The structure we have described is a p-n-p bipolar junction transistor. The forward-biased junction which injects holes into the center n region is called the emitter junction, and the reverse-biased junction which collects the injected holes is called the collector junction. The np region, which serves as the source of injected holes, is called the emitter, and the p region into which the holes are swept by the reverse-biased junction is called the collector. The center n region is called the base, for reasons which will become clear in Section 7.3, when we discuss the historical development of transistor fabrication. The biasing arrangement in Fig. 7-2 is called the common base configuration, since the base electrode B is common to the emitter and collector circuit.

To have a good p-n-p transistor, we would prefer that almost all the holes injected by the emitter into the base be collected. Thus the n-type base region should be narrow, and the hole lifetime $\tau_p$ should be long. This requirement is summed up by specifying $W_n \ll L_p$, where $W_n$ is the length of the neutral n material of the base (measured between the depletion regions of the emitter and collector junctions), and $L_p$ is the diffusion length for holes in the base ($D_p \tau_p^{1/2}$). With this requirement satisfied, an average hole injected at the emitter junction will diffuse to the depletion region of the collector junction without recombination in the base. A second requirement is that the current $I_k$ crossing the emitter junction should be composed almost entirely of holes injected into the base, rather than electrons crossing from base to emitter. This requirement is satisfied by doping the base region lightly compared with the emitter, so that the p-n emitter junction of Fig. 7-2 results.

It is clear that current $I_k$ flows into the emitter of a properly biased p-n-p transistor and that $I_k$ flows out at the collector, since the direction of hole flow is from emitter to collector. However, the base current $I_k$ requires a bit more thought. In a good transistor the base current will be very small since $I_k$ is essentially hole current, and the collected hole current $I_C$ is almost equal to $I_k$. There must be some base current, however, due to requirements

In this section we shall discuss rather simply the various factors involved in transistor amplification. Basically, the transistor is useful in amplifiers because the currents at the emitter and collector are controllable by the relatively small base current. The essential mechanisms are easy to understand if various secondary effects are neglected. We shall use total current (dc currents only) and total current gain (dc gain only).
plus a-c) in this discussion, with the understanding that the simple analysis applies only to d-e and to small-signal a-c at low frequencies. We can relate the terminal currents of the transistor $i_E, i_B,$ and $i_C$ by several important factors. In this introduction we shall neglect the saturation current at the collector (Fig. 7-3, component 3) and such effects as recombination in the transition regions. Under these assumptions, the collector current is made up entirely of those holes injected at the emitter which are not lost to recombination in the base. Thus $i_C$ is proportional to the hole component of the emitter current $i_{Ep}$:

$$i_C = B i_{Ep},$$

(7-1)

The proportionality factor $B$ is simply the fraction of injected holes which make it across the base to the collector; $B$ is called the base transport factor. The total emitter current $i_E$ is made up of the hole component $i_{Ep}$ and the electron component $i_{En}$ due to electrons injected from base to emitter (component 5 in Fig. 7-3). The emitter injection efficiency $\gamma$ is

$$\gamma = \frac{i_{Ep}}{i_{En} + i_{Ep}}$$

(7-2)

For an efficient transistor we would like $B$ and $\gamma$ to be very near unity, that is, the emitter current should be due mostly to holes ($\gamma = 1$), and most of the injected holes should eventually participate in the collector current ($B = 1$). The relation between the collector and emitter currents is

$$\frac{i_C}{i_E} = \frac{B i_{Ep}}{i_{En} + i_{Ep}} = B \gamma = \alpha$$

(7-3)

The product $B \gamma$ is defined as the factor $\alpha$, called the current transfer ratio, which represents the emitter-to-collector current amplification. There is no real amplification between these currents, since $\alpha$ is smaller than unity. On the other hand, the relation between $i_C$ and $i_B$ is more promising for amplification.

In accounting for the base current, we must include the rates at which electrons are lost from the base by injection across the emitter junction ($i_{En}$) and the rate of electron recombination with holes in the base. In each case, the lost electrons must be resupplied through the base current $i_B$. If the fraction of injected holes making it across the base without recombination is $B$, then it follows that $(1 - B)$ is the fraction recombining in the base. Thus the base current is

$$i_B = i_{Ep} + (1 - B)i_{Ep}$$

(7-4)

neglecting the collector saturation current. The relation between the collector and base currents is found from Eqs. (7-1) and (7-4):

$$\frac{i_C}{i_B} = \frac{B i_{Ep}}{i_{En} + (1 - B)i_{Ep}} = \frac{B(i_{Ep}/(i_{Ep} + i_{Ep}))}{1 - B(i_{Ep}/(i_{Ep} + i_{Ep}))}$$

(7-5)

The factor $B$ relating the collector current to the base current is the base-to-collector current amplification factor. Since $\alpha$ is near unity, it is clear that $B$ can be large for a good transistor, and the collector current is large compared with the base current.

