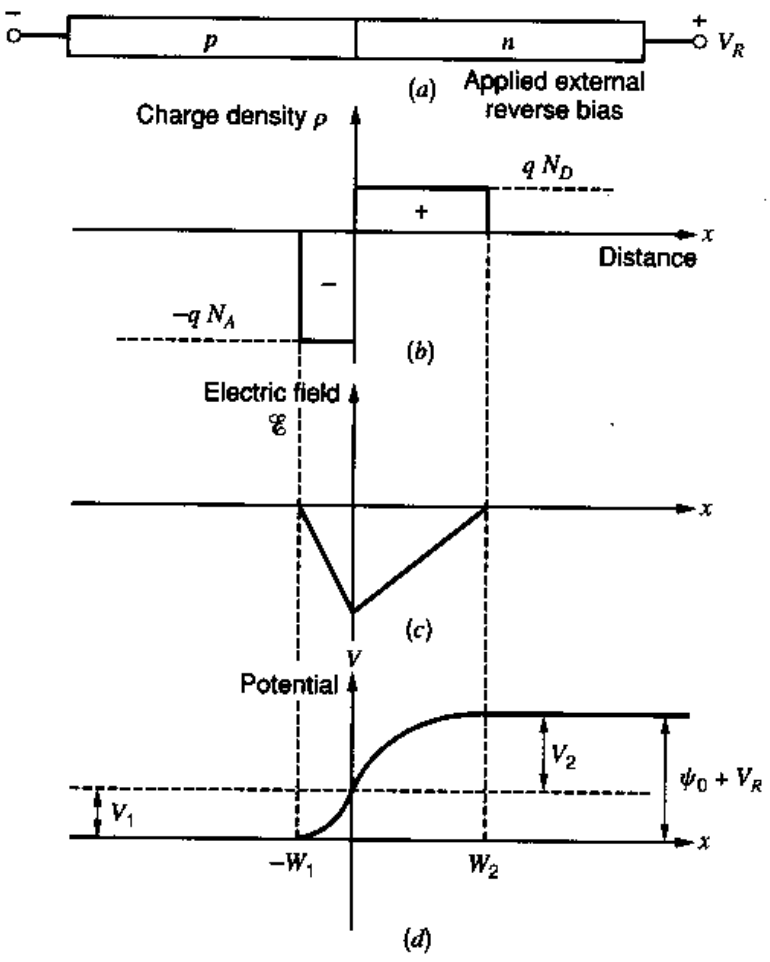


Figure 1.1 The abrupt junction under reverse bias V_R. (*a*) Schematic. (*b*) Charge density. (*c*) Electric field. (*d*) Electrostatic potential.

the *depletion* region or *space-charge* region. It is assumed that the edges of the depletion region are sharply defined as shown in Fig. 1.1, and this is a good approximation in most cases.

For zero applied bias, there exists a voltage ψ_0 across the junction called the *built-in potential*. This potential opposes the diffusion of mobile holes and electrons across the junction in equilibrium and has a value[1]

$$\psi_0 = V_T \ln \frac{N_A N_D}{n_i^2} \tag{1.1}$$

where

$$V_T = \frac{kT}{q} \simeq 26 \text{ mV} \quad \text{at} \quad 300^\circ\text{K}$$

the quantity n_i is the intrinsic carrier concentration in a pure sample of the semiconductor and $n_i \simeq 1.5 \times 10^{10}\text{cm}^{-3}$ at 300°K for silicon.

In Fig. 1.1 the built-in potential is augmented by the applied reverse bias, V_R, and the total voltage across the junction is $(\psi_0 + V_R)$. If the depletion region penetrates a distance W_1 into the *p*-type region and W_2 into the *n*-type region, then we require

$$W_1 N_A = W_2 N_D \tag{1.2}$$

because the total charge per unit area on either side of the junction must be equal in magnitude but opposite in sign.

Poisson's equation in one dimension requires that

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\epsilon} = \frac{qN_A}{\epsilon} \quad \text{for} \quad -W_1 < x < 0 \tag{1.3}$$

where ρ is the charge density, q is the electron charge (1.6×10^{-19} coulomb), and ϵ is the permittivity of the silicon (1.04×10^{-12} farad/cm). The permittivity is often expressed as

$$\epsilon = K_S \epsilon_0 \tag{1.4}$$

where K_S is the dielectric constant of silicon and ϵ_0 is the permittivity of free space (8.86×10^{-14} F/cm). Integration of (1.3) gives

$$\frac{dV}{dx} = \frac{qN_A}{\epsilon} x + C_1 \tag{1.5}$$

where C_1 is a constant. However, the electric field $\mathscr{E}$ is given by

$$\mathscr{E} = -\frac{dV}{dx} = -\left(\frac{qN_A}{\epsilon} x + C_1\right) \tag{1.6}$$

Since there is zero electric field outside the depletion region, a boundary condition is

$$\mathscr{E} = 0 \quad \text{for} \quad x = -W_1$$

and use of this condition in (1.6) gives

$$\mathscr{E} = -\frac{qN_A}{\epsilon}(x + W_1) = -\frac{dV}{dx} \quad \text{for} \quad -W_1 < x < 0 \tag{1.7}$$

Thus the dipole of charge existing at the junction gives rise to an electric field that varies linearly with distance.

Integration of (1.7) gives

$$V = \frac{qN_A}{\epsilon}\left(\frac{x^2}{2} + W_1 x\right) + C_2 \tag{1.8}$$

If the zero for potential is arbitrarily taken to be the potential of the neutral p-type region, then a second boundary condition is

$$V = 0 \quad \text{for} \quad x = -W_1$$

and use of this in (1.8) gives

$$V = \frac{qN_A}{\epsilon}\left(\frac{x^2}{2} + W_1 x + \frac{W_1^2}{2}\right) \quad \text{for} \quad -W_1 < x < 0 \tag{1.9}$$

At $x = 0$, we define $V = V_1$, and then (1.9) gives

$$V_1 = \frac{qN_A}{\epsilon}\frac{W_1^2}{2} \tag{1.10}$$

If the potential difference from $x = 0$ to $x = W_2$ is V_2, then it follows that

$$V_2 = \frac{qN_D}{\epsilon}\frac{W_2^2}{2} \tag{1.11}$$

and thus the total voltage across the junction is

$$\psi_0 + V_R = V_1 + V_2 = \frac{q}{2\epsilon}(N_A W_1^2 + N_D W_2^2) \tag{1.12}$$

When the surface potential in the silicon reaches a critical value equal to twice the Fermi level ϕ_f, a phenomenon known as *inversion* occurs.[16] The Fermi level ϕ_f is defined as

$$\phi_f = \frac{kT}{q} \ln\left[\frac{N_A}{n_i}\right] \tag{1.135}$$

where k is Boltzmann's constant. Also, n_i is the intrinsic carrier concentration, which is

$$n_i = \sqrt{N_c N_v}\, \exp\left(-\frac{E_g}{2kT}\right) \tag{1.136}$$

where E_g is the band gap of silicon at $T = 0°\text{K}$, N_c is the density of allowed states near the edge of the conduction band, and N_v is the density of allowed states near the edge of the valence band, respectively. The Fermi level ϕ_f is usually about 0.3 V. After the potential in the silicon reaches $2\phi_f$, further increases in gate voltage produce no further changes in the depletion-layer width but instead induce a thin layer of electrons in the depletion layer at the surface of the silicon directly under the oxide. Inversion produces a continuous n-type region with the source and drain regions and forms the conducting channel between source and drain. The conductivity of this channel can be modulated by increases or decreases in the gate-source voltage. In the presence of an inversion layer, and without substrate bias, the depletion region contains a fixed charge density

$$Q_{b0} = \sqrt{2qN_A\epsilon 2\phi_f} \tag{1.137}$$

If a substrate bias voltage V_{SB} (positive for n-channel devices) is applied between the source and substrate, the potential required to produce inversion becomes $(2\phi_f + V_{SB})$, and the charge density stored in the depletion region in general is

$$Q_b = \sqrt{2qN_A\epsilon(2\phi_f + V_{SB})} \tag{1.138}$$

The gate-source voltage V_{GS} required to produce an inversion layer is called the threshold voltage V_t and can now be calculated. This voltage consists of several components. First, a voltage $[2\phi_f + (Q_b/C_{ox})]$ is required to sustain the depletion-layer charge Q_b, where C_{ox} is the gate oxide capacitance per unit area. Second, a work-function difference ϕ_{ms} exists between the gate metal and the silicon. Third, positive charge density Q_{ss} always exists in the oxide at the silicon interface. This charge is caused by crystal discontinuities at the $\text{Si} - \text{SiO}_2$ interface and must be compensated by a gate-source voltage contribution of $-Q_{ss}/C_{ox}$. Thus we have a threshold voltage

$$V_t = \phi_{ms} + 2\phi_f + \frac{Q_b}{C_{ox}} - \frac{Q_{ss}}{C_{ox}} \tag{1.139}$$

$$= \phi_{ms} + 2\phi_f + \frac{Q_{b0}}{C_{ox}} - \frac{Q_{ss}}{C_{ox}} + \frac{Q_b - Q_{b0}}{C_{ox}}$$

$$= V_{t0} + \gamma\left(\sqrt{2\phi_f + V_{SB}} - \sqrt{2\phi_f}\right) \tag{1.140}$$

where (1.137) and (1.138) have been used, and V_{t0} is the threshold voltage with $V_{SB} = 0$. The parameter γ is defined as

$$\gamma = \frac{1}{C_{ox}}\sqrt{2q\epsilon N_A} \tag{1.141}$$

1.2.1 Depletion-Region Capacitance

Since there is a *voltage-dependent charge* Q associated with the depletion region, we can calculate a small-signal capacitance C_j as follows:

$$C_j = \frac{dQ}{dV_R} = \frac{dQ}{dW_1}\frac{dW_1}{dV_R} \tag{1.16}$$

Now

$$dQ = AqN_A dW_1 \tag{1.17}$$

where A is the cross-sectional area of the junction. Differentiation of (1.14) gives

$$\frac{dW_1}{dV_R} = \left[\frac{\epsilon}{2qN_A\left(1 + \frac{N_A}{N_D}\right)(\psi_0 + V_R)}\right]^{1/2} \tag{1.18}$$

Use of (1.17) and (1.18) in (1.16) gives

$$C_j = A\left[\frac{q\epsilon N_A N_D}{2(N_A + N_D)}\right]^{1/2}\frac{1}{\sqrt{\psi_0 + V_R}} \tag{1.19}$$

The above equation was derived for the case of reverse bias V_R applied to the diode. However, it is valid for positive bias voltages as long as the forward current flow is small. Thus, if V_D represents the bias on the junction (positive for forward bias, negative for reverse bias), then (1.19) can be written as

$$C_j = A\left[\frac{q\epsilon N_A N_D}{2(N_A + N_D)}\right]^{1/2}\frac{1}{\sqrt{\psi_0 - V_D}} \tag{1.20}$$

$$= \frac{C_{j0}}{\sqrt{1 - \frac{V_D}{\psi_0}}} \tag{1.21}$$

where C_{j0} is the value of C_j for $V_D = 0$.

Equations 1.20 and 1.21 were derived using the assumption of constant doping in the p-type and n-type regions. However, many practical diffused junctions more closely approach a *graded* doping profile as shown in Fig. 1.2. In this case a similar calculation yields

$$C_j = \frac{C_{j0}}{\sqrt[3]{1 - \frac{V_D}{\psi_0}}} \tag{1.22}$$

Note that both (1.21) and (1.22) predict values of C_j approaching infinity as V_D approaches ψ_0. However, the current flow in the diode is then appreciable and the equations no longer valid. A more exact analysis[2,3] of the behavior of C_j as a function of V_D gives the result shown in Fig. 1.3. For forward bias voltages up to about $\psi_0/2$, the values of C_j predicted by (1.21) are very close to the more accurate value. As an approximation, some computer programs approximate C_j for $V_D > \psi_0/2$ by a linear extrapolation of (1.21) or (1.22).

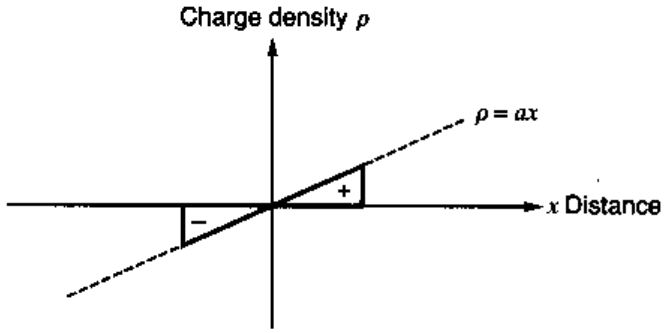


Figure 1.2 Charge density versus distance in a graded junction.

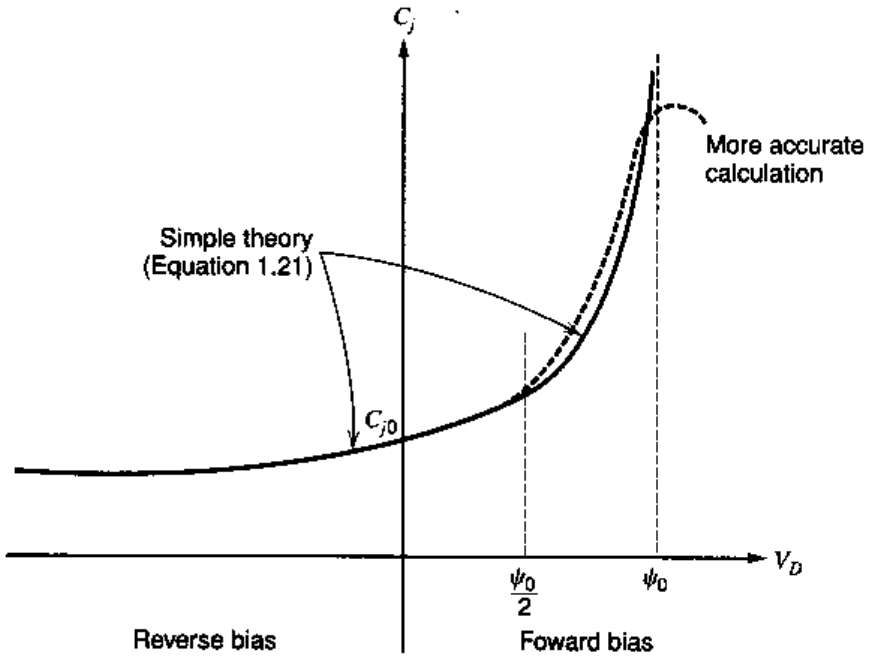


Figure 1.3 Behavior of *pn* junction depletion-layer capacitance C_j as a function of bias voltage V_D.

■ **EXAMPLE**

If the zero-bias capacitance of a diffused junction is 3 pF and $\psi_0 = 0.5$ V, calculate the capacitance with 10 V reverse bias. Assume the doping profile can be approximated by an abrupt junction.

