Chapter 8
Homojunction and Heterojunction Bipolar Transistors

Bipolar junction transistors (BJTs) and hereojunction bipolar transistors (HBTs) represent two of the most interesting devices that have been introduced over the past five decades as each of the two devices account for a noble prize in Physics (Brattain and Shockley, 1956 and Kroemer 2000). Although bipolar transistors' transfer characteristics are exponential, they are extremely linear when fed with an input current signal. As a result, they are often preferred in linear power amplifiers. Their main advantage of CMOS and other unipolar transistors such as MESFETs and HEMTs is that they do not require sub-micron technology to operate in multi-GHz frequency range.

In this chapter, we first introduce basic operation of bipolar junction transistors. Then we discuss the physics of uniformly doped BJTs and briefly discuss high current effects in these devices. We then discuss the promises of HBTs and why they are so much advantageous over their homojunction counterparts. We then discuss different material systems that are used to build HBTs and noise and power performance of HBTs. We conclude this chapter by presenting advantages of HBTs over MESFETs, HEMTS and CMOS transistors.

A Simplified Picture of BJT Operation

BJT uses a small input current (Base Current) to control a larger output current (Collector current). BJT comes in both NPN and PNP configurations each with three terminals known as base, emitter and collector. Direction of the arrows in the symbols show the direction of positive current.

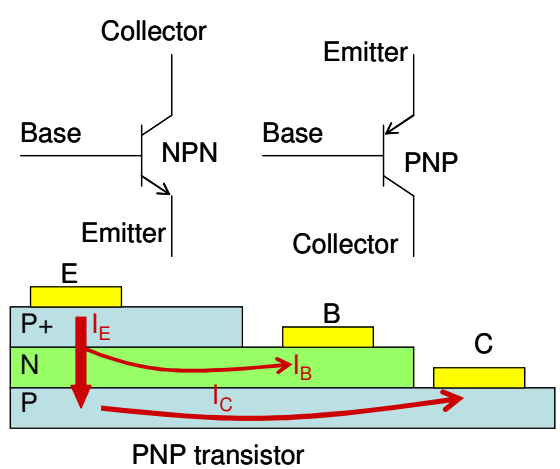
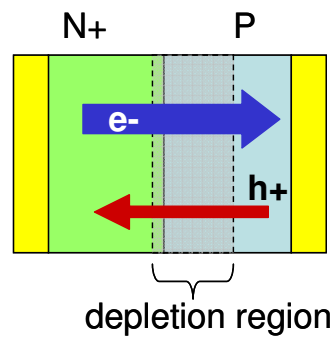


Fig. 8.1. NPN and PNP transistor Symbols. A simple PNP transistor cross section.

Under normal operation condition, the base-emitter junction is forward biased while the collector-base junction is reversed biased. Note that if you have two separate diodes one forward and one

reverse biased back to back, you will not get any current through. In BJTs the two closely placed PN junctions have current because the reversed biased junction steals the current of the forward biased junction. The following explains what really happens. Imagine that you have an N^+P junction. When you forward bias this diode, current that is formed is mainly due to electrons being injected from the N+ region to P region. Only small part of the current is due to holes being injected to N+ region. In 1954, Brattain and Shockley decided to steal some of the electron currents and send it to a third terminal using a structure similar to the one shown in Fig. 8.2. Majority of the electron current (I_{EN}) in the forward biased N^+P junction would make it to the third terminal known as collector (I_{CN}). The electrons lost in the base region (I_{BN}) would recombine in the base with majority carriers (holes) due to less than ideal base transport factor (BTF). The small number of holes injected from base to emitter (I_{EP}) is responsible for the other part of the base current contributing to an emitter injection efficiency (EIE) of less than 100%.



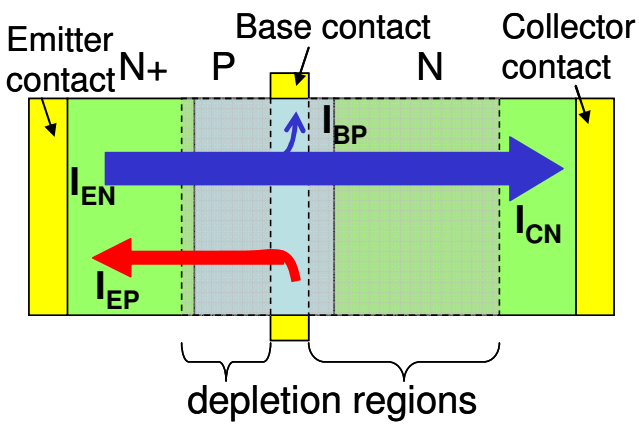


Fig. 8.2. Various currents in an NPN transistor

$$EIE = \frac{I_E - I_{EP}}{I_E} = \frac{I_{EN}}{I_E}$$
$$BTF = \frac{I_{CN}}{I_{EN}} \tag{8.1}$$

Common collector current gain α is defined as

$$\alpha = EIE \times BTF = \frac{I_{CN}}{I_E} \tag{8.2}$$

Common emitter current gain β is defined as

$$\beta = \frac{I_{CN}}{I_{EP} + I_{BN}} = \frac{I_C}{I_B} = \frac{1}{\frac{I_E}{I_C} - 1} = \frac{\alpha}{1-\alpha} \tag{8.3}$$

Since α is usually close to 1, β can be very large. An interesting property of bipolar transistors is that to the first degree, $\beta = \frac{I_C}{I_B}$ is constant which means current gain in bipolar transistors is very linear, a clear advantage over their MOS counterparts.

During normal operation of an NPN Si bipolar junction transistor, V_{BE} of 0.6V to 0.75V is applied. A positive V_{CB} value keeps the Base-Collector diode in reversed bias mode. Therefore, V_{CE} of V_{BE} + V_{CB} = 0.7V or higher is needed. It turns out that for a well designed bipolar transistor, if the base collector diode is even forward biased but not turned on, the transistor works properly. Therefore for V_{CB} = -0.4V to -0.5V, we can calculate the minimum V_{CE} of 0.2V to 0.3V. Above this V_{CE} value known as $V_{CE\text{-}sat}$, the transistor works properly (constant β). Below this $V_{CE\text{-}sat}$, there will be charges accumulated in the base region which results in low base transport factor as well as low emitter injection efficiency. For normal transistor operation (active mode, $V_{CE} > V_{CE\text{-}sat}$), one can write:

$$\begin{aligned} I_B &= I_S\left(e^{V_{BE}/nV_{th}} - 1\right) \\ I_C &= \beta I_B = \beta I_S\left(e^{V_{BE}/nV_{th}} - 1\right) \end{aligned} \tag{8.4}$$

where I_S is the reverse saturation current of base emitter diode. When $V_{CE} < V_{CE\text{-}sat}$ and V_{BE} is forward biased, the transistor is in saturation mode as shown in Fig. 8.3.