It remains to be shown that the collector current $i_C$ can be controlled by variations in the small current $i_B$. In the discussion to this point, we have indicated control of $i_C$ by the emitter current $i_E$, with the base current characterized as a small side effect. In fact, we can show from space charge neutrality arguments that $i_B$ can indeed be used to determine the magnitude of $i_C$. Let us consider the transistor of Fig. 7-4, in which $i_B$ is determined by a biasing circuit. For simplicity, we shall assume unity emitter injection efficiency and negligible collector saturation current. Since the n-type base region is electrostatically neutral between the two transition regions, the presence of excess holes in transit from emitter to collector calls for compensating excess electrons from the base contact. However, there is an important difference in the times which electrons and holes spend in the base. The average excess

![Figure 7-4](image-url)

Example of amplification in a common-emitter transistor circuit.
(a) biasing circuit.
(b) addition of ac variation of base current $i_B$ resulting in an ac component $i_C$.\[\text{Figure 7-4}\]
Bipolar Junction Transistors

If a small a-c current $i_C$ is superimposed on the steady state base current of Fig. 7-4a, a corresponding a-c current $i_C$ appears in the collector circuit. The time-varying portion of the collector current will be $i_C$ multiplied by the factor $\beta$, and current gain results.

We have neglected a number of important properties of the transistor in this introductory discussion, and many of these properties will be treated in detail below. We have established, however, the fundamental basis of operation for the bipolar transistor and have indicated in a simplified way how it can be used to produce current gain in an electronic circuit.

\[
\frac{i_C}{i_B} = \beta = \frac{\tau_p}{\tau_r} \quad (7-7)
\]

for $\gamma = 1$ and negligible collector saturation current.

If the electron supply to the base $(i_B)$ is restricted, the traffic of holes from emitter to base is correspondingly reduced. This can be argued simply by supposing that the hole injection does continue despite the restriction on electrons from the base contact. The result would be a net buildup of positive charge in the base and a loss of forward bias (and therefore a loss of hole injection) at the emitter junction. Clearly, the supply of electrons through $i_B$ can be used to raise or lower the hole flow from emitter to collector.

The base current is controlled independently in Fig. 7-4. This is called a common-emitter circuit, since the emitter electrode is common to the base and collector circuits. The emitter junction is clearly forward biased by the battery in the base circuit. The voltage drop in the forward-biased emitter junction is small, however, so that almost all of the voltage from collector to emitter appears across the reverse-biased collector junction. Since $V_{CE}$ is small for the forward-biased junction, we can neglect it and approximate the base current as $5 V / 50 k\Omega = 0.1 mA$. If $\tau_p = 10 \mu s$ and $\tau_r = 0.1 \mu s$, $\beta$ for the transistor is 100 and the collector current $I_C$ is 10 mA. It is important to note that $i_C$ is determined by $\beta$ and the base current, rather than by the battery and resistor in the collector circuit (as long as these are of reasonable values to maintain a reverse-biased collector junction). In this example 5 V of the collector circuit battery voltage appears across the 500 $\Omega$ resistor, and 5 V serves to reverse bias the collector junction.

\[\text{Example 7-1}\]

(a) Show that Eq. (7-7) is valid from arguments of the steady state replacement of stored charge. Assume that $\tau_n = \tau_r$.

(b) What is the steady state charge $Q_n = Q_p$ due to excess electrons and holes in the neutral base region for the transistor of Fig. 7-4?

(a) In steady state there are excess electrons and holes in the base. The charge in the electron distribution $Q_e$, is replaced every $\tau_r$ seconds. Thus $i_p = Q_e/\tau_r$. The charge in the hole distribution $Q_p$, is collected every $\tau_r$ seconds, and $i_C = Q_p/\tau_r$. For space charge neutrality, $Q_n = Q_p$, and

\[
\frac{i_C}{i_p} = \frac{Q_n/\tau_r}{Q_p/\tau_r} = \frac{\tau_p}{\tau_r}
\]

(b) $Q_n = Q_p = i_C \tau_r = i_B \tau_p = 10^{-9} C$.

