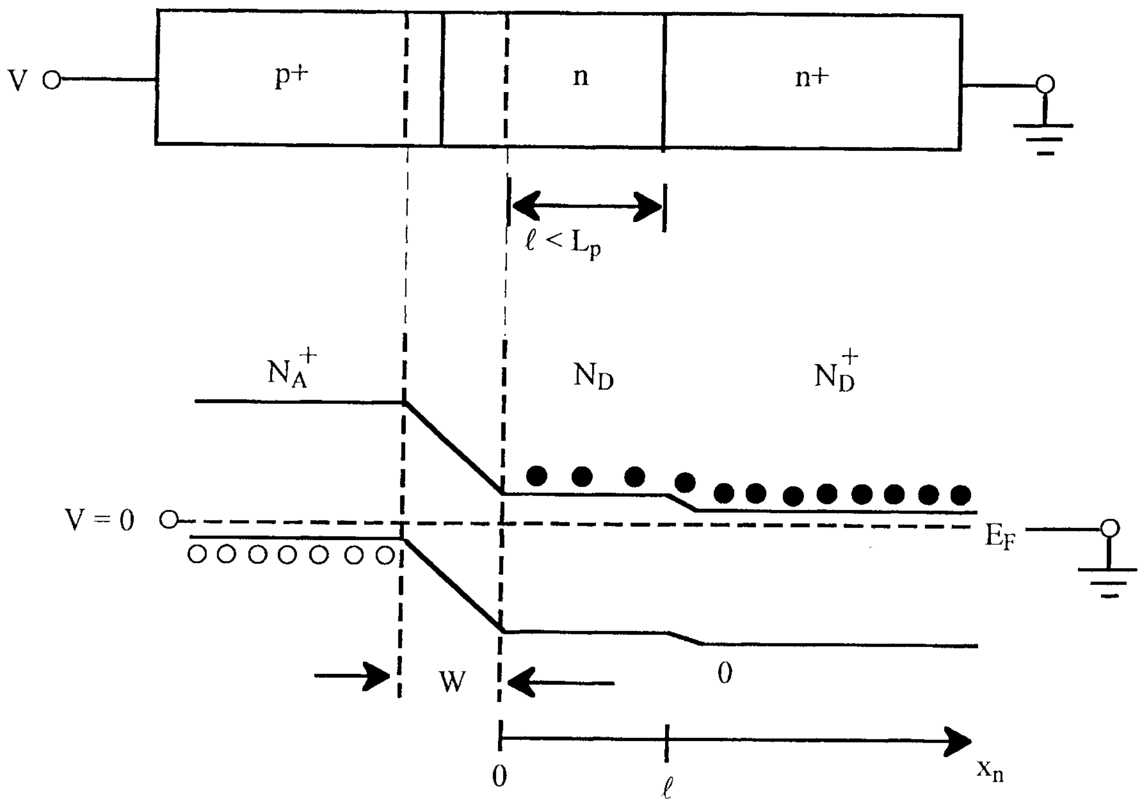
Narrow-Base Diode

A narrow-base p^+-n diode is sketched below along with the equilibrium band diagram. Diffusion of minority holes within the n-region will be characterized very simply through use of a 'straight-line' approximation. These results are then compared to exact expressions obtained from solving the 1D diffusion equation. A thorough understanding of this material is particularly important, as it will become the basis for a highly simplified and intuitive understanding of bipolar junction transistors.



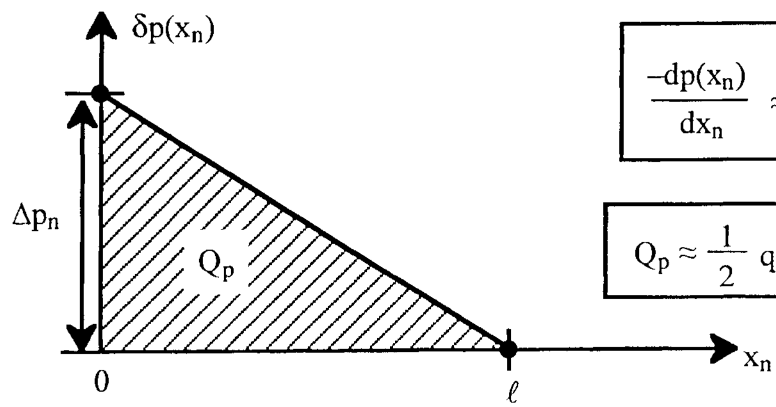
In the narrow-base p^+-n diode, a heavily doped n^+ contact is placed close to the junction boundary, much less than an average hole diffusion length away. The neutral portion of the lightly doped n-region thus has a width $\ell \ll L_P$. This is common in modern integrated circuits, since typical diffusion lengths $L_p \sim 1\text{-}10\ \mu m$ are much greater than most layer thicknesses. Current flow will still be dominated by holes as in a semi-infinite p^+-n diode, but the boundary condition for hole diffusion in the n-region is significantly altered. Minority holes that cross $x_n = \ell$ are assumed to recombine immediately upon entering the n^+ contact. With $N_D{}^+ \sim 10^4 N_D$, for example, the average hole lifetime, τ_p, is expected to be $\sim 10^4$ times shorter in the n^+ contact than in the lightly doped n-region. Thus, to a very good approximation, the excess hole population will vanish, $\delta p(x_n) = 0$, for all $x_n > \ell$. The boundary conditions for minority hole diffusion across the neutral n-region then become:

$$\delta p(x_n = 0) = \Delta p_n = p_n(e^{qV/kT} - 1)$$

$$\delta p(x_n = \ell) \approx 0$$

Straight-line approximation

Most of the injected minority holes will diffuse across the n-type base, $\ell << L_P$, without recombining until they hit the n^+ contact. The hole diffusion current will then be almost as large at the contact, $x_n = \ell$, as at the junction boundary, $x_n = 0$. This implies that, for constant cross sectional area, the hole gradient, $dp(x_n)/dx_n \approx -\Delta p/\ell$, must also be nearly the same everywhere inside the n-region; and we can approximate the excess carrier profile by a straight line.



$$\frac{-dp(x_n)}{dx_n} \approx \frac{\Delta p_n}{\ell}$$

$$Q_p \approx \frac{1}{2} qA\ell\Delta p_n$$

Minority holes contribute to current flow only by diffusion (for low-level injection), so the hole current density in the n-region is approximately given by:

$$J_p(x_n) \approx J_p(\text{diff}) = -qD_p \frac{dp(x_n)}{dx_n} \approx qD_p \frac{\Delta p_n}{\ell}$$

In this straight-line approximation, the total hole current diffusing across the narrow n-type base region is:

$$I_p(x_n) \approx A\, J_p(\text{diff})$$

$$\approx qA\frac{D_p}{\ell} p_n(e^{qV/kT} - 1)$$

Note that the diffusion current in the narrow-base structure is much larger than in an ordinary p^+-n diode, for the same voltage, since $\ell << L_P$ replaces L_P in the denominator of the diode equation. Conceptually, all excess holes recombine immediately at $x_n = \ell$, reducing $\delta p(x_n = \ell)$ to zero — far below the density of minority holes, $\Delta p\exp(-\ell/L_P)$, that would have been present at that point under the same conditions in a semi-infinite n-region. The much larger change in hole concentration across the neutral n-region then sets up a much larger diffusion current through random thermal motion, as half of the holes within every small subsection (roughly a mean-free path) go left or right with equal probability after each collision.