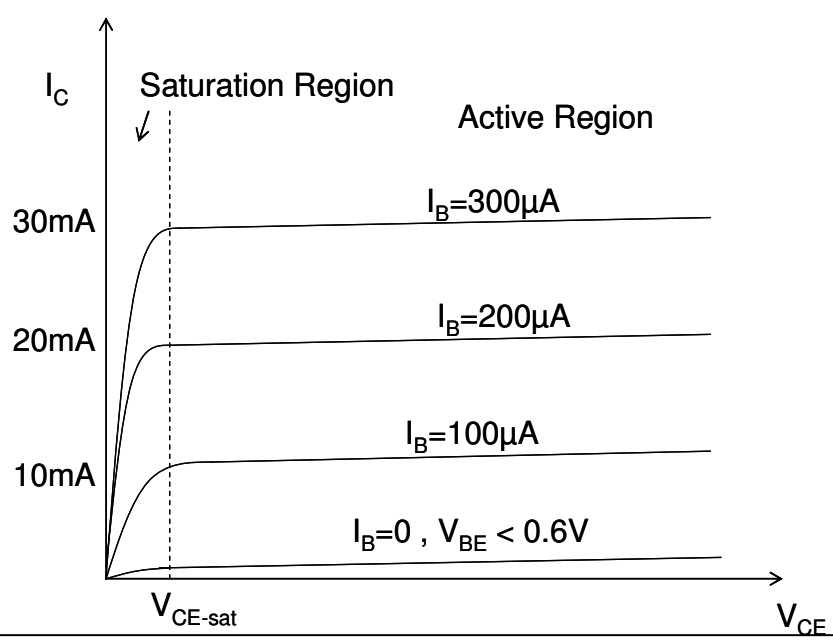


Fig. 8.3. V_{CE}-I_C characteristics of a bipolar transistor indicating the active region (constant current gain) and saturation region.

Bipolar transistors suffer from output conductance phenomena known as Early effect which is due to the projection of the base-collector space charge region into the neutral base region (see Fig. 8.4).

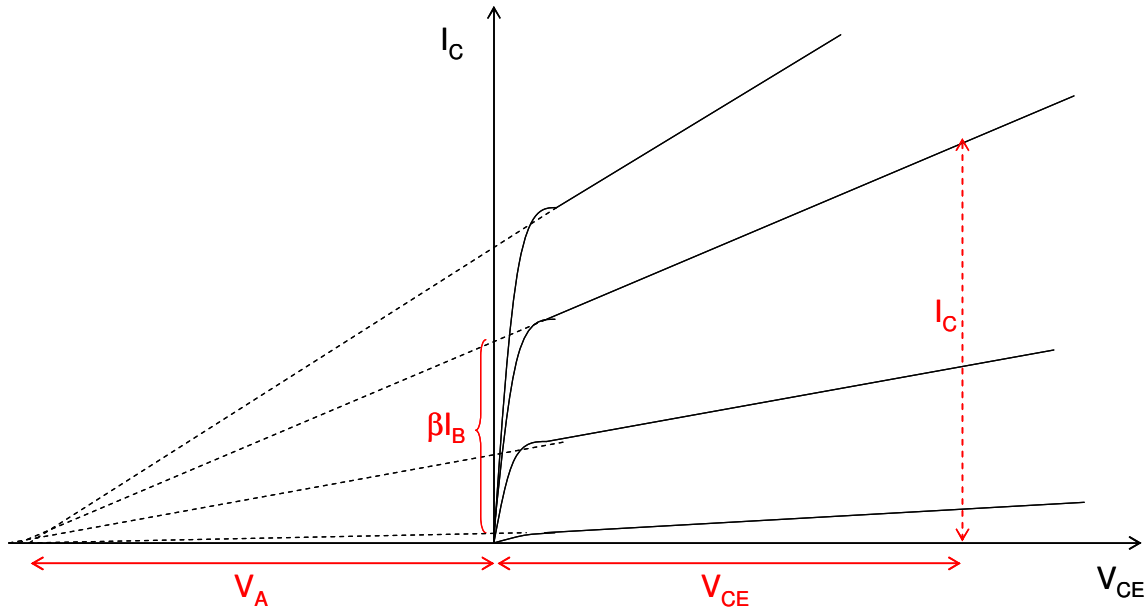


Fig. 8.4. V_{CE}-I_C characteristics of a bipolar transistor including the Early Effect

Usually the Early voltage V_A is larger than V_{CE}, so the effect is a second order effect. The collector current is modified according to

$$I_C = \beta I_B \left(1 + \frac{V_{CE}}{|V_A|}\right) \tag{8.5}$$

A simple small signal model for a BJT is shown Fig. 8.5. Some of these parameters are easy to extract, without knowing much about the physics of the transistor. Recall the DC current equations for base current of equation (8.4) and collector current of equation (8.5). The transconductance of the device defined as the derivative of the collector current with respect to the base-emitter voltage can be found from equation (8.5) according to:

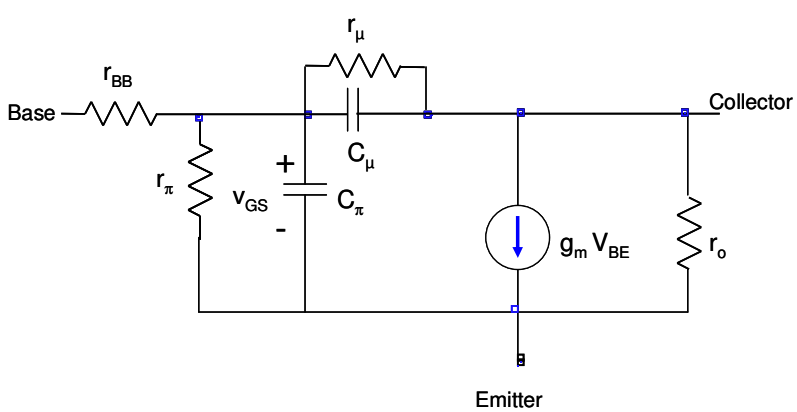


Fig. 8.5. A simplified Small-signal model for bipolar transistors.

$$g_m = \frac{dI_C}{dV_{BE}} = \frac{\beta I_S}{nV_{th}} e^{\frac{V_{BE}}{nV_{th}}} \left(1 + \frac{V_{CE}}{|V_A|}\right) = \frac{I_C}{nV_{th}} \tag{8.6}$$

The output conductance is calculated according to

$$r_o = \frac{1}{dI_C / dV_{CE}} = \frac{1}{\frac{1}{|V_A|}\beta I_B} = \frac{|V_A|}{\beta I_B} \tag{8.7}$$

The incremental base resistance is calculated according to:

$$r_\pi = \frac{1}{dI_B / dV_{BE}} = \frac{1}{\frac{I_S}{nV_{th}} e^{\frac{V_{BE}}{nV_{th}}}} = \frac{nV_{th}}{I_B} \tag{8.8}$$

