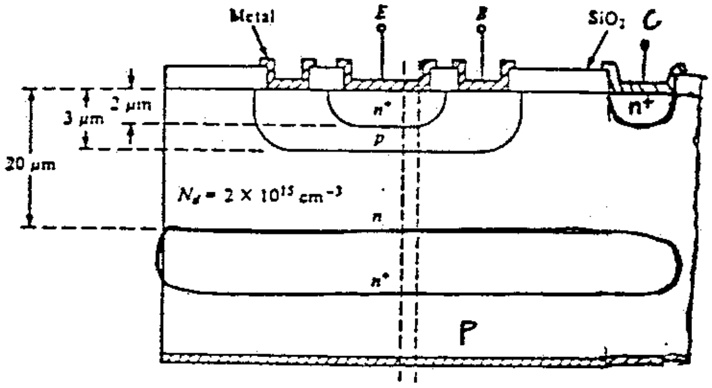
Bipolar Transistors

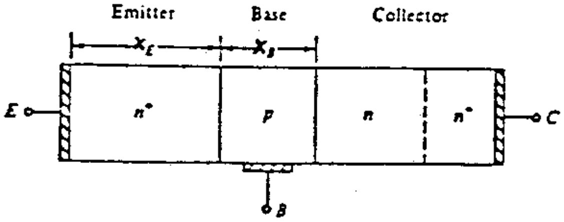
Both MOSFET and JFET devices utilize reverse-biased PN junctions in their operation. The bipolar transistor is unique in that it utilizes forward biased junctions and therefore injection of carriers across the junction. This results in:

1. Higher g_m than either MOSFETs or JFETs.
2. More complex fabrication in general.
3. Lower input impedance since BJTs are current controlled devices.
4. Higher current capability in BJTs because g_m is higher.
5. Generally faster switching times when non-saturating circuit configurations are used or when other means are provided to remove stored minority carriers.
6. Generally more complex device physics because all of the physical mechanisms we have discussed are often important.
7. Larger physical sizes than MOSFETs; comparable sizes to JFETs.

Basic Structure



(a)



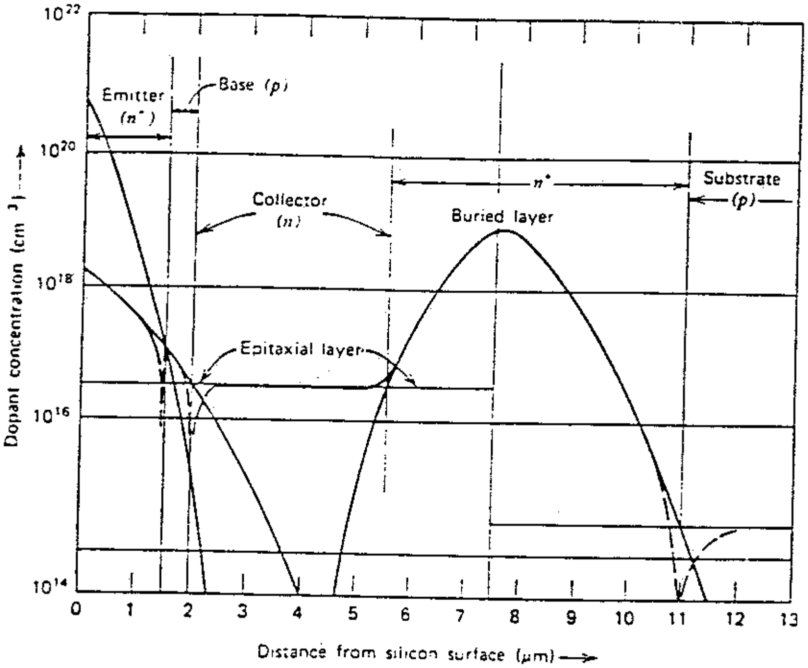
Sequential P and N diffusions form base and emitter regions.

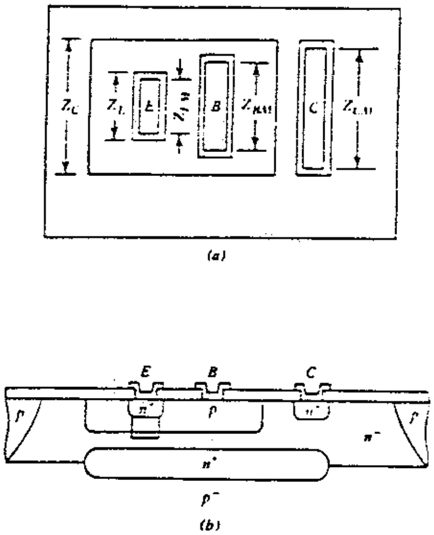
Vertical and horizontal scales are not the same.

The intrinsic portion of the transistor is represented by the section through the dashed lines.

The rest of the lateral structure is not intrinsic and may be regarded as parasitic.

Typical IC vertical and horizontal structure.





Note the following about the typical IC structure:

1. The base region is non-uniformly doped. This results in a built-in $\mathcal{E}$ field across the base which aids the transport of electrons from emitter to collector.
2. Parasitics exists in the structure. For example

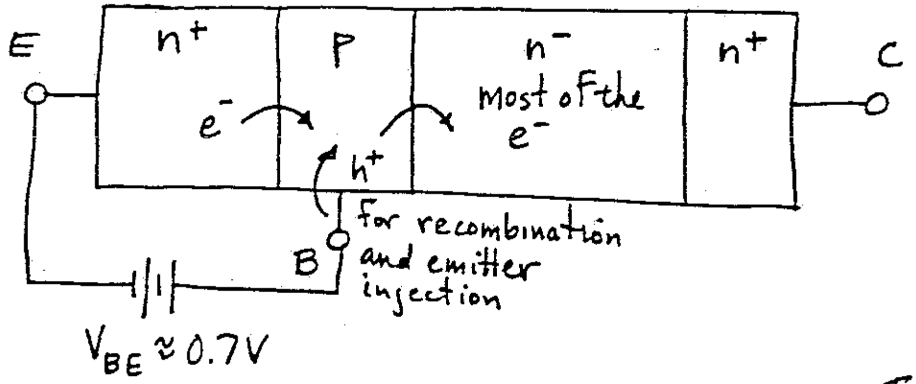
   R_B: base resistance from base contact to active base area.

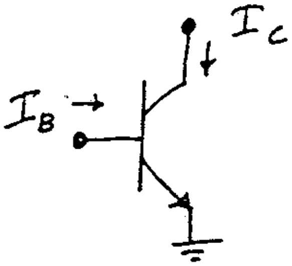
   R_C: collector resistance (predominantly through N^- layer).
3. Isolation must be provided between adjacent devices. Typically reverse biased PN junctions or insulating oxide regions are used.
4. The N^- collector region adjacent to the base reduces C_{BC}, improves BV_{CB} and decreases base width modulation by the collector voltage but adds series resistance to the device.

Basic Operation

The BJT operates basically as follows:

1. An external voltage is applied across the E-B junction to forward bias it (~0.7 volts).
2. Electrons are injected into the base by the emitter (holes are also injected into the emitter by the base but their numbers are much smaller because of relative values of N_a, N_d.
3. If $W_B \ll L_n$ in the base, most of the injected electrons get to the collector without recombining. A few do recombine; the holes necessary for this are supplied as base current.
4. The electrons reaching the collector are accelerated across the C-B depletion region by the electric field in that region.





Since most of the injected electrons reach the collector and only a few holes are injected into the emitter, the required $I_B \ll I_C$. Therefore the device has substantial current gain.

⇒ Internally, the controlling parameter is V_{BE} (determines injection level). Usually the device is considered a current controlled device however with I_B provided externally, producing I_C.

To derive the basic relationship for electron current flowing between the emitter and the collector, we begin by assuming the device current gain is high. Therefore, $I_B \cong 0$ and

$$J_p = \text{hole current in base} \cong 0$$

but

$$J_p = q\mu_p p \mathcal{E}_x - qD_p \frac{dp}{dx} \tag{1}$$

Therefore,

$$\mathcal{E}_x = \frac{D_p}{\mu_p} \frac{1}{p} \frac{dp}{dx} \tag{2}$$

Thus for uniform doping in the base, $\mathcal{E}_x = 0$ and electrons travelling through the base will **move by diffusion only**. In real IC transistors,

$$\frac{dp}{dx} \neq 0 \quad \Rightarrow \quad \mathcal{E}_x \neq 0$$

The direction of this field in an npn transistor aids electrons flow $E \rightarrow C$ and retards electron flow $C \rightarrow E$, thereby increasing the device performance.

The electron flow between E and C is given by

$$J_n = q\mu_n n \mathcal{E}_x + qD_n \frac{dn}{dx} \tag{3}$$

which, using (2) becomes

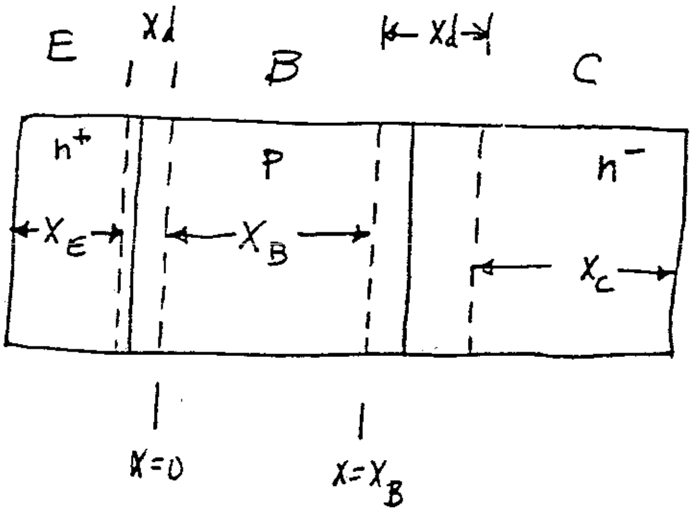
$$J_n = \underbrace{kT\mu_n \frac{n}{p} \frac{dp}{dx}}_{\text{drift}} + \underbrace{qD_n \frac{dn}{dx}}_{\text{diffusion}} \tag{4}$$