From (1.21)

$$C_j = \frac{3}{\sqrt{1 + \dfrac{10}{0.5}}}\text{pF} = 0.65\text{ pF}$$

■

1.2.2 Junction Breakdown

From Fig. 1.1*c* it can be seen that the maximum electric field in the depletion region occurs at the junction, and for an abrupt junction (1.7) yields a value

$$\mathscr{E}_{\max} = -\frac{qN_A}{\epsilon}W_1 \tag{1.23}$$

Substitution of (1.14) in (1.23) gives

$$|\mathscr{E}_{\max}| = \left[\frac{2qN_AN_DV_R}{\epsilon\,(N_A + N_D)}\right]^{1/2} \tag{1.24}$$

where ψ_0 has been neglected. Equation 1.24 shows that the maximum field increases as the doping density increases and the reverse bias increases. Although useful for indicating the functional dependence of $\mathscr{E}_{\max}$ on other variables, this equation is strictly valid for an ideal plane junction only. Practical junctions tend to have edge effects that cause somewhat higher values of $\mathscr{E}_{\max}$ due to a concentration of the field at the curved edges of the junction.

Any reverse-biased *pn* junction has a small reverse current flow due to the presence of minority-carrier holes and electrons in the vicinity of the depletion region. These are swept across the depletion region by the field and contribute to the leakage current of the junction. As the reverse bias on the junction is increased, the maximum field increases and the carriers acquire increasing amounts of energy between lattice collisions in the depletion region. At a critical field $\mathscr{E}_{\text{crit}}$ the carriers traversing the depletion region acquire sufficient energy to create new hole-electron pairs in collisions with silicon atoms. This is called the *avalanche process* and leads to a sudden increase in the reverse-bias leakage current since the newly created carriers are also capable of producing avalanche. The value of $\mathscr{E}_{\text{crit}}$ is about 3×10^5 V/cm for junction doping densities in the range of 10^{15} to 10^{16} atoms/cm^3, but it increases slowly as the doping density increases and reaches about 10^6 V/cm for doping densities of 10^{18} atoms/cm^3.

A typical *I-V* characteristic for a junction diode is shown in Fig. 1.4, and the effect of avalanche breakdown is seen by the large increase in reverse current, which occurs as the reverse bias approaches the breakdown voltage *BV*. This corresponds to the maximum field $\mathscr{E}_{\max}$ approaching $\mathscr{E}_{\text{crit}}$. It has been found empirically[4] that if the normal reverse bias current of the diode is I_R with no avalanche effect, then the actual reverse current near the breakdown voltage is

$$I_{RA} = MI_R \tag{1.25}$$

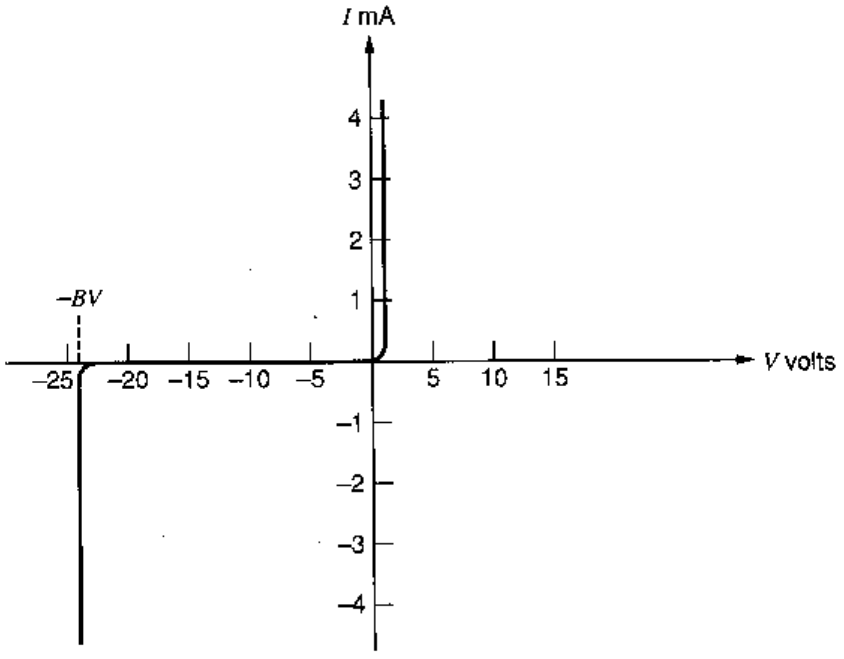


Figure 1.4 Typical *I-V* characteristic of a junction diode showing avalanche breakdown.

where M is the *multiplication factor* defined by

$$M = \frac{1}{1 - \left(\frac{V_R}{BV}\right)^n} \tag{1.26}$$

In this equation, V_R is the reverse bias on the diode and n has a value between 3 and 6.

The operation of a *pn* junction in the breakdown region is not inherently destructive. However, the avalanche current flow must be limited by external resistors in order to prevent excessive power dissipation from occurring at the junction and causing damage to the device. Diodes operated in the avalanche region are widely used as voltage references and are called *Zener diodes*. There is another, related process called *Zener breakdown*,[5] which is different from the avalanche breakdown described above. Zener breakdown occurs only in very heavily doped junctions where the electric field becomes large enough (even with small reverse-bias voltages) to strip electrons away from the valence bonds. This process is called *tunneling*, and there is no multiplication effect as in avalanche breakdown. Although the Zener breakdown mechanism is important only for breakdown voltages below about 6 V, all breakdown diodes are commonly referred to as Zener diodes.

The calculations so far have been concerned with the breakdown characteristic of plane abrupt junctions. Practical diffused junctions differ in some respects from these results and the characteristics of these junctions have been calculated and tabulated for use by designers.[5] In particular, edge effects in practical diffused junctions can result in breakdown voltages as much as 50 percent below the value calculated for a plane junction.

■ **EXAMPLE**

An abrupt plane *pn* junction has doping densities $N_A = 5 \times 10^{15}$ atoms/cm^3 and $N_D = 10^{16}$ atoms/cm^3. Calculate the breakdown voltage if $\mathscr{E}_{crit} = 3 \times 10^5$ V/cm.

The breakdown voltage is calculated using $\mathscr{E}_{max} = \mathscr{E}_{crit}$ in (1.24) to give

$$\begin{aligned} BV &= \frac{\epsilon (N_A + N_D)}{2qN_AN_D}\mathscr{E}_{crit}^2 \\ &= \frac{1.04 \times 10^{-12} \times 15 \times 10^{15}}{2 \times 1.6 \times 10^{-19} \times 5 \times 10^{15} \times 10^{16}} \times 9 \times 10^{10} \text{ V} \\ &= 88 \text{ V} \end{aligned}$$

■

1.3 Large-Signal Behavior of Bipolar Transistors

In this section, the large-signal or dc behavior of bipolar transistors is considered. Large-signal models are developed for the calculation of total currents and voltages in transistor circuits, and such effects as breakdown voltage limitations, which are usually not included in models, are also considered. Second-order effects, such as current-gain variation with collector current and Early voltage, can be important in many circuits and are treated in detail.

The sign conventions used for bipolar transistor currents and voltages are shown in Fig. 1.5. All bias currents for both *npn* and *pnp* transistors are assumed positive going into the device.

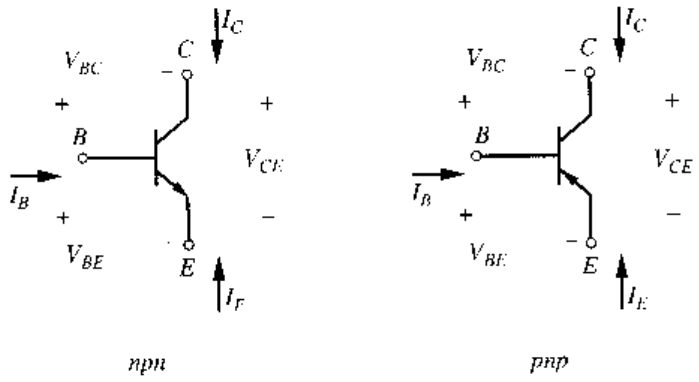


Figure 1.5 Bipolar transistor sign convention.

1.3.1 Large-Signal Models in the Forward-Active Region

A typical *npn* planar bipolar transistor structure is shown in Fig. 1.6*a*, where collector, base, and emitter are labeled *C*, *B*, and *E*, respectively. The method of fabricating such transistor structures is described in Chapter 2. It is shown there that the impurity doping density in the base and the emitter of such a transistor is not constant but varies with distance from the top surface. However, many of the characteristics of such a device can be predicted by analyzing the idealized transistor structure shown in Fig. 1.6*b*. In this structure the base and emitter doping densities are assumed constant, and this is sometimes called a *uniform-base* transistor. Where possible in the following analyses, the equations for the uniform-base analysis are expressed in a form that applies also to nonuniform-base transistors.

A cross section AA' is taken through the device of Fig. 1.6*b* and carrier concentrations along this section are plotted in Fig. 1.6*c*. Hole concentrations are denoted by p and electron concentrations by n with subscripts p or n representing p-type or n-type regions. The n-type emitter and collector regions are distinguished by subscripts E and C, respectively. The carrier concentrations shown in Fig. 1.6*c* apply to a device in the *forward-active region*. That is, the base-emitter junction is forward biased and the base-collector junction is reverse biased. The minority-carrier concentrations in the base at the edges of the depletion regions can be calculated from a Boltzmann approximation to the Fermi-Dirac distribution function to give[6]

$$n_p(0) = n_{po} \exp \frac{V_{BE}}{V_T} \tag{1.27}$$

$$n_p(W_B) = n_{po} \exp \frac{V_{BC}}{V_T} \simeq 0 \tag{1.28}$$

where W_B is the width of the base from the base-emitter depletion layer edge to the base-collector depletion layer edge and n_{po} is the equilibrium concentration of electrons in the base. Note that V_{BC} is negative for an *npn* transistor in the forward-active region and thus $n_p(W_B)$ is very small. Low-level injection conditions are assumed in the derivation of (1.27) and (1.28). This means that the minority-carrier concentrations are always assumed much smaller than the majority-carrier concentration.

If *recombination* of holes and electrons in the base is small, it can be shown that[7] the minority-carrier concentration $n_p(x)$ in the base varies *linearly* with distance. Thus a straight line can be drawn joining the concentrations at $x = 0$ and $x = W_B$ in Fig. 1.6*c*.

For charge neutrality in the base, it is necessary that

$$N_A + n_p(x) = p_p(x) \tag{1.29}$$

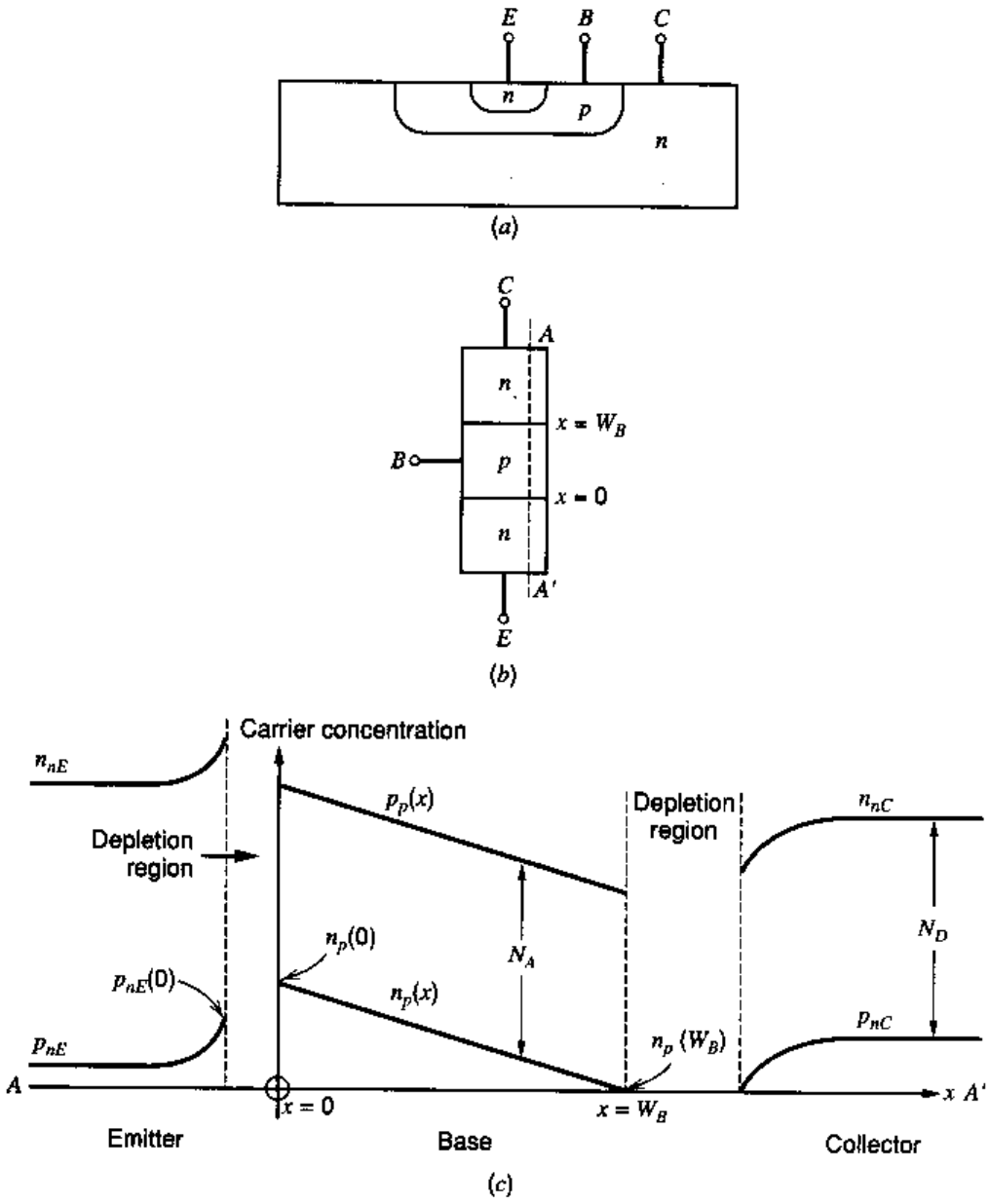


Figure 1.6 (*a*) Cross section of a typical *npn* planar bipolar transistor structure. (*b*) Idealized transistor structure. (*c*) Carrier concentrations along the cross section AA' of the transistor in (*b*). Uniform doping densities are assumed. (Not to scale.)

and thus

$$p_p(x) - n_p(x) = N_A \tag{1.30}$$

where $p_p(x)$ is the hole concentration in the base and N_A is the base doping density that is assumed constant. Equation 1.30 indicates that the hole and electron concentrations are separated by a constant amount and thus $p_p(x)$ also varies linearly with distance.