<u>Revisiting the Physics of Semiconductors</u>

In this section, we briefly discuss some of the important equations that come handy when dealing with BJTs. We start with Fermi-dirac distribution function:

$$F(E) = \frac{1}{1 + \exp[(E - E_F)/kT]} \tag{8.9}$$

where E is the energy of the electrons and E_F is the fermi level. The concentration of electrons in conduction band follows the fermi-Dirac distribution when the device is working close to the equilibrium condtion (no or small external force). Therefore electron concentration is calculated according to:

$$n = N_C \exp\left(\frac{E_F - E_C}{kT}\right) \tag{8.10}$$

where E_C is the bottom of the conduction band and N_C is the effective denisty of states for electrons given by

$$N_C = \frac{4\sqrt{2}}{h^3}\left(\pi m_e^* kT\right)^{3/2} \tag{8.11}$$

where h is the plank constant and m_e^* is the electron effective mass in the material. One can calculate the concentrations of holes and effective density of states for holes according to

$$\begin{aligned} p &= N_V \exp\left(\frac{E_V - E_F}{kT}\right) \\ N_V &= \frac{4\sqrt{2}}{h^3}\left(\pi m_h^* kT\right)^{3/2} \end{aligned} \tag{8.12}$$

where m_h^* is the hole effective mass in the material. Therefore the product of n and p in any material at equilibrium is given by

$$np = N_C N_V \exp\left(-\frac{E_g}{kT}\right) = n_i^2 \tag{8.13}$$

where E_g is the bandgap of the material and n_i is the intrinsic carrier concentration the heavily depends on the bandgap and temperature. From the above equations, we can find the Fermi level from the concentration of electrons and holes in the material according to

$$E_F = \frac{E_C + E_V}{2} + \frac{kT}{2}\ln\left(\frac{N_V}{N_C}\right) + \frac{kT}{2}\ln\left(\frac{n}{p}\right) \tag{8.14}$$

As N_V is usually larger than N_C (look at their dependence on effective masses), the fermi level of intrinisic semiconductor ($n=p=n_i$) is slightly above the bandgap.

When a semiconductor is doped with impurity atoms, it is important to know how many of those atoms are ionized (give up their electrons or accept an extra electron). The concentration of the ionized donors N_D^+ can be found from

$$N_D^+ = N_D\left[1 - \frac{1}{1 + \frac{1}{2}\exp\left[(E_D - E_F)/kT\right]}\right] \tag{8.15}$$

where E_D is the energy level of the donor atoms and typically around 10meV below the conduction band. The concentration of the ionized acceptors is given by

$$N_A^+ = \frac{N_A}{1 + 4\exp\left[(E_A - E_F)/kT\right]} \tag{8.16}$$

where E_A is the energy level of the acceptor atoms and typically around 10meV above the valence band. For typical doping levels and temperature encountered in bipolar transistors, it is customary to assume 100 percent ionization of donor and acceptor atoms.

Current flow in semiconductors is due to drift and diffusion of electrons and holes. The equations for such components are given by

$$\begin{aligned} J_{e-drift} &= qn\mu_n \mathrm{E}_{field} \\ J_{h-drift} &= qp\mu_p \mathrm{E}_{field} \\ J_{e-diffusion} &= qD_n \frac{dn}{dx} \\ J_{h-diffusion} &= -qD_p \frac{dp}{dx} \end{aligned} \tag{8.17}$$

where Dn and Dp are the electron and hole diffusion coefficents, respectively. For non-degenerate semiconductors Einstein relationship holds:

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{q} = V_{th} \tag{8.18}$$

During the development of bipolar transistors, researchers realized that heavily doping of a semiconductor is not as effective as the theory predicts. As a result, the current gain of a homojunction bipolar transistor is less than the expected value. The reason for the difference is attributed to the bandgap narrowing (BGN) effect. Therefore, the intrinsic carrier concentration is modified according to its effective value n_{ie}

$$n_{ie}^2 = n_i^2 \exp\left(\frac{\Delta E_g}{kT}\right) \tag{8.19}$$

where ΔE_g is the effective bandgap narrowing in N material given by

$$\Delta E_g = 9meV\left\{\ln\left(\frac{N}{10^{17}}\right) + \sqrt{\left[\ln\left(\frac{N}{10^{17}}\right)\right]^2 + 0.5}\right\} \tag{8.20}$$

The following two equations provide simplified formula for bandgap narrowing effect for both P- and N-doped materials

$$\begin{aligned} \Delta E_g &= 17meV \ln\left(\frac{N_D}{10^{17}}\right) \qquad N-doped \\ \Delta E_g &= 2\times10^{-11}\sqrt{N_A} \qquad P-doped \end{aligned} \tag{8.21}$$

BJT Basic Operation

In this section we assume that a vertical NPN transistor has constant doping profiles in emitter, base and collector regions as shown in Fig. 8.6. Same argument can be applied to PNP transistors. Fig. 8.6 also shows thicknesses of different regions of the device. First, we start from the collector region. We assume that the BC junction is reversed biased and the electric field in the space charge layer (W_{scl}) is

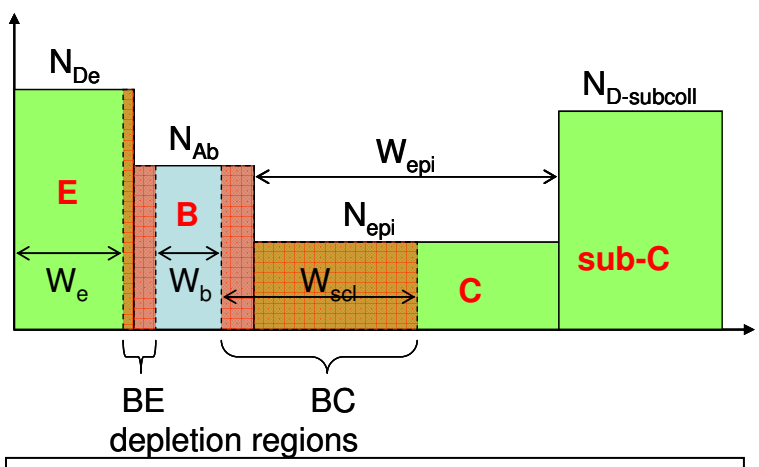


Fig. 8.6. Basic BJT structure

high enough such that the electrons achieve their drift velocity. The current due to electrons in the collector (I_{nc}) can be written as:

$$I_{nc} = qA_E v_d n_c \tag{8.22}$$

where A_E is the emitter area, v_d is the electron drift velocity and n_c is thefinite value of electron density within the space charge layer. Assuming no recombination in the base region, the electron current injected from the emitter to the base which is merely due to diffusion can be written as

$$I_{nc} = \frac{qA_E D_n (n_b(0) - n_c)}{W_b} \approx \frac{qA_E D_n n_b(0)}{W_b} = \frac{qA_E D_n n_{b0} e^{V_{BE}/V_{th}}}{W_b} \tag{8.23}$$