Therefore

$$J_n = \frac{qD_n}{p} \left[n\frac{dp}{dx} + p\frac{dn}{dx} \right]$$

or

$$J_n = \frac{qD_n}{p} \frac{d(pn)}{dx} \tag{5}$$



(Neglecting depletion regions, the effective base width x_B is just the metallurgical base width.) Therefore, rearranging (5) and integrating across the base region,

$$J_n \int_0^{x_B} \frac{p}{qD_n} dx = \int_0^{x_B} \frac{d(pn)}{dx} dx \tag{6}$$

J_n is pulled outside the integral by assuming no recombination of electrons in the base (i.e. J_n = constant).

$$J_n \int_0^{x_B} \frac{p}{qD_n} dx = pn(x_B) - pn(0) \tag{7}$$

From our diode analysis, we know that the pn products at the edge of the depletion regions are given by

$$pn(0) = n_i{}^2 \exp \frac{qV_{BE}}{kT} \tag{8}$$

$$pn(x_B) = n_i{}^2 \exp \frac{qV_{BC}}{kT} \tag{9}$$

Therefore

$$J_n = \frac{qn_i{}^2 \left[\exp \frac{qV_{BC}}{kT} - \exp \frac{qV_{BE}}{kT}\right]}{\int_0^{x_B} \frac{p}{D_n} dx} \tag{10}$$

But

$$\frac{Q'_B}{q} = \int_0^{x_B} p\, dx = \text{integrated dose per area in undepleted base region.}$$

Therefore,

$$I_n = I_s \left[\exp \frac{qV_{BC}}{kT} - \exp \frac{qV_{BE}}{kT} \right] \qquad (11)$$

where

$$I_s = \frac{q^2 A n_i{}^2 \tilde{D}_n}{Q'_B} \qquad (12)$$

This is an extremely important result. Note that:

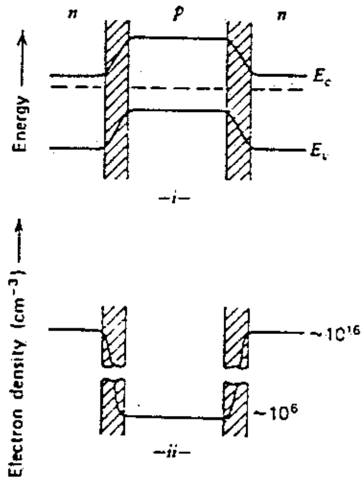
1. Usually only one of the two exponential terms is important because of the fact that one junction is typically reverse biased. (When the device is in saturation, both junctions are forward biased and both terms must be included.)

2. The quantity

$$\frac{Q'_B}{q} = \int_0^{x_B} N_a(x)\, dx$$

is called the base Gummel number. It is the total integrated base charge (atoms/cm^2). Since $I \propto 1/Q'_B$, it is important to minimize Q'_B. Therefore, IC transistors use low doping levels in the base.

As an example, consider the situations shown below



Equilibrium

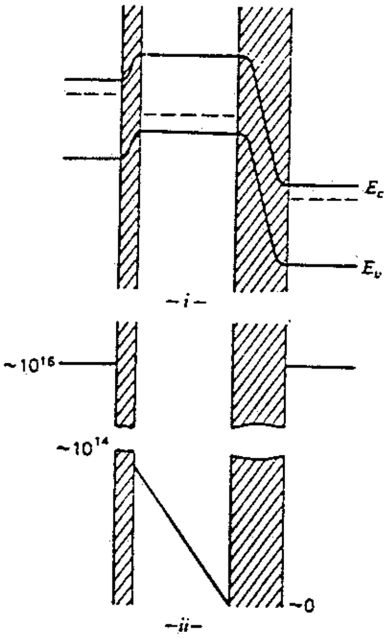
E_F = constant

no applied bias

V_{BE} +

Injection

E → B



V_{BC} −

Collection

B → C

Active mode

If the base is uniformly doped, $\mathcal{E}_x = 0$ and our analysis of the short base diode told us that the electron concentration in the base is linear and that at the edges of the base-emitter depletion region,

$$n_p(x_p) = n_{p0} \exp \frac{qV_{BE}}{kT}, \quad n_{p0} = \frac{n_i{}^2}{N_a}$$

$$p_n(-x_n) = p_{n0} \exp \frac{qV_{BE}}{kt}, \quad p_{n0} = \frac{n_i{}^2}{N_d}$$

Note that $p_n \ll n_p$ because of the relative doping levels.

If the base doping level is constant (N_a), then $Q'_B = qN_a x_B$ and we have directly from (11) and (12) that in the forward active mode,

$$I_n = -\frac{qD_n A n_i{}^2}{N_a x_B} \exp \frac{qV_{BE}}{kT} \tag{13}$$

(The $\exp(qV_{BC}/kT)$ term is negligible because the BC junction is reverse biased.)

Alternatively, this may be written as

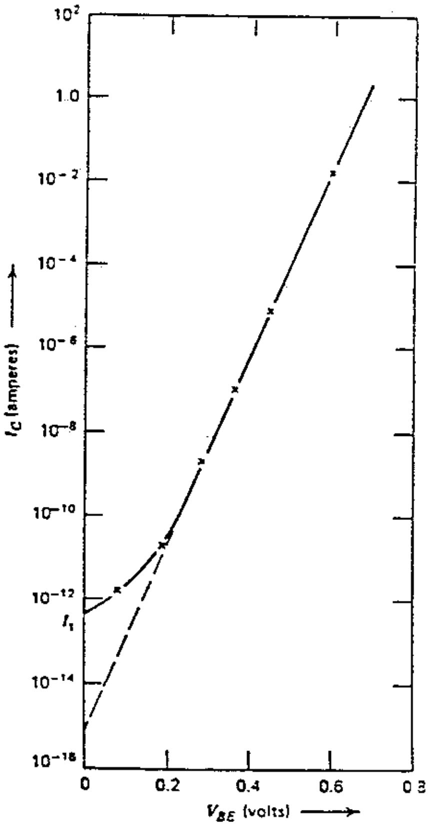
$$I_n = -\frac{q^2 A \tilde{D}_n n_i{}^2}{Q'_B} \exp \frac{qV_{BE}}{kT} \cong I_E \cong -I_C \tag{14}$$

This equation predicts an exponential relationship between I_C and V_{BE}.

$$\text{slope} = \frac{60\,\text{mV}}{\text{decade I}}$$

This relationship holds extremely well for IC transistors over many decades of current.

In general, Q'_B is obtained by integration over the base region. Q'_B is typically well controlled to $\approx 10^{12}\ \text{cm}^{-2}$ to give high I_C for a given V_{BE}.



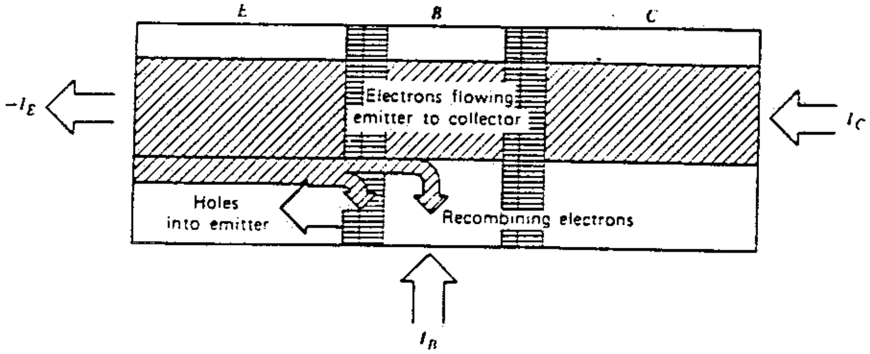
Current Gain

A number of factors can contribute to base current in a BJT.

1. Recombination in neutral base region.
2. Minority carrier injection into the emitter.
3. Recombination in the base-emitter depletion region.

Recombination In the Neutral Base Region

In the general case, some of the electrons traversing the base will recombine with majority carrier holes. (This is usually unimportant in modern IC BJTs.)



If we assume that the base is uniformly doped so that $\mathcal{E}_x = 0$, then the electron current (transport) and continuity equations are

$$I_n = qAD_n \frac{dn_p}{dx} \tag{15}$$

$$D_n \frac{d^2 n_p}{dx^2} - \frac{n_p - n_{p0}}{\tau_n} = 0 \tag{16}$$

As we saw in the case of the PN junction, the general solution of these equations is

$$n_p - n_{p0} = K_1 \exp \frac{-x}{L_n} + K_2 \exp \frac{x}{L_n} \tag{17}$$

The appropriate boundary conditions are

$$n_p(x = 0) = n_{p0} \exp \frac{qV_{BE}}{kT}$$

$$n_p(x = x_B) = n_{p0} \exp \frac{qV_{BC}}{kT} = n_{p0}, \quad \text{if } V_{BC} = 0$$

With these boundary conditions, (17) reduces to

$$\Delta n_p = n_{p0} \left(\exp \frac{qV_{BE}}{kT} - 1 \right) \frac{\sinh \left[\dfrac{x_B - x}{L_n} \right]}{\sinh \left[\dfrac{x_B}{L_n} \right]} \tag{18}$$

Substitution of (18) into (15) yields for the emitter and collector electron currents,

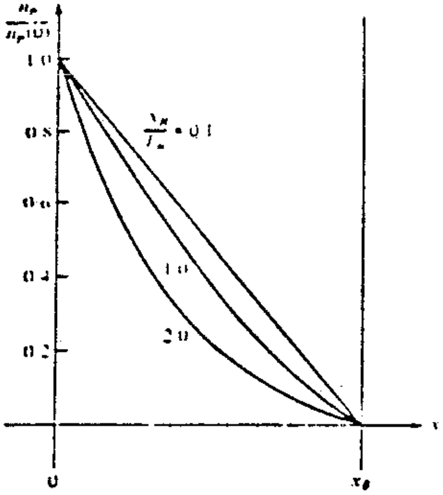
$$I_{nE} = \frac{qAD_n n_{p0}}{L_n} \left(\exp \frac{qV_{BE}}{kT} - 1 \right) \coth \frac{x_B}{L_n} \tag{19}$$

$$I_{nC} = \frac{qAD_n n_{p0}}{L_n} \left(\exp \frac{qV_{BE}}{kT} - 1 \right) \operatorname{csch} \frac{x_B}{L_n} \tag{20}$$

The ratio of these two currents is defined as the base transport factor

$$\alpha_T \equiv \frac{I_{nC}}{I_{nE}} = \operatorname{sech} \frac{x_B}{L_n} \tag{21}$$

In modern IC BJTs, $x_B \ll L_{nB}$ and there is little recombination in the base region.