Although the hole diffusion current is the same throughout the n-region in straight-line approximation, there is actually a small decrease that we can estimate by looking at the recombination rate for stored minority hole charge. Only a few of the holes recombine inside the narrow n-region, but each time this happens another electron must flow in from the n^+ contact to preserve space-charge neutrality:

$$I_n(\text{recomb.}) = \frac{Q_p}{\tau_p} \approx \frac{\frac{1}{2}qA\ell\Delta p_n}{\tau_p}$$

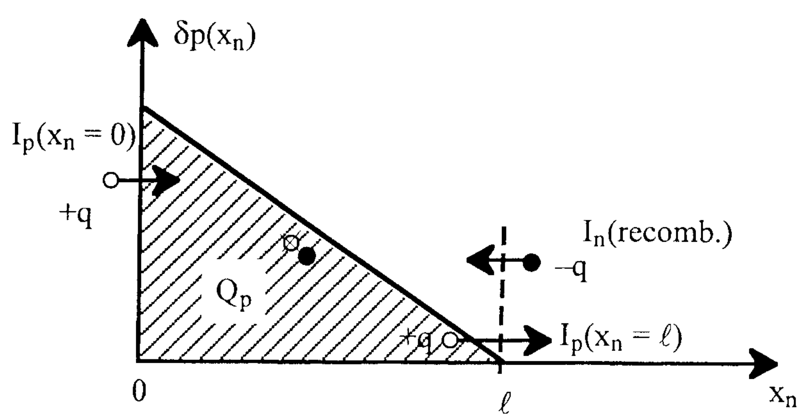
$$\approx \frac{qA\ell}{2\tau_p} p_n (e^{qV/kT} - 1)$$

This majority electron current flowing into the n-region at $x_n = \ell$ compensates the small decrease in hole diffusion current due to recombination within the base region:

$$I_n(\text{recomb.}) = I_p(x_n = 0) - I_p(x_n = \ell)$$

$$\approx I_p(x_n = 0)\left[\frac{\ell^2}{2L_p^2}\right]$$

By writing the result in this form, the difference term in brackets is seen to be very small for $\ell << L_P$, so the hole diffusion current and $dp(x_n)/dx_n \approx -\Delta p_n/\ell$ will remain nearly constant and accurately characterized by the straight-line approximation up to the n^+ contact. These results are sketched below. For $x_n > \ell$ inside the n^+ contact, the entire current will be carried by majority electrons according to this model with $\delta p(x_n > \ell) = 0$ throughout.



$$I_p(x_n = 0) = I_p(x_n = \ell) + I_n(\text{recomb.})$$

Exact solution to the 1D diffusion equation

Assuming constant cross sectional area, an exact solution to the 1D diffusion equation can be easily obtained as a linear combination of $\exp(-x_n/L_P)$ and $\exp(x_n/L_P)$ satisfying the above boundary conditions:

$$\delta p(x_n) = \Delta p_n \frac{[e^{(\ell - x_n)/L_p} - e^{(x_n - \ell)/L_p}]}{e^{\ell/L_p} - e^{-\ell/L_p}}$$

$$= \Delta p_n \,, \qquad \text{for } x_n = 0$$

$$= 0 \,, \qquad \text{for } x_n = \ell$$

The hole diffusion current at any point within the n-region is then given by:

$$I_p(x_n) = -qAD_p \frac{d}{dx_n} \delta p(x_n)$$

$$= qA \frac{D_p}{L_p} \Delta p_n \frac{[e^{(\ell - x_n)/L_p} + e^{(x_n - \ell)/L_p}]}{e^{\ell/L_p} - e^{-\ell/L_p}} .$$

Evaluating this expression at $x_n = 0$ yields the hole current injected at the junction boundary:

$$I_p(x_n = 0) = qA \frac{D_p}{L_p} \Delta p_n \, \mathrm{ctnh}\, (\ell/L_p)$$

$$\approx qA \frac{D_p}{\ell} \Delta p_n \left[1 + \frac{\ell^2}{3L_p^2} \right], \qquad \ell << L_p$$

The small argument expansion $\mathrm{ctnh}(y) \sim y^{-1}[1 + y^2/3 + ...]$ has been employed here to obtain the lowest correction to the straight-line approximation in the narrow-base limit. Note that $\mathrm{ctnh}(y) \to 1$ for $\ell >> L_P$ yielding the standard diode equation for long base width.

At $x_n = \ell$, the exact solution gives hole diffusion current at the n^+ contact in the form:

$$I_p(x_n = \ell) = qA\frac{D_p}{L_p}\Delta p_n \operatorname{csch}(\ell/L_p)$$

$$\approx qA\frac{D_p}{\ell}\Delta p_n\left[1-\frac{\ell^2}{6L_p^2}\right], \qquad \ell << L_p$$

where $\operatorname{csch}(y) \sim y^{-1}[1 - y^2/6 + ...]$ is used for small argument. Notice that this is slightly less than $I_p(x_n = 0)$ due to a difference in the first correction to the straight-line approximation. Since the total current must be the same at all points, the difference must correspond to the electron current flowing into the base from the n^+ contact to offset recombination of holes:

$$I_n(\text{recomb.}) = I_p(x_n = 0) - I_p(x_n = \ell)$$

$$= qA\frac{D_p}{L_p}\Delta p_n \tanh(\ell/2L_p)$$

$$\approx qA\frac{D_p}{\ell}\Delta p_n\left[\frac{\ell^2}{2L_p^2}\right], \qquad \ell << L_p$$

Summary

An exact solution to the 1D diffusion equation can be obtained for the narrow-base diode, but that approach yields messy results involving hyperbolic functions that obscure the simplicity of this device. Fortunately, for our purposes, all base widths ℓ less than ~0.5 L_P can be adequately characterized using the straight-line approximation. The simple expressions that it gives for hole concentration gradient:

$$\frac{dp(x_n)}{dx_n} \approx -\frac{\Delta p_n}{\ell}$$

and stored minority hole charge;

$$Q_p \approx \frac{1}{2}qA\ell\Delta p_n$$

lead to reasonable numerical estimates that can be readily evaluated. More importantly, this approach will permit us to focus on achieving an intuitive grasp of the more complex device physics of the bipolar transistor, based on the results presented here. For the BJT, however, we will need to include one additional element.

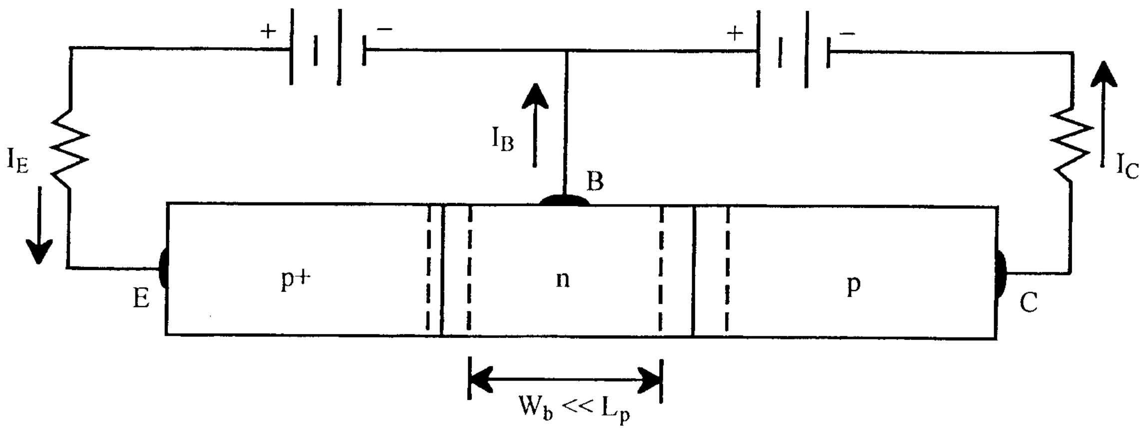
The electron component of current crossing the p^+-n junction has thus far been neglected in our analysis of the narrow-base diode. Assuming the p^+ region to be much wider than a minority electron diffusion length, this additional electron component of the junction current injected into the p^+ region will be given by the standard diode equation:

$$I_n(\text{inj.}) = qA\frac{D_n}{L_n}n_p(e^{qV/kT} - 1)$$

The equilibrium concentration of minority electrons, $n_p = n_i^2/N_A^+$, diffusion coefficient, D_n, and average diffusion length, L_n, are to be evaluated in the heavily doped p^+ region. Because the doping level, N_A^+, in the p^+ region is far greater than the donor density, N_D, in the lightly doped n region, this term represents a very small correction to the much larger current, $I_p(x_n = 0)$, due to holes in the narrow-base diode. However, in the three-terminal configuration of a p^+-n-p transistor, a small electron base current corresponding to the sum of I_n(inj.) and I_n(recomb.) can be independently adjusted to control the much larger current due to holes crossing the neutral n-region, leading to current amplification.