The hole diffusion current I_{pe} due to injected holes from the base to the emitter is

$$I_{pe} = qA_E D_p \frac{(p_{ne}(0) - p_{ne0})}{W_e} = qA_E D_p p_{ne0} \frac{\exp(V_{BE}/V_{th}) - 1}{W_e} \approx \frac{qA_E D_p p_{ne0} e^{V_{BE}/V_{th}}}{W_e} \tag{8.24}$$

Therefore, the current gain β can be calculated according to

$$\beta = \frac{I_C}{I_B} = \frac{I_{nc}}{I_{pe}} = \frac{W_e D_n n_{b0}}{W_b D_p p_{e0}} = \frac{W_e D_n N_{De}}{W_b D_p N_{Ab}} \tag{8.25}$$

If the base recombination is considered, the above equation is modified according to

$$\frac{1}{\beta} = \frac{W_b D_p N_{Ab}}{W_e D_n N_{De}} + \frac{t_{bb}}{\tau_B} \tag{8.26}$$

where τ_B is the recombination lifetime in the base region and $t_{bb} = W_b^2 / 2D_n$ is the base transit time.

Let us now derive the rest of small signal parameters of the BJT model shown in Fig. 8.5. The capacitances are calculated according to

$$C_\mu = C_{jc} = \frac{C_{jc0}}{\left(1 - V_{BC}/V_{bic}\right)^{\gamma_C}}$$
$$C_\pi = C_{je} + C_{diff} = \frac{C_{je0}}{\left(1 - V_{BE}/V_{bie}\right)^{\gamma_E}} + g_m t_{bb} \tag{8.27}$$

is the where γ_C and γ_E are the exponent with a value of ½ for abrupt junction and1/3 for graded junction , C_{jc0} and C_{je0} are the collector-base and emitter base junction capacittances at zero vol bias, repectively and V_{bic} and V_{bie} are built-in voltages of the two junctions.

$C_{diff} = dQ_b / dV_{BE}$ represents the diffusion capacitance due to the modulation of charges inside the base region with base-emitter voltage.

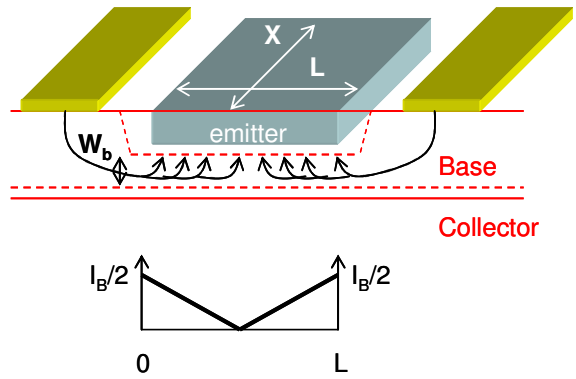


Fig. 8.7. Calculation of the base spreading resistance due to triangular shape base current in the device.

Next, we will calculate the base resistance r_{bb} due to the way the current flows from two side base contacts to the emitter region. Note that this calculation does not include semiconductor metal contact resistance which should be added if significant. The value of base resistance r_{bb} can be calculated from Fig. 8.7 as

$$r_{bb} = \frac{1}{12}\frac{L/X}{q\mu_p N_{Ab} W_b} \tag{8.28}$$

where X is the width of the emitter finger and L is its length.

We will assume r_μ is large for RF BJT transistors and heterojunction bipolar transistors.

Cut-off frequency and Maximum Oscillation Frequency

In this section we calculate both cut-off frequency f_T and maximum oscillation frequency f_{max} of a BJT transistor. A simple first-order calculation for cut-off frequency yields:

$$\frac{1}{2\pi f_T} = \frac{C_\pi + C_\mu}{g_m} = \frac{V_{th}}{I_C}\left(C_{je} + C_{jc}\right) + t_{bb} \tag{8.29}$$

A more rigorous analysis that considers the delay in collector space charge layer as well as the effect of finite access resistances results in

$$\frac{1}{2\pi f_T} = t_E + t_{bb} + t_{scl} + t_C = \frac{C_{je}}{g_m} + \frac{W_b^2}{2D_n} + \frac{W_{scl}}{2v_d} + C_{jc}\left(R_E + R_C + \frac{1}{g_m}\right) \tag{8.30}$$

where R_E and R_C and emitter and collector access resistance, respectively and W_{scl} is the thickness of the collector space charge region in which the carriers are drifting at a saturated velocity of v_d.

For the calculation of maximum oscillation frequency we stick to the simple relationship of

$$f_{\max} = \sqrt{\frac{f_T}{8\pi C_\mu r_{bb}}} \tag{8.31}$$

Note that f_{max} is roughly proportional to $\sqrt{\mu_n \mu_p}$ due to its dependence on f_T and r_{bb}. One can plot f_T, f_{max} and β as a function of collector current. Such a plot is shown in Fig. 8.8. Note that according to equation 8.29, the cut-off frequency has a maximum ideal value of $\frac{1}{2\pi t_{bb}}$, although high current effects and device parasitics will result in the degradation of cut-off frequency at higher I_C values. For bipolar transistors, it is possible to find a range of collector currents that result in optimum f_T, f_{max} and β.

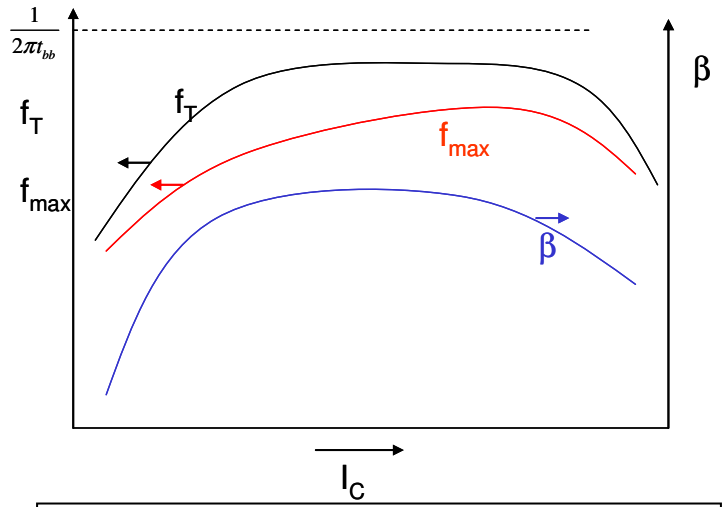


Fig. 8.8. Variation of f_T, f_{max} and β with collector current indicating high current effect around high I_C and large recombination dominated by surface velocity recombination at low I_C.