Normalized plots of Equation (18) for different values of x_B/L_{nB}. Note that if $x_B \ll L_{nB}$ then the distribution is approximately linear.

If $x_B \ll L_{nB}$ then (20) reduces to our earlier expression for electron current in a short base diode and the base transport factor becomes

$$\alpha_T \cong 1 - \frac{{x_B}^2}{2{L_{nB}}^2}; \quad x_B \ll L_{nB} \tag{22}$$

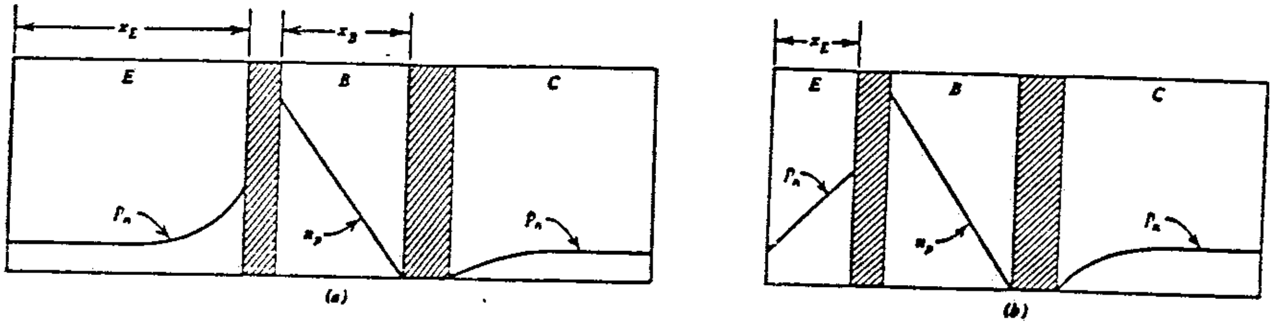
In a typical device, $x_B \cong 0.3\mu$m and $L_{nB} \cong 30\mu$m so that $\alpha_T \cong 0.99995$. If this was the limiting factor, then the current gain would be $I_C/I_B = \beta \cong 20,000$ which is higher than normally observed. in devices with 0.3μmbase widths. Therefore, α_T is usually **not** a limiting factor in the current gain.

The base current due to recombination in in the neutral base is

$$I_{B\,rec} = \frac{Q_{nB}}{\tau_{nB}} = \frac{qA_E \int_0^{x_B} n_p(x)\,dx}{\tau_{nB}} = \frac{qA_E {n_i}^2 x_B}{2N_{aB}\tau_{nB}} \left[\exp\frac{qV_{BE}}{kT} - 1\right] \tag{23}$$

Hole Injection into the Emitter

The dominant mechanism in limiting β in modern IC BJTs is hole injection into the emitter from the base. By forward biasing the base-emitter junction, the barrier to holes from B to E is reduced at the same time the barrier for electrons from E to B is reduced.



$x_E \gg L_{pE}$ $x_E \ll L_{pE}$

The injected hole currents in each case can be calculated directly from our analysis of the long and short base diode, respectively.

$$x_E \gg L_{pE} \quad \Rightarrow \quad I_{pE} = \frac{qAn_i^2 D_{pE}}{N_{dE} L_{pE}} \left(\exp \frac{qV_{BE}}{kT} - 1 \right) \tag{24}$$

$$x_E \ll L_{pE} \quad \Rightarrow \quad I_{pE} = \frac{qAn_i^2 D_{pE}}{N_{dE} x_E} \left(\exp \frac{qV_{BE}}{kT} - 1 \right) \tag{25}$$

The emitter injection efficiency is defined as

$$\gamma \equiv \frac{I_{nE}}{I_E} = \frac{I_{nE}}{I_{nE} + I_{pE}} \tag{26}$$

From (25) and (26), we can calculate directly that

$$\gamma = \frac{1}{1 + \dfrac{x_B N_{aB} D_{pE}}{x_E N_{dE} D_{nB}}} \tag{27}$$

(If $x_B \gg L_{nB}$ or $x_E \gg L_{pE}$, then the long diode approximations replace x_B or x_E with L_{nB} or L_{pE}.)

Note that this is only an approximate expression in real structures because the doping in not constant.

γ can be made close to 1 by:

1. Making $N_{dE} \gg N_{aB}$.

2. Making x_E large or preventing hole recombination at the emitter contact.

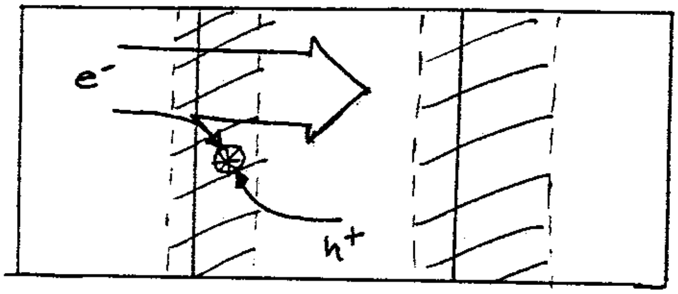
3. Making x_B small. This is also desirable as we saw for increasing α_T.

Typically, $\gamma \sim 0.99$ to 0.999, primarily due to the doping ratio. This would result in a current gain of $\beta = 100$ to 1000 as observed in IC BJTs.

Recombination in the Base-Emitter Depletion Region

Note that both α_T and γ are independent of V_{BE}, which would imply that the ratio of collector to base current (β) is constant independent of the biasing conditions.

In practice, β is not independent of biasing. In particular, for small values of V_{BE}, recombination in the E-B depletion region can become important.



In our discussion of the PN junction, we saw that some recombination occurs in the depletion region, given by

$$I_{\text{rec}} = \frac{qAn_i x_d'}{2\tau_0} \exp\frac{qV_{BE}}{2kT} \tag{28}$$

where τ_0 is the lifetime in the depletion region.

- This current flows between the base and emitter. The electron injection and thus collector current is unchanged. Therefore, as I_{rec} becomes important, $\beta = I_C/I_B$ will drop.
- The recombination current varies as $\exp(qV_{BE}/2kT)$ and will dominate at low current levels.

We can define the depletion region efficiency to account for depletion region recombination.

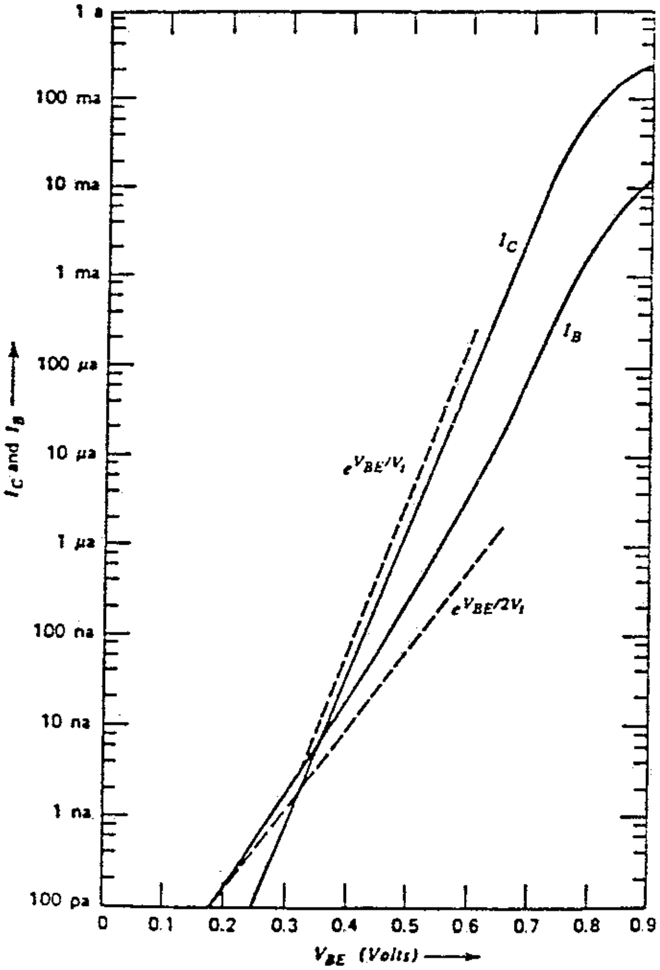
$$\delta \equiv \frac{I_{nE} + I_{pE}}{I_{nE} + I_{pE} + I_{rec}} \cong \frac{1}{1 + \frac{N_{aB} x_B x_d}{2n_i D_{nB} \tau_0} \exp\left(-\frac{qV_{BE}}{2kT}\right)} \tag{29}$$

To summarize, considering all three components of base current:

$$\alpha_F = \frac{I_{nC}}{I_{nE}} \frac{I_{nE}}{I_{nE}+I_{pE}} \frac{I_{nE}+I_{pE}}{I_E} = \alpha_T \gamma \delta \qquad \text{and} \qquad \beta_F = \frac{\alpha_F}{1-\alpha_F}$$

For $\beta_F \gg 1$,

$$\frac{1}{\beta_F} = \frac{I_B}{I_C} \cong \underbrace{\frac{{x_B}^2}{2{L_{nB}}^2}}_{1-\alpha_T} + \underbrace{\frac{x_B N_{aB} D_{pE}}{x_E N_{dE} D_{nB}}}_{1-\gamma} + \underbrace{\frac{N_{aB} x_B x_d}{2 n_i D_{nB} \tau_0} \exp - \frac{q V_{BE}}{2kT}}_{1-\delta} \tag{30}$$

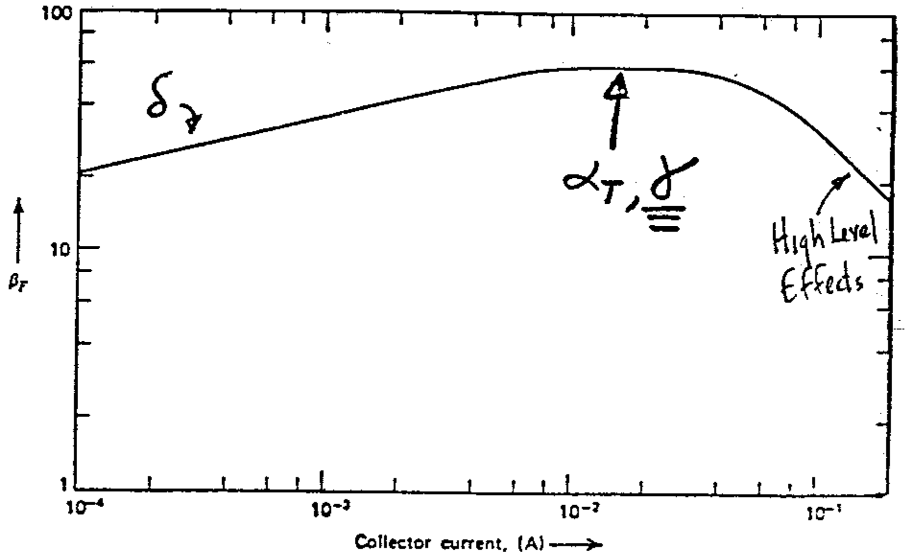


$I_C \propto \exp \dfrac{V_{BE}}{kT}$ over a very wide range of I_C

$I_B \propto \exp \dfrac{V_{BE}}{kT}$ at moderate currents

$I_B \propto \exp \dfrac{V_{BE}}{2kT}$ at low levels $\Rightarrow \beta \downarrow$

High level effects will be discussed later.