BJT Fundamentals: **p-n-p transistor normal (active) mode**

DC bias circuit (shown here in common base configuration for simplicity):



$I_E = I_B + I_C$

Emitter junction forward biased: $V_{EB} > 0$.

Collector junction reverse biased: $V_{CB} < 0$.

(1) Holes are injected as minority carriers across the forward biased emitter p^+-n junction into the neutral portion of an n-type base. Emitter injection efficiency, γ, is the fraction of total emitter current due to holes:

$$I_{Ep} = \gamma \, I_E$$

$$\gamma = \frac{I_{Ep}}{I_{Ep} + I_{En}} \qquad (\gamma \gtrsim 0.99 \text{ typical})$$

(2) A large fraction B (base transport factor) of these injected holes diffuses across the narrow base width $W_b << L_P$, small compared to an average diffusion length, and gets swept into the reverse-biased collector:

$$I_{Cp} = B \, I_{Ep} = B\gamma \, I_E$$

$$= \alpha \, I_E \qquad (\alpha \gtrsim 0.98 \text{ typical})$$

Here $\alpha = B\gamma$ is called the current transfer ratio. Thermally generated electron currents crossing the reverse biased collector junction are negligible in normal mode operation.

(3) In order to preserve space-charge neutrality, the base current must supply electrons to replace those which:

(a) recombine with a small fraction (1-B) of the holes transiting the base.

(b) are injected into the emitter across the forward biased p+-n junction.

$$I_B = I_B(\text{recomb.}) + I_B(\text{inj.})$$
$$= (1-B)\,\gamma\, I_E + (1-\gamma)\, I_E$$
$$= (1-\alpha)\, I_E \qquad [(1-\alpha) \lesssim 0.02 \text{ typical}]$$

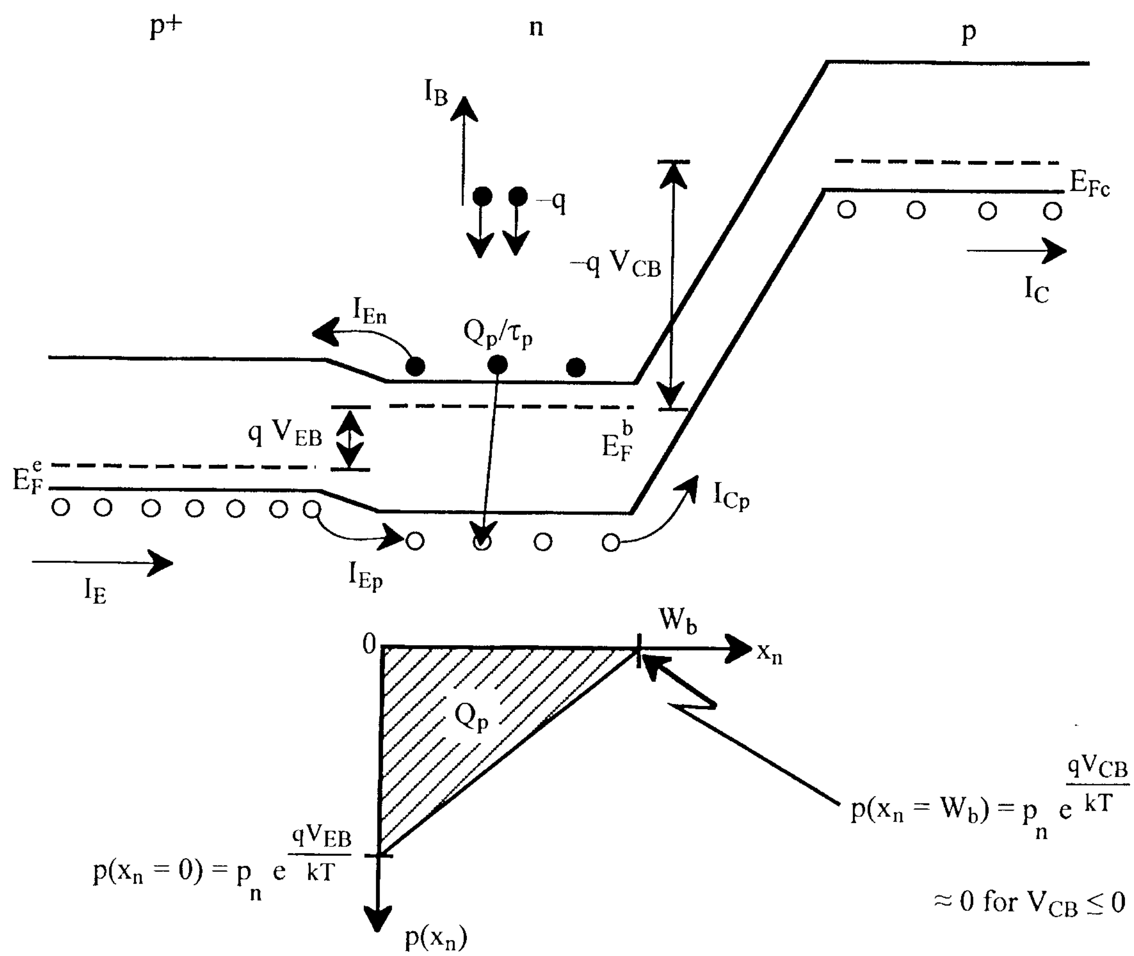
Note that electrons entering through the base (ohmic) contact carry negative charge, representing current flow out of the base in the positive direction.

(4) The most important parameter for a bipolar transistor is the normal mode current gain (as realized in a well-designed amplifier circuit):

$$\beta = \left(\frac{I_C}{I_B}\right)_{\text{normal mode}} = \frac{\alpha}{1-\alpha}$$

Typical values are $\alpha \sim 0.99$ and $\beta \sim 100$. Notice that these parameters are determined by the product of the emitter injection efficiency and base transport factor.

Band diagram: p-n-p normal mode



Boundary conditions for hole diffusion are set by the junction voltages:

$$\Delta p_E = p_n(e^{qV_{EB}/kT} - 1) \approx p_n e^{qV_{EB}/kT}$$

$$\Delta p_C = p_n(e^{qV_{CB}/kT} - 1) \approx -p_n$$

Diffusion of minority holes across the base yields a collector current given by the gradient of hole concentration at the junction boundary, W_b:

$$I_C \approx I_{Cp} = -q\ A\ D_p \left.\frac{dp(x_n)}{dx_n}\right|_{x_n = W_b}$$

Straight-line approximation:

When the base width, W_b, is narrow compared to the average diffusion length, L_P, the vast majority of minority holes injected at the emitter will diffuse across to the reverse biased collector junction in the normal mode. For constant cross sectional area, A, the diffusion current will be directly proportional to $-dp(x_n)/dx$; and the slope of the minority hole concentration must therefore be nearly the same at $x_n = W_b$ as at $x_n = 0$ (to within ~1% typically). For normal mode operation, the total current crossing to the reverse biased collector junction is determined by the base transport factor, $I_C = BI_{Ep}$, with B ~ 0.99.