The first transistor invented by Bardeen and Brattain in 1947 was the point contact transistor. In this device two sharp metal wires, or “cat’s whiskers,” formed an “emitter” of carriers and a “collector” of carriers. These wires were simply pressed onto a slab of Ge which provided a “base” or mechanical support, through which the injected carriers flowed. This basic invention rapidly led to the BJT, in which charge injection and collection was achieved using two p-n junctions in close proximity to each other. The p-n junctions in BJTs can be formed in a variety of ways using thermal diffusion, but modern devices are generally made using ion implantation (Section 5.1.4).

Let us review a simplified version of how to make a double polysilicon, self-aligned n-p-n Si BJT. This is the most commonly used, state-of-the-art technique for making BJTs for use in an IC. Use of n-p-n transistors is much more popular than p-n-p devices because of the higher mobility of electrons compared to holes. The process steps are shown in cross-sectional view in Fig. 7-5. A p-type Si substrate is oxidized, windows are defined using photolithography and etched in the oxide. Using the photoresist and oxide as an implant mask, a donor with very small diffusivity in Si, such as As or Sb, is implanted into the open window to form a highly conductive n' layer (Fig. 7-5a). Subsequently,
the photoresist and the oxide are removed, and a lightly doped n-type epitaxial layer is grown. During this high temperature growth, the implanted n' layer diffuses only slightly towards the surface and becomes a conductive buried collector (also called a sub-collector). The n' sub-collector layer guarantees a low collector series resistance when it is connected subsequently to the collector ohmic contact, sometimes through the use of an optional, masked deep n' 'sinker' implant or diffusion only in the collector contact region (Fig. 7-5c). The lightly doped n-type collector region above the n' sub-collector in the part of the BJT where the base and emitter are formed ensures a high base-collector reverse breakdown voltage. (It turns out that wherever the sub-collector is formed, and subsequently the epitaxial layer is grown on top, there is a notch or step in the substrate surface. This notch is not explicitly shown in Fig. 7-5a. This notch is very useful as a marker of the location of the sub-collectors because subsequently, we have to align the LOCOS isolation mask with respect to the sub-collector.)

For integrated circuits involving not just discrete BJTs, but many interconnected transistors, there are issues involving electrical isolation of adjacent BJTs in order to ensure that there is no electrical cross-talk between them. As described in Section 6.4.1, such isolation can be achieved by LOCOS to form field or isolation oxides after a B channel stop implant (Fig. 7-5b). Another isolation scheme that is particularly well suited for high density bipolar circuits involves the formation of shallow trenches by RIE, backfilled with oxide and polysilicon (Section 9.3.1). In this process a nitride layer is patterned and used as an etch mask for an anisotropic etch of the silicon to form the trench. Using reactive ion etching, a narrow trench about 1 μm deep can be formed with very straight sidewalls. Oxidation inside the trench forms an insulating layer, and the trench is then filled with oxide by Low-Pressure Chemical Vapor Deposition (LPCVD).

A polysilicon layer is deposited by LPCVD, and doped heavily p' with B either during deposition or subsequently by ion implantation. An oxide layer is deposited next by LPCVD. Using photolithography with the base/emitter mask, a window is etched in the polysilicon/oxide stack by RIE (Fig. 7-5c). A heavily doped “extrinsic” p' base is formed by diffusion of B from the doped polysilicon layer into the substrate in order to provide a low resistance, high speed base ohmic contact. An oxide layer is then deposited by LPCVD, which has the effect of closing up the base window that was etched previously, and B is implanted into this window (Fig. 7-5d). This base implant forms a more lightly p doped “intrinsical” base through which most of the current flows from the emitter to the collector. The more heavily doped extrinsic base forms a collar around the intrinsic base, and serves to reduce the base series resistance. It is critical that the base be enclosed well within the collector because otherwise it would be shorted to the p' substrate. Finally, another LPCVD oxide layer is deposited to close up the base window further, and the oxide is etched all the way to the Si substrate by RIE, leaving oxide spacers on the sidewalls. Heavily n' doped (typically with As) polysilicon is then deposited on the substrate, patterned and etched...
forming polysilicon emitter (polyemitter) and collector contacts, as shown in Fig. 7-5c. The use of two LPCVD polysilicon layers explains why this process is referred to as the double-polysilicon process. Arsenic from the polysilicon is diffused into the substrate to form the n-emitter region, resists within the base in a self-aligned manner, as well as the n'-collector contact. Self-alignment refers to the fact that a separate lithography step is not required to form the n-emitter region. We cleverly make use of the oxide sidewall spacers to ensure that the n-emitter region lies within the intrinsic p-type base. This is critical because otherwise the emitter gets shorted to the collector; we also want a gap between the n-emitter and the p'-extrinsic base, because otherwise the emitter-base junction capacitance becomes too high. In the vertical direction, the difference between the emitter-base junction and the base-collector junction determines the base width. This is made very narrow in high gain, high speed BJTs.