Ebers-Moll Model

Recall from page 6 that in an npn transistor with constant base doping,

$$I_{nE} = \frac{qAD_{nB}{n_i}^2}{x_B N_{aB}} \left[\left(\exp \frac{qV_{BE}}{kT} - 1 \right) - \left(\exp \frac{qV_{BC}}{kT} - 1 \right) \right] \tag{31}$$

The hole current injected into the emitter is (from (26))

$$I_{pE} = \frac{qAD_{pE}{n_i}^2}{x_E N_{dE}} \left[\left(\exp \frac{qV_{BE}}{kT} - 1 \right) \right] \tag{32}$$

The total emitter current is the sum of these two components:

$$I_E = -(I_{pE} + I_{nE}) = a_{11} \left(\exp \frac{qV_{BE}}{kT} - 1 \right) + a_{12} \left(\exp \frac{qV_{BC}}{kT} - 1 \right) \tag{33}$$

where the $-$ sign is added since we define the emitter current to be into the emitter and

$$a_{11} = -qA{n_i}^2 \left(\frac{D_{nB}}{x_B N_{aB}} + \frac{D_{pE}}{x_E N_{dE}} \right)$$

$$a_{12} = \frac{qAD_{nB}{n_i}^2}{x_B N_{aB}}$$

Similarly, the total collector current is given by

$$I_C = a_{21} \left(\exp \frac{qV_{BE}}{kT} - 1 \right) + a_{22} \left(\exp \frac{qV_{BC}}{kT} - 1 \right) \tag{34}$$

where

$$a_{21} = \frac{qAD_{nB}{n_i}^2}{x_B N_{aB}}$$

$$a_{22} = -qA{n_i}^2 \left(\frac{D_{nB}}{x_B N_{aB}} + \frac{D_{pC}}{L_{pC} N_{dC}} \right)$$

where D_{pC}, L_{pC} and N_{dC} are the minority carrier diffusion coefficient, minority carrier diffusion length and doping, respectively, in the collector.

Note that $a_{12} = a_{21}$. This relation, known as **reciprocity**, is true for an arbitrary transistor structure.

We can rewrite equations (35) and (36) in a more intuitively useful form:

$$I_E = -I_F + \alpha_R I_R \tag{35}$$

$$I_C = \alpha_F I_F - I_R \tag{36}$$

$$I_B = -(I_E + I_C) = (1 - \alpha_F) I_F + (1 - \alpha_R) I_R \tag{37}$$

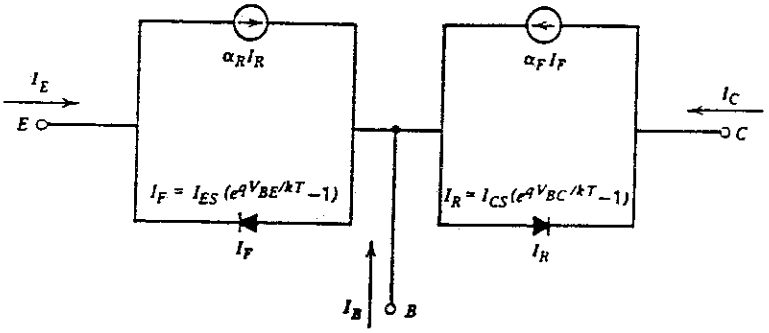
where

$$I_F = I_{ES}\left[\exp\frac{qV_{BE}}{kT} - 1\right]$$

$$I_R = I_{CS}\left[\exp\frac{qV_{BC}}{kT} - 1\right]$$

Note that the reciprocity relation becomes

$$a_{12} = \alpha_F I_{ES} = \alpha_R I_{CS} = a_{21}. \tag{38}$$



This is the Ebers-Moll transistor model. The terms are defined as

$$\alpha_F \equiv \text{forward alpha (gain)} \cong \frac{I_C}{I_E}, \quad \text{if } V_{BE} > 0, V_{BC} = 0$$

$$\alpha_R \equiv \text{reverse alpha (gain)} \cong \frac{I_E}{I_C}, \quad \text{if } V_{BE} = 0, V_{BC} > 0$$

$$I_{ES} \equiv \text{emitter junction reverse saturation current}$$

$$I_{CS} \equiv \text{collector junction reverse saturation current}$$

These are the equations for npn transistors, for pnp transistors, the coupled diodes are forward biased with opposite voltages applied.

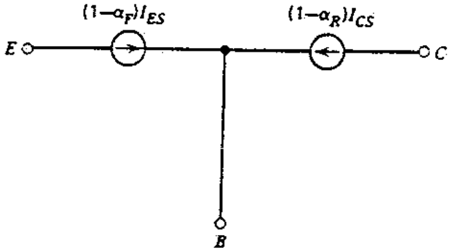
The Ebers-Moll model is extremely useful for modeling the static operation of the bipolar transistor. The charge-control model which is an extension of the E-B model is used for transistor switching.

There are four modes of operation for a BJT:

	V_{BE}	V_{BC}
forward active	$+$	$-$
reverse active	$-$	$+$
cutoff	$-$	$-$
saturation	$+$	$+$

In each of these modes (assuming $|V_{BE}|, |V_{BC}| > 4kT/q$) the Ebers-Moll model can be simplified.

For example, in the cutoff mode only reverse saturation currents flow and the model reduces to:

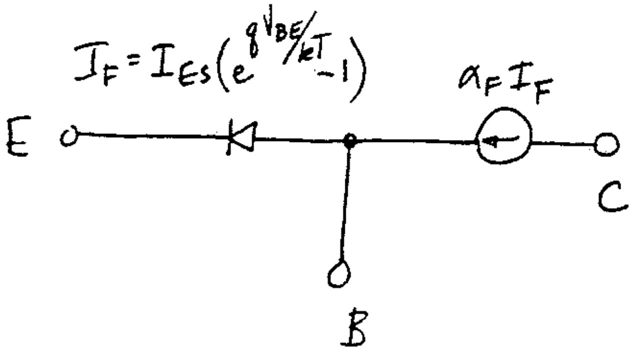


$$I_E \cong (1-\alpha_F)I_{ES}$$

$$I_C \cong (1-\alpha_R)I_{CS}$$

$$I_B = -(1-\alpha_F)I_{ES} - (1-\alpha_R)I_{CS}$$

In the forward active mode (the normal amplifying mode), $I_R \cong -I_{CS}$ and the model reduces to



$$I_E \cong -I_F - \alpha_R I_{CS} \cong -I_{ES}\exp\frac{qV_{BE}}{kT} - \alpha_R I_{CS}$$

$$I_C \cong \alpha_F I_{ES}\exp\frac{qV_{BE}}{kT} + I_{CS}$$

Ignoring the reverse saturation current

$$I_E \cong -I_{ES}\exp\frac{qV_{BE}}{kT}$$

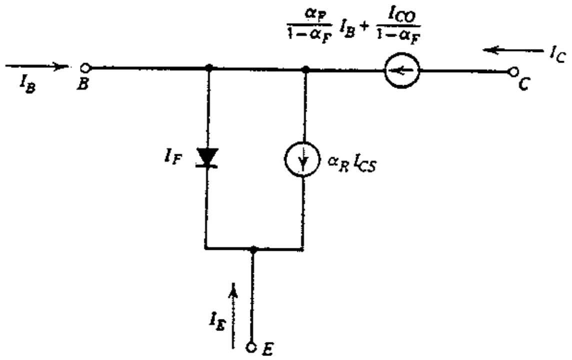
$$I_C \cong \alpha_F I_{ES}\exp\frac{qV_{BE}}{kT}$$

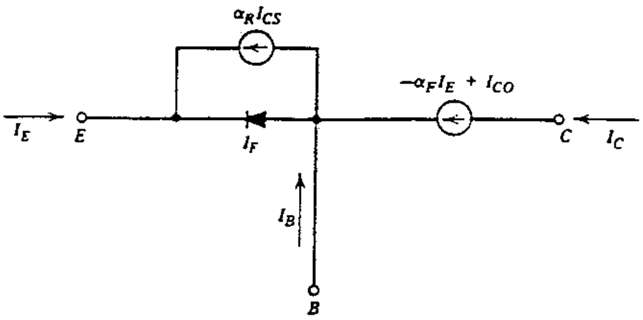
$$I_B \cong (1-\alpha_F)I_{ES}\exp\frac{qV_{BE}}{kT}$$

Equations (35-37) can also be rewritten to highlight the relations between the currents.

$$I_C = \frac{\alpha_F}{1-\alpha_F} I_B - \frac{1-\alpha_F\alpha_R}{1-\alpha_F} I_{CS} = \beta_F I_B + (\beta_F + 1) I_{C0} \tag{39}$$

where $I_{C0} = I_{CS}(1-\alpha_F\alpha_R)$ is the collector current with no emitter current.





Alternatively, in terms of emitter current,

$$I_C = -\alpha_F I_E - I_R(1-\alpha_F\alpha_R) \tag{40}$$

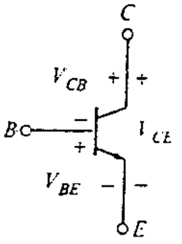
which in forward active mode becomes

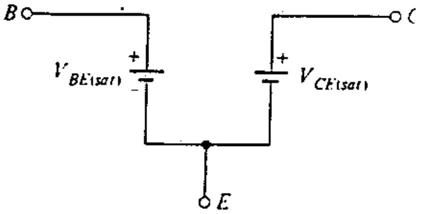
$$I_C = -\alpha_F I_E + I_{C0}$$

Similar equations can be derived for use in the reverse active mode.