Collector current is produced by minority-carrier electrons in the base diffusing in the direction of the concentration gradient and being swept across the collector-base depletion region by the field existing there. The diffusion current density due to electrons in the base is

$$J_n = qD_n \frac{dn_p(x)}{dx} \tag{1.31}$$

where D_n is the diffusion constant for electrons. From Fig. 1.6*c*

$$J_n = -qD_n \frac{n_p(0)}{W_B} \tag{1.32}$$

If I_C is the collector current and is taken as positive flowing *into* the collector, it follows from (1.32) that

$$I_C = qAD_n \frac{n_p(0)}{W_B} \tag{1.33}$$

where A is the cross-sectional area of the emitter. Substitution of (1.27) into (1.33) gives

$$I_C = \frac{qAD_n n_{po}}{W_B} \exp \frac{V_{BE}}{V_T} \tag{1.34}$$

$$= I_S \exp \frac{V_{BE}}{V_T} \tag{1.35}$$

where

$$I_S = \frac{qAD_n n_{po}}{W_B} \tag{1.36}$$

and I_S is a constant used to describe the transfer characteristic of the transistor in the forward-active region. Equation 1.36 can be expressed in terms of the base doping density by noting that[8] (see Chapter 2)

$$n_{po} = \frac{n_i^2}{N_A} \tag{1.37}$$

and substitution of (1.37) in (1.36) gives

$$I_S = \frac{qAD_n n_i^2}{W_B N_A} = \frac{qA\bar{D}_n n_i^2}{Q_B} \tag{1.38}$$

where $Q_B = W_B N_A$ is the number of doping atoms in the base per unit area of the emitter and n_i is the intrinsic carrier concentration in silicon. In this form (1.38) applies to both uniform- and nonuniform-base transistors and D_n has been replaced by $\bar{D}_n$, which is an average effective value for the electron diffusion constant in the base. This is necessary for nonuniform-base devices because the diffusion constant is a function of impurity concentration. Typical values of I_S as given by (1.38) are from 10^{-14} to 10^{-16} A.

Equation 1.35 gives the collector current as a function of base-emitter voltage. The base current I_B is also an important parameter and, at moderate current levels, consists of two major components. One of these (I_{B1}) represents recombination of holes and electrons in the base and is proportional to the minority-carrier charge Q_e in the base. From Fig. 1.6*c*, the minority-carrier charge in the base is

$$Q_e = \frac{1}{2} n_p(0) W_B q A \tag{1.39}$$

and we have

$$I_{B1} = \frac{Q_e}{\tau_b} = \frac{1}{2} \frac{n_p(0) W_B q A}{\tau_b} \tag{1.40}$$

where τ_b is the minority-carrier lifetime in the base. I_{B1} represents a flow of majority holes from the base lead into the base region. Substitution of (1.27) in (1.40) gives

$$I_{B1} = \frac{1}{2}\frac{n_{po}W_B qA}{\tau_b}\exp\frac{V_{BE}}{V_T} \tag{1.41}$$

The second major component of base current (usually the dominant one in integrated-circuit npn devices) is due to injection of holes from the base into the emitter. This current component depends on the gradient of minority-carrier holes in the emitter and is[9]

$$I_{B2} = \frac{qAD_p}{L_p}p_{nE}(0) \tag{1.42}$$

where D_p is the diffusion constant for holes and L_p is the diffusion length (assumed small) for holes in the emitter. $p_{nE}(0)$ is the concentration of holes in the emitter at the edge of the depletion region and is

$$p_{nE}(0) = p_{nEo}\exp\frac{V_{BE}}{V_T} \tag{1.43}$$

If N_D is the donor atom concentration in the emitter (assumed constant), then

$$p_{nEo} \simeq \frac{n_i^2}{N_D} \tag{1.44}$$

The emitter is deliberately doped much more heavily than the base, making N_D large and p_{nEo} small, so that the base-current component, I_{B2}, is minimized.

Substitution of (1.43) and (1.44) in (1.42) gives

$$I_{B2} = \frac{qAD_p}{L_p}\frac{n_i^2}{N_D}\exp\frac{V_{BE}}{V_T} \tag{1.45}$$

The total base current, I_B, is the sum of I_{B1} and I_{B2}:

$$I_B = I_{B1} + I_{B2} = \left(\frac{1}{2}\frac{n_{po}W_B qA}{\tau_b} + \frac{qAD_p}{L_p}\frac{n_i^2}{N_D}\right)\exp\frac{V_{BE}}{V_T} \tag{1.46}$$

Although this equation was derived assuming uniform base and emitter doping, it gives the correct functional dependence of I_B on device parameters for practical double-diffused nonuniform-base devices. Second-order components of I_B, which are important at low current levels, are considered later.

Since I_C in (1.35) and I_B in (1.46) are both proportional to $\exp(V_{BE}/V_T)$ in this analysis, the base current can be expressed in terms of collector current as

$$I_B = \frac{I_C}{\beta_F} \tag{1.47}$$

where β_F is the forward current gain. An expression for β_F can be calculated by substituting (1.34) and (1.46) in (1.47) to give

$$\beta_F = \frac{\dfrac{qAD_n n_{po}}{W_B}}{\dfrac{1}{2}\dfrac{n_{po}W_B qA}{\tau_b} + \dfrac{qAD_p n_i^2}{L_p N_D}} = \frac{1}{\dfrac{W_B^2}{2\tau_b D_n} + \dfrac{D_p}{D_n}\dfrac{W_B}{L_p}\dfrac{N_A}{N_D}} \tag{1.48}$$

where (1.37) has been substituted for n_{po}. Equation 1.48 shows that β_F is maximized by minimizing the base width W_B and maximizing the ratio of emitter to base doping

densities N_D/N_A. Typical values of β_F for npn transistors in integrated circuits are 50 to 500, whereas lateral pnp transistors (to be described in Chapter 2) have values 10 to 100. Finally, the emitter current is

$$I_E = -(I_C + I_B) = -\left(I_C + \frac{I_C}{\beta_F}\right) = -\frac{I_C}{\alpha_F} \tag{1.49}$$

where

$$\alpha_F = \frac{\beta_F}{1 + \beta_F} \tag{1.50}$$

The value of α_F can be expressed in terms of device parameters by substituting (1.48) in (1.50) to obtain

$$\alpha_F = \frac{1}{1 + \dfrac{1}{\beta_F}} = \frac{1}{1 + \dfrac{W_B^2}{2\tau_b D_n} + \dfrac{D_p}{D_n}\dfrac{W_B}{L_p}\dfrac{N_A}{N_D}} \simeq \alpha_T \gamma \tag{1.51}$$

where

$$\alpha_T = \frac{1}{1 + \dfrac{W_B^2}{2\tau_b D_n}} \tag{1.51a}$$

$$\gamma = \frac{1}{1 + \dfrac{D_p}{D_n}\dfrac{W_B}{L_p}\dfrac{N_A}{N_D}} \tag{1.51b}$$

The validity of (1.51) depends on $W_B^2/2\tau_b D_n \ll 1$ and $(D_p/D_n)(W_B/L_p)(N_A/N_D) \ll 1$, and this is always true if β_F is large [see (1.48)]. The term γ in (1.51) is called the *emitter injection efficiency* and is equal to the ratio of the electron current (npn transistor) injected into the base from the emitter to the total hole and electron current crossing the base-emitter junction. Ideally $\gamma \to 1$, and this is achieved by making N_D/N_A large and W_B small. In that case very little reverse injection occurs from base to emitter.

The term α_T in (1.51) is called the *base transport factor* and represents the fraction of carriers injected into the base (from the emitter) that reach the collector. Ideally $\alpha_T \to 1$ and this is achieved by making W_B small. It is evident from the above development that fabrication changes that cause α_T and γ to approach unity also maximize the value of β_F of the transistor.

The results derived above allow formulation of a large-signal model of the transistor suitable for bias-circuit calculations with devices in the forward-active region. One such circuit is shown in Fig. 1.7 and consists of a base-emitter diode to model (1.46) and a controlled collector-current generator to model (1.47). Note that the collector voltage ideally has no influence on the collector current and the collector node acts as a high-impedance current source. A simpler version of this equivalent circuit, which is often useful, is shown in Fig. 1.7*b*, where the input diode has been replaced by a battery with a value $V_{BE(on)}$, which is usually 0.6 to 0.7 V. This represents the fact that in the forward-active region the base-emitter voltage varies very little because of the steep slope of the exponential characteristic. In some circuits the temperature coefficient of $V_{BE(on)}$ is important, and a typical value for this is -2 mV/°C. The equivalent circuits of Fig. 1.7 apply for npn transistors. For pnp devices the corresponding equivalent circuits are shown in Fig. 1.8.

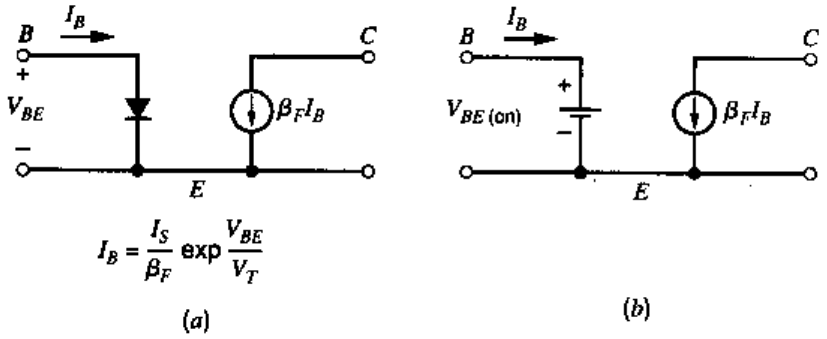


Figure 1.7 Large-signal models of *npn* transistors for use in bias calculations. (*a*) Circuit incorporating an input diode. (*b*) Simplified circuit with an input voltage source.

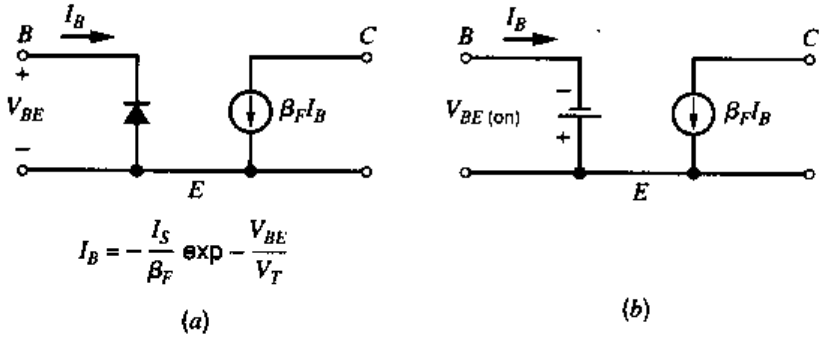


Figure 1.8 Large-signal models of *pnp* transistors corresponding to the circuits of Fig. 1.7.

1.3.2 Effects of Collector Voltage on Large-Signal Characteristics in the Forward-Active Region

In the analysis of the previous section, the collector-base junction was assumed reverse biased and ideally had no effect on the collector currents. This is a useful approximation for first-order calculations, but is not strictly true in practice. There are occasions where the influence of collector voltage on collector current is important, and this will now be investigated.

The collector voltage has a dramatic effect on the collector current in two regions of device operation. These are the saturation (V_{CE} approaches zero) and breakdown (V_{CE} very large) regions that will be considered later. For values of collector-emitter voltage V_{CE} between these extremes, the collector current increases slowly as V_{CE} increases. The reason for this can be seen from Fig. 1.9, which is a sketch of the minority-carrier concentration in the base of the transistor. Consider the effect of changes in V_{CE} on the carrier concentration for constant V_{BE}. Since V_{BE} is constant, the change in V_{CB} equals the change in V_{CE} and this causes an increase in the collector-base depletion-layer width as shown. The change in the base width of the transistor, ΔW_B, equals the change in the depletion-layer width and causes an increase ΔI_C in the collector current.

From (1.35) and (1.38) we have

$$I_C = \frac{qA\bar{D}_n n_i^2}{Q_B} \exp \frac{V_{BE}}{V_T} \tag{1.52}$$

Differentiation of (1.52) yields

$$\frac{\partial I_C}{\partial V_{CE}} = -\frac{qA\bar{D}_n n_i^2}{Q_B^2} \left(\exp \frac{V_{BE}}{V_T} \right) \frac{dQ_B}{dV_{CE}} \tag{1.53}$$

and substitution of (1.52) in (1.53) gives

$$\frac{\partial I_C}{\partial V_{CE}} = -\frac{I_C}{Q_B} \frac{dQ_B}{dV_{CE}} \tag{1.54}$$

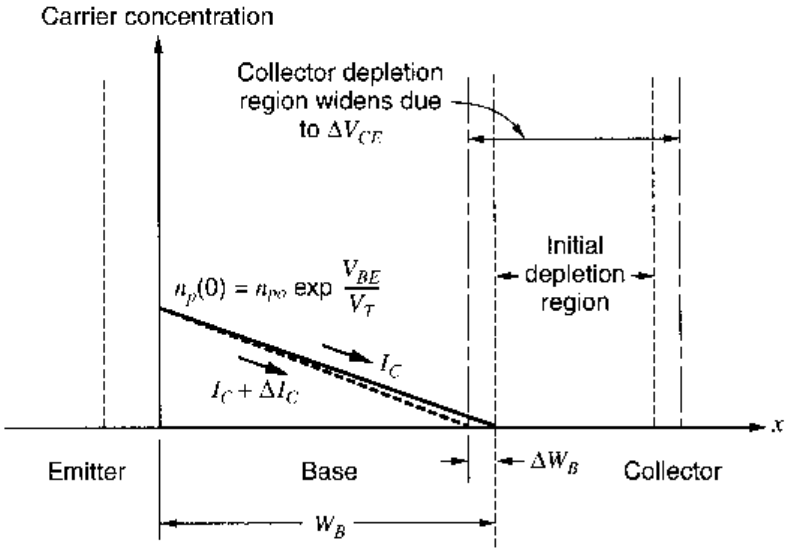


Figure 1.9 Effect of increases in V_{CE} on the collector depletion region and base width of a bipolar transistor.

For a uniform-base transistor $Q_B = W_B N_A$, and (1.54) becomes

$$\frac{\partial I_C}{\partial V_{CE}} = -\frac{I_C}{W_B}\frac{dW_B}{dV_{CE}} \tag{1.55}$$

Note that since the base width *decreases* as V_{CE} increases, dW_B/dV_{CE} in (1.55) is negative and thus $\partial I_C/\partial V_{CE}$ is positive. The magnitude of dW_B/dV_{CE} can be calculated from (1.18) for a uniform-base transistor. This equation predicts that dW_B/dV_{CE} is a function of the bias value of V_{CE}, but the variation is typically small for a reverse-biased junction and dW_B/dV_{CE} is often assumed constant. The resulting predictions agree adequately with experimental results.