Finally, an oxide layer is deposited by CVD, windows are etched in it corresponding to the emitter (E), base (B) and collector (C) contacts, and a suitable contact metal such as Al is sputter deposited to form the ohmic contact. The Al is patterned photolithographically using the interconnect mask, and etched using RIE. The many ICS that are made simultaneously on the wafer are then separated into individual dies by sawing, mounted on suitable packages, and the various contacts are wire bonded to the external leads of the package.

7.4 Solution of the Diffusion Equation in the Base Region

In later sections we shall consider the implications of imperfect injection efficiency, drift due to nonuniform doping in the base, structural effects such as different areas for the emitter and collector junctions, and capacitance and transit time effects in a-c operation.

7.4.1 Solution of the Diffusion Equation in the Base Region

Since the injected holes are assumed to flow from emitter to collector by diffusion, we can evaluate the currents crossing the two junctions by techniques used in Chapter 5. Neglecting recombination in the two depletion regions, the hole current entering the base at the emitter junction is the current \( I_E \), and the hole current leaving the base at the collector is \( I_C \). If we can solve for the distribution of excess holes in the base region, it is simple to evaluate the gradient of the distribution at the two ends of the base to find the currents. We shall consider the simplified geometry of Fig. 7-6, in which the base width is \( W_b \), between the two depletion regions, and the uniform cross-sectional area is \( A \). The excess hole concentration at the edge of the emitter depletion region \( \Delta p_E \) and the corresponding concentration on the collector side of the base \( \Delta p_C \) are found from Eq. (5-29):

\[
\Delta p_E = p_n e^{QW_{em}/kT} - 1 \quad \text{(7.8a)}
\]

\[
\Delta p_C = p_n e^{QW_{em}/kT} - 1 \quad \text{(7.8b)}
\]

If the emitter junction is strongly forward biased \( (V_{EB} \gg kT/q) \) and the collector junction is strongly reverse biased \( (V_{CB} \ll 0) \), these excess concentrations simplify to

\[
\Delta p_E = p_n e^{QW_{em}/kT} \quad \text{(7.9a)}
\]

\[
\Delta p_C = -p_n \quad \text{(7.9b)}
\]

We can solve for the excess hole concentration as a function of distance in the base \( \delta p(x) \) by using the proper boundary conditions in the diffusion equation, Eq. (4-34b):

![Figure 7-6](Image)

**Figure 7-6**

Simplified p-n-p transistor geometry used in the calculation.
\[
\frac{d^2 \delta p(x_n)}{dx_n^2} = \frac{\delta p(x_n)}{L_p^2}
\]  

(7-10)

The solution of this equation is
\[
\delta p(x_n) = C_1 e^{x_n/L_p} + C_2 e^{-x_n/L_p}
\]  

(7-11)

where \(L_p\) is the diffusion length of holes in the base region. Unlike the simple problem of injection into a long \(n\) region, we cannot eliminate one of the constants by assuming the excess holes disappear for large \(x_n\). In fact, since \(W_b < L_p\) in a properly designed transistor, most of the injected holes reach the collector at \(W_b\). The solution is very similar to that of the narrow base diode problem. In this case the appropriate boundary conditions are
\[
\delta p(x_n = 0) = C_1 + C_2 = \Delta p_E
\]  

(7-12a)

\[
\delta p(x_n = W_b) = C_1 e^{W_b/L_p} + C_2 e^{-W_b/L_p} = \Delta p_C
\]  

(7-12b)

