7. **The Bipolar Junction Transistor**

The operation of bipolar junction transistors relies on the coupling between two p-n diodes which share a common n or p-type region called the base. This coupling is due to minority carriers injected across one of the p-n junctions, which then travel through this common region and contribute to the current in the other p-n diode. A typical mode of operation, called the active forward mode, is obtained when the p-n diode which injects minority carriers into the middle region is forward biased. The other diode is reversed biased so that the current through that diode is primarily due to the minority carriers which were injected from the other diode. The result is that the first p-n diode controls the current through both, while the current supplied through the middle contact is small. This causes both voltage and current gain in the device when applying a signal to the base contact.

### 7.1 Emitter efficiency and current gain

The structure is shown in the figure below.

![Fig.7.1 Structure of a bipolar transistor.](image)

It consists of three regions labeled emitter, base and collector. The collector and emitter can be both n-type or p-type, while the base is doped with the opposite type impurity. An
NPN structure was chosen for the discussion below. The figure also shows the currents which flow under active forward bias conditions including the currents due to recombination in the quasi-neutral base region.

The emitter efficiency, $\gamma_E$, and current gain, $\beta$, are given by:

$$\gamma_E = \frac{J_n}{J_n + J_p} = \frac{1}{1 + \frac{J_p}{J_n}} \quad [7.1.1]$$

$$\beta = \frac{J_n - J_{rec}}{J_p + J_{rec}} \quad \text{or} \quad \beta = \frac{J_n}{J_p} \quad \text{for} \quad J_{rec} = 0 \quad [7.1.2]$$

Both depend on $J_n/J_p$, which can be obtained from equations [4.5.5] and [4.5.6], yielding:

$$\frac{J_n}{J_p} = \frac{D_n n_{ip}^2 L_E N_d}{D_p n_{in}^2 W_B' N_a} \quad [7.1.3]$$

where we assume that the emitter diffusion length, $L_E$, is much smaller than the width of the quasi-neutral region in the emitter, $W_E'$, and that the base diffusion length, $L_B$, is much larger than the width of the quasi-neutral region in the emitter, $W_B'$. This expression for the current ratio can be used to illustrate the difference between homojunction and the heterojunction transistors: The intrinsic carrier concentration is constant throughout a homojunction device so that the current ratio is dominated by the ratio of the emitter doping concentration to the base doping concentration. The intrinsic carrier concentration in a heterojunction devices depends exponentially on the energy bandgap, so that the current ratio can be made very large by using a wide bandgap material for the emitter, even if the emitter doping concentration is lower than that of the base. The current gain in the absence of base recombination is given by:

$$\beta = \frac{D_n}{D_p} \frac{N_{cp}^2 N_{vp}}{N_{cn} N_{vn}} \frac{L_E}{W_B'} \frac{N_E}{N_B} \exp\left(-\frac{\Delta E_g}{kT}\right) \quad [7.1.4]$$
Where $\Delta E_g$ is the difference in energy bandgap of the emitter material versus that of the base material, and $N_E$ and $N_B$ are the doping concentrations of the emitter and base.

The use of a heterojunction is the solution to the traditional trade-off in a silicon homojunction bipolar transistor: to improve the emitter efficiency one has to increase the emitter doping concentration, but this also decreases the energy bandgap in the emitter which increases the intrinsic concentration and in turn reduces the emitter efficiency. The bandgap reduction in silicon is given by:\[ \Delta E_g = \frac{3}{16\pi\varepsilon_s} \left( \frac{q^2 N_E}{\varepsilon_s kT} \right)^{1/2} = 225 \left( \frac{N_E}{10^{18}\text{ cm}^{-3}} \right)^{1/2} \text{ at } T = 300 \text{ K} \] \[ 7.1.5 \]

This problem can be avoided by using a wider bandgap amorphous-silicon emitter or a narrower bandgap silicon-germanium base. Compound semiconductor material systems provide a much larger energy bandgap range and the freedom to chose different materials throughout the device which greatly enhances the ability to optimize the device performance.

Another parameter which influences the current gain is the base transport factor, $\alpha_{tr}$, which is the fraction of the injected current which makes it through the base region. Whereas the current in most silicon bipolar transistors is limited by the emitter efficiencies, one finds the current gain of heterojunction bipolar transistors to be limited by the carrier recombination in the base. This is especially the case for base regions consisting of a direct bandgap material, such as GaAs, which tends to have a short minority carrier lifetime. The base transport factor, $\alpha_{tr}$, is given by.