Equation 1.55 shows that $\partial I_C/\partial V_{CE}$ is proportional to the collector-bias current and inversely proportional to the transistor base width. Thus narrow-base transistors show a greater dependence of I_C on V_{CE} in the forward-active region. The dependence of $\partial I_C/\partial V_{CE}$ on I_C results in typical transistor output characteristics as shown in Fig. 1.10. In accordance with the assumptions made in the foregoing analysis, these characteristics are shown for constant values of V_{BE}. However, in most integrated-circuit transistors the base current is dependent only on V_{BE} and not on V_{CE}, and thus constant-base-current characteristics can often be used in the following calculation. The reason for this is that the base current is usually dominated by the I_{B2} component of (1.45), which has no dependence on V_{CE}. Extrapolation of the characteristics of Fig. 1.10 back to the V_{CE} axis gives an intercept V_A called the Early voltage, where

$$V_A = \frac{I_C}{\dfrac{\partial I_C}{\partial V_{CE}}} \tag{1.56}$$

Substitution of (1.55) in (1.56) gives

$$V_A = -W_B \frac{dV_{CE}}{dW_B} \tag{1.57}$$

which is a constant, independent of I_C. Thus all the characteristics extrapolate to the same point on the V_{CE} axis. The variation of I_C with V_{CE} is called the Early effect, and V_A is a common model parameter for circuit-analysis computer programs. Typical values of V_A

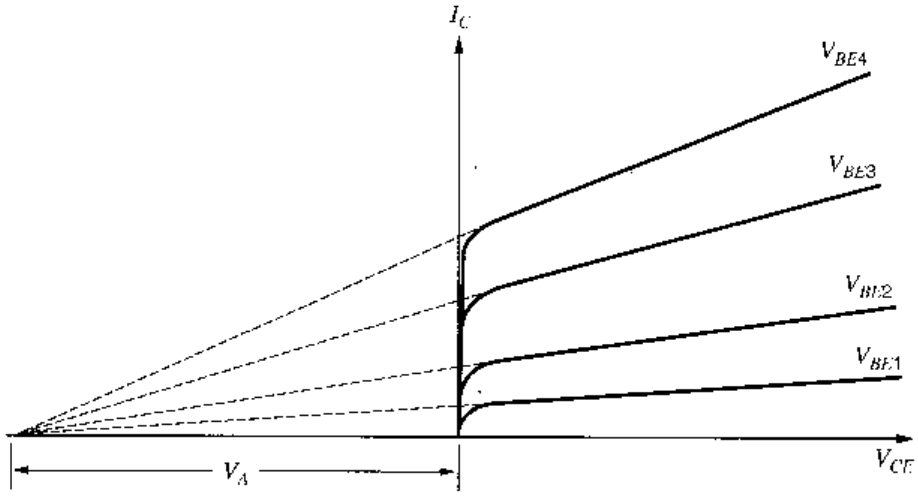


Figure 1.10 Bipolar transistor output characteristics showing the Early voltage, V_A.

for integrated-circuit transistors are 15 to 100 V. The inclusion of Early effect in dc bias calculations is usually limited to computer analysis because of the complexity introduced into the calculation. However, the influence of the Early effect is often dominant in small-signal calculations for high-gain circuits and this point will be considered later.

Finally, the influence of Early effect on the transistor large-signal characteristics in the forward-active region can be represented approximately by modifying (1.35) to

$$I_C = I_S\left(1 + \frac{V_{CE}}{V_A}\right)\exp\frac{V_{BE}}{V_T} \tag{1.58}$$

This is a common means of representing the device output characteristics for computer simulation.

1.3.3 Saturation and Inverse-Active Regions

Saturation is a region of device operation that is usually avoided in analog circuits because the transistor gain is very low in this region. Saturation is much more commonly encountered in digital circuits, where it provides a well-specified output voltage that represents a logic state.

In saturation, both emitter-base and collector-base junctions are forward biased. Consequently, the collector-emitter voltage V_{CE} is quite small and is usually in the range 0.05 to 0.3 V. The carrier concentrations in a saturated *npn* transistor with uniform base doping are shown in Fig. 1.11. The minority-carrier concentration in the base at the edge of the depletion region is again given by (1.28) as

$$n_p(W_B) = n_{po}\exp\frac{V_{BC}}{V_T} \tag{1.59}$$

but since V_{BC} is now positive, the value of $n_p(W_B)$ is no longer negligible. Consequently, changes in V_{CE} with V_{BE} held constant (which cause equal changes in V_{BC}) directly affect $n_p(W_B)$. Since the collector current is proportional to the slope of the minority-carrier concentration in the base [see (1.31)], it is also proportional to $[n_p(0) - n_p(W_B)]$ from Fig. 1.11. Thus changes in $n_p(W_B)$ directly affect the collector current, and the collector node of the transistor appears to have a *low impedance*. As V_{CE} is decreased in saturation with V_{BE} held constant, V_{BC} increases, as does $n_p(W_B)$ from (1.59). Thus from Fig. 1.11 the collector

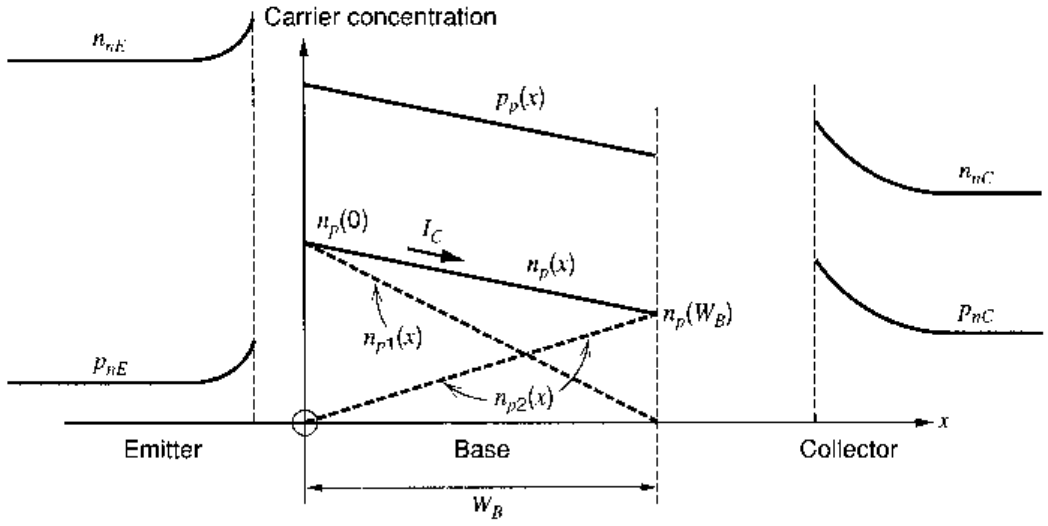


Figure 1.11 Carrier concentrations in a saturated *npn* transistor. (Not to scale.)

current decreases because the slope of the carrier concentration decreases. This gives rise to the saturation region of the $I_C - V_{CE}$ characteristic shown in Fig. 1.12. The slope of the $I_C - V_{CE}$ characteristic in this region is largely determined by the resistance in series with the collector lead due to the finite resistivity of the n-type collector material. A useful model for the transistor in this region is shown in Fig. 1.13 and consists of a fixed voltage source to represent $V_{BE(\text{on})}$, and a fixed voltage source to represent the collector-emitter voltage $V_{CE(\text{sat})}$. A more accurate but more complex model includes a resistor in series with the collector. This resistor can have a value ranging from 20 to 500 Ω, depending on the device structure.

An additional aspect of transistor behavior in the saturation region is apparent from Fig. 1.11. For a given collector current, there is now a much larger amount of stored charge in the base than there is in the forward-active region. Thus the base-current contribution represented by (1.41) will be larger in saturation. In addition, since the collector-base junction is now forward biased, there is a new base-current component due to injection of carriers from the base to the collector. These two effects result in a base current I_B in saturation, which is larger than in the forward-active region for a given collector current I_C. Ratio I_C/I_B in saturation is often referred to as the *forced* β and is always less than β_F. As the forced β is made lower with respect to β_F, the device is said to be more *heavily saturated*.

The minority-carrier concentration in saturation shown in Fig. 1.11 is a straight line joining the two end points, assuming that recombination is small. This can be represented as a linear superposition of the two dotted distributions as shown. The justification for this is that the terminal currents depend *linearly* on the concentrations $n_p(0)$ and $n_p(W_B)$. This picture of device carrier concentrations can be used to derive some general equations describing transistor behavior. Each of the distributions in Fig. 1.11 is considered separately and the two contributions are combined. The *emitter* current that would result from $n_{p1}(x)$ above is given by the classical diode equation

$$I_{EF} = -I_{ES}\left(\exp\frac{V_{BE}}{V_T} - 1\right) \tag{1.60}$$

where I_{ES} is a constant that is often referred to as the *saturation current* of the junction (no connection with the transistor saturation previously described). Equation 1.60 predicts that the junction current is given by $I_{EF} \simeq I_{ES}$ with a reverse-bias voltage applied. However,

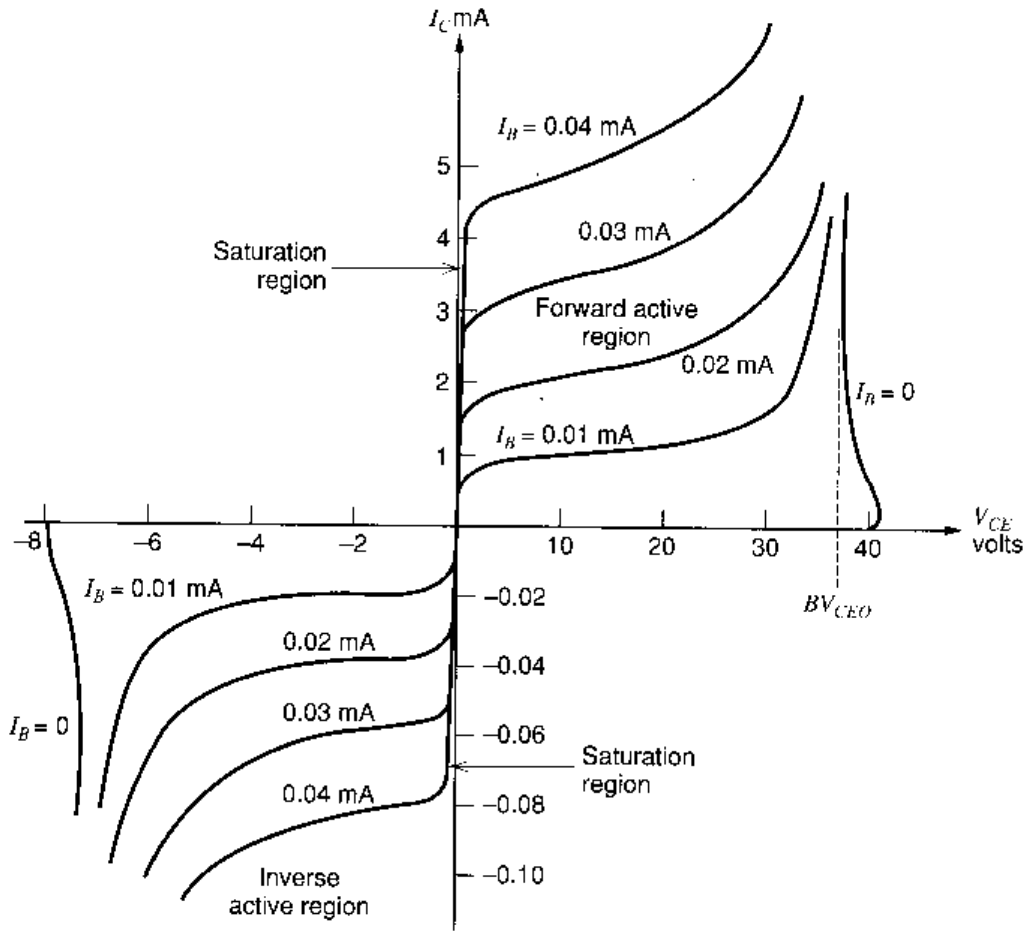


Figure 1.12 Typical I_C-V_{CE} characteristics for an *npn* bipolar transistor. Note the different scales for positive and negative currents and voltages.

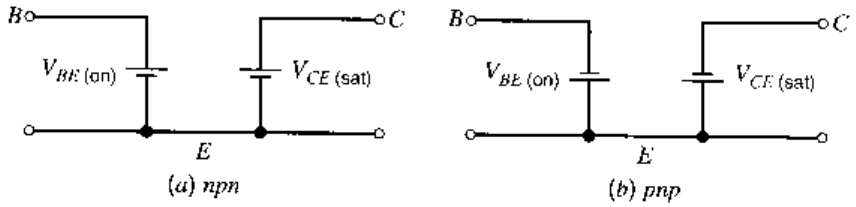


Figure 1.13 Large-signal models for bipolar transistors in the saturation region.

in practice (1.60) is applicable only in the forward-bias region, since second-order effects dominate under reverse-bias conditions and typically result in a junction current several orders of magnitude larger than I_{ES}. The junction current that flows under reverse-bias conditions is often called the *leakage current* of the junction.

Returning to Fig. 1.11, we can describe the *collector* current resulting from $n_{p2}(x)$ alone as

$$I_{CR} = -I_{CS}\left(\exp\frac{V_{BC}}{V_T} - 1\right) \tag{1.61}$$

where I_{CS} is a constant. The total collector current I_C is given by I_{CR} plus the fraction of I_{EF} that reaches the collector (allowing for recombination and reverse emitter injection). Thus

$$I_C = \alpha_F I_{ES}\left(\exp\frac{V_{BE}}{V_T} - 1\right) - I_{CS}\left(\exp\frac{V_{BC}}{V_T} - 1\right) \tag{1.62}$$

where α_F has been defined previously by (1.51). Similarly, the total emitter current is composed of I_{EF} plus the fraction of I_{CR} that reaches the emitter with the transistor acting in an inverted mode. Thus

$$I_E = -I_{ES}\left(\exp\frac{V_{BE}}{V_T} - 1\right) + \alpha_R I_{CS}\left(\exp\frac{V_{BC}}{V_T} - 1\right) \tag{1.63}$$

where α_R is the ratio of emitter to collector current with the transistor operating *inverted* (i.e., with the collector-base junction forward biased and emitting carriers into the base and the emitter-base junction reverse biased and collecting carriers). Typical values of α_R are 0.5 to 0.8. An inverse current gain β_R is also defined

$$\beta_R = \frac{\alpha_R}{1 - \alpha_R} \tag{1.64}$$

and has typical values 1 to 5. This is the current gain of the transistor when operated inverted and is much lower than β_F because the device geometry and doping densities are designed to maximize β_F. The inverse-active region of device operation occurs for V_{CE} negative in an *npn* transistor and is shown in Fig. 1.12. In order to display these characteristics adequately in the same figure as the forward-active region, the negative voltage and current scales have been expanded. The inverse-active mode of operation is rarely encountered in analog circuits.

Equations 1.62 and 1.63 describe *npn* transistor operation in the saturation region when V_{BE} and V_{BC} are both positive, and also in the forward-active and inverse-active regions. These equations are the *Ebers-Moll* equations. In the forward-active region, they degenerate into a form similar to that of (1.35), (1.47), and (1.49) derived earlier. This can be shown by putting V_{BE} positive and V_{BC} negative in (1.62) and (1.63) to obtain

$$I_C = \alpha_F I_{ES}\left(\exp\frac{V_{BE}}{V_T} - 1\right) + I_{CS} \tag{1.65}$$

$$I_E = -I_{ES}\left(\exp\frac{V_{BE}}{V_T} - 1\right) - \alpha_R I_{CS} \tag{1.66}$$

Equation 1.65 is similar in form to (1.35) except that leakage currents that were previously neglected have now been included. This minor difference is significant only at high temperatures or very low operating currents. Comparison of (1.65) with (1.35) allows us to identify $I_S = \alpha_F I_{ES}$, and it can be shown[10] in general that

$$\alpha_F I_{ES} = \alpha_R I_{CS} = I_S \tag{1.67}$$

where this expression represents a reciprocity condition. Use of (1.67) in (1.62) and (1.63) allows the Ebers-Moll equations to be expressed in the general form

$$I_C = I_S\left(\exp\frac{V_{BE}}{V_T} - 1\right) - \frac{I_S}{\alpha_R}\left(\exp\frac{V_{BC}}{V_T} - 1\right) \tag{1.62a}$$

$$I_E = -\frac{I_S}{\alpha_F}\left(\exp\frac{V_{BE}}{V_T} - 1\right) + I_S\left(\exp\frac{V_{BC}}{V_T} - 1\right) \tag{1.63a}$$

This form is often used for computer representation of transistor large-signal behavior.