Under these conditions, the solution of the 1D diffusion equation can be approximated by a straight line with slope:

$$-\frac{d\delta p(x_n)}{dx_n} \approx \frac{\Delta p_E - \Delta p_C}{W_b}$$

neglect for normal mode

The collector current is then:

$$I_C \approx q\,A\,\frac{D_p}{W_b}\Delta p_E = q\,A\,\frac{D_p}{W_b}p_n(e^{qV_{EB}/kT} - 1)$$

Note that W_b replaces L_P in the hole term of the diode equation, just as in the narrow-base diode. A tiny reverse current due to thermally generated electrons crossing from collector to base can be neglected. However, in determining the emitter current, it is important to include the small (but non-negligible) component of electron current injected into the p^+ region across the forward biased emitter junction to obtain the total emitter current:

$$I_E = I_{Ep} + I_{En}$$

$$\approx qA\left[\frac{D_p}{W_b}p_n + \frac{D_n^E}{L_n^E}n_p^E\right](e^{qV_{EB}/kT} - 1)$$

The electron term here is much smaller than the hole current because: (1) the p^+ region is much more heavily doped than the n-type base, so $n_p^E << p_n$, and (2) a smaller base width, W_b, has replaced the hole diffusion length, L_P, in the denominator of the hole term. The ratio of these terms gives the emitter injection efficiency:

$$\gamma = \left[1 + \frac{I_{En}}{I_{Ep}}\right]^{-1} \approx \left[1 + \frac{D_n^E}{L_n^E}\frac{W_b}{D_p}\frac{n_p^E}{p_n}\right]^{-1}$$

$$\approx \frac{1}{1 + \frac{\mu_n^E}{\mu_p^B}\frac{W_b}{L_n^E}\frac{N_D^B}{N_A^E}}$$

The stored minority hole charge in the base can be easily estimated using the straight-line approximation:

$$Q_p = q\,A\,\frac{1}{2}W_b(\Delta p_E + \Delta p_C)$$

neglect for normal mode

The base current component required to offset recombination with excess holes is then:

$$I_B(\text{recomb.}) = \frac{Q_p}{\tau_p}$$

$$\approx \frac{q\,A\,W_b}{2\tau_p} p_n(e^{qV_{EB}/kT} - 1)$$

The base current needed to replace the electrons injected into the p^+ emitter is:

$$I_B(\text{inj.}) = I_{En}$$

$$= \frac{q\,A\,D_n^E}{L_n^E} n_p^E(e^{qV_{EB}/kT} - 1)$$

and the total base current is given by:

$$I_B = I_B(\text{recomb.}) + I_B(\text{inj.})$$

Sometimes, for conceptual purposes, the authors of textbooks like to discuss a fictitious 'ideal' bipolar transistor model having perfect emitter injection efficiency, $\gamma = 1$. In that case, I_B(inj.) is far less than I_B(recomb.) and makes a negligible contribution to the base current. Under this assumption, the normal mode current gain can be written:

$$\beta \overset{\gamma\to 1}{=} \frac{I_C}{I_B(\text{recomb.})} = \frac{B}{1-B}$$

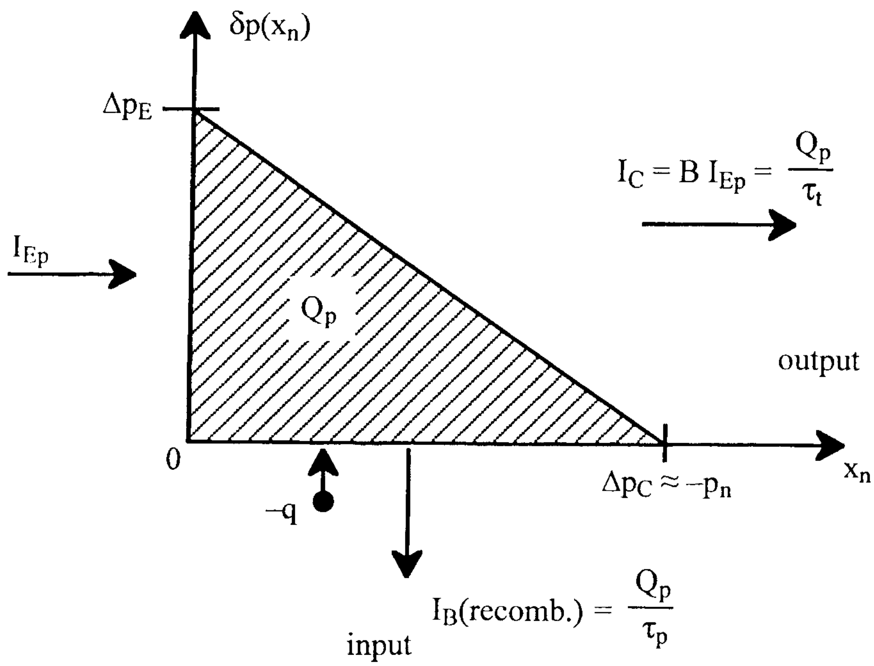
$$\approx \frac{\dfrac{q\,A\,D_p}{W_b}\Delta p_E}{\dfrac{q\,A\,W_b}{2\tau_p}\Delta p_E} = \frac{2L_p^2}{W_b^2}, \qquad \text{for } W_b \ll L_p$$

where $L_P^2 = D_p\tau_p$. Note that, for $\gamma = 1$, a current gain $\beta = 100$ corresponds to $W_b \sim L_P/7$. In most BJTs, however, the small difference between γ and 1 must be taken into account; and the base width for $\beta = 100$ would thus be somewhat smaller. What this 'ideal' calculation really gives us is an expression for the base transport factor correct to lowest order in $(W_b/L_P)^2$:

$$B \approx 1 - \frac{1}{2}\left(\frac{W_b}{L_p}\right)^2, \qquad \text{for } W_b << L_p$$

Current amplification:

We have seen that the parameter $\beta = I_C/I_B$ can be very large. When appropriate dc bias conditions are supplied in the common emitter configuration, small ac variations on the input current, I_B, will be amplified by β on the output current, I_C. The detailed circuit arrangements for a common emitter amplifier will be described later on, as the culmination of our introductory description of BJTs. Here we will try to illustrate the essential features of the current amplification effect in terms of the following sketch.



Once again, minority holes are shown diffusing across the neutral n-type base region with a forward biased emitter junction and reverse biased collector. For simplicity, the component of electron current injected into the emitter is neglected. In a properly designed amplifier circuit, the base current I_B can be modulated independently. The junction voltages are not clamped and will adjust themselves automatically to changes in I_B. Under these conditions, the stored minority hole charge, Q_P, is determined by a small base current through the requirement for space-charge neutrality:

$$Q_p = I_B(\text{recomb.})\,\tau_p$$

In turn, Q_P determines the much larger current of holes, I_C, diffusing from emitter to collector in an average transit time, $\tau_t << \tau_p$, much less than the average recombination lifetime:

$$I_C = \frac{Q_P}{\tau_t}$$

For this 'ideal' transistor with emitter injection efficiency $\gamma = 1$, we can now express the current gain as a simple ratio of lifetime divided by transit time:

$$\beta \overset{\gamma \to 1}{=} \frac{I_C}{I_B(\text{recomb.})} = \frac{\tau_p}{\tau_t} >> 1$$

$$\approx \frac{2L_p^2}{W_b^2} \qquad \text{(for } W_b << L_p \text{ as estimated above)}$$

This result yields a mean transit time for diffusion across the base width given by:

$$\tau_t = \frac{W_b^2}{2D_p}$$

Basic diffusion theory predicts that the average distance diffused is equal to the square root of diffusion coefficient multiplied by time, to within a factor of ~2 or so depending upon the detailed boundary conditions. In this case, $W_b = \sqrt{2D_p\tau_t}$, so we could have guessed the result to within a factor of $\sqrt{2}$.

Transient response:

An excess majority electron charge, $Q_n = -Q_P$, must be present within the base to screen the positive charges due to stored holes diffusing toward the collector. In steady state, a small current I_B of electrons entering via the base contact exactly compensates the loss of electrons due to recombination with holes. Now, imagine the following sequence of events.