In saturation, fewer simplifications are possible. Of particular interest is $V_{CE\,sat}$ which can be calculated by solving for V_{BC} and V_{BE} in terms of the currents.

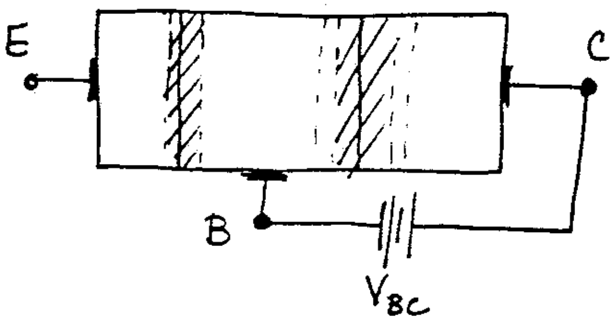
$$V_{CE\,sat} = \frac{kT}{q} \ln \left\{ \frac{\left[1 + \frac{I_C}{I_B}(1-\alpha_R)\right]}{\alpha_R \left[1 - \frac{I_C}{I_B}\left(\frac{1-\alpha_F}{\alpha_F}\right)\right]} \right\} \tag{41}$$





Early Effect (Base Width Modulation)

So far we have assumed that the behavior of a BJT in the forward active mode is independent of the reverse bias on the base-collector junction. In practice, that voltage will change the width of the base-collector depletion region, modulating the base width.



$$x_B = f(V_{BC})$$

We saw earlier that

$$I_C = \frac{q\tilde{D}_{nB}{n_i}^2 A_E \exp\dfrac{qV_{BE}}{kT}}{\displaystyle\int_0^{x_B} p\,dx} \tag{42}$$

where

$$\int_0^{x_B} p\,dx \equiv \text{base Gummel number} = N_a x_B \text{ for constant doping}$$

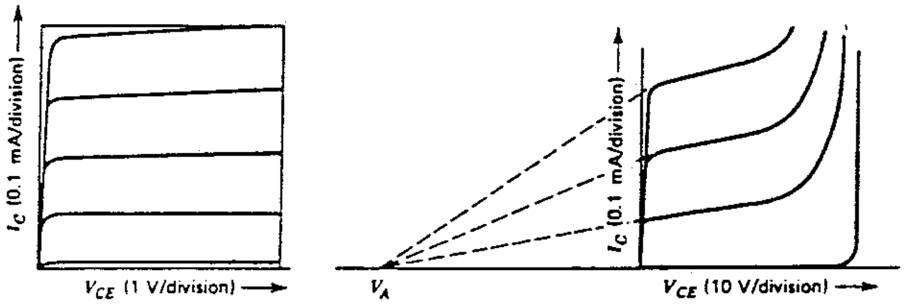
Any influence of the output voltage (V_{CE} with V_{BE} constant) on the output current results in a finite output resistance.

The inverse of the output resistance is given by

$$\begin{aligned}\frac{1}{r_o} = \left.\frac{\partial I_C}{\partial V_{CE}}\right|_{V_{BE}} &= \left.\frac{\partial I_C}{\partial V_{CB}}\right|_{V_{BE}} \\ &= -I_C \frac{p(x_B)}{\displaystyle\int_0^{x_B} p\,dx}\,\frac{dx_B}{dV_{CB}} \\ &= \left|\frac{I_C}{V_A}\right|\end{aligned} \tag{43}$$

The Early voltage is defined to be

$$V_A \equiv -\frac{\displaystyle\int_0^{x_B} p\,dx}{p(x_B)\dfrac{dx_B}{dV_{CB}}} \tag{44}$$



Note that the numerator multiplied by the electron charge is the total base charge per unit area and the denominator times the electron charge is just the collector junction capacitance per area C_{jc}/A_C. Therefore,

$$V_A = \frac{Q_B/A_E}{C_{jc}/A_C} = \frac{Q_B A_C}{C_{jc} A_E} \tag{45}$$

For maximum output resistance, V_A, the Early voltage, should be as large as possible.

- Want Q_B large $\Rightarrow$ heavy base doping, wide base region.
- Want C_{jc} small $\Rightarrow$ one side of junction lightly doped.

For large gain, we saw that the base doping and base width should be small, so in order to increase the output resistance, the collector should be made very lightly doped. This makes the collector resistance large, but a buried layer can be used to reduce the collector resistance.

High Level Effects

Several factors can contribute to a reduction of I_C and β at high current levels.

High Level Injection at the Emitter

As the forward bias on the base-emitter junction is increased our assumption of low level injection where the majority carrier concentrations are unaffected becomes no longer valid. High level injection will occur first in the more lightly doped base region and we must consider the effect of minority carrier injection on the majority carrier concentration.

The quasi-neutrality condition in the base requires that

$$p(x) \cong N_a(x) + n(x)$$

In our analysis of the pn junction we saw that the pn product at the edges of the depletion region depend on the junction bias.

$$p(0)n(0) = n_i{}^2 \exp \frac{qV_{BE}}{kT}$$

Now when $n(0) \ll N_a(0)$, the majority concentration was essentially unchanged and

$$n(0) = \frac{n_i{}^2}{N_a} \exp \frac{qV_{BE}}{kT}$$

For high level injection, $p(0)$ increases along with $n(0)$ and the collector current no longer increases as fast as $\exp(qV_{BE}/kT)$.

In the limit of very high level injection where $n(0) \gg N_a$ and therefore $p(0) \cong n(0)$,

$$n(0) = n_i \exp \frac{qV_{BE}}{2kT} \quad \Rightarrow \quad I_C \propto \exp \frac{qV_{BE}}{2kT} \tag{46}$$

The base current is nearly unchanged by this effect since the injection into the much more heavily doped emitter is still low level.

$$I_C \downarrow \;\Rightarrow\; \frac{I_C}{I_B} \downarrow \;\Rightarrow\; \beta \downarrow$$

High Level Injection Effects at the Collector (Kirk Effect)

At high current levels we must reevaluate our assumptions underlying the depletion approximation at the base-collector junction.

If the current is high enough, the numbers of electrons collected by the B-C junction can be large enough to affect the charge density in the depletion region.

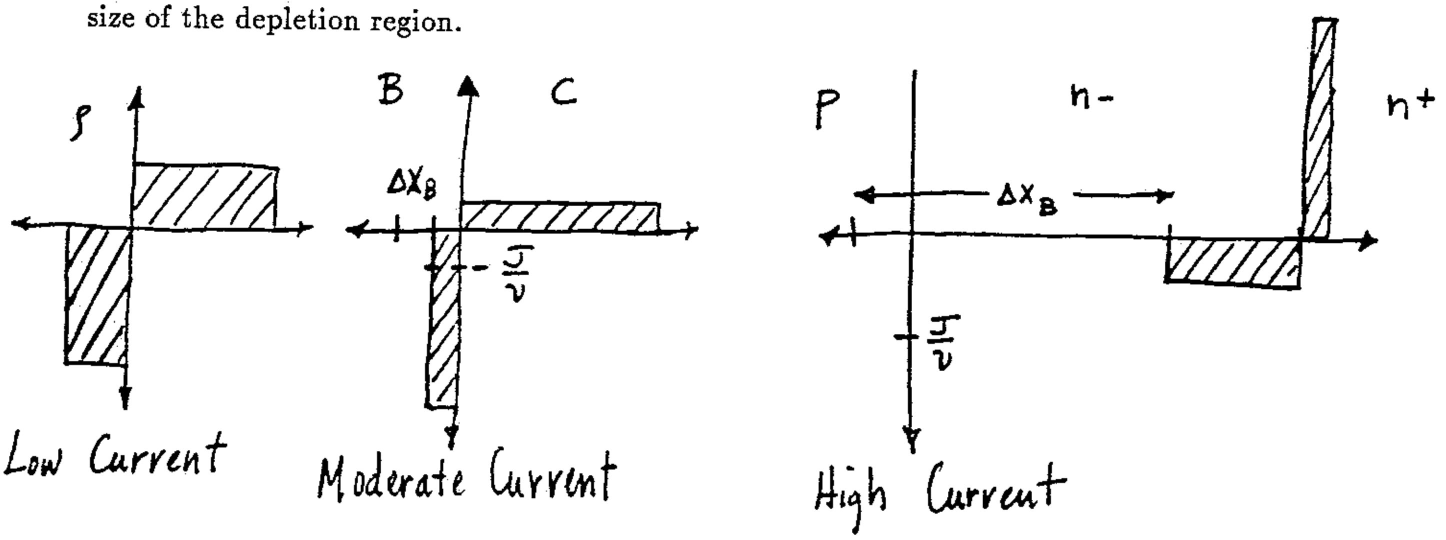
If the electrons are travelling at an average velocity v, then at any time, the electron density in the depletion region is given by J/qv and

$$\rho = q(N_d(x) - N_a(x)) - \frac{|J|}{v} \tag{47}$$

This phenomenon is most important at the collector side because it is so lightly doped, but near the collector side the base is also lightly doped and will be affected.

If J/qv is comparable to the net doping, then some of the $\mathcal{E}$ field lines will terminate on the electrons rather than on the fixed impurity ions.

This change in the charge density in the depletion region will change the position and size of the depletion region.



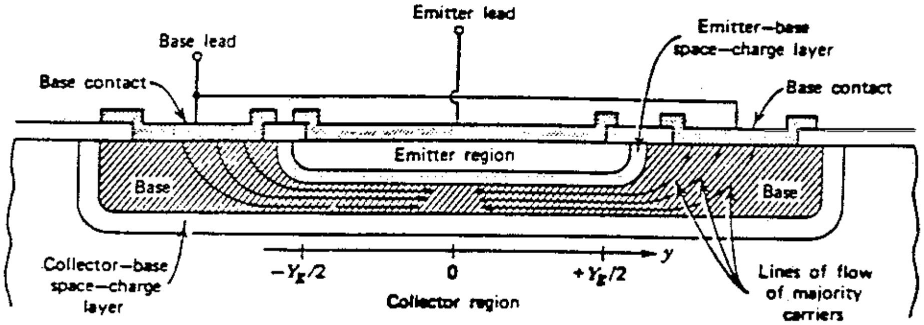
Low Current Moderate Current High Current

The net result will be to widen the neutral base region since the depletion region moves into the lightly doped collector In extreme cases the depletion region moves all the way to the buried contact.