The effect of leakage currents mentioned above can be further illustrated as follows. In the forward-active region, from (1.66)

$$I_{ES}\left(\exp\frac{V_{BE}}{V_T} - 1\right) = -I_E - \alpha_R I_{CS} \tag{1.68}$$

Substitution of (1.68) in (1.65) gives

$$I_C = -\alpha_F I_E + I_{CO} \tag{1.69}$$

where

$$I_{CO} = I_{CS}(1 - \alpha_R \alpha_F) \tag{1.69a}$$

and I_{CO} is the collector-base leakage current with the emitter open. Although I_{CO} is given theoretically by (1.69a), in practice, surface leakage effects dominate when the collector-base junction is reverse biased and I_{CO} is typically several orders of magnitude larger than the value given by (1.69a). However, (1.69) is still valid if the appropriate measured value for I_{CO} is used. Typical values of I_{CO} are from 10^{-10} to 10^{-12} A at 25°C, and the magnitude doubles about every 8°C. As a consequence, these leakage terms can become very significant at high temperatures. For example, consider the base current I_B. From Fig. 1.5 this is

$$I_B = -(I_C + I_E) \tag{1.70}$$

If I_E is calculated from (1.69) and substituted in (1.70), the result is

$$I_B = \frac{1 - \alpha_F}{\alpha_F} I_C - \frac{I_{CO}}{\alpha_F} \tag{1.71}$$

But from (1.50)

$$\beta_F = \frac{\alpha_F}{1 - \alpha_F} \tag{1.72}$$

and use of (1.72) in (1.71) gives

$$I_B = \frac{I_C}{\beta_F} - \frac{I_{CO}}{\alpha_F} \tag{1.73}$$

Since the two terms in (1.73) have opposite signs, the effect of I_{CO} is to *decrease* the magnitude of the external base current at a given value of collector current.

■ **EXAMPLE**

If I_{CO} is 10^{-10} A at 24°C, estimate its value at 120°C.

Assuming that I_{CO} doubles every 8°C, we have

$$I_{CO}(120°\text{C}) = 10^{-10} \times 2^{12}$$
$$= 0.4\ \mu\text{A}$$

■

1.3.4 Transistor Breakdown Voltages

In Section 1.2.2 the mechanism of avalanche breakdown in a *pn* junction was described. Similar effects occur at the base-emitter and base-collector junctions of a transistor and these effects limit the maximum voltages that can be applied to the device.

First consider a transistor in the common-base configuration shown in Fig. 1.14*a* and supplied with a constant emitter current. Typical $I_C - V_{CB}$ characteristics for an *npn* transistor in such a connection are shown in Fig. 1.14*b*. For $I_E = 0$ the collector-base junction breaks down at a voltage BV_{CBO}, which represents collector-base breakdown with the emitter open. For finite values of I_E, the effects of avalanche multiplication are apparent for values of V_{CB} below BV_{CBO}. In the example shown, the effective common-base current

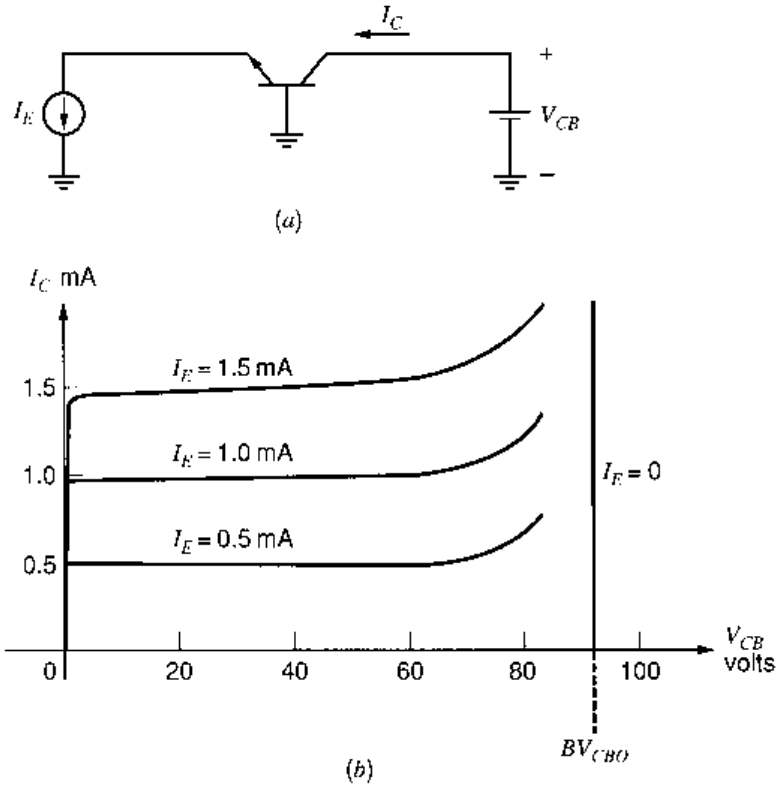


Figure 1.14 Common-base transistor connection. (*a*) Test circuit. (*b*) $I_C - V_{CB}$ characteristics.

gain $\alpha_F = I_C/I_E$ becomes larger than unity for values of V_{CB} above about 60 V. Operation in this region (but below BV_{CBO}) can, however, be safely undertaken if the device power dissipation is not excessive. The considerations of Section 1.2.2 apply to this situation, and neglecting leakage currents, we can calculate the collector current in Fig. 1.14*a* as

$$I_C = -\alpha_F I_E M \tag{1.74}$$

where M is defined by (1.26) and thus

$$I_C = -\alpha_F I_E \frac{1}{1 - \left(\dfrac{V_{CB}}{BV_{CBO}}\right)^n} \tag{1.75}$$

One further point to note about the common-base characteristics of Fig. 1.14*b* is that for low values of V_{CB} where avalanche effects are negligible, the curves show very little of the Early effect seen in the common-emitter characteristics. Base widening still occurs in this configuration as V_{CB} is increased, but unlike the common-emitter connection, it produces little change in I_C. This is because I_E is now fixed instead of V_{BE} or I_B, and in Fig. 1.9, this means the slope of the minority-carrier concentration at the emitter edge of the base is fixed. Thus the collector current remains almost unchanged.

Now consider the effect of avalanche breakdown on the common-emitter characteristics of the device. Typical characteristics are shown in Fig. 1.12, and breakdown occurs at a value BV_{CEO}, which is sometimes called the sustaining voltage LV_{CEO}. As in previous cases, operation near the breakdown voltage is destructive to the device only if the current (and thus the power dissipation) becomes excessive.

The effects of avalanche breakdown on the common-emitter characteristics are more complex than in the common-base configuration. This is because hole-electron *pairs* are produced by the avalanche process and the holes are swept into the base, where they effectively contribute to the base current. In a sense the avalanche current is then *amplified*

by the transistor. The base current is still given by

$$I_B = -(I_C + I_E) \tag{1.76}$$

Equation 1.74 still holds, and substitution of this in (1.76) gives

$$I_C = \frac{M\alpha_F}{1 - M\alpha_F} I_B \tag{1.77}$$

where

$$M = \frac{1}{1 - \left(\dfrac{V_{CB}}{BV_{CBO}}\right)^n} \tag{1.78}$$

Equation 1.77 shows that I_C approaches infinity as $M\alpha_F$ approaches unity. That is, the effective β approaches infinity because of the additional base-current contribution from the avalanche process itself. The value of BV_{CEO} can be determined by solving

$$M\alpha_F = 1 \tag{1.79}$$

If we assume that $V_{CB} \simeq V_{CE}$, this gives

$$\frac{\alpha_F}{1 - \left(\dfrac{BV_{CEO}}{BV_{CBO}}\right)^n} = 1 \tag{1.80}$$

and this results in

$$\frac{BV_{CEO}}{BV_{CBO}} = \sqrt[n]{1 - \alpha_F}$$

and thus

$$BV_{CEO} \simeq \frac{BV_{CBO}}{\sqrt[n]{\beta_F}} \tag{1.81}$$

Equation 1.81 shows that BV_{CEO} is less than BV_{CBO} by a substantial factor. However, the value of BV_{CBO}, which must be used in (1.81), is the *plane* junction breakdown of the collector-base junction, neglecting any edge effects. This is because it is only collector-base avalanche current actually under the emitter that is amplified as described in the previous calculation. However, as explained in Section 1.2.2, the measured value of BV_{CBO} is usually determined by avalanche in the curved region of the collector, which is remote from the active base. Consequently, for typical values of $\beta_F = 100$ and $n = 4$, the value of BV_{CEO} is about one-half of the measured BV_{CBO} and not 30 percent as (1.81) would indicate.

Equation 1.81 explains the shape of the breakdown characteristics of Fig. 1.12 if the dependence of β_F on collector current is included. As V_{CE} is increased from zero with $I_B = 0$, the initial collector current is approximately $\beta_F I_{CO}$ from (1.73); since I_{CO} is typically several picoamperes, the collector current is very small. As explained in the next section, β_F is small at low currents, and thus from (1.81) the breakdown voltage is high. However, as avalanche breakdown begins in the device, the value of I_C increases and thus β_F increases. From (1.81) this causes a *decrease* in the breakdown voltage and the characteristic bends back as shown in Fig. 1.12 and exhibits a negative slope. At higher collector currents, β_F approaches a constant value and the breakdown curve with $I_B = 0$ becomes perpendicular to the V_{CE} axis. The value of V_{CE} in this region of the curve is

usually defined to be BV_{CEO}, since this is the maximum voltage the device can sustain. The value of β_F to be used to calculate BV_{CEO} in (1.81) is thus the *peak* value of β_F. Note from (1.81) that high-β transistors will thus have low values of BV_{CEO}.

The base-emitter junction of a transistor is also subject to avalanche breakdown. However, the doping density in the emitter is made very large to ensure a high value of β_F [N_D is made large in (1.45) to reduce I_{B2}]. Thus the base is the more lightly doped side of the junction and determines the breakdown characteristic. This can be contrasted with the collector-base junction, where the collector is the more lightly doped side and results in typical values of BV_{CBO} of 20 to 80 V or more. The base is typically an order of magnitude more heavily doped than the collector, and thus the base-emitter breakdown voltage is much less than BV_{CBO} and is typically about 6 to 8 V. This is designed BV_{EBO}. The breakdown voltage for inverse-active operation shown in Fig. 1.12 is approximately equal to this value because the base-emitter junction is reverse biased in this mode of operation.

The base-emitter breakdown voltage of 6 to 8 V provides a convenient reference voltage in integrated-circuit design, and this is often utilized in the form of a *Zener* diode. However, care must be taken to ensure that all other transistors in a circuit are protected against reverse base-emitter voltages sufficient to cause breakdown. This is because, unlike collector-base breakdown, base-emitter breakdown *is* damaging to the device. It can cause a large degradation in β_F, depending on the duration of the breakdown-current flow and its magnitude.[11] If the device is used purely as a Zener diode, this is of no consequence, but if the device is an amplifying transistor, the β_F degradation may be serious.

■ **EXAMPLE**

If the collector doping density in a transistor is 2×10^{15} atoms/cm^3 and is much less than the base doping, calculate BV_{CEO} for $\beta = 100$ and $n = 4$. Assume $\mathscr{E}_{crit} = 3 \times 10^5$ V/cm.

The plane breakdown voltage in the collector can be calculated from (1.24) using $\mathscr{E}_{max} = \mathscr{E}_{crit}$:

$$BV_{CBO} = \frac{\epsilon (N_A + N_D)}{2qN_AN_D}\mathscr{E}_{crit}^2$$

Since $N_D \ll N_A$, we have

$$BV_{CBO}|_{plane} = \frac{\epsilon}{2qN_D}\mathscr{E}_{crit}^2 = \frac{1.04 \times 10^{-12}}{2 \times 1.6 \times 10^{-19} \times 2 \times 10^{15}} \times 9 \times 10^{10} \text{ V} = 146 \text{ V}$$

From (1.81)

$$BV_{CEO} = \frac{146}{\sqrt[4]{100}}\text{V} = 46 \text{ V}$$

■

1.3.5 Dependence of Transistor Current Gain β_F on Operating Conditions

Although most first-order analyses of integrated circuits make the assumption that β_F is constant, this parameter does in fact depend on the operating conditions of the transistor. It was shown in Section 1.3.2, for example, that increasing the value of V_{CE} increases I_C while producing little change in I_B, and thus the effective β_F of the transistor increases. In Section 1.3.4 it was shown that as V_{CE} approaches the breakdown voltage, BV_{CEO}, the collector current increases due to avalanche multiplication in the collector. Equation 1.77 shows that the effective current gain approaches infinity as V_{CE} approaches BV_{CEO}.

In addition to the effects just described, β_F also varies with both temperature and transistor collector current. This is illustrated in Fig. 1.15, which shows typical curves of β_F versus I_C at three different temperatures for an *npn* integrated circuit transistor. It is evident that β_F increases as temperature increases, and a typical temperature coefficient for β_F is +7000 ppm/°C (where ppm signifies *parts per million*). This temperature dependence of β_F is due to the effect of the extremely high doping density in the emitter,[12] which causes the emitter injection efficiency γ to increase with temperature.

The variation of β_F with collector current, which is apparent in Fig. 1.15, can be divided into three regions. Region I is the low-current region, where β_F decreases as I_C decreases. Region II is the midcurrent region, where β_F is approximately constant. Region III is the high-current region, where β_F decreases as I_C increases. The reasons for this behavior of β_F with I_C can be better appreciated by plotting base current I_B and collector current I_C on a log scale as a function of V_{BE}. This is shown in Fig. 1.16, and because of the log scale on the vertical axis, the value of ln β_F can be obtained directly as the distance between the two curves.

At moderate current levels represented by region II in Figs. 1.15 and 1.16, both I_C and I_B follow the ideal behavior, and

$$I_C = I_S \exp \frac{V_{BE}}{V_T} \tag{1.82}$$

$$I_B = \frac{I_S}{\beta_{FM}} \exp \frac{V_{BE}}{V_T} \tag{1.83}$$

where β_{FM} is the maximum value of β_F and is given by (1.48).

At low current levels, I_C still follows the ideal relationship of (1.82), and the decrease in β_F is due to an additional component in I_B, which is mainly due to recombination of carriers in the base-emitter depletion region and is present at any current level. However, at higher current levels the base current given by (1.83) dominates, and this additional component has little effect. The base current resulting from recombination in the depletion region is[5]

$$I_{BX} \simeq I_{SX} \exp \frac{V_{BE}}{mV_T} \tag{1.84}$$

where

$$m \simeq 2$$

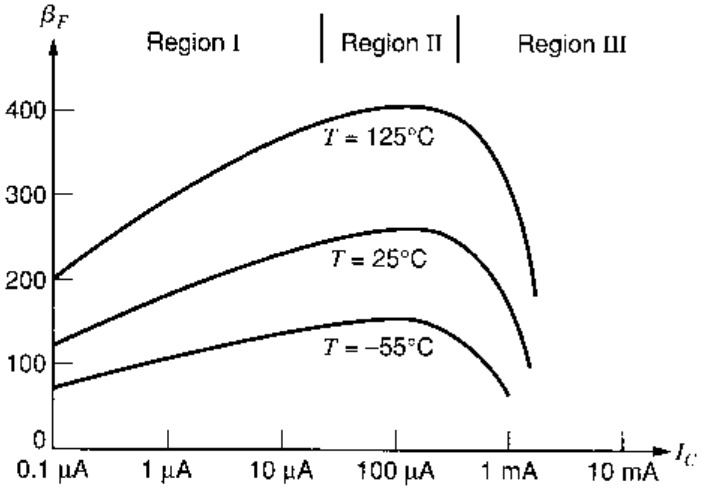


Figure 1.15 Typical curves of β_F versus I_C for an *npn* integrated-circuit transistor with 6 μm^2 emitter area.