Solving for the parameters \(C_1\) and \(C_2\) we obtain
\[
C_1 = \frac{\Delta p_C - \Delta p_E e^{-W_b/L_p}}{e^{W_b/L_p} - e^{-W_b/L_p}}
\]  

(7-13a)

\[
C_2 = \frac{\Delta p_E e^{W_b/L_p} - \Delta p_C}{e^{W_b/L_p} - e^{-W_b/L_p}}
\]  

(7-13b)

These parameters applied to Eq. (7-11) give the full expression for the excess hole distribution in the base region. For example, if we assume that the collector junction is strongly reverse biased [Eq. (7-9b)] and the equilibrium hole concentration \(p_h\) is negligible compared with the injected concentration \(\Delta p_E\), the excess hole distribution simplifies to
\[
\delta p(x_n) = \Delta p_E \frac{e^{W_b/L_p} - e^{-W_b/L_p}}{e^{W_b/L_p} - e^{-W_b/L_p}}
\]  

(for \(\Delta p_C = 0\))

(7-14)

The various terms in Eq. (7-14) are sketched in Fig. 7-7, and the corresponding excess hole distribution in the base region is demonstrated for a moderate value of \(W_b/L_p\). Note that \(\delta p(x_n)\) varies almost linearly between the emitter and collector junction depletion regions. As we shall see, the slight deviation from linearity of the distribution indicates the small value of \(I_b\) caused by recombination in the base region.

### 7.4.2 Evaluation of the Terminal Currents

Having solved for the excess hole distribution in the base region, we can evaluate the emitter and collector currents from the gradient of the hole concentration at each depletion region edge. From Eq. (4-22b) we have:
\[
I_p(x_n) = -q A D_p \frac{d\delta p(x_n)}{dx_n}
\]  

(7-15)

\[
\delta p(x_n) = M_1 \Delta p_E e^{-x_n/W_b} - M_2 \Delta p_C e^{x_n/W_b}
\]

where
\[
M_1 = \frac{e^{W_b/L_p}}{e^{W_b/L_p} - e^{-W_b/L_p}}
\]
\[
M_2 = \frac{e^{-W_b/L_p}}{e^{W_b/L_p} - e^{-W_b/L_p}}
\]

This expression evaluated at \(x_n = 0\) gives the hole component of the emitter current,
\[
I_{E_p} = I_p(x_n = 0) = qA D_p \frac{\Delta p_E}{L_p} (C_2 - C_1)
\]  

(7-16)

Similarly, if we neglect the electrons crossing from collector to base in the collector reverse saturation current, \(I_C\) is made up entirely of holes entering the collector depletion region from the base. Evaluating Eq. (7-15) at \(x_n = W_b\) we have the collector current
\[
I_C = I_p(x_n = W_b) = qA D_p \frac{\Delta p_E}{L_p} (C_1 e^{W_b/L_p} - C_2 e^{-W_b/L_p})
\]  

(7-17)

When the parameters \(C_1\) and \(C_2\) are substituted from Eqs. (7-13), the emitter and collector currents take a form that is most easily written in terms of hyperbolic functions:
\[
I_{E_p} = qA D_p \frac{\Delta p_E [e^{W_b/L_p} + e^{-W_b/L_p}] - 2\Delta p_C}{e^{W_b/L_p} - e^{-W_b/L_p}}
\]  

(7-18a)

\[
I_{E_p} = qA D_p \frac{\Delta p_E \cosh W_b/L_p - \Delta p_C \sinh W_b/L_p}{L_p}
\]

(7-18b)

\[
I_C = qA D_p \frac{\Delta p_E \sinh W_b/L_p - \Delta p_C \cosh W_b/L_p}{L_p}
\]

(7-18b)

Now we can obtain the value of \(I_b\) by a current summation, noting that the sum of the base and collector currents leaving the device must equal the emitter current entering. If \(I_b = I_{E_p}\) for \(\gamma = 1\),
\[ I_E = I_{IC} = qA \frac{D_p}{L_p} \left[ \left( \Delta p_E + \Delta p_C \right) \left( \tanh \frac{W_b}{L_p} - \text{csch} \frac{W_b}{L_p} \right) \right] \]

\[ I_B = qA \frac{D_p}{L_p} \left[ \left( \Delta p_E + \Delta p_C \right) \tanh \frac{W_b}{2L_p} \right] \tag{7-19} \]

By using the techniques of Chapter 5 we have evaluated the three terminal currents of the transistor in terms of the material parameters, the base width, and the excess concentrations \( \Delta p_E \) and \( \Delta p_C \). Furthermore, since these excess concentrations are related in a straightforward way to the emitter and collector junction bias voltages by Eq. (7-8), it should be simple to evaluate the transistor performance under various biasing conditions. It is important to note here that Eqs. (7-18) and (7-19) are not restricted to the case of the usual transistor biasing. For example, \( \Delta p_e \), may be \( -p_e \) for a strongly reverse-biased collector, or it may be a significant positive number if the collector is positively biased. The generality of these equations will be used in Section 7.5 in considering the application of transistors to switching circuits.