Therefore at high values of I_C,

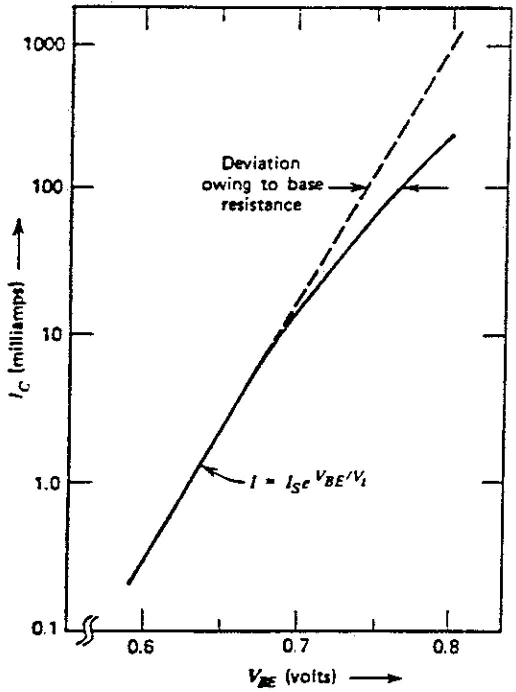
$$x_B \uparrow \;\Rightarrow\; \beta \downarrow$$

Base Resistance



In the integrated circuit structure there is a parasitic resistance R_B between the base contact and the active base region. Although, due to large values of β, base currents are generally small, any voltage drops across R_B reduces the collector current exponentially. The effective V_{BE} in the active portion of the device is not as great as the externally applied voltage.

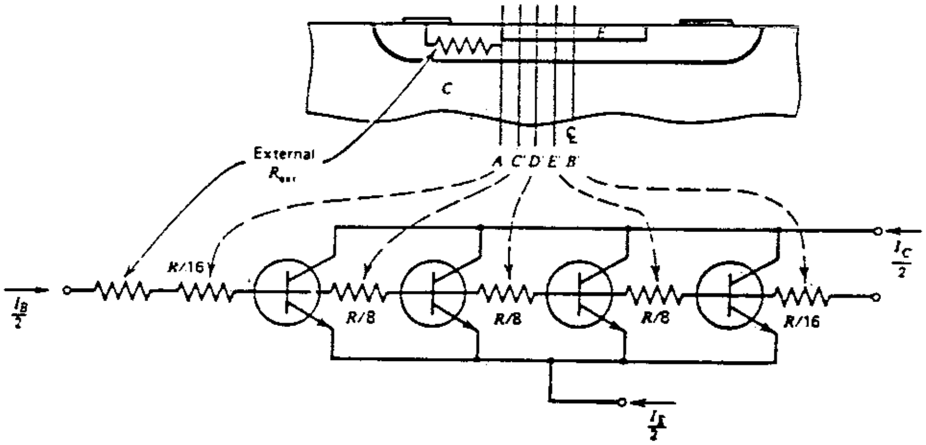
$$I_C = I_s \exp \frac{q(V_{BE} - I_B R_B)}{kT} \tag{48}$$



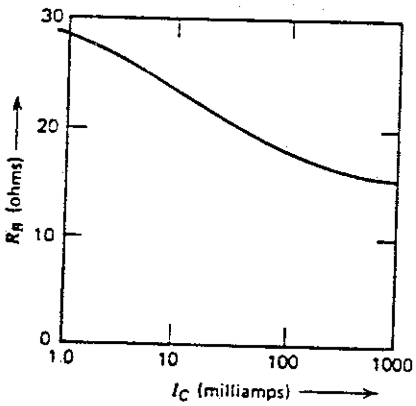
Therefore, the collector current decreases below the ideal $\exp(qV_{BE}/kT)$ behavior.

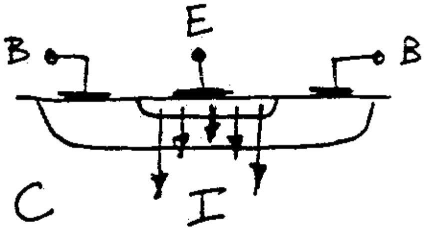
The base resistance also reduces the base current proportionately and therefore doesn't have a strong effect on the β.

This is essentially a distributed effect, since the resistance to the active base region increases towards the center. The transistor could be thought of as many transistors in parallel, connected by base resistances.



As the current increases the difference in V_{BE} between the center and the edges of the active region increases and the current becomes dominated by the edges of the active region.





In most IC structures, the Kirk effect is the dominant factor in β rolloff at high currents, while both R_B and the Kirk effect are often significant in the reduction of I_C.

Frequency Limitations

Base Transit Time

In the absence of $\mathcal{E}$ fields in the base (N_A constant, low level injection), the electron concentration varies linearly across the base.

The total electron charge in the base is simply

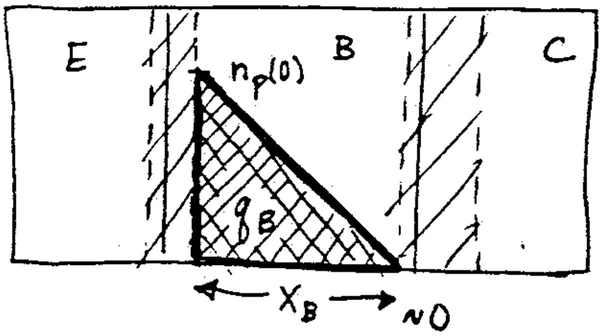
$$q_B = \frac{1}{2} q n_p(0) x_B A_E \tag{49}$$

The transit time across the base is just

$$\tau_B = \frac{q_B}{I_C} \tag{50}$$

Using our equations for I_C and q_B with constant base doping,

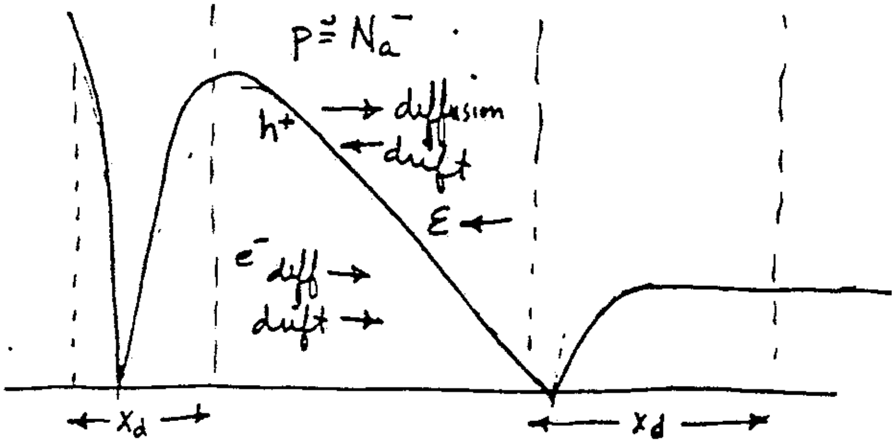
$$\tau_B \cong \frac{{x_B}^2}{2D_{nB}} \tag{51}$$



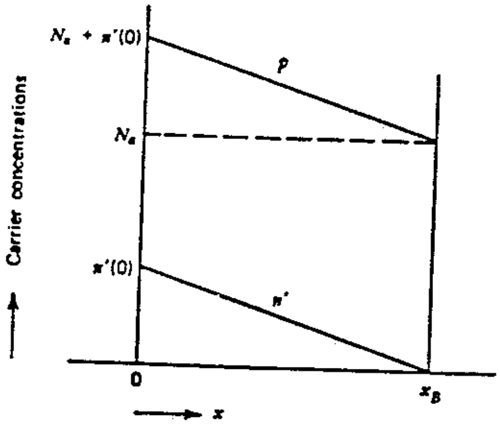
Example: If $x_B = 1\mu\text{m}$ and the base is lightly doped so that $D_{nB} = 38\,\text{cm}^2/\text{s}$, then

$$\tau_B = 132\,\text{ps}$$

If instead the base doping is graded (as is usually the case in IC transistors), the resulting electric field can cause the transit time to be reduced by up to an order of magnitude or so, limited by the need for $N_a(0)$ to be less than the emitter doping for high gain. The electric field is increased with smaller base widths (larger doping gradients) thus making the transit time depend more strongly than the square of x_B.



Also, under high level injection, as we saw in the analysis of β rolloff, in order to maintain quasi-neutrality the excess hole concentration follows the electron concentration, resulting in an additional gradient and increasing the electric field that aids the movement of electrons across the base. This effect (Webster effect) reduces the transit time by a factor of 2 in a transistor with a uniformly doped base.



The base transit time τ_B is usually **not** the dominant factor in limiting the frequency response of modern BJTs.

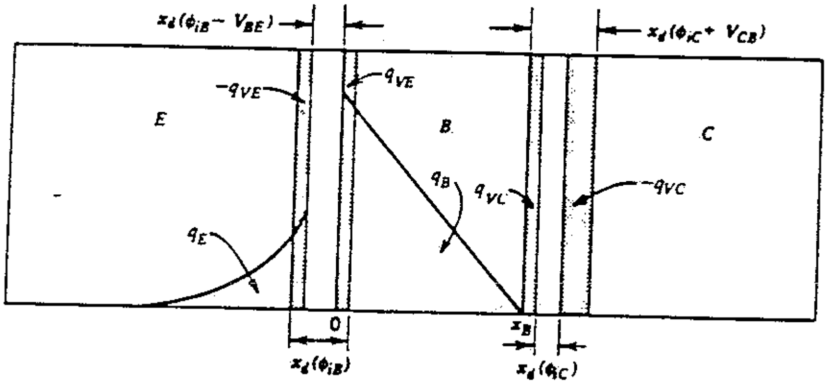
Charge Control Model

We will consider the BJT as a charge-controlled device, rather than a voltage or current controlled device.

In our analysis of the base transit time we derived a basic charge-control relationship, $I_C = q_B/\tau_B$, between the minority charge stored in the base, transit time across the base and the collector current.

This expression, however, considers only minority carrier transport across the base and thus is just a portion of the charge control model.

Let us reexamine the bipolar transistor in the forward active mode.