(1) I_B is increased to a larger value I_B'.
(2) a momentary charge imbalance, $Q_n > Q_P$, occurs due to the larger influx of electrons, making the base more negative and increasing the forward bias across the emitter junction.
(3) the injected hole current, I_{Ep}, increases in response to the increased forward bias, and Q_P increases accordingly until the recombination rate for stored holes becomes large enough to precisely balance the new value of base current:

$$Q'_p = I'_B \tau_p$$

(4) I_C increases along with Q_P, because the vast majority of injected holes manage to diffuse into the collector before recombining, stabilizing at:

$$I'_C = \frac{Q'_p}{\tau_t} = \beta I'_B$$

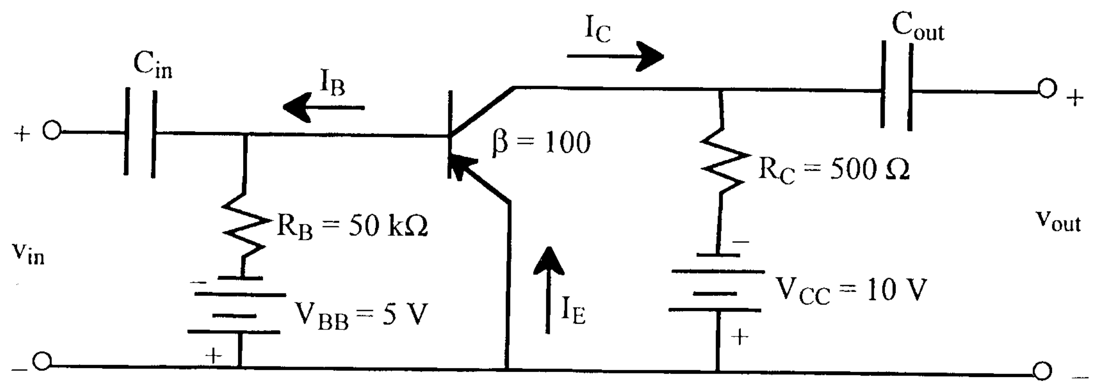
In this context, it may be useful to think of the base current, I_B, as a pump that determines the amount of minority stored hole charge, Q_P, which is allowed to be in transit inside the neutral base region at any given time. Since almost all of these holes diffuse from emitter to collector in a transit time much less than the average lifetime for recombination, only a very small fraction of holes recombine with electrons inside the base; yet this is the tail that wags the dog!

At the typical carrier densities found in semiconductor devices, even small deviations from space-charge neutrality generate large electric fields and voltage differences. In amplifier circuits, V_{EB} is allowed to adjust in response to changes in the base current I_B while nearly all of the injected holes wind up in the reverse biased collector. In this way, a small current of electrons controls a much larger minority hole diffusion current, because the condition of space-charge neutrality is automatically satisfied to very high precision in any region (in this case the neutral n-type base) containing a significant density of mobile carriers.

Common-Emitter Amplifier

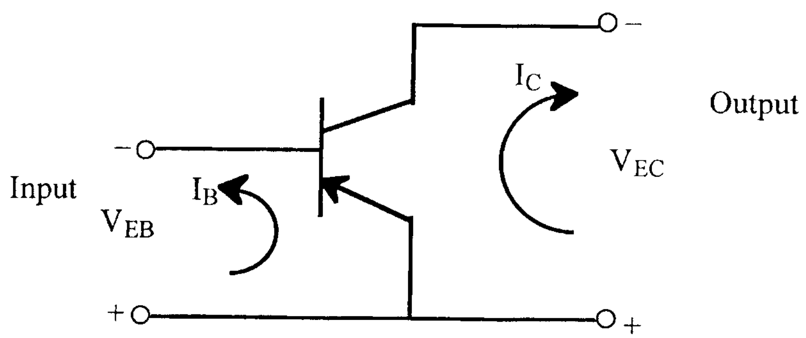
Most applications of bipolar transistors involve current amplification. The classic example of a common-emitter amplifier circuit is illustrated in Streetman, Fig. 7-4, p. 327. Here we take a more detailed look at this important topic as an extension of the brief discussion given in the text. The main purpose is to firmly establish the relationship between physical principles and circuit function for bipolar junction transistors.

The circuit diagram of Fig. 7-4 is redrawn below with two small alterations. First, coupling capacitors, C_{in} and C_{out}, have been added to accommodate ac signals at the input and output terminals. These capacitors are large enough to function as short circuits at the ac frequencies of interest (i.e., $1/i\omega C$ is negligible compared to other impedances in the circuit). On the other hand, the coupling capacitors present an open circuit to dc current flow and, because of this, the bias conditions set up by dc voltage sources and resistors in the input and output loops will remain unchanged. It is common practice to represent the small ac components of current and voltage with lower case letters and the larger dc currents and voltages with capitals.



In redrawing Fig. 7-4, the circuit symbol for a p-n-p transistor has been inserted in place of a schematic physical illustration. The emitter is identified by an arrow that points in the direction of majority carrier flow when current is injected into the emitter contact. The emitter region is p-type in this case, so holes will flow in the same direction as the injected current. For an n-p-n transistor, the arrow would be reversed.

In a common-emitter circuit, input is applied to the base and output is taken at the collector. Standard notation for voltage and current in the input and output loops is defined for a p-n-p transistor in the following sketch.



with $I_E = I_B + I_C$.

DC bias levels

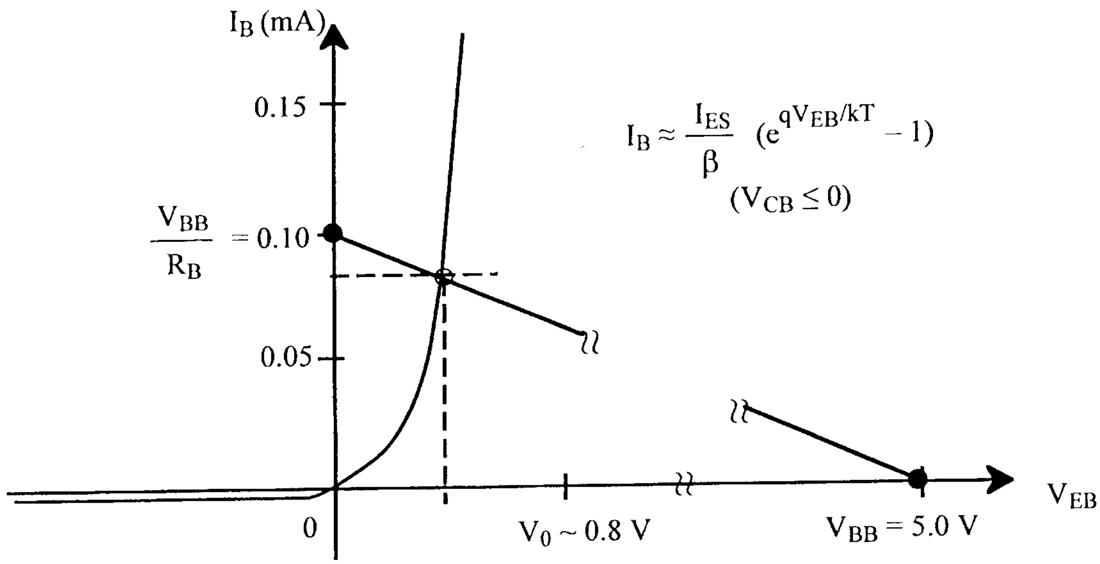
The dc portion of the amplifier circuit is designed to bias the transistor in the normal mode with the emitter junction forward biased ($V_{EB} > 0$) and the collector junction reverse biased ($V_{CB} < 0$). Under these conditions, the collector and base currents are effectively independent of the magnitude of the reverse bias across the collector junction:

$$I_C = \beta I_B \approx I_{ES} \left(e^{qV_{EB}/kT} - 1\right)$$

where we found:

$$I_{ES} \approx qA\left(\frac{D_p}{W_b} p_n + \frac{D_n^E}{L_n^E} n_p^E\right)$$

for $W_b << L_p$ and constant A within the straight-line approximation. The input characteristic is thus approximately independent of the output, so long as $V_{CB} \leq 0$:



Input Characteristic

The symbol "I_{ES}" denotes the emitter junction saturation current with the collector junction shorted.