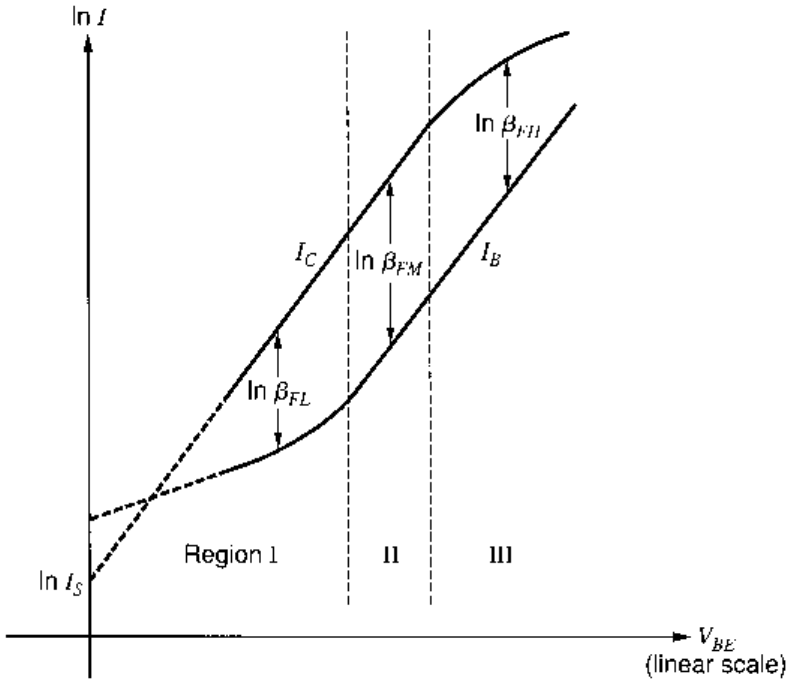


Figure 1.16 Base and collector currents of a bipolar transistor plotted on a log scale versus V_{BE} on a linear scale. The distance between the curves is a direct measure of ln β_F.

At very low collector currents, where (1.84) dominates the base current, the current gain can be calculated from (1.82) and (1.84) as

$$\beta_{FL} \simeq \frac{I_C}{I_{BX}} = \frac{I_S}{I_{SX}} \exp \frac{V_{BE}}{V_T} \left(1 - \frac{1}{m}\right) \tag{1.85}$$

Substitution of (1.82) in (1.85) gives

$$\beta_{FL} \simeq \frac{I_S}{I_{SX}} \left(\frac{I_C}{I_S}\right)^{[1-(1/m)]} \tag{1.86}$$

If $m \simeq 2$, then (1.86) indicates that β_F is proportional to $\sqrt{I_C}$ at very low collector currents.

At high current levels, the base current I_B tends to follow the relationship of (1.83), and the decrease in β_F in region III is due mainly to a decrease in I_C below the value given by (1.82). (In practice the measured curve of I_B versus V_{BE} in Fig. 1.16 may also deviate from a straight line at high currents due to the influence of voltage drop across the base resistance.) The decrease in I_C is due partly to the effect of high-level injection, and at high current levels the collector current approaches[7]

$$I_C \simeq I_{SH} \exp \frac{V_{BE}}{2V_T} \tag{1.87}$$

The current gain in this region can be calculated from (1.87) and (1.83) as

$$\beta_{FH} \simeq \frac{I_{SH}}{I_S} \beta_{FM} \exp\left(-\frac{V_{BE}}{2V_T}\right) \tag{1.88}$$

Substitution of (1.87) in (1.88) gives

$$\beta_{FH} \simeq \frac{I_{SH}^2}{I_S} \beta_{FM} \frac{1}{I_C}$$

Thus β_F decreases rapidly at high collector currents.

In addition to the effect of high-level injection, the value of β_F at high currents is also decreased by the onset of the Kirk effect,[13] which occurs when the minority-carrier concentration in the collector becomes comparable to the donor-atom doping density. The base region of the transistor then stretches out into the collector and becomes greatly enlarged.

1.4 Small-Signal Models of Bipolar Transistors

Analog circuits often operate with signal levels that are small compared to the bias currents and voltages in the circuit. In these circumstances, *incremental* or *small-signal* models can be derived that allow calculation of circuit gain and terminal impedances without the necessity of including the bias quantities. A hierarchy of models with increasing complexity can be derived, and the more complex ones are generally reserved for computer analysis. Part of the designer's skill is knowing which elements of the model can be omitted when performing hand calculations on a particular circuit, and this point is taken up again later.

Consider the bipolar transistor in Fig. 1.17*a* with bias voltages V_{BE} and V_{CC} applied as shown. These produce a quiescent collector current, I_C, and a quiescent base current, I_B, and the device is in the *forward-active region*. A *small-signal* input voltage v_i is applied in series with V_{BE} and produces a small variation in base current i_b and a small variation in collector current i_c. Total values of base and collector currents are I_b and I_c, respectively, and thus $I_b = (I_B + i_b)$ and $I_c = (I_C + i_c)$. The carrier concentrations in the base of the transistor corresponding to the situation in Fig. 1.17*a* are shown in Fig. 1.17*b*. With only

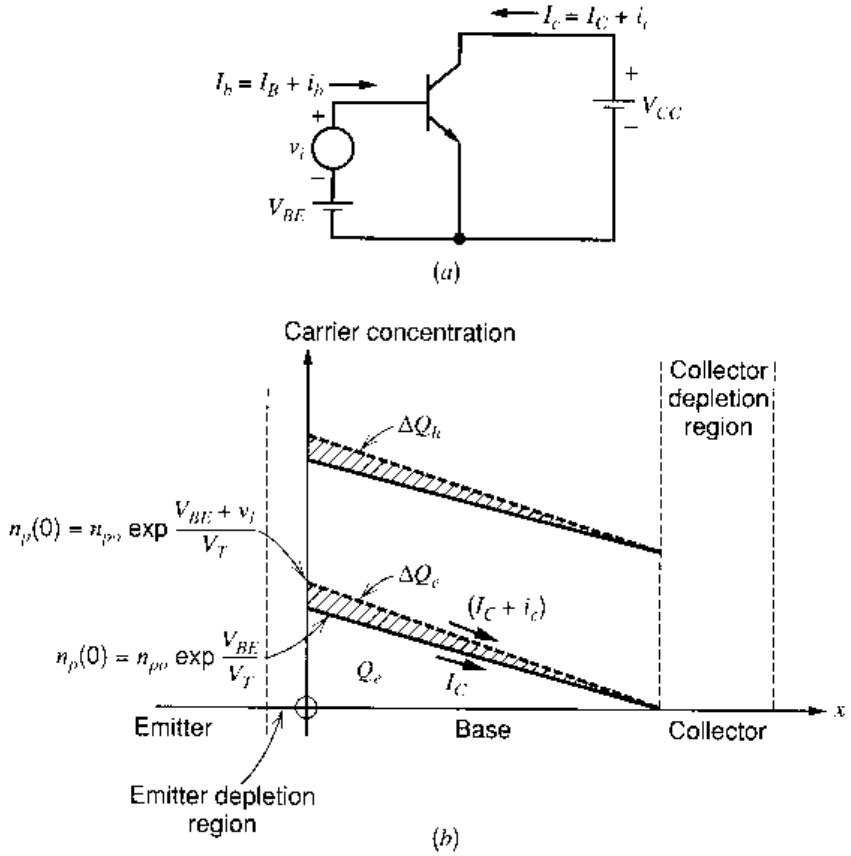


Figure 1.17 Effect of a small-signal input voltage applied to a bipolar transistor. (*a*) Circuit schematic. (*b*) Corresponding changes in carrier concentrations in the base when the device is in the forward-active region.

bias voltages applied, the carrier concentrations are given by the solid lines. Application of the small-signal voltage v_i causes $n_p(0)$ at the emitter edge of the base to increase, and produces the concentrations shown by the dotted lines. These pictures can now be used to derive the various elements in the small-signal equivalent circuit of the bipolar transistor.

1.4.1 Transconductance

The transconductance is defined as

$$g_m = \frac{dI_C}{dV_{BE}} \tag{1.89}$$

Since

$$\Delta I_C = \frac{dI_C}{dV_{BE}} \Delta V_{BE}$$

we can write

$$\Delta I_C = g_m \Delta V_{BE}$$

and thus

$$i_c = g_m v_i \tag{1.90}$$

The value of g_m can be found by substituting (1.35) in (1.89) to give

$$g_m = \frac{d}{dV_{BE}} I_S \exp \frac{V_{BE}}{V_T} = \frac{I_S}{V_T} \exp \frac{V_{BE}}{V_T} = \frac{I_C}{V_T} = \frac{qI_C}{kT} \tag{1.91}$$

The transconductance thus depends linearly on the bias current I_C and is 38 mA/V for $I_C = 1$ mA at 25°C for any bipolar transistor of either polarity (*npn* or *pnp*), of any size, and made of any material (Si, Ge, GaAs).

To illustrate the limitations on the use of small-signal analysis, the foregoing relation will be derived in an alternative way. The total collector current in Fig. 1.17*a* can be calculated using (1.35) as

$$I_c = I_S \exp \frac{V_{BE} + v_i}{V_T} = I_S \exp \frac{V_{BE}}{V_T} \exp \frac{v_i}{V_T} \tag{1.92}$$

But the collector bias current is

$$I_C = I_S \exp \frac{V_{BE}}{V_T} \tag{1.93}$$

and use of (1.93) in (1.92) gives

$$I_c = I_C \exp \frac{v_i}{V_T} \tag{1.94}$$

If $v_i < V_T$, the exponential in (1.94) can be expanded in a power series,

$$I_c = I_C \left[1 + \frac{v_i}{V_T} + \frac{1}{2} \left(\frac{v_i}{V_T} \right)^2 + \frac{1}{6} \left(\frac{v_i}{V_T} \right)^3 + \dots \right] \tag{1.95}$$

Now the incremental collector current is

$$i_c = I_c - I_C \tag{1.96}$$

and substitution of (1.96) in (1.95) gives

$$i_c = \frac{I_C}{V_T} v_i + \frac{1}{2}\frac{I_C}{V_T^2} v_i^2 + \frac{1}{6}\frac{I_C}{V_T^3} v_i^3 + \dots \tag{1.97}$$

If $v_i \ll V_T$, (1.97) reduces to (1.90), and the small-signal analysis is valid. The criterion for use of small-signal analysis is thus $v_i = \Delta V_{BE} \ll 26$ mV at 25°C. In practice, if ΔV_{BE} is less than 10 mV, the small-signal analysis is accurate within about 10 percent.

1.4.2 Base-Charging Capacitance

Figure 1.17*b* shows that the change in base-emitter voltage $\Delta V_{BE} = v_i$ has caused a change $\Delta Q_e = q_e$ in the minority-carrier charge in the base. By charge-neutrality requirements, there is an equal change $\Delta Q_h = q_h$ in the majority-carrier charge in the base. Since majority carriers are supplied by the base lead, the application of voltage v_i requires the supply of charge q_h to the base, and the device has an apparent input capacitance

$$C_b = \frac{q_h}{v_i} \tag{1.98}$$

The value of C_b can be related to fundamental device parameters as follows. If (1.39) is divided by (1.33), we obtain

$$\frac{Q_e}{I_C} = \frac{W_B^2}{2D_n} = \tau_F \tag{1.99}$$

The quantity τ_F has the dimension of time and is called the base transit time in the forward direction. Since it is the ratio of the charge in transit (Q_e) to the current flow (I_C), it can be identified as the average time per carrier spent in crossing the base. To a first order it is independent of operating conditions and has typical values 10 to 500 ps for integrated *npn* transistors and 1 to 40 ns for lateral *pnp* transistors. Practical values of τ_F tend to be somewhat lower than predicted by (1.99) for diffused transistors that have nonuniform base doping.[14] However, the functional dependence on base width W_B and diffusion constant D_n is as predicted by (1.99).

From (1.99)

$$\Delta Q_e = \tau_F \Delta I_C \tag{1.100}$$

But since $\Delta Q_e = \Delta Q_h$, we have

$$\Delta Q_h = \tau_F \Delta I_C \tag{1.101}$$

and this can be written

$$q_h = \tau_F i_c \tag{1.102}$$

Use of (1.102) in (1.98) gives

$$C_b = \tau_F \frac{i_c}{v_i} \tag{1.103}$$

and substitution of (1.90) in (1.103) gives

$$C_b = \tau_F g_m \tag{1.104}$$

$$= \tau_F \frac{qI_C}{kT} \tag{1.105}$$

Thus the small-signal, base-charging capacitance is proportional to the collector bias current.

In the inverse-active mode of operation, an equation similar to (1.99) relates stored charge and current via a time constant τ_R. This is typically orders of magnitude larger than τ_F because the device structure and doping are optimized for operation in the forward-active region. Since the saturation region is a combination of forward-active and inverse-active operation, inclusion of the parameter τ_R in a SPICE listing will model the large charge storage that occurs in saturation.

1.4.3 Input Resistance

In the forward-active region, the base current is related to the collector current by (1.47) as

$$I_B = \frac{I_C}{\beta_F} \tag{1.47}$$

Small changes in I_B and I_C can be related using (1.47):

$$\Delta I_B = \frac{d}{dI_C}\left(\frac{I_C}{\beta_F}\right)\Delta I_C \tag{1.106}$$

and thus

$$\beta_0 = \frac{\Delta I_C}{\Delta I_B} = \frac{i_c}{i_b} = \left[\frac{d}{dI_C}\left(\frac{I_C}{\beta_F}\right)\right]^{-1} \tag{1.107}$$

where β_0 is the *small-signal* current gain of the transistor. Note that if β_F is constant, then $\beta_F = \beta_0$. Typical values of β_0 are close to those of β_F, and in subsequent chapters little differentiation is made between these quantities. A single value of β is often assumed for a transistor and then used for both ac and dc calculations.

Equation 1.107 relates the change in base current i_b to the corresponding change in collector current i_c, and the device has a small-signal input resistance given by

$$r_\pi = \frac{v_i}{i_b} \tag{1.108}$$

Substitution of (1.107) in (1.108) gives

$$r_\pi = \frac{v_i}{i_c}\beta_0 \tag{1.109}$$

and use of (1.90) in (1.109) gives

$$r_\pi = \frac{\beta_0}{g_m} \tag{1.110}$$

Thus the small-signal input shunt resistance of a bipolar transistor depends on the current gain and is inversely proportional to I_C.