The injected minority charge due to the bias on the base-emitter junction q_F consists of the charge injected into the base plus that injected into the emitter.

$$q_F = q_{BF} + q_{EF} \tag{52}$$

Since the injected charge varies proportionally to the excess minority carrier density at the edges of the depletion region.

$$q_F = q_{F0}\left[\exp\frac{qV_{BE}}{kT} - 1\right] \tag{53}$$

Therefore we can write

$$I_C = \frac{q_F}{\tau_F} \tag{54}$$

with τ_F a constant dependent on the transistor structure. Since q_F is larger than the q_B, τ_F must be larger than τ_B.

The steady-state base current is proportional to the rate at which q_B recombines in the base region plus the rate at which holes are injected into the emitter. Both of these rates again depend on the excess minority charge density at the edges of the depletion region and thus are proportional to q_F.

$$I_B = \frac{q_F}{\tau_{BF}} \tag{55}$$

We can express the current gain therefore as

$$\frac{I_C}{I_B} = \beta_F = \frac{\tau_{BF}}{\tau_F} \tag{56}$$

Clearly, based on our analysis of the emitter efficiency and recombination in the base and emitter, we could calculate expressions for τ_F and τ_{BF} for the case of constant doping in each region. For a real device, τ_F and τ_{BF} can be obtained from measurements.

Example Assume we have a transistor with constant doping in each region and with a very high emitter efficiency. In that case,

$$q_F \approx q_B = \frac{1}{2} q \Delta n(0) x_B A_E$$

and

$$\tau_F \approx \tau_B = \frac{{x_B}^2}{2D_{nB}}$$

Since base recombination dominates the base current, $\tau_{BF} = \tau_{nB}$ and

$$\beta_F = \frac{2{L_{nB}}^2}{{x_B}^2}$$

This is the same result as would be calculated based on our analysis of the base transport factor.

In order to develop a full model, we must also consider changes in stored charges over time, including charges stored in the depletion region, since the model must be useful outside of steady state. We will use q_{VE} and q_{VC} to represent the charges stored in the emitter and collector depletion regions respectively.

We can therefore write the charge control expression for the base current in the forward active mode as

$$i_B = \frac{q_F}{\tau_{BF}} + \frac{dq_F}{dt} + \frac{dq_{VE}}{dt} + \frac{dq_{VC}}{dt} \tag{57}$$

The first term represents the number of holes which must be supplied by the base in order to replenish those lost to recombination and injection into the emitter. The rest of the current is available to increase the injected minority charge or depletion region charge.

All but the the last term flows in the emitter circuit. Therefore,

$$i_C = \frac{q_F}{\tau_F} - \frac{dq_{VC}}{dt} \tag{58}$$

and

$$i_E = -(i_B + i_C) = -q_F\left(\frac{1}{\tau_F} + \frac{1}{\tau_{BF}}\right) - \frac{dq_F}{dt} - \frac{dq_{VE}}{dt} \tag{59}$$

(57)-(59) make up a set of linear equations relating the currents and charges in the bipolar transistor as opposed to the nonlinear relationships developed between voltages and currents.

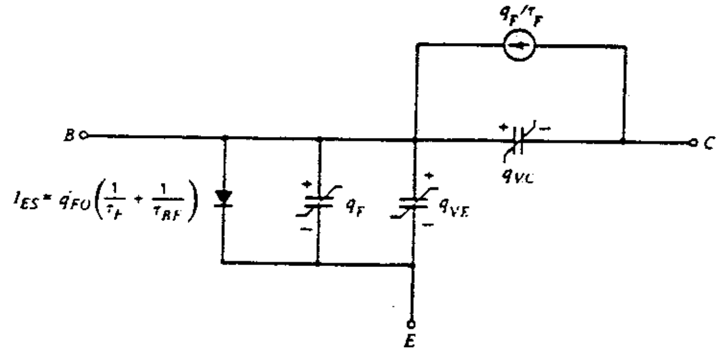
We can identify the time independent portions of the charge control model with the Ebers-Moll parameters. For example, we can see that the

$$I_F = q_F\left(\frac{1}{\tau_F} + \frac{1}{\tau_{BF}}\right)$$

and therefore

$$I_{ES} = q_{F0}\left(\frac{1}{\tau_F} + \frac{1}{\tau_{BF}}\right)$$

We can also draw a model for the forward active mode similar to the Ebers-Moll model, but including the voltage-dependent charge storage elements.

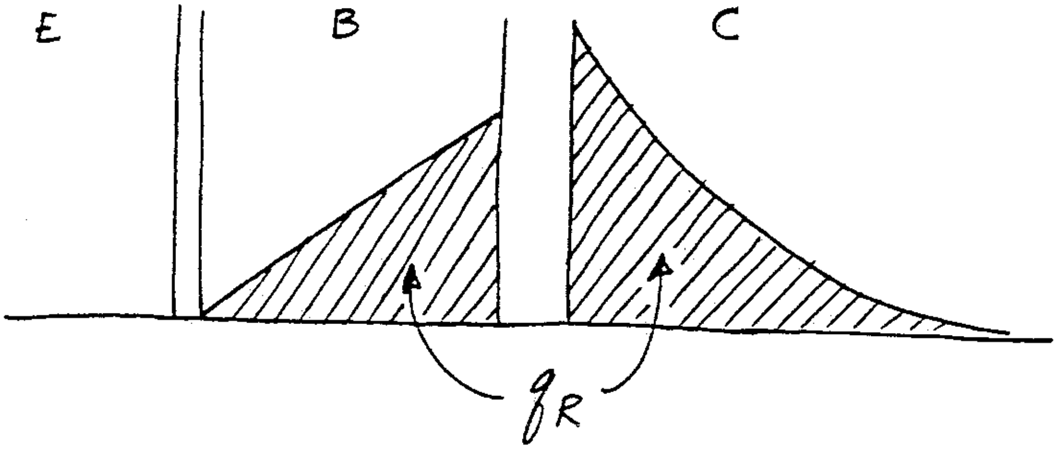


The underlying assumption of the charge control model is that the injected charges maintain their steady-state distributions. The main limitation arising from this is that the time periods of interest must be much longer than the base transit time.

Large Signal Model

The charge-control model is most useful for large-signal switching between the different transistor modes. In order to develop the full large signal model, we must reverse active operation as well. The number of injected minority carriers due to the base-collector junction is

$$q_R = q_{R0}\left[\exp\frac{qV_{BC}}{kT} - 1\right] \tag{60}$$



In analogy to the forward active mode, in the reverse active mode in steady state,

$$I_E = \frac{q_R}{\tau_R} \tag{61}$$

$$I_B = \frac{q_R}{\tau_{BR}} \tag{62}$$

$$\frac{I_E}{I_B} = \beta_R = \frac{\tau_{BR}}{\tau_R} \tag{63}$$

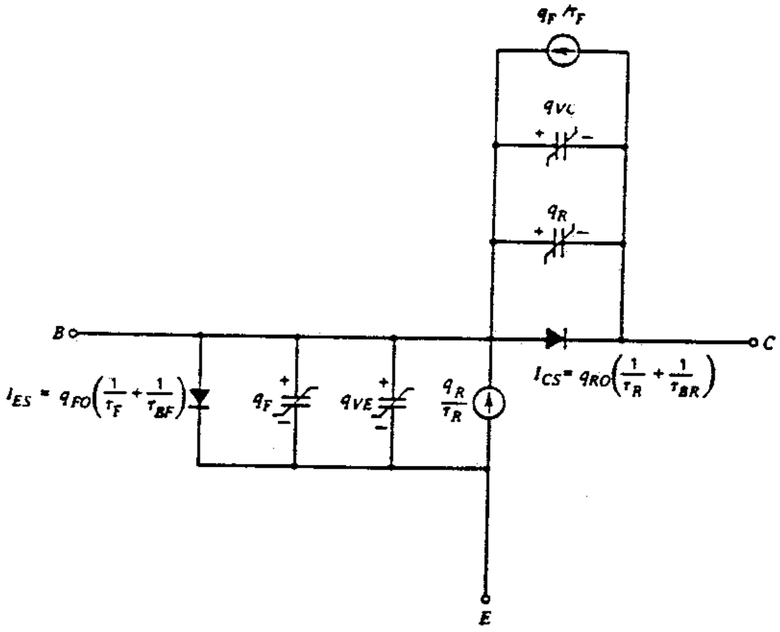
The full charge-control model therefore is

$$i_E = -q_F\left(\frac{1}{\tau_F}+\frac{1}{\tau_{BF}}\right) - \frac{dq_F}{dt} + \frac{q_R}{\tau_R} - \frac{dq_{VE}}{dt} \tag{64}$$

$$i_C = \frac{q_F}{\tau_F} - q_R\left(\frac{1}{\tau_R}+\frac{1}{\tau_{BR}}\right) - \frac{dq_R}{dt} - \frac{dq_{VC}}{dt} \tag{65}$$

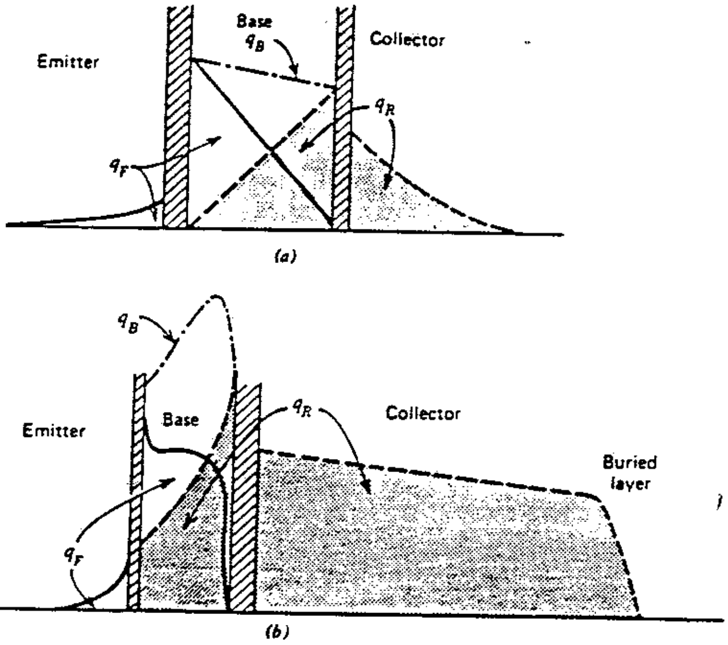
$$i_B = \frac{q_F}{\tau_{BF}} + \frac{dq_F}{dt} + \frac{q_R}{\tau_{BR}} + \frac{dq_R}{dt} + \frac{dq_{VE}}{dt} + \frac{dq_{VC}}{dt} \tag{66}$$

This model is valid in all four transistor modes and therefore is very useful for switching calculations.