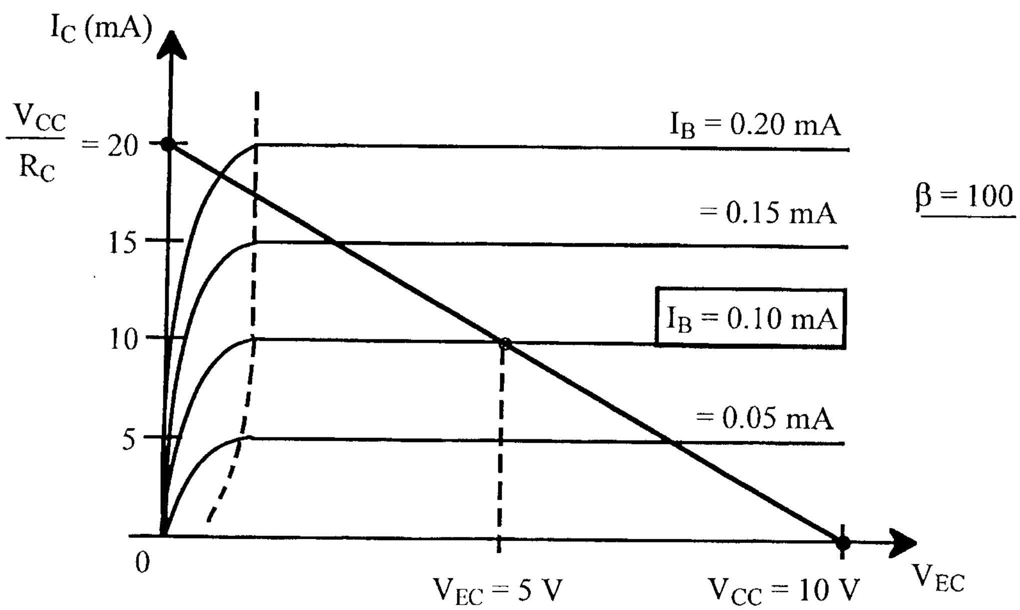
The input load line has been plotted on the figure above to determine the dc operating point (open circle denoting intersection of the load line with the transistor's input characteristic). Summing dc voltages around the input loop gives $V_{BB} = V_{EB} + I_B R_B$, so:

$$I_B = \frac{V_{BB} - V_{EB}}{R_B}$$

This relation is a straight line with the two intercepts V_{BB}/R_B and V_{BB} as shown. The horizontal voltage scale is broken here, since $V_{BB} = 5$ V is much larger than the contact potential, $V_0 \sim 0.8$ V, which limits forward bias across the emitter junction. Because V_{EB} is relatively small in forward bias, the dc operating point corresponds to:

$$I_B \approx \frac{V_{BB}}{R_B} = \frac{5\text{ V}}{50\text{ k}\Omega} = 0.10\text{ mA}$$

The output characteristics for common emitter are plotted using the input base current, I_B, as a parameter.



Output Characteristics

The voltage across the output can be expressed as the difference between the emitter and collector junction bias voltages:

$$V_{EC} = V_{EB} - V_{CB}$$

Except within ~1 Volt of the origin, the output characteristics are constant: $I_C = \beta\, I_B$ with $\beta = 100$ for operation in the normal mode. As $V_{EC} \rightarrow 0$, however, the collector junction voltage V_{CB} must approach the positive emitter voltage V_{EB}. Reverse bias on the collector is then lost, and the output characteristics deviate sharply from $I_C = \beta\, I_B$ to the left of the dotted line as I_C falls toward zero.

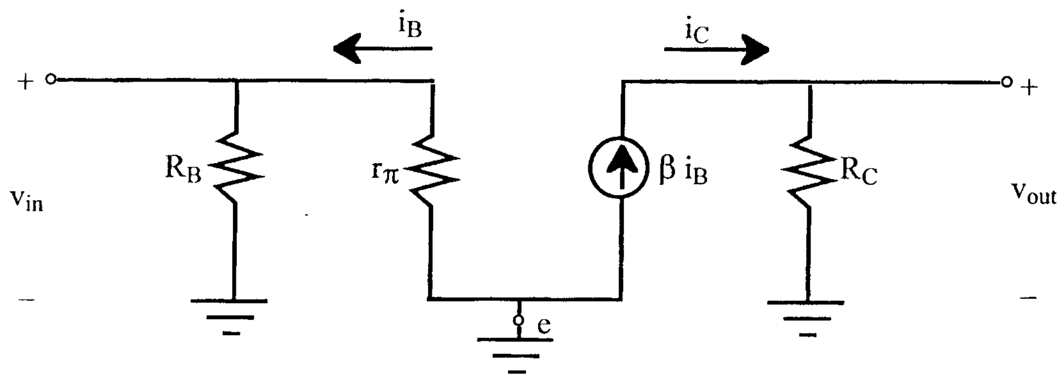
The load line for the output circuit is obtained by summing voltages $V_{CC} = V_{EC} + I_C R_C$:

$$I_C = \frac{V_{CC} - V_{EC}}{R_C}$$

Its intersection with the transistor characteristic corresponding to the dc base current, $I_B \approx 0.10$ mA, yields the indicated operating point with $V_{EC} \approx 5$ V. By positioning this point about halfway between the ends of the load line, the circuit design insures that the transistor will remain in the normal mode with $I_C = \beta I_B$ over a large range about the dc operating point.

ac small signal amplification

The circuit diagram at the top of this note shows the ac input voltage, v_{in}, applied across the emitter-base junction of the p-n-p transistor, in parallel with the base bias resistor R_B. The coupling capacitor, C_{in}, and dc voltage source, V_{BB}, act as ac shorts. Note that the polarity of the ac component of emitter-base voltage will oppose the dc bias, $V_{EB} > 0$, when v_{in} is positive. To characterize the small-signal amplification, a simplified ac equivalent circuit can be constructed as follows.



The differential conductance of the forward biased emitter junction, for small applied ac signals < kT/q ≈ 26 mV, is given by the slope of the input characteristic at the dc bias point:

$$\frac{1}{r_\pi} = \frac{dI_B}{dV_{EB}} \approx \frac{q}{kT} I_B$$

The input voltage, v_{in}, in the above equivalent circuit then produces a small ac modulation on the base current:

$$i_B = -\frac{v_{in}}{r_\pi}$$

In this example:

$$r_\pi = \frac{kT/q}{I_B} \approx \frac{0.026 \text{ V}}{0.10 \text{ mA}}$$

$$\approx 260\ \Omega$$

and good power matching at the input requires $r_\pi << R_B$.

Since the dc operating point is placed securely in the normal mode on the output characteristics, the induced ac component of collector current is:

$$i_C = \beta\, i_B = -\beta \frac{v_{in}}{r_\pi}$$

For simplicity, we will evaluate the output voltage under open-circuit conditions with no load resistance:

$$v_{out} = i_C R_C = -\beta \frac{R_C}{r_\pi} v_{in}$$

The open-circuit voltage gain is therefore given by:

$$\frac{v_{out}}{v_{in}} = -\beta \frac{R_C}{r_\pi}$$

$$\approx -100 \times \frac{500\ \Omega}{260\ \Omega}$$

$$\approx -192$$

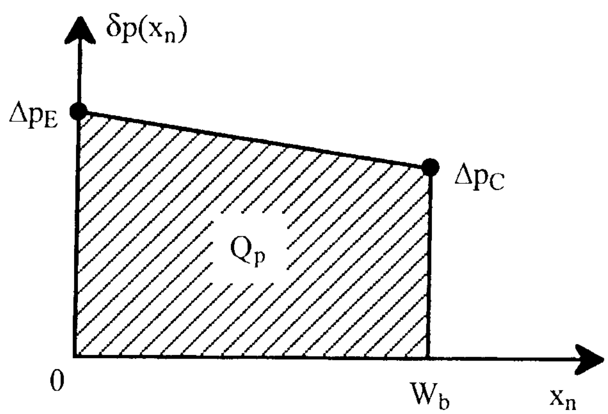
The π-resistance, r_π, presumably derives its name from the shape of the central portion of the ac equivalent circuit.

A useful way of envisioning the current amplification process is to examine the changes taking place about the dc bias point on the output characteristics. As I_B is modulated between 0.08 mA and 0.12 mA, for example, the instantaneous operating point will move up and down along the load line, oscillating about the dc set point. All of the characteristics, $I_C = \beta\, I_B$, are constant in this region, so the ac current will also be given by $i_C = \beta\, i_B$. In order to appreciate this more fully, you are encouraged to add collector currents for $I_B = 0.08$ mA and $I_B = 0.12$ mA onto the output characteristics shown above and note their intersections with the load line. You can also obtain a visual impression of the small signal ac conductance for the emitter-base junction, $1/r_\pi$, by drawing a tangent to the input characteristic at the dc bias point. Finally, in order to round out your physical understanding of the common-emitter amplifier, it is well to remember that the small ac changes in base current, $i_B = -v_{in}/r_\pi$, induce proportional changes in stored minority hole charge, δQ_p, and thereby a much larger change, $i_C = \delta Q/\tau_t$, on the ac collector current due to holes diffusing across the n-type base region.