(a) Find the expression for the current \( I \) for the transistor connection shown if \( \gamma = 1 \).

(b) How does the current \( I \) divide between the base lead and the collector lead?

(a) Since \( V_{CB} = 0 \), Eq. (7-8b) gives \( \Delta p_C = 0 \). Thus from Eq. (7-18a),

\[ I_E = I = qA \frac{D_p}{L_p} \Delta p_E \tanh \frac{W_b}{L_p} \]

(b) \[ I_C = qA \frac{D_p}{L_p} \Delta p_E \text{csch} \frac{W_b}{L_p} \]

\[ I_B = qA \frac{D_p}{L_p} \Delta p_E \tanh \frac{W_b}{2L_p} \]

Series expansions of the hyperbolic functions are given in Table 7-1. For small values of \( W_b/L_p \), we can neglect terms above the first order of the argument. It is clear from this table and Eq. (7-20) that \( I_C \) is only slightly smaller than \( I_E \), as expected. The first-order approximation of \( \tanh y \) is simply \( y \), so that the base current is

\[ I_B = qA \frac{D_p}{L_p} \Delta p_E \frac{W_b}{2L_p} = \frac{qA W_b \Delta p_E}{2\tau_p} \tag{7-21} \]

where \( I_E \) and \( I_B \) are the components in the collector lead and base lead, respectively. Note that these results are analogous to those of Probs. 5.35 and 5.36 for the narrow base diode.

7.4.3 Approximations of the Terminal Currents

The general equations of the previous section can be simplified for the case of normal transistor biasing, and such simplification allows us to gain insight into the current flow. For example, if the collector is reverse biased, \( \Delta p_C = -p_e \) from Eq. (7-9b). Furthermore, if the equilibrium hole concentration \( p_e \) is small (Fig. 7-8a), we can neglect the terms involving \( \Delta p_C \). For \( \gamma = 1 \), the terminal currents reduce to those of Example 7-2:

\[ I_E = qA \frac{D_p}{L_p} \Delta p_E \tanh \frac{W_b}{L_p} \tag{7-20a} \]

\[ I_C = qA \frac{D_p}{L_p} \Delta p_E \text{csch} \frac{W_b}{L_p} \tag{7-20b} \]

\[ I_B = qA \frac{D_p}{L_p} \Delta p_E \tanh \frac{W_b}{2L_p} \tag{7-20c} \]

Figure 7-8
Approximate excess hole distributions in the base: (a) forward-biased emitter, reverse-biased collector; (b) triangular distribution for \( V_{CB} = 0 \) or for negligible \( p_e \).
Table 7-1 Expansion of several pertinent hyperbolic functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>sech y</td>
<td>1 - (\frac{y^2}{2} + \frac{5y^4}{24} + \ldots)</td>
</tr>
<tr>
<td>tanh y</td>
<td>(\frac{y}{3} + \frac{y^3}{45} + \ldots)</td>
</tr>
<tr>
<td>csch y</td>
<td>(\frac{y}{6} + \frac{7y^3}{360} + \ldots)</td>
</tr>
</tbody>
</table>

The same approximate expression for the base current is found from the difference in the first-order approximations to \(I_E\) and \(I_C\):

\[
I_B = I_E - I_C
\]

\[
= qA \frac{\alpha_p}{L_p} \Delta p_E \left[ \left( \frac{1}{W_b/L_p} + \frac{W_b/L_p}{3} \right) - \left( \frac{1}{W_b/L_p} - \frac{W_b/L_p}{6} \right) \right]
\]

\[
= \frac{qA \alpha_p W_b \Delta p_E}{2L_p} = \frac{q A W_b \Delta p_E}{2 \tau_p}
\]

(7-22)

This expression for \(I_B\) accounts for recombination in the base region. We must include injection into the emitter in many BJT devices, as discussed in Section 7.4.4.