1.4.4 Output Resistance

In Section 1.3.2 the effect of changes in collector-emitter voltage V_{CE} on the large-signal characteristics of the transistor was described. It follows from that treatment that small changes ΔV_{CE} in V_{CE} produce corresponding changes ΔI_C in I_C, where

$$\Delta I_C = \frac{\partial I_C}{\partial V_{CE}}\Delta V_{CE} \tag{1.111}$$

Substitution of (1.55) and (1.57) in (1.111) gives

$$\frac{\Delta V_{CE}}{\Delta I_C} = \frac{V_A}{I_C} = r_o \tag{1.112}$$

where V_A is the Early voltage and r_o is the small-signal output resistance of the transistor. Since typical values of V_A are 50 to 100 V, corresponding values of r_o are 50 to 100 kΩ for $I_C = 1$mA. Note that r_o is inversely proportional to I_C, and thus r_o can be related to g_m, as are many of the other small-signal parameters.

$$r_o = \frac{1}{\eta g_m} \tag{1.113}$$

where

$$\eta = \frac{kT}{qV_A} \tag{1.114}$$

If $V_A = 100$ V, then $\eta = 2.6 \times 10^{-4}$ at 25°C. Note that $1/r_o$ is the slope of the output characteristics of Fig. 1.10.

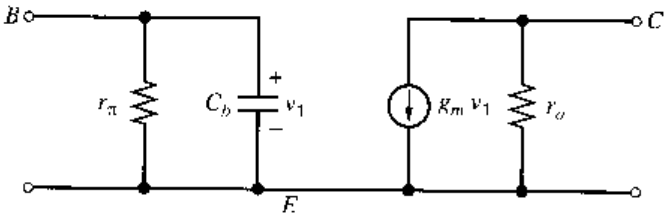
1.4.5 Basic Small-Signal Model of the Bipolar Transistor

Combination of the above small-signal circuit elements yields the small-signal model of the bipolar transistor shown in Fig. 1.18. This is valid for both *npn* and *pnp* devices in the forward-active region and is called the *hybrid-π* model. Collector, base, and emitter nodes are labeled *C*, *B* and *E*, respectively. The elements in this circuit are present in the equivalent circuit of *any* bipolar transistor and are specified by relatively few parameters (β, τ_F, η, I_C). Note that in the evaluation of the small-signal parameters for *pnp* transistors, the *magnitude only* of I_C is used. In the following sections, further elements are added to this model to account for parasitics and second-order effects.

1.4.6 Collector-Base Resistance

Consider the effect of variations in V_{CE} on the minority charge in the base as illustrated in Fig. 1.9. An increase in V_{CE} causes an increase in the collector depletion-layer width and consequent reduction of base width. This causes a reduction in the total minority-carrier charge stored in the base and thus a reduction in base current I_B due to a reduction in I_{B1} given by (1.40). Since an increase ΔV_{CE} in V_{CE} causes a *decrease* ΔI_B in I_B, this effect can be modeled by inclusion of a resistor r_μ from collector to base of the model of Fig. 1.18. If V_{BE} is assumed held constant, the value of this resistor can be determined as follows.

$$r_\mu = \frac{\Delta V_{CE}}{\Delta I_{B1}} = \frac{\Delta V_{CE}}{\Delta I_C}\frac{\Delta I_C}{\Delta I_{B1}} \tag{1.115}$$



$$r_\pi = \frac{\beta}{g_m},\ r_o = \frac{1}{\eta g_m},\ g_m = \frac{qI_C}{kT},\ C_b = \tau_F g_m$$

Figure 1.18 Basic bipolar transistor small-signal equivalent circuit.

Substitution of (1.112) in (1.115) gives

$$r_\mu = r_o \frac{\Delta I_C}{\Delta I_{B1}} \tag{1.116}$$

If the base current I_B is composed entirely of component I_{B1}, then (1.107) can be used in (1.116) to give

$$r_\mu = \beta_0 r_o \tag{1.117}$$

This is a lower limit for r_μ. In practice, I_{B1} is typically less than 10 percent of I_B [component I_{B2} from (1.42) dominates] in integrated npn transistors, and since I_{B1} is very small, the change ΔI_{B1} in I_{B1} for a given ΔV_{CE} and ΔI_C is also very small. Thus a typical value for r_μ is greater than $10\beta_0 r_o$. For lateral pnp transistors, recombination in the base is more significant, and r_μ is in the range $2\beta_0 r_o$ to $5\beta_0 r_o$.

1.4.7 Parasitic Elements in the Small-Signal Model

The elements of the bipolar transistor small-signal equivalent circuit considered so far may be considered basic in the sense that they arise directly from essential processes in the device. However, technological limitations in the fabrication of transistors give rise to a number of parasitic elements that must be added to the equivalent circuit for most integrated-circuit transistors. A cross section of a typical npn transistor in a junction-isolated process is shown in Fig. 1.19. The means of fabricating such devices is described in Chapter 2.

As described in Section 1.2, all pn junctions have a voltage-dependent capacitance associated with the depletion region. In the cross section of Fig. 1.19, three depletion-region capacitances can be identified. The base-emitter junction has a depletion-region capacitance C_{je} and the base-collector and collector-substrate junctions have capacitances C_μ and C_{cs}, respectively. The base-emitter junction closely approximates an abrupt junction due to the steep rise of the doping density caused by the heavy doping in the emitter. Thus the variation of C_{je} with bias voltage is well approximated by (1.21). The collector-base junction behaves like a graded junction for small bias voltages since the doping density is a function of distance near the junction. However, for larger reverse-bias values (more than about a volt), the junction depletion region spreads into the collector, which is

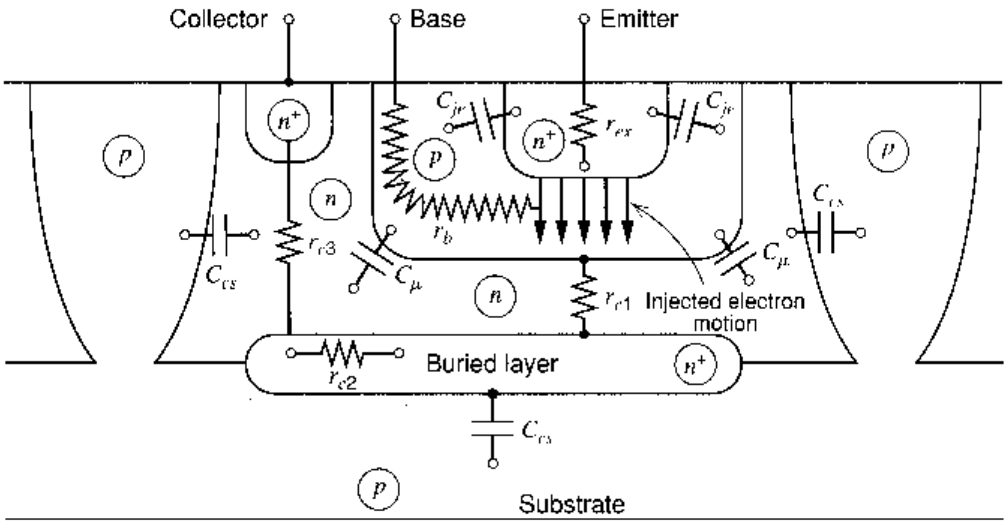


Figure 1.19 Integrated-circuit npn bipolar transistor structure showing parasitic elements. (Not to scale.)

uniformly doped, and thus for devices with thick collectors the junction tends to behave like an abrupt junction with uniform doping. Many modern high-speed processes, however, have very thin collector regions (of the order of one micron), and the collector depletion region can extend all the way to the buried layer for quite small reverse-bias voltages. When this occurs, both the depletion region and the associated capacitance vary quite slowly with bias voltage. The collector-base capacitance C_μ thus tends to follow (1.22) for very small bias voltages and (1.21) for large bias voltages in thick-collector devices. In practice, measurements show that the variation of C_μ with bias voltage for most devices can be approximated by

$$C_\mu = \frac{C_{\mu 0}}{\left(1 - \dfrac{V}{\psi_o}\right)^n} \tag{1.117a}$$

where V is the forward bias on the junction and n is an exponent between about 0.2 and 0.5. The third parasitic capacitance in a monolithic npn transistor is the collector-substrate capacitance C_{cs}, and for large reverse bias voltages this varies according to the abrupt junction equation (1.21) for junction-isolated devices. In the case of oxide-isolated devices, however, the deep p diffusions used to isolate the devices are replaced by oxide. The sidewall component of C_{cs} then consists of a fixed oxide capacitance. Equation 1.117a may then be used to model C_{cs}, but a value of n less than 0.5 gives the best approximation. In general, (1.117a) will be used to model all three parasitic capacitances with subscripts e, c, and s on n and ψ_0 used to differentiate emitter-base, collector-base, and collector-substrate capacitances, respectively. Typical zero-bias values of these parasitic capacitances for a minimum-size npn transistor in a modern oxide-isolated process are $C_{je0} \simeq 10$ fF, $C_{\mu 0} \simeq 10$ fF, and $C_{cs0} \simeq 20$ fF. Values for other devices are summarized in Chapter 2.

As described in Chapter 2, lateral pnp transistors have a parasitic capacitance C_{bs} from base to substrate in place of C_{cs}. Note that the substrate is always connected to the most negative voltage supply in the circuit in order to ensure that all isolation regions are separated by reverse-biased junctions. Thus the substrate is an ac ground, and all parasitic capacitance to the substrate is connected to ground in an equivalent circuit.

The final elements to be added to the small-signal model of the transistor are resistive parasitics. These are produced by the finite resistance of the silicon between the top contacts on the transistor and the active base region beneath the emitter. As shown in Fig. 1.19, there are significant resistances r_b and r_c in series with the base and collector contacts, respectively. There is also a resistance r_{ex} of several ohms in series with the emitter lead that can become important at high bias currents. (Note that the collector resistance r_c is actually composed of three parts labeled r_{c1}, r_{c2}, and r_{c3}.) Typical values of these parameters are $r_b = 50$ to 500 Ω, $r_{ex} = 1$ to 3 Ω, and $r_c = 20$ to 500 Ω. The value of r_b varies significantly with collector current because of *current crowding*.[15] This occurs at high collector currents where the dc base current produces a lateral voltage drop in the base that tends to forward bias the base-emitter junction preferentially around the edges of the emitter. Thus the transistor action tends to occur along the emitter periphery rather than under the emitter itself, and the distance from the base contact to the active base region is reduced. Consequently, the value of r_b is reduced, and in a typical npn transistor, r_b may decrease 50 percent as I_C increases from 0.1 mA to 10 mA.

The value of these parasitic resistances can be reduced by changes in the device structure. For example, a large-area transistor with multiple base and emitter stripes will have a smaller value of r_b. The value of r_c is reduced by inclusion of the low-resistance buried n^+ layer beneath the collector.

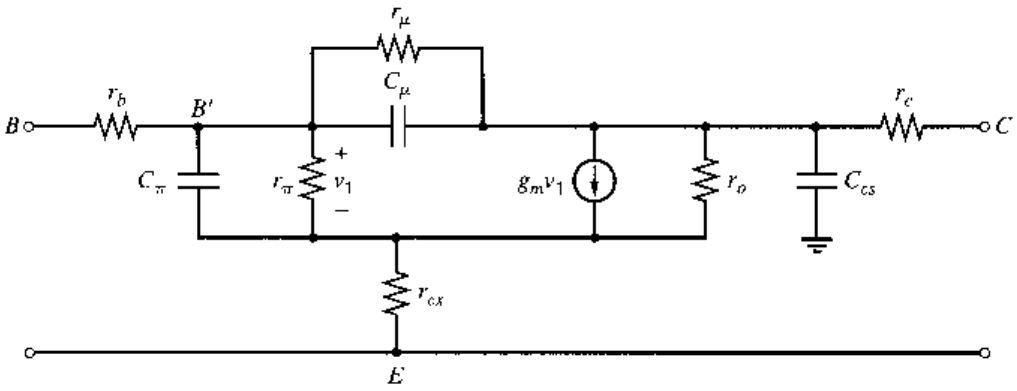


Figure 1.20 Complete bipolar transistor small-signal equivalent circuit.

The addition of the resistive and capacitive parasitics to the basic small-signal circuit of Fig. 1.18 gives the complete small-signal equivalent circuit of Fig. 1.20. The internal base node is labeled B' to distinguish it from the external base contact B. The capacitance C_π contains the base-charging capacitance C_b and the emitter-base depletion layer capacitance C_{je}.

$$C_\pi = C_b + C_{je} \tag{1.118}$$

Note that the representation of parasitics in Fig. 1.20 is an approximation in that lumped elements have been used. In practice, as suggested by Fig. 1.19, C_μ is distributed across r_b and C_{cs} is distributed across r_c. This lumped representation is adequate for most purposes but can introduce errors at very high frequencies. It should also be noted that while the parasitic resistances of Fig. 1.20 can be very important at high bias currents or for high-frequency operation, they are usually omitted from the equivalent circuit for low-frequency calculations, particularly for collector bias currents less than 1 mA.

■ EXAMPLE

Derive the complete small-signal equivalent circuit for a bipolar transistor at $I_C = 1$ mA, $V_{CB} = 3$ V, and $V_{CS} = 5$ V. Device parameters are $C_{je0} = 10$ fF, $n_e = 0.5$, $\psi_{0e} = 0.9$ V, $C_{\mu 0} = 10$ fF, $n_c = 0.3$, $\psi_{0c} = 0.5$ V, $C_{cs0} = 20$ fF, $n_s = 0.3$, $\psi_{0s} = 0.65$ V, $\beta_0 = 100$, $\tau_F = 10$ ps, $V_A = 20$ V, $r_b = 300\ \Omega$, $r_c = 50\ \Omega$, $r_{ex} = 5\ \Omega$, $r_\mu = 10\ \beta_0 r_o$.

Since the base-emitter junction is forward biased, the value of C_{je} is difficult to determine for reasons described in Section 1.2.1. Either a value can be determined by computer or a reasonable estimation is to double C_{je0}. Using the latter approach, we estimate

$$C_{je} = 20 \text{ fF}$$

Using (1.117a) gives, for the collector-base capacitance,

$$C_\mu = \frac{C_{\mu 0}}{\left(1 + \dfrac{V_{CB}}{\psi_{0c}}\right)^{n_c}} = \frac{10}{\left(1 + \dfrac{3}{0.5}\right)^{0.3}} = 5.6 \text{ fF}$$

The collector-substrate capacitance can also be calculated using (1.117a)

$$C_{cs} = \frac{C_{cs0}}{\left(1 + \dfrac{V_{CS}}{\psi_{0s}}\right)^{n_s}} = \frac{20}{\left(1 + \dfrac{5}{0.65}\right)^{0.3}} = 10.5 \text{ fF}$$

From (1.91) the transconductance is

$$g_m = \frac{qI_C}{kT} = \frac{10^{-3}}{26 \times 10^{-3}} \text{A/V} = 38 \text{ mA/V}$$

From (1.104) the base-charging capacitance is

$$C_b = \tau_F g_m = 10 \times 10^{-12} \times 38 \times 10^{-3} \text{ F} = 0.38 \text{ pF}$$

The value of C_π from (1.118) is

$$C_\pi = 0.38 + 0.02 \text{ pF} = 0.4 \text{ pF}$$

The input resistance from (1.110) is

$$r_\pi = \frac{\beta_0}{g_m} = 100 \times 26 \ \Omega = 2.6 \text{ k}\Omega$$

The output resistance from (1.112) is

$$r_o = \frac{20}{10^{-3}} \ \Omega = 20 \text{ k}\Omega$$

and thus the collector-base resistance is

$$r_\mu = 10\beta_0 r_o = 10 \times 100 \times 20 \text{ k}\Omega = 20 \text{ M}\Omega$$

■ The equivalent circuit with these parameter values is shown in Fig. 1.21.