Large-Signal Modeling

In this section we briefly discuss the fully analytical Ebers-Moll large-signal model of BJTs. More advanced models such as VBIC and MEXTRAM provide more accuracy, but include some fitting parameters. Ebers-Moll model can be simply extracted from current equations. For instance the two components of collector current following equation 8.23 can be written as

$$I_{nc} = \frac{qA_E D_n (n_b(0) - n_c)}{W_b} = \frac{qA_E D_n n_i^2}{W_b N_{Ab}} \left[e^{V_{BE}/V_{th}} - e^{V_{BC}/V_{th}} \right]$$

$$I_{pc} = \frac{qA_E p_{nc0} W_{epi}}{\tau_{epi}} \left[e^{V_{BC}/V_{th}} - 1 \right] \tag{8.32}$$

$$I_C = I_{nc} - I_{pc}$$

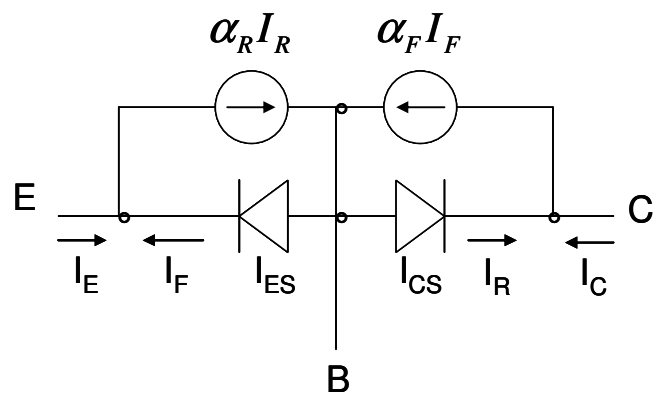
With BC junction properly reversed biased, one can write the following equations for collector and Emitter currents:

$$I_E = -I_{ES}\left[e^{V_{BE}/V_{th}} - 1\right] - \alpha_R I_{CS}$$
$$I_C = \alpha_F I_{ES}\left[e^{V_{BE}/V_{th}} - 1\right] + I_{CS} \tag{8.33}$$

where

$$I_{ES} = \frac{qA_E D_n n_{b0}}{W_b} + \frac{qA_E D_p p_{e0}}{W_e}$$
$$I_{CS} = \frac{qA_E D_n n_{b0}}{W_b} + \frac{qA_E p_{nc0} W_{epi}}{\tau_{epi}}$$
$$\alpha_F = \frac{D_n n_{b0}/W_b}{D_n n_{b0}/W_b + D_p p_{e0}/W_e} \tag{8.34}$$
$$\alpha_R = \frac{D_n n_{b0}/W_b}{D_n n_{b0}/W_b + p_{nc0} W_{epi}/\tau_{epi}}$$

Fig. 8.9 demonstrate Ebers-Moll large-signal model.



$$I_F = I_{ES}\left(e^{V_{BE}/V_{th}} - 1\right)$$
$$I_R = I_{CS}\left(e^{V_{BC}/V_{th}} - 1\right)$$

Fig. 8.9. Ebers-Moll large-signal model for bipolar transistors.

Bandgap Narrowing, Poly-Si Emitter and High Current Effects in BJTs

Early on, researchers found that the quest to improve transistor gain by increasing the doping in the emitter region is impeded by the bandgap narrowing effect in the emitter. In order to explain gain stagnation despite increasing emitter doping, the simple model of gain derived in equation (8.25) is modified with the bandgap narrowing effect modeled by equations (8.20) and (8.21). Therefore

$$\beta = \frac{I_C}{I_B} = \frac{W_e D_n N_{De}}{W_b D_p N_{Ab}} e^{\Delta E_g/kT} \tag{8.35}$$

where ΔE_g is the bandgap difference between emitter and base regions and therefore is negative of the value calculated by equations (8.20) and (8.21) when emitter is heavily doped but base is lightly doped. Note that the above equation points out to the usefulness of using a heterojunction at the emitter base junction with wider emitter bandgap than that of base.

Let us also point out to poly-Si BJTs where the emitter metal contact is replaced by poly-Si material with finite surface recombination velocity followed by emitter metal. In this case the density of holes at the semiconductor/poly-Si interface does not reach the equilibrium of p_{ne0}, resulting in smaller hole diffusion current in the emitter. This reduces the base current and helps increasing the current gain which can be traded off with speed.

Before we consider HBTs, let us point out to a few high current effects that are typically dominating the performance of bipolar transistors, but are pushed back or eliminated in HBTs. The first mechanism is called Emitter Current Crowding which causes an increase in the base access resistance and an uneven current density across the emitter area. Fig. 8.10 shows how the finite drop of base emitter voltage across the base spreading resistance results in uneven current across the emitter finger length. The center of emitter finger is essentially less "on" than its edges as the current increases (blue curve vs. black curve in Fig. 8.10) due to extra drop of base emitter voltage across the base resistance. As the current crowding takes place, base current has shorter distance to travel as it does not need to reach the center of emitter finger, resulting in smaller base resistance under high current injection modeled with;

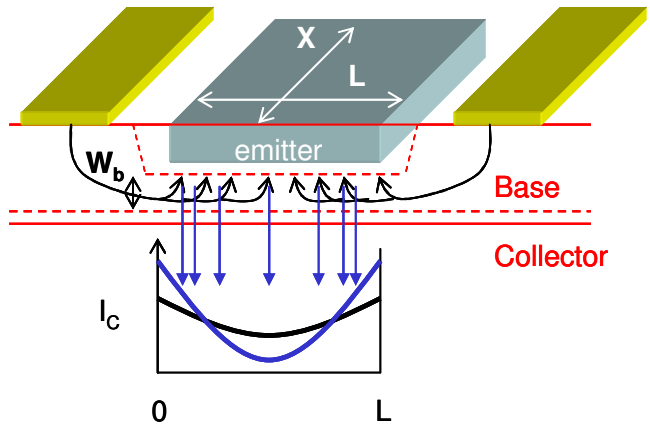


Fig. 8.10. Emitter Current Crowding due to r_{bb} and its effect on non-uniform collector current across the length of the emitter finger.

$$\frac{r_{bb}(I)}{r_{bb}} = \frac{1}{1+\frac{I_B r_{bb}}{2V_{th}}} \tag{8.36}$$

where $r_{bb}(I)$ is the base resistance under high level injection. As we shall see later, emitter current crowding is of minor importance in HBTs as base is highly doped resulting in insignificant base emitter voltage drop across the base spreading resistance.

Another high current effect known as Kirk effect of Base push out happens as the high number of carriers injected from base-emitter junction reach the collector space charge region. As the

carriers have limited drift velocity, increasing the current increase the number of carriers in the depletion region according to $n_C = I_C/(qA_E v_d)$. The increase can screen out the background doping density of ionized dopants in the collector region (donor atoms in the collector doped at the level of N_{epi}) resulting in the modification of the electric field profile at the collector base junction. This can decrease the length of the space charge region W_{scl} and increase the base width W_b at high collector current injection. The onset of Kirk effect is at current densities that result in carrier density inside the space charge region given by

$$n_{C-onset} = N_{epi} + \frac{2\varepsilon(V_{CB} + V_{bic}}{qW_{scl}^2} \tag{8.37}$$

Again by using HBTs with materials presenting velocity overshoot effect, the onset of Kirk effect can be pushed to much higher current densities as the density of carriers in the space charge region is inversely proportional to the velocity of the carriers in this region.