If recombination in the base dominates the base current, \(I_B\) can be obtained from the charge control model, assuming an essentially straight-line hole distribution in the base (Fig. 7-8b). Since the hole distribution diagram appears as a triangle in this approximation, we have

\[
Q_p = \frac{1}{2} q A \Delta p_E W_b
\]

(7-23)

If we consider that this stored charge must be replaced every \(\tau_p\) seconds and relate the recombination rate to the rate at which electrons are supplied by the base current, \(I_B\) becomes

\[
I_B = \frac{Q_p}{\tau_p} = \frac{q A W_b \Delta p_E}{2 \tau_p}
\]

(7-24)

which is the same as that found in Eqs. (7-21) and (7-22).

Since we have neglected the collector saturation current and have assumed \(\gamma = 1\) in these approximations, the difference between \(I_E\) and \(I_C\) is accounted for by the requirements of recombination in the base. In Eq. (7-24) we have a clear demonstration that the base current is reduced for small \(W_b\) and large \(\tau_p\). We can increase \(\tau_p\) by using light doping in the base region, which of course also improves the emitter injection efficiency.

The straight-line approximation of the excess hole distribution (Fig. 7-8) is fairly accurate in calculating the base current. On the other hand, it does not give a valid picture of \(I_E\) and \(I_C\). If the distribution were perfectly straight, the slope would be the same at each end of the base region. This would imply zero base current, which is not the case. There must be some “droop” to the distribution, as in the more accurate curve of Fig. 7-7. This slight deviation from linearity gives a steeper slope at \(x_n = 0\) than at \(x_n = W_n\), and the value of \(I_E\) is larger than \(I_C\) by the amount \(I_B\). The reason we can use the straight-line approximation in the charge control calculation of base current is that the area under the hole distribution curve is essentially the same in the two cases.

7.4.4 Current Transfer Ratio

The value of \(I_C\) calculated thus far in this section is more properly designated \(I_{Eg}\), since we have assumed that \(\gamma = 1\) (the emitter current due entirely to hole injection). Actually, there is always some electron injection across the forward-biased emitter junction in a real transistor, and this effect is important in calculating the current transfer ratio. It is easy to show that the emitter injection efficiency of a p-n-p transistor can be written in terms of the emitter and base properties:

\[
\gamma = \left[ 1 + \frac{L_p^2 \alpha_p \mu_p^2}{L_n^2 \mu_n^2} \tan \left( \frac{W_b}{L_n} \right) \right]^{-1} = \left[ 1 + \frac{L_p^2 \alpha_p \mu_p^2}{L_n^2 \mu_n^2} \right]^{-1}
\]

(7-25)

In this equation we use superscripts to indicate which side of the emitter-base junction is referred to. For example, \(L_n^2\) is the hole diffusion length in the n-type base region and \(\mu_n^2\) is the electron mobility in the p-type emitter region. In a n-p-n the superscripts and subscripts would be changed along with the majority carrier symbols. Using Eq. (7-20a) for \(I_{Eg}\) and Eq. (7-20b) for \(I_C\), the base transport factor \(B\) is

\[
B = \frac{I_C}{I_{Eg}} = \frac{\text{sech} \left( W_b/L_p \right)}{\text{ctnh} \left( W_b/L_p \right)} = \text{sech} \left( \frac{W_b}{L_p} \right)
\]

(7-26)

and the current transfer ratio \(\alpha\) is the product of \(B\) and \(\gamma\) as in Eq. (7-3).

Assume that a p-n-p transistor is doped such that the emitter doping is ten times that in the base, the minority carrier mobility in the emitter is one-half that in the base, and the base width is one-tenth the minority carrier diffusion length. The carrier lifetimes are equal. Calculate \(\alpha\) and \(\beta\) for this transistor.

From Eqs. (7-25) and (7-26), we have

\[
\alpha = B \gamma = \left[ \cosh \frac{W_b}{L_p} + \frac{L_p^2 \alpha_p \mu_p^2}{L_n^2 \mu_n^2} \right]^{-1}
\]

Using the values given, and taking \(L_p^2/L_n^2 = \sqrt{2}(0.1)\) for equal lifetimes,

\[
\alpha = \left[ \cosh 0.1 + \sqrt{2}(0.1) \right]^{-1}
\]

\[
\approx 0.988
\]