Note that the parameters of this model are bias dependent. For large signal switching, the time delay can be approximated by $\tau \cong \Delta Q/\langle I\rangle$. Accurate large signal or switching behavior of BJTs is usually accomplished with computer techniques (see text).

Of particular interest is the saturation region in which both the junctions are forward biased. In saturation, the base current increases, while the collector current remains constant since the forward and reverse currents partially cancel each other out.

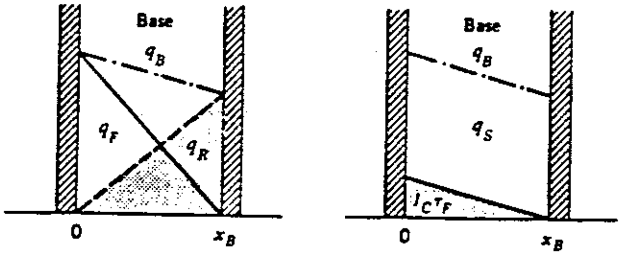


Constant Doping

Epitaxial Diffused Transistor

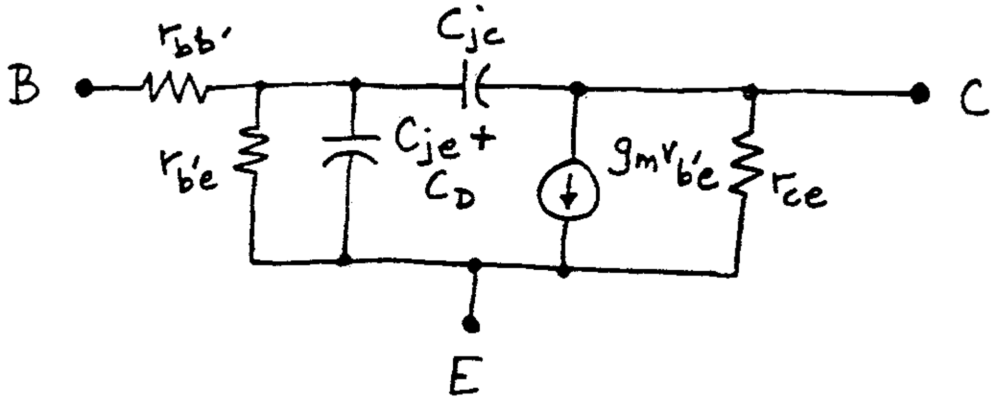
Since the collector is both lightly doped and relatively long, in saturation, the stored charge is often dominated by the injection into the collector, which can be quite large even for small forward base-collector biases.

In order to switch the transistor out of saturation that stored charge must be eliminated, which can result in large time delays. In early transistors, the collector region was doped heavily with gold (providing recombination centers) in order to reduce the minority carrier lifetime. Modern devices use nonsaturating transistors (Schottky-clamped for example) to reduce the switching time by never allowing the base-collector junction to become significantly forward biased.



Hybrid Pi Equivalent Circuit

A useful small signal ($v_{be} < kT/q$), AC equivalent circuit for the BJT is shown below.



The parameters are defined as follows

$$r_{bb'} = \text{base spreading resistance} = \frac{d(I_B R_B)}{dI_B} = R_B + I_C \frac{dR_B}{dI_C} \tag{67}$$

$$g_m = \text{transconductance} = \frac{dI_C}{dV_{BE}}$$

But,

$$I_C \cong \frac{qAD_n {n_i}^2}{x_B N_a} \exp \frac{qV_{BE}}{kT}$$

Therefore,

$$g_m \cong \frac{q}{kT} I_C \tag{68}$$

$$r_{b'e} \equiv \frac{dV_{BE}}{dI_B} = \frac{\beta_F}{g_m} \tag{69}$$

since $I_B \cong I_C/\beta_F$.

$C_{je} =$ capacitance associated with BE depletion region

$C_D \equiv$ diffusion capacitance of BE junction (due to stored minority carriers) $= \dfrac{dq_B}{dV_{BE}}$

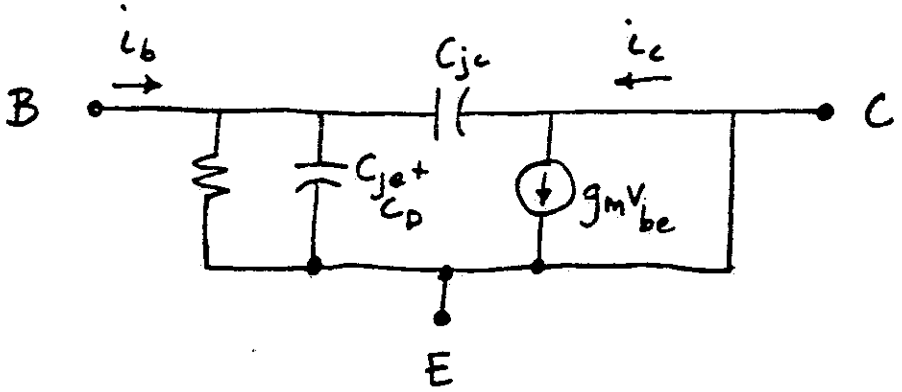
From (39), $q_B = (1/2)qn_{pB}(0)x_B A_E$. Therefore,

$$C_D \cong \frac{q^2 A_E x_B n_{pB}(0)}{2kT} = g_m \frac{{x_B}^2}{2D_n} = g_m \tau_B \tag{70}$$

$C_{jc} =$ depletion layer capacitance of collector base junction

$$r_{ce} = \text{ output resistance due to Early effect} = \frac{|V_A|}{I_C} = \frac{q|V_A|}{kTg_m} \tag{71}$$

The cutoff frequency is defined to be the frequency f_T at which the gain is reduced to 1. We may considering a simplified model ignoring the base resistance and output resistance.



The short-circuit current gain is therefore

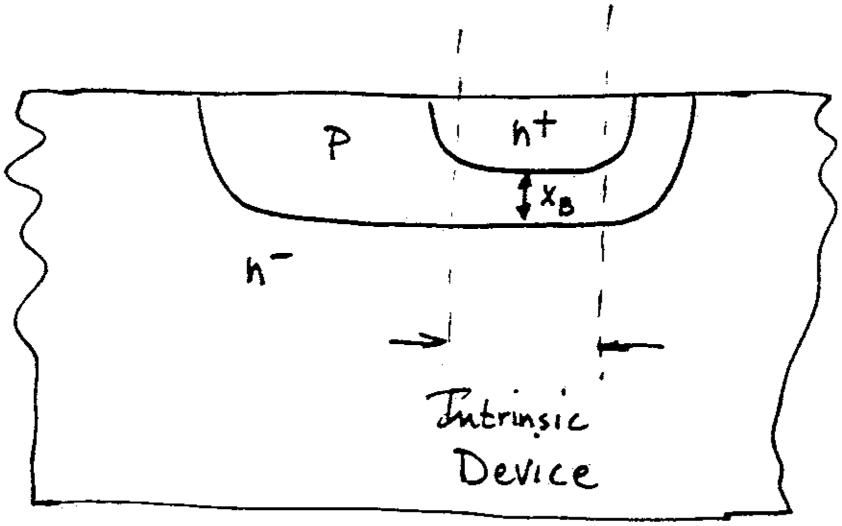
$$\frac{i_C}{i_B} = \frac{r_{b'e} g_m (1 - j\omega C_{jc}/g_m)}{1 + j\omega r_{b'e}(C_{jc} + C_{je} + C_D)} \tag{72}$$

and the unity gain bandwidth is

$$f_T \cong \frac{g_m}{2\pi(2C_{jc} + C_{je} + C_D)} \tag{73}$$

Above this frequency the device is not useful as an amplifier. Typical f_T of modern IC BJTs range from 500 MHz to several GHz.

Scaling Bipolar Transistors

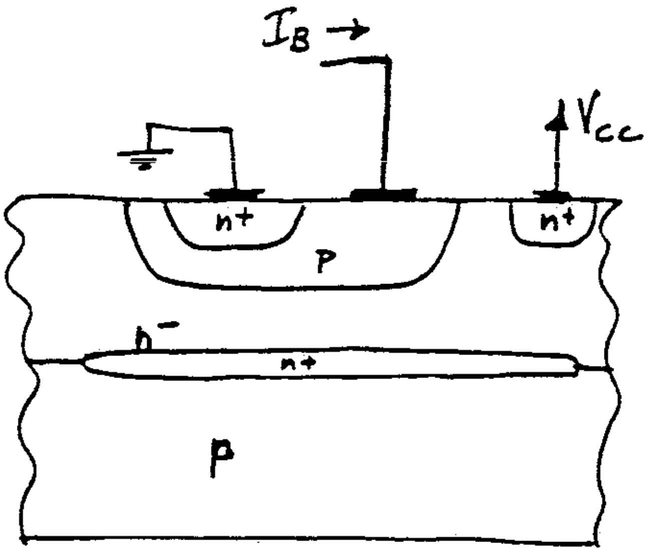


⇒ The intrinsic device is vertical. Therefore, scaling of lateral dimensions as we considered for MOS transistors does not improve the intrinsic device.

⇒ Scaling of the intrinsic device implies $x_B \downarrow$, but, x_B is already ~ 1000-3000 Åin modern devices. Fundamental limits occur around 300 Å. Scaling to these dimensions will require extremely tight process control.

⇒ Lateral scaling is employed in bipolar IC's, however. This reduces device size (therefore more devices / unit area) and reduces parasitic capacitances.

Summary



1. Bipolar transistors are low input impedance, current controlled devices

2. Transconductance ($\partial I_C/\partial V_{BE}$) is always higher in a BJT than in a MOSFET. Bipolar devices typically achieve the physical limit of 60 mV/decade of current. MOS devices approach this limit in scaled down structures ($x_{ox} \rightarrow 0$)

3. Because of exponential dependence of current on input voltage, fast, high-current devices can be made without increasing size.

4. Bipolar devices are typically used in the fastest logic circuits. They are faster than MOS because of (2) and (3).

5. Because they are larger physically and consume more power, BJT IC's are generally restricted to less complex functions than MOS IC's.