Other high current effects observed in bipolar transistors are lateral base widening and quasi saturation are of minor interest here. As we shall see next, HBTs by their nature are more suitable for higher current densities and higher speed resulting in their overwhelming implementation as high power and high speed circuits.

Heterojunction Bipolar Transistors

Fig. 8.11 shows the band alignment for a BJT, a single heterojunction bipolar transistor (S-HBT) in equilibrium (no bias applied) and under active conditions.in an S-HBT, the emitter is a widebandgap material (i.e. AlGaAs) while base and collector remain as small bandgap material (i.e. GaAs). Equation 8.35 can be used to estimate the current gain of the transistor with ΔE_g being a positive value due to larger bandgap of the emitter compared to that of the base. The boost that can be achieved in the current gain due to the heterostructure is often traded off for higher f_T and f_{max}. For instance by doping the base region at very high levels (10^{20}~10^{21}cm^{-3} in HBTs compared to ~10^{18}cm^{-3} in BJTs), base spreading resistance r_{bb} can be significantly lowered resulting in much higher f_{max} for the device. On the other hand emitter doping can be lowered to achieve lower base emitter junction capacitance, thus higher f_T. In the following we list the

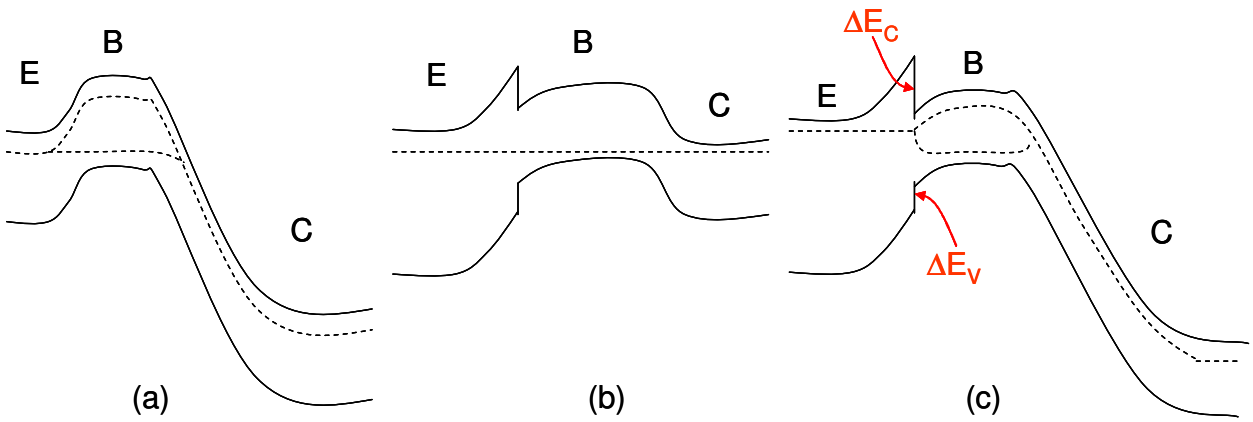


Fig. 8.11. Band diagram for (a) a BJT in normal operation mode (b) a single HBT in equilibrium and (c) a single HBT in normal operation mode.

modifications that the HBT devices often take advantage of and their consequence on the performance of the transistor. Let us start from the emitter region.

Emitter Modifications:

The emitter region often uses a wide-bandgap material that can boost the current gain according to equation (8.35) with ΔE_g being the difference between the bandgap in the emitter and base regions. Note that you may need to consider bandgap narrowing in the base which tends to increase the effective ΔE_g as base is often doped at a higher density than emitter. A large ΔE_C although inconsequential in determining the current gain has an impact on large turn on voltage (as electrons need to overcome the ΔE_C barrier of HBTs). The requirement of low ΔE_C yet large ΔE_g has pushed to use of newer materials such as InGaP/GaAs HBTs (ΔE_C 33% of ΔE_g for lattice matched structure) instead of AlGaAs/GaAs HBTs (ΔE_C 62% of ΔE_g). Notice that the small ΔE_C requirement is the opposite trend required for HEMT devices where large ΔE_C barrier resulted in better confinement of carriers in 2-DEG. Another effective way to taper off the effect of large ΔE_C is by using a graded bandgap emitter layer as shown in Fig. 8.12. In this case the Al mole fraction in the AlGaAs emitter layer is lowered in the vicinity of the base region.

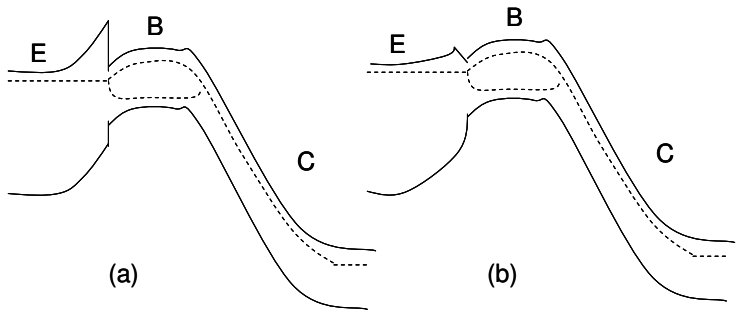


Fig. 8.12. Band diagram for (a) an abrupt heterostructure S-HBT and (b) a graded heterostructure S-HBT

Another consequence of large ΔE_C barrier is its effect in lunching carriers to the base with large kinetic energy. This results in premature velocity saturation as electrons with high energy occupy higher valleys and assume higher effective mass and lower effective mobilities. Note that a small ΔE_C barrier tends to optimize the energy of electrons being injected to base and continue to drift to the collector space charge region. Such small ΔE_C barrier is used in ballistic HBTs that demonstrate significant velocity overshoot effect in the neutral base as well as base collector space charge region. Unfortunately ballistic transport in these devices can be observed for only low base collector voltages and disappear as the bias voltage is changed.

In addition to the bandgap, one usually lowers the doping in the emitter region N_{De} to ~$10^{18}cm^{-3}$ values. By doing so, the extent of depletion region of the forward biased base emitter junction into the base is increased resulting in lower base-emitter junction capacitance and its transit time. This effectively increases the cut-off frequency of the device. To accommodate for lower emitter doping an emitter cap layer with higher doping concentration and sometimes with low bandgap material such as InGaAs is added on top of the heterostructure stack to reduce the emitter access resistance.