Ebers-Moll: the terminal equations

A complete set of terminal equations will be presented here for a p-n-p transistor of constant cross sectional area, based on the 1D diffusion equation. The exact expressions contain a plethora of hyperbolic functions that are not readily transparent, even in this model, and algebraic details will not be discussed. Our main purpose is to make contact with the generalized descriptions of BJTs found in most textbooks, and then to build on the intuitive understanding of transistor function based on the straight-line approximation. The exact results for constant area will be put into perspective by first discussing them as a special case of the Ebers-Moll equations that are used to characterize any bipolar transistor under all bias conditions, and then by reducing them to the limit $W_b << L_P$ of narrow base width in the straight-line approximation for normal mode operation.

In order to model a bipolar transistor under any circuit conditions, the terminal equations must permit all values of voltage —both positive and negative— to be applied across the emitter and collector junctions. An example, called "saturation," is sketched below with both junctions in forward bias — a condition that commonly occurs in BJT switching circuits.



The two junction voltages, generalized here to include all possible bias conditions, give boundary conditions for diffusion of minority holes across the base:

$$\Delta p_E = p_n\left(e^{qV_{EB}/kT} - 1\right)$$

$$\Delta p_C = p_n\left(e^{qV_{CB}/kT} - 1\right)$$

In our simplified 1D model, the solution corresponding to these two boundary conditions can be obtained by solving for the coefficients, C_1 and C_2, in the general solution:

$$\delta p(x_n) = C_1 e^{x_n/L_p} + C_2 e^{-x_n/L_p}$$

$$= \Delta p_E, \qquad x_n = 0$$

$$= \Delta p_C, \qquad x_n = W_b$$

The resulting expressions for $C_1$and C_2 are given in Streetman, Eq. (7-13) p. 334. The text following this equation should be studied carefully to appreciate the manner in which the straight-line approximation emerges from the exact solution in the normal mode, as illustrated in Fig. 7-7.

Having obtained the two coefficients in terms of junction voltages, the minority hole diffusion current can now be calculated at any point within the neutral base region:

$$I_p(x_n) = -qAD_p \, d\delta p(x_n)/dx_n$$

$$= qA\frac{D_p}{L_p}\left[C_2 e^{-x_n/L_p} - C_1 e^{x_n/L_p}\right]$$

Evaluating this expression at the boundaries of the emitter junction, $x_n = 0$, and the collector junction, $x_n = W_b$, then yields the following results (after much algebra) as given in Streetman Eqs. (7-18) and (7-19):

$$I_{Ep} = I_p(x_n = 0) = qA\frac{D_p}{L_p}\left\{\Delta p_E \operatorname{ctnh}\left(\frac{W_b}{L_p}\right) - \Delta p_C \operatorname{csch}\left(\frac{W_b}{L_p}\right)\right\}$$

$$I_{Cp} = I_p(x_n = W_b) = qA\frac{D_p}{L_p}\left\{\Delta p_E \operatorname{csch}\left(\frac{W_b}{L_p}\right) - \Delta p_C \operatorname{ctnh}\left(\frac{W_b}{L_p}\right)\right\}$$

$$I_B(\text{recomb.}) = \frac{Q_p}{\tau_p} = I_{Ep} - I_{Cp}$$

$$= qA\frac{D_p}{L_p}\left(\Delta p_E + \Delta p_C\right)\tanh\left(\frac{W_b}{2L_p}\right)$$

Note that there are two terms in each terminal equation. The terms proportional to Δp_E represent the normal mode with a forward biased emitter junction and a grounded (or reverse biased) collector. The terms involving Δp_C represent the inverted mode in which the transistor is run backwards with a forward biased collector junction and a grounded (or reverse biased) emitter. The inverted mode terms appear with a negative sign in I_{Ep} and I_{Cp}, because minority holes injected across the collector junction will diffuse in the negative direction under this condition. The base current has a positive sign for both modes, however, since electrons will need to run down the base wire to offset hole recombination in either case. The diffusion equation is linear, thus all possible 1D solutions can be represented as a superposition of these two modes. As an example, the 'saturation' condition sketched above can be decomposed into its normal and inverted mode components as shown in Streetman Fig. 7-9. There is one additional but potentially confusing detail. The *mathematical* definition of these modes is specified by grounding the opposite junction: for example, the Δp_C terms will be *exactly* zero in the normal mode only for $V_{CB} = 0$. In practice these inverted mode terms will be negligible for any reverse bias on the collector, $\Delta p_C \sim -p_n$, so the normal mode is usually understood to include all $V_{CB} < 0$ when the emitter junction is forward biased.

Thus far, we have dealt only with the component of emitter and collector current carried by minority hole diffusion across the base, together with the much smaller flow of electrons into the neutral n-region needed to offset hole recombination. To complete this constant area 1D model, the electron components of current injected across the emitter and collector junctions must also be included. Assuming the widths of both p-type regions to be much greater than a minority electron diffusion length, the electron current components of I_E and I_C will have the same form as for a semi-infinite p-n junction (with negative sign in the inverted mode), because the boundary condition $\delta n \to 0$ far inside these p-regions admits only the exponentially decaying solution.

$$I_{En} = qA \frac{D_n^E}{L_n^E} n_p^E \left(e^{qV_{EB}/kT} - 1 \right)$$

$$I_{Cn} = -qA \frac{D_n^C}{L_n^C} n_p^C \left(e^{qV_{CB}/kT} - 1 \right)$$

Superscripts indicate that the equilibrium minority electron density, $n_p = n_i^2/N_A$, diffusion coefficient D_n and average diffusion length L_n are to be evaluated for p-type material doped to the density of the emitter and collector, respectively.

Finally, the complete set of terminal equations for the p-n-p transistor with constant area is obtained by combining all of the foregoing results:

$$I_E = I_{Ep} + I_{En}$$

$$I_C = I_{Cp} + I_{Cn}$$

$$I_B = \frac{Q_p}{\tau_p} + \left(I_{En} - I_{Cn} \right)$$

$$= I_B(\text{recomb.}) + I_B(\text{inj.})$$

The first term in I_B represents the rate at which electrons enter the base to offset recombination with injected holes. The second term accounts for the additional current needed to replace electrons that are injected across the emitter and base junctions under forward bias. Note that I_B(inj.) sums the total electron current leaving via both junctions, due to the minus sign convention in defining I_{Cn}. The total base current maintains space-charge neutrality within the neutral n-region, and the dominant contribution can be either of these two terms depending upon the relative values of base transport factor vs. the emitter and collector injection efficiencies.

Ebers-Moll equations

The terminal equations for the 1D model can be easily put into a general form derived by Ebers and Moll at Bell Laboratories in 1954 [Streetman, Eq. (7-32), p.342]:

$$I_E = I_{ES}\left(e^{qV_{EB}/kT} - 1\right) - \alpha_I I_{CS}\left(e^{qV_{CB}/kT} - 1\right)$$

$$I_C = \alpha_N I_{ES}\left(e^{qV_{EB}/kT} - 1\right) - I_{CS}\left(e^{qV_{CB}/kT} - 1\right)$$

Notice that these equations take the form of two coupled p-n diodes, with positive and negative signs for the normal and inverted modes respectively. The notation here is based upon the standard two-terminal diode measurements that are typically performed to experimentally extract the parameters for any BJT device. For example, I_{ES} represents the emitter diode saturation current with the collector junction *shorted*, experimentally defined by fitting current data obtained in forward bias, $V_{EB} > 0$, with $V_{CB} = 0$. Similarly, I_{CS} represents the collector saturation current in the inverted mode with the emitter junction shorted. In an alternative form of the Ebers-Moll equations, given in Streetman Eq. (7-37), I_{EO} represents the emitter saturation current measured with the collector junction *open*. Besides the saturation currents for both junctions (with opposite junction either shorted or open), the only additional parameters needed are the two current transfer ratios, α_N and α_I, for the normal and inverted modes as experimentally determined by very accurate current measurements. Individual components of the current transfer ratio, $\alpha = B\gamma$, cannot be measured separately.

In this introductory course, we will have little to say about generalized bias conditions and the inverted mode. Nevertheless, a basic exposure to this subject is required because the Ebers-Moll equations provide the starting point for simulating bipolar circuits.