1.4.8 Specification of Transistor Frequency Response

The high-frequency gain of the transistor is controlled by the capacitive elements in the equivalent circuit of Fig. 1.20. The frequency capability of the transistor is most often specified in practice by determining the frequency where the magnitude of the short-circuit, common-emitter current gain falls to unity. This is called the *transition frequency*, f_T, and is a measure of the maximum useful frequency of the transistor when it is used as an amplifier. The value of f_T can be measured as well as calculated, using the ac circuit of Fig. 1.22. A small-signal current i_i is applied to the base, and the output current i_o is measured with the collector short-circuited for ac signals. A small-signal equivalent circuit can be formed for this situation by using the equivalent circuit of Fig. 1.20 as shown in Fig. 1.23, where r_{ex} and r_μ have been neglected. If r_c is assumed small, then r_o and C_{cs} have no

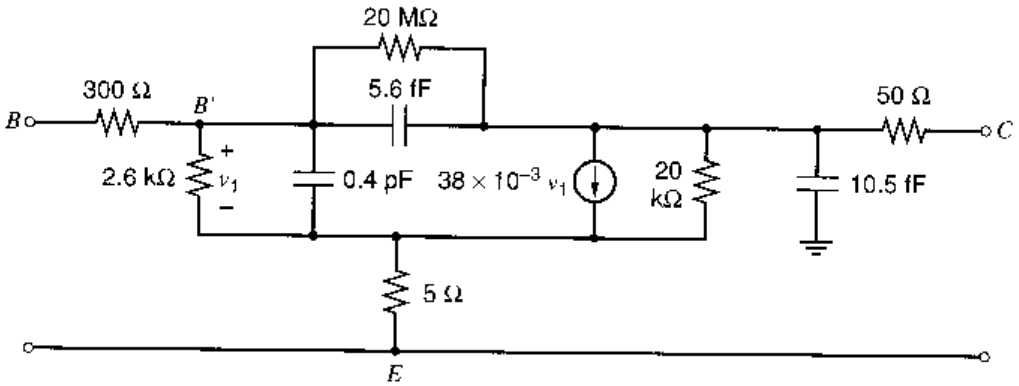


Figure 1.21 Complete small-signal equivalent circuit for a bipolar transistor at $I_C = 1$ mA, $V_{CB} = 3$ V, and $V_{CS} = 5$ V. Device parameters are $C_{je0} = 10$ fF, $n_e = 0.5$, $\psi_{0e} = 0.9$ V, $C_{\mu 0} = 10$ fF, $n_c = 0.3$, $\psi_{0c} = 0.5$ V, $C_{cs0} = 20$ fF, $n_s = 0.3$, $\psi_{0s} = 0.65$ V, $\beta_0 = 100$, $\tau_F = 10$ ps, $V_A = 20$ V, $r_b = 300\ \Omega$, $r_c = 50\ \Omega$, $r_{ex} = 5\ \Omega$, $r_\mu = 10\beta_0 r_0$.

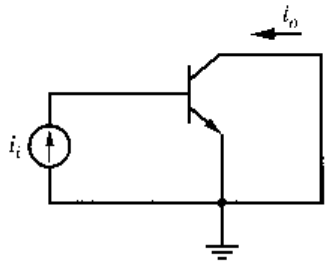


Figure 1.22 Schematic of ac circuit for measurement of f_T.

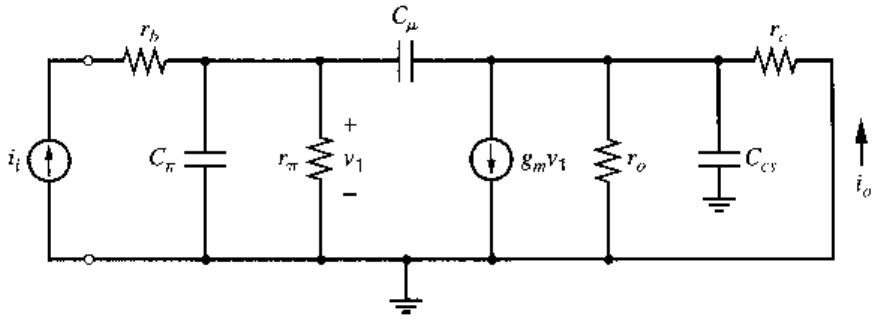


Figure 1.23 Small-signal equivalent circuit for the calculation of f_T.

influence, and we have

$$v_1 \simeq \frac{r_\pi}{1 + r_\pi(C_\pi + C_\mu)s} i_i \tag{1.119}$$

If the current fed forward through C_μ is neglected,

$$i_o \simeq g_m v_1 \tag{1.120}$$

Substitution of (1.119) in (1.120) gives

$$i_o \simeq i_i \frac{g_m r_\pi}{1 + r_\pi(C_\pi + C_\mu)s}$$

and thus

$$\frac{i_o}{i_i}(j\omega) = \frac{\beta_0}{1 + \beta_0 \dfrac{C_\pi + C_\mu}{g_m} j\omega} \tag{1.121}$$

using (1.110).

Now if $i_o/i_i(j\omega)$ is written as $\beta(j\omega)$ (the high-frequency, small-signal current gain), then

$$\beta(j\omega) = \frac{\beta_0}{1 + \beta_0 \dfrac{C_\pi + C_\mu}{g_m} j\omega} \tag{1.122}$$

At high frequencies the imaginary part of the denominator of (1.122) is dominant, and we can write

$$\beta(j\omega) \simeq \frac{g_m}{j\omega\,(C_\pi + C_\mu)} \tag{1.123}$$

From (1.123), $|\beta(j\omega)| = 1$ when

$$\omega = \omega_T = \frac{g_m}{C_\pi + C_\mu} \tag{1.124}$$

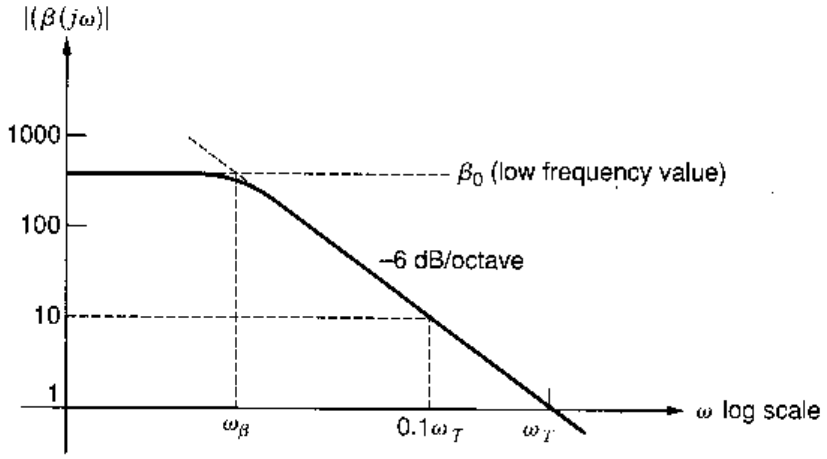


Figure 1.24 Magnitude of small-signal ac current gain $|\beta(j\omega)|$ versus frequency for a typical bipolar transistor.

and thus

$$f_T = \frac{1}{2\pi}\frac{g_m}{C_\pi + C_\mu} \tag{1.125}$$

The transistor behavior can be illustrated by plotting $|\beta(j\omega)|$ using (1.122) as shown in Fig. 1.24. The frequency ω_β is defined as the frequency where $|\beta(j\omega)|$ is equal to $\beta_0/\sqrt{2}$ (3 dB down from the low-frequency value). From (1.122) we have

$$\omega_\beta = \frac{1}{\beta_0}\frac{g_m}{C_\pi + C_\mu} = \frac{\omega_T}{\beta_0} \tag{1.126}$$

From Fig. 1.24 it can be seen that ω_T can be determined by measuring $|\beta(j\omega)|$ at some frequency ω_x where $|\beta(j\omega)|$ is falling at 6 dB/octave and using

$$\omega_T = \omega_x|\beta(j\omega_x)| \tag{1.127}$$

This is the method used in practice, since deviations from ideal behavior tend to occur as $|\beta(j\omega)|$ approaches unity. Thus $|\beta(j\omega)|$ is typically measured at some frequency where its magnitude is about 5 or 10, and (1.127) is used to determine ω_T.

It is interesting to examine the time constant, τ_T, associated with ω_T. This is defined as

$$\tau_T = \frac{1}{\omega_T} \tag{1.128}$$

and use of (1.124) in (1.128) gives

$$\tau_T = \frac{C_\pi}{g_m} + \frac{C_\mu}{g_m} \tag{1.129}$$

Substitution of (1.118) and (1.104) in (1.129) gives

$$\tau_T = \frac{C_b}{g_m} + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m} = \tau_F + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m} \tag{1.130}$$

Equation 1.130 indicates that τ_T is dependent on I_C (through g_m) and approaches a constant value of τ_F at high collector bias currents. At low values of I_C, the terms involving C_{je} and C_μ dominate, and they cause τ_T to *rise* and f_T to *fall* as I_C is decreased. This behavior is illustrated in Fig. 1.25, which is a typical plot of f_T versus I_C for an integrated-circuit *npn* transistor. The decline in f_T at high collector currents is not predicted by this simple theory and is due to an increase in τ_F caused by high-level injection and Kirk effect at high

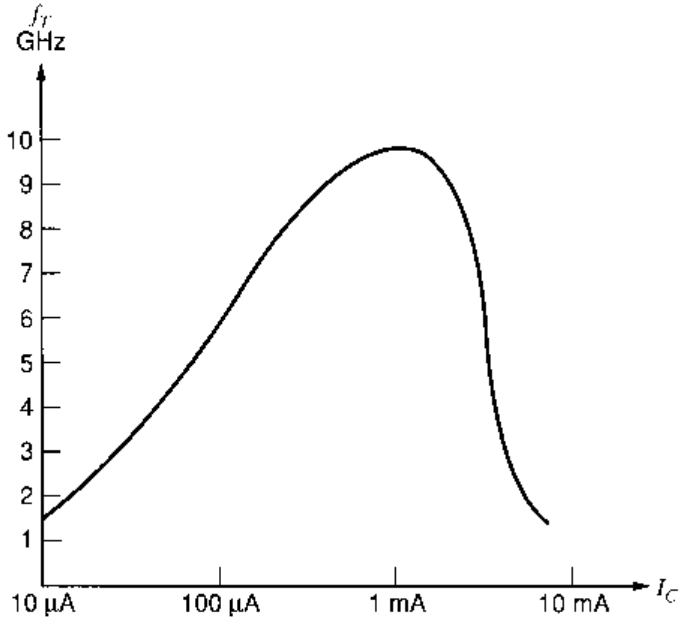


Figure 1.25 Typical curve of f_T versus I_C for an npn integrated-circuit transistor with 6 μm^2 emitter area in a high-speed process.

currents. These are the same mechanisms that cause a decrease in β_F at high currents as described in Section 1.3.5.

■ EXAMPLE

A bipolar transistor has a short-circuit, common-emitter current gain at 1 GHz of 8 with $I_C = 0.25$ mA and 9 with $I_C = 1$ mA. Assuming that high-level injection effects are negligible, calculate C_{je} and τ_F, assuming both are constant. The measured value of C_μ is 10 fF.

From the data, values of f_T are

$$f_{T1} = 8 \times 1 = 8 \text{ GHz} \quad \text{at} \quad I_C = 0.25 \text{ mA}$$
$$f_{T2} = 9 \times 1 = 9 \text{ GHz} \quad \text{at} \quad I_C = 1 \text{ mA}$$

Corresponding values of τ_T are

$$\tau_{T1} = \frac{1}{2\pi f_{T1}} = 19.9 \text{ ps}$$

$$\tau_{T2} = \frac{1}{2\pi f_{T2}} = 17.7 \text{ ps}$$

Using these data in (1.130), we have

$$19.9 \times 10^{-12} = \tau_F + 104(C_\mu + C_{je}) \tag{1.131}$$

at $I_C = 0.25$ mA. At $I_C = 1$ mA we have

$$17.7 \times 10^{-12} = \tau_F + 26(C_\mu + C_{je}) \tag{1.132}$$

Subtraction of (1.132) from (1.131) yields

$$C_\mu + C_{je} = 28.2 \text{ fF}$$

Since C_μ was measured as 10 fF, the value of C_{je} is given by

$$C_{je} \simeq 18.2 \text{ fF}$$

Substitution in (1.131) gives

$$\tau_F = 17 \text{ ps}$$

This is an example of how basic device parameters can be determined from high-frequency current-gain measurements. Note that the assumption that C_{je} is constant is a useful approximation in practice because V_{BE} changes by only 36 mV as I_C increases from 0.25 mA to 1 mA. ■

1.5 Large-Signal Behavior of Metal-Oxide-Semiconductor Field-Effect Transistors

Metal-oxide-semiconductor field-effect transistors (MOSFETs) have become dominant in the area of digital integrated circuits because they allow high density and low power dissipation. In contrast, bipolar transistors still provide many advantages in stand-alone analog integrated circuits. For example, the transconductance per unit bias current in bipolar transistors is usually much higher than in MOS transistors. So in systems where analog techniques are used on some integrated circuits and digital techniques on others, bipolar technologies are often preferred for the analog integrated circuits and MOS technologies for the digital. To reduce system cost and increase portability, both increased levels of integration and reduced power dissipation are required, forcing the associated analog circuits to use MOS-compatible technologies. One way to achieve these goals is to use a processing technology that provides both bipolar and MOS transistors, allowing great design flexibility. However, all-MOS processes are less expensive than combined bipolar and MOS processes. Therefore, economic considerations drive integrated-circuit manufacturers to use all-MOS processes in many practical cases. As a result, the study of the characteristics of MOS transistors that affect analog integrated-circuit design is important.

1.5.1 Transfer Characteristics of MOS Devices

A cross section of a typical enhancement-mode n-channel MOS (NMOS) transistor is shown in Fig. 1.26. Heavily doped n-type source and drain regions are fabricated in a p-type substrate (often called the body). A thin layer of silicon dioxide is grown over the

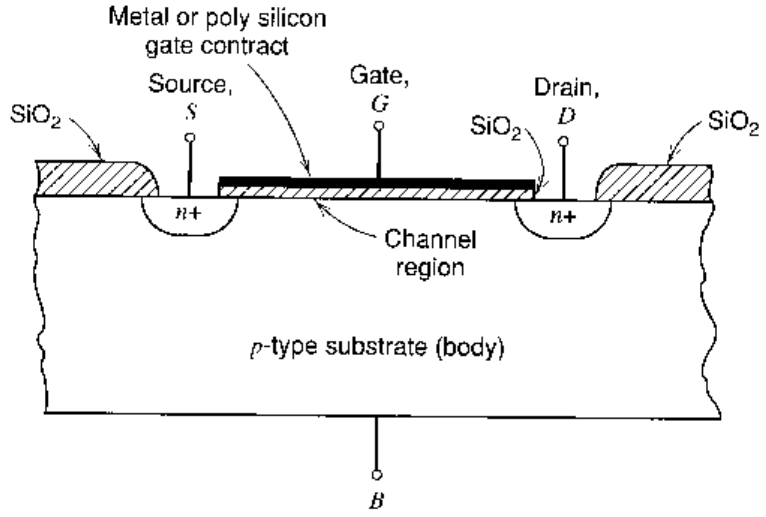


Figure 1.26 Typical enhancement-mode NMOS structure.