Base Modifications:

The base material is usually a low bandgap material such as GaAs in the case of AlGaAs emitter or InGaAs in the case of AlGaAs, InAlAs or InP emitters. Sometimes quarternary materials are used in the base to help varying the bandgap in the base region to generate a drift component that assists the diffusion of carriers and reduces the base transit time t_{bb}. The drift component can be generated using a graded doping concentration in the base. The effective electric field that generate a drift component is calculated according to

$$E_{field-drift} = \frac{kT}{q}\left(\frac{dN_{Ab}(x)/dx}{N_{Ab}(x)}\right) - \frac{1}{q}\frac{dE_g(x)}{dx} \tag{8.38}$$

where the variable acceptor doping in the base is denoted as $N_{Ab}(x)$.

In general doping in the base is at very high levels typically in the range of 10^{20}~$10^{21}cm^{-3}$ values to reduce the base spreading resistance given by equation (8.28). Higher base doping also helps eliminating the Early effect resulting in very large V_A for HBTs. This is due to the fact that the

projection of the base-collector space charge region into the neutral base is significantly reduced due to high doping in the base. The base thickness W_b is therefore less affected by the variation in the base collector voltage resulting in less early effect and thus higher output conductance. High doping in the base and its resulting lower base spreading resistance puts off the onset of the emitter current crowding to much higher collector current essentially increasing the high current capability of HBTs. Note that bandgap narrowing effect in the base helps increase the overall ΔE_g that can be calculated from equation (8.21) for p-doped materials resulting in higher current gain.

Now that the base region is doped very high its overall thickness can be lowered to values in the range of 300 to 1000Å. This lowers the transit time in the base resulting in higher f_T and β of HBTs. Narrow base transistors also suffer less from recombination in the base that is dominated by Auger recombination due to high doping of the base. f_{max} can be maximized by a tradeoff between high f_T (small W_b) and low r_{bb} (large W_b).

Collector Modifications:

If the collector region in an HBT has the same bandgap as base region, the device is called a single-HBT. If the collector has wider bandgap than base, the device is referred to as a double HBT (D-HBT). The reason to use a wider bandgap material as is done in D-HBTs are two-fold. Larger bandgap material typically have higher breakdown voltage. Therefore D-HBTs are superior devices compared to S-HBT for power applications. Another advantage of using wide-bandgap material in collector is to reduce the V_{CEsat} of the device to a lower value. As discussed earlier in this chapter minimum $V_{CE} = V_{BE} + V_{CB}$ that can be used ina transistor before it enters the saturation regime is calculated from a forward biased and on base-emitter junction (positive V_{BE}) and a forward biased but not on collector base junction (negative and smaller magnitude V_{CB}). Therefore a V_{CE} of 0.2V to 0.3V is expected in Si BJTs. In S-HBTs especially when emitter bandgap is not graded, V_{BEon} is increased due to the ΔE_C barrier. With essential the same V_{CB}, the value of V_{CEsat} of the S-HBT is increased to 0.5 to 0.6V. This is not attractive for low-power applications or power applications that high linearity and high power added efficiencies are sought. D-HBTs essentially increase the V_{CEsat} to values comparable to BJTs at a cost of a

new barrier for electrons at collector base junction, which may affect the ideality factor of the transistor in the collector.

The tradeoff in collector doping N_{epi} is as following. For too high doping N_{epi}, low breakdown voltage and large feedback capacitance (C_μ) are achieved. Too low of N_{epi} results in a wide space charge region W_{scl}, a large carrier transit time t_{scl} in this region and degraded f_T of the device.

The thickness of the collector region W_{epi} also needs to be optimized. Depending on the application it ranges from 2000Å for high speed HBTs to 10000 Å for high power HBTs. Important considerations in determining the thickness is collector punch-through as well as excess resistance due to undepleted W_{epi} layer.

Sub-collector layers are typically doped very high and are thick to reduce the collector access resistance. If possible, they are formed with low bandgap materials that help reduce the resistance.

Material Systems Used in HBTs

In this section we discuss advantages and diadvantages of three important material systems used for HBTs.

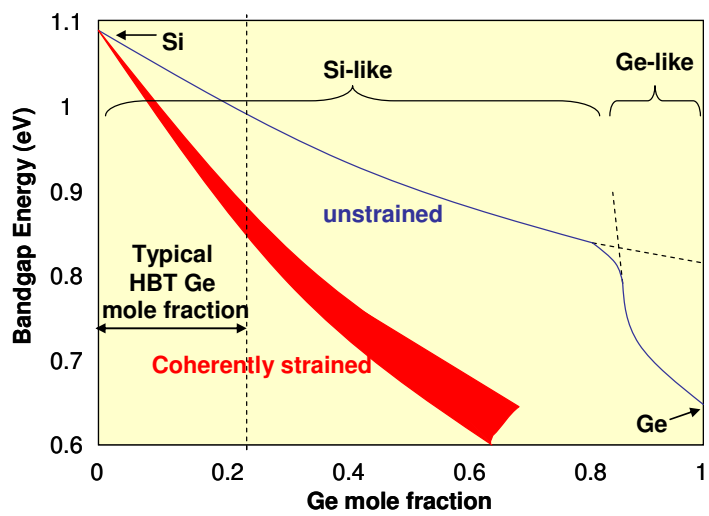


Fig. 8.13. Strained and unstrained SiGe band gap

SiGe HBTs:

Si/SiGe/Si HBTs by their nature are double HBTs. They do not take advantage of velocity overshoot as Si does not show this phenomena due to its indirect bandgap. SiGe is not lattice matched to Si so the resulting device is a pseudomorphic device. The strain in the base actually helps increase the band separation beyond the value of relaxed SiGe material as shown in Fig. 8.13. Therefore for smaller Ge mole fractions in the base

larger ΔE_g can be achieved. Typical Ge mole fraction is limited to about 15% to 25% depending on the thickness of the base layer that is limited by the critical thickness. NPN SiGe HBT transistors are nowadays an integral part of BiCMOS processes offered from various foundries over the world. The availability of SiGe HBT in an essentially CMOS process allows the designers to take advantage of higher speed and transconductance of these devices when it is needed. Since Si does not have much surface recombination, SiGe HBTs can be scaled down to submicron emitter and base finger dimensions resulting in extremely high f_T, and f_{max} at very low collector currents. Such devices have found tremendous RF and mixed signal applications and are essentially competing with InP and GaAs based transistors for low noise, medium power and microwave and mm-wave applications. Si substrate suffer from eddy current induced by EM fields penetrating the substrate resulting in relatively low quality factors of passive components such as inductors and transmission lines.