Normal mode transistor parameters

Transistor parameters for the normal mode will now be evaluated for the 1D constant cross-sectional area model. The base transport factor is determined by taking the ratio $B = I_{Cp}/I_{Ep}$ in the normal mode:

$$B = \frac{I_{Cp}}{I_{Ep}} = \frac{\mathrm{csch}(W_b/L_p)}{\mathrm{ctnh}(W_b/L_p)} = \mathrm{sech}(W_b/L_p)$$

$$\approx 1 - \frac{1}{2}\left(\frac{W_b}{L_p}\right)^2, \qquad \text{for } W_b \ll L_p$$

The emitter saturation current I_{ES} may be obtained by combining I_{En} and I_{Ep} with $\Delta p_C = 0$:

$$I_{ES} = qA\frac{D_p}{L_p}p_n \mathrm{ctnh}(W_b/L_p) + qA\frac{D_n^E}{L_n^E}n_p^E$$

$$\approx qA\left(\frac{D_p}{W_b}p_n + \frac{D_n^E}{L_n^E}n_p^E\right), \qquad \text{for } W_b \ll L_p$$

The emitter injection efficiency, $\gamma = I_{Ep}/(I_{Ep} + I_{En})$, can then be calculated by taking the ratio of the electron and hole components in the above expression:

$$\gamma^{-1} = 1 + \frac{I_{En}}{I_{Ep}}$$

$$= 1 + \frac{D_n^E n_p^E}{L_n^E}\frac{L_p}{D_p p_n}\frac{1}{\mathrm{ctnh}(W_b/L_p)}$$

By employing the Einstein relation to replace the diffusion coefficients with mobilities, and expressing the equilibrium minority carrier densities in terms of the corresponding majority (doping) densities, this result can be rewritten in the form given by Streetman Eq. (7-25):

$$\gamma = \left[1 + \frac{L_p \mu_n^E N_D^B}{L_n^E \mu_p^B N_A^E}\tanh(W_b/L_p)\right]^{-1}$$

$$\approx \frac{1}{1 + \dfrac{\mu_n^E W_b N_D^B}{\mu_p^B L_n^E N_A^E}}, \qquad \text{for } W_b \ll L_p$$

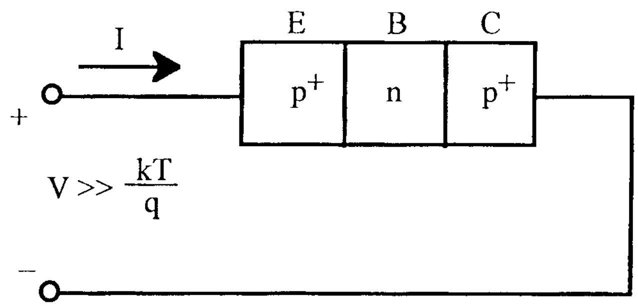
The normal mode current transfer ratio, $\alpha = B\gamma$, is the product of these two parameters. For typical BJTs the current transfer ratio $\alpha \sim 0.99$ is very close to unity, yielding a current amplification factor $\beta = \alpha/(1-\alpha) \sim 100$. If deviation of the base transport factor B from unity is primarily responsible for determining the values of α and β, then it is natural to think of the underlying gain mechanism in terms of a small electron recombination current, $I_B \approx Q_P/\tau_p$, dictating the much larger collector current, $I_C \approx Q_P/\tau_t$, of holes diffusing across the base in an average transit time $\tau_t << \tau_p$. On the other hand, if the deviation of emitter injection efficiency γ from unity is numerically more important in determining α and β, then it is appropriate to think of the current gain in terms of a very large ratio of hole current induced into the base region in response to a much smaller electron current, $I_B \approx I_{En}$, injected into the emitter. In either case, the fundamental requirement for space-charge neutrality within the non-depleted base region allows a small electron base current, I_B, to dictate the emitter-base junction voltage, V_{EB}, and a much larger hole current, I_C, diffusing toward the reverse biased collector in a properly designed amplifier circuit. Pumping more (fewer) electrons into the base by increasing (decreasing) I_B will momentarily cause the base to become more (less) negative with respect to the emitter. The forward bias on the emitter junction will therefore increase (decrease) until the stored hole charge, Q_P, and electron emitter current, I_{En}, come into balance with the new value of steady-state base current, $I_B = Q_P/\tau_p + I_{En}$. The end result is an amplification of ac changes in base current given by $i_C = \beta i_B$.

Ebers-Moll: example problems

Symmetric p^+-n-p^+, $I_{ES} = I_{CS}$, $\alpha_N = \alpha_I$

(7.34) $$I_E = \frac{I_{ES}}{P_n}(\Delta P_E - a\Delta P_C),\ I_C = \frac{I_{ES}}{P_n}(\alpha\Delta P_E - \Delta P_C)$$

7.7(a)



$I = I_E + I_C$, $\quad I_B = 0$

emitter jcn. forward biased

collector jcn. reverse biased

$\Rightarrow \Delta P_C \approx -P_n$

$I_E = I_C \quad \Rightarrow \quad \Delta P_E \approx +P_n$

Most of applied voltage is dropped across reverse biased collector.

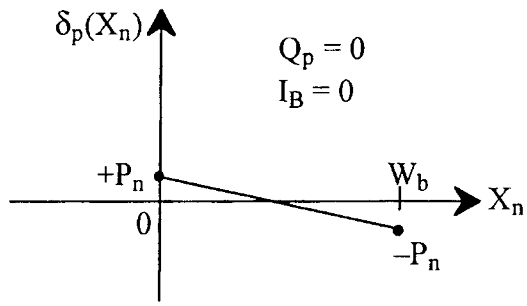
Very small forward voltage on emitter:

$$\Delta P_E = P_n\left(e^{\frac{qV_{EB}}{kT}} - 1\right) \approx +P_n$$

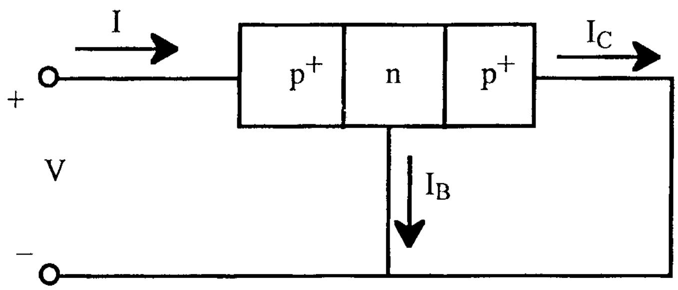
$$\Rightarrow \quad V_{EB} \simeq \frac{kT}{q}\ln 2$$

$I \approx I_{ES}(1 + \alpha)$ (very small)

No stored minority hole charge:



7.7(b)



Collector jcn. shorted: $V_{CB} = 0$

$$\Rightarrow \qquad \Delta P_C = 0$$

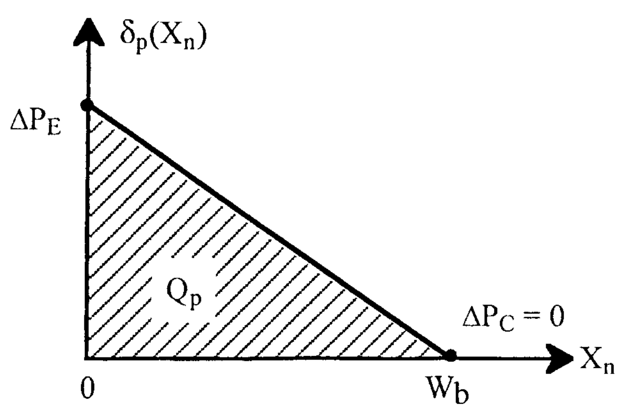
Emitter jcn. voltage: $V_{EB} = V$

$$\Rightarrow \qquad \Delta P_E = P_n\left(e^{\frac{qV}{kT}} - 1\right)$$

Current:

$$I = I_E = I_{ES}\left(e^{\frac{qV}{kT}} - 1\right)$$

(narrow base diode)



$$Q_P = q\,A\,\frac{1}{2}\,W_b\,\Delta P_E \qquad\qquad I_B = \frac{Q_P}{\tau_P} + I_E^n$$