---

\[ \alpha_{tr} = \frac{J_n - J_{rec}}{J_n} \quad [7.1.6] \]

For a uniformly doped base (using equation [4.1.19a]) the base transport factor becomes:

\[ \alpha_{tr} = \frac{1}{\cosh\left(\frac{W_B}{L_n}\right)} = 1 - \frac{W_B^2}{2 L_n^2} = 1 - \frac{\tau_{tr}}{\tau_0} \quad [7.1.7] \]

where the transit time, \( \tau_{tr} \), is given by \( \tau_{tr} = \frac{W_B^2}{2 D_n} \) and \( \tau_0 \) is the carrier lifetime.

This problem does not exist when using the GaInAs/InP lattice matched material system because of the large minority carrier lifetime in GaInAs. Heterojunction bipolar transistors made using this material system therefore have much larger current gains. The highest reported current gain (10,000,000) has been obtained for a GaN/SiC heterojunction bipolar transistor\(^2\). Such a device combines the large emitter efficiency of a heterojunction device with an indirect material as base to provide a large base transport factor.

**7.2 Transit time of a bipolar transistor**

The transit time of a bipolar transistor is directly related to the maximum unity current gain frequency and has become an important measure for the high-speed capability of the device. It can be calculated using the time moment method by summing the emitter charging time, the base transit time and the collector transit time (including the collector charging time)

\[ \tau_{\text{total}} = \tau_E + \tau_{\text{tr}} + \tau_C \]  

(7.2.1)

with

\[ \tau_E = r_e (C_e + C_c + C_p) \]  

(7.2.2)

where \( r_e \) is the emitter resistance, \( C_e \) is the emitter capacitance, \( C_c \) is the collector capacitance and \( C_p \) is the parasitic capacitance between the base and the collector. Under forward bias conditions one finds:

\[ r_e = \frac{V_t}{I_E} \]  

(7.2.3)

The base transit time generalized to include the effect of a built-in field, \( \varepsilon_{\text{bi}} \), is given by:

\[ \tau_{\text{tr}} = \frac{W_B^2}{\eta D_p} \]  

(7.2.4)

\[ \text{with } \eta = 2 \left( 1 + \left( \frac{\varepsilon_{\text{bi}}}{\varepsilon_0} \right)^{3/2} \right) \]  

with \( \varepsilon_0 = \frac{V_t}{W_B} \)

The built-in electric field depends on the doping profile as well as the bandgap grading in the base layer. A discussion of the built-in field in a graded bandgap p-n diode can be found in section 4.5.6, yielding for a linearly graded bandgap:

\[ \varepsilon_{\text{bi}} = -\frac{1}{q} \frac{dE_{\text{bi}}}{dx} \]  

(7.2.5)

For a material with constant bandgap and exponential doping profile one finds

\[ \varepsilon_{\text{bi}} = \frac{V_t}{W_B} \ln \left( \frac{N_{\text{max}}}{N_{\text{min}}} \right) \]  

(7.2.6)

The collector transit time (including the collector charging time) is given by

---

\[ \tau_c = \frac{x_{nC} + x_{pB}}{2 v_s} + r_c (C_c + C_p) \] \[7.2.7\]

where \( x_{nC} \) and \( x_{pB} \) are the depletion layer widths of the base-collector junction and \( v_s \) is the saturation velocity of the carriers. The transit frequency, \( f_T \), becomes:

\[ \frac{1}{f_T} = 2\pi \tau_{\text{total}} = 2\pi \left[ \frac{V_t}{I_E} (C_c + C_c + C_p) + \frac{W_B^2}{\eta \mathcal{D}_p} + \frac{x_{nC} + x_{pB}}{2 v_s} + r_c (C_c + C_p) \right] \] \[7.2.8\]

### 7.3 Microwave equivalent circuit

The microwave equivalent circuit is shown in its simplest form in the figure below.

![Microwave equivalent circuit of a bipolar transistor](image)

From this equivalent circuit one finds that the transit frequency described in [7.2.8], corresponds to the unity current gain frequency as measured with a shorted collector in a common emitter configuration.\(^4\)

The maximum oscillation frequency (power gain = 1) is given by

\(^4\)It should be noted that the equivalent circuit does not include the effects of carrier drift through the base and the base-collector depletion region.
\[ f_{\text{max}} = \frac{1}{2} \left( \frac{f_T}{2\pi r_b (C_c + C_p)} \right)^{1/2} \]  

[7.2.1]

### 7.4 Bipolar transistor design

For optimal circuit performance one has to design a bipolar transistor so that it has a large enough DC current gain, a large transit frequency and a large maximum oscillation frequency. To satisfy the above requirements one needs: for large gain a wide bandgap emitter, a highly doped emitter, and a thin base. For large transit frequency \( f_T \), a low doped emitter and base, a large emitter current, a thin base and a high doped collector. For large \( f_{\text{max}} \), a thick base and a low doped collector.