AlGaAs HBTs:

AlGaAs/GaAs HBTs are the first HBTs to be fabricated. These single-HBTs suffer from DX-centers of AlGaAs and large ΔE_C for abrupt heterostructures. The velocity overshoot in GaAs is an important aspect that needs to be taken advantage by proper design. Large relative bandgap of GaAs translates into low intrinsic doping level (n_i=2.3×10^{6}cm^{-3} at room temperature) which means a semi-insulating material for RF and microwave application. This is a big advantage for GaAs (and InP) based devices as passive components such as inductors, transmission lines and interconnects present high quality factor due to insignificant dielectric loss of the semi-insulating material. A major drawback of GaAs substrate is its high thermal resistance (almost 3× that of Si), which shows in the device I_C-V_{CE} characteristics through a negative output resistance known as self-heating effect (see Fig. 8.14). Note that the negative resistance is only seen at Dc and low frequency sweeps and can be avoided by performing pulse IV measurement. Modern

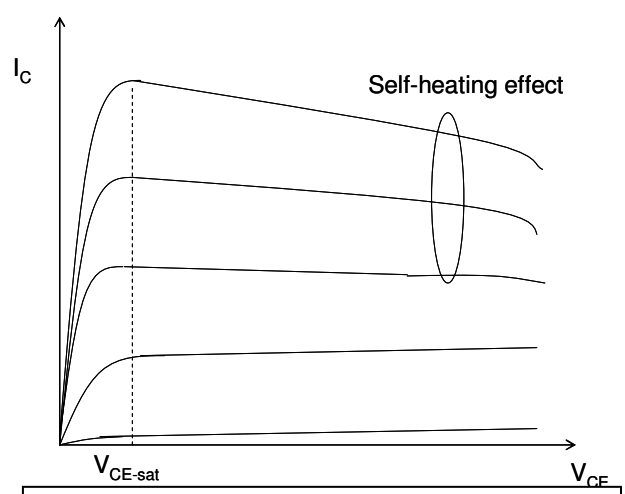


Fig. 8.14. High VCE-sat and self heating effect in AlGaAs/GaAs S-HBTs

AlGaAs/GaAs HBTs may use a pseudomorphic InGaAs base layer (with In mole fraction of ~15%) and GaAs collector. This D-HBT structure helps achieving higher ΔE_g as well as better transport properties in the base due to InGaAs layer. InGaAs base can be bandgap engineered to yield a drift electric field in the base and achieve higher speed.

AlGaAs/GaAs HBTs cannot be effectively scaled down to submicron structure due to high surface recombination of GaAs. Therefore, often a large emitter finger is needed to achieve high speed. This translates to burning many mWatts of power for RF and microwave devices based on GaAs material. This problem is less severe in InP and Si based HBTs as they can be effectively scaled down to submicron dimensions.

InAlAs/InGaAs and InP/InGaAs HBTs:

These InP based HBTs are extremely fast due to extremely high mobility of InGaAs base region with high In mole fractoion. Both InGaAs and InP have higher critical electric field and larger overshoot velocity compared to GaAs which are helpful in achieving better velocity overshoot in base and collector region. The band alignment of InAlAs/InGaAs typically grown with MBE and InP/InGaAs grown with MOCVD are shown in Fig. 8.15. InP has lower thermal resistance than GaAs. InP HBTs can be easily converted to D-HBT for better power performance.InP based HBTs have been scaled down to submicron dimensions with f_T and f_{max} over 500GHz at collector currents in the range of only a few mA. InGaAs-based HBTs can be integrated with 1.3-1.55µm laser sources that are also fabricated based on InAlAs/InGaAs materials.

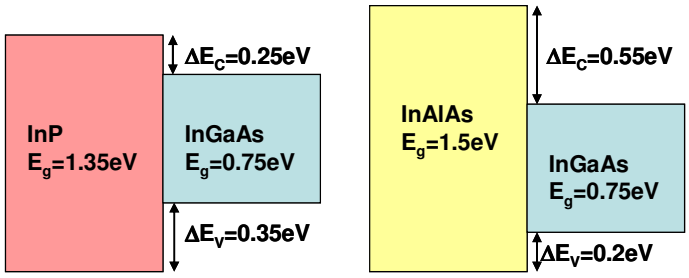


Fig. 8.15. Lattice-matched InP/InGaAs and InAlAs/InGaAs band alignment.

Noise in HBTs

HBT devices typically have lower 1/f noise than their HEMT and MESFET counterparts. This is due to the fact that the conduction in these devices occurs away from the surface as opposed to unipolar transistors. 1/f noise corner frequencies as low as 100Hz have been observed even in

AlGaAs/GaAs HBTs, which due to their nature are prone to DX-centers. Therefore, in this section we only consider high frequency noise performance of HBTs. Unlike FETs, the noise optimization of HBTs is particularly challenging, especially when considered in conjunction with suitable bias conditions for high-frequency performance. Generally, it is desired to design HBTs with optimum high-frequency performance at small bias currents where the shot noise sources do not dominate the high-frequency noise characteristics. As mentioned before the noise figure of a transistor is given according to

$$F = F_{\min} + 4\frac{R_n}{Z_O} \cdot \frac{|\Gamma_S - \Gamma_{opt}|^2}{|1+\Gamma_{opt}|^2 (1-|\Gamma_S|^2)} \tag{8.39}$$

Fig. 8.16 shows the T-equivalent noise model of an HBT. The model parameters can be directly extracted from S-parameter measurements and from noise data. This model includes base, emitter and collector series resistance and their noise sources, base and collector shot noise source ($\overline{i_B^2}$, $\overline{i_C^2}$) and parasitic inductances to achieve better accuracy at high frequencies. Unlike FETs, it is not necessary to consider cross-correlated noise sources to represent the noise performance of the HBT.

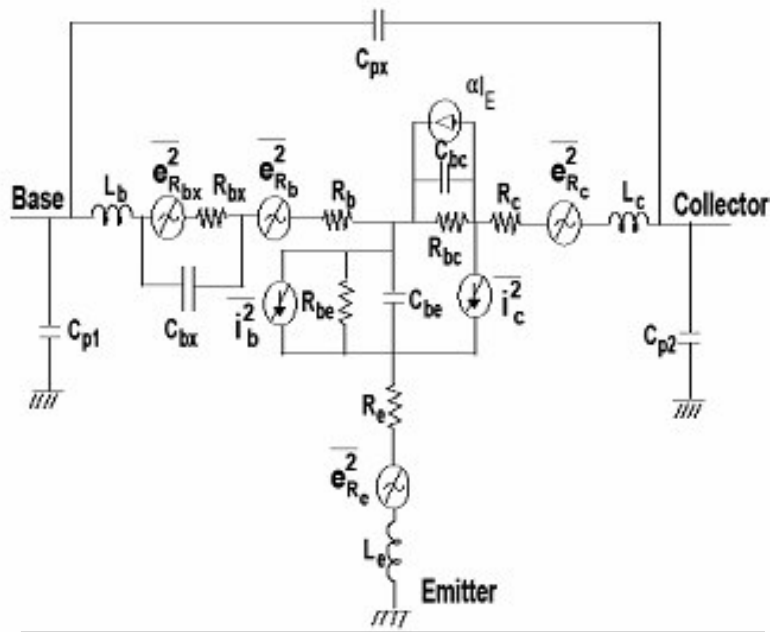


Fig.8.16. HBT noise equivalent circuit model based on a T-model.

As reported in the literature minimum noise figure of bipolar transistors increases in a quadratic manner with frequency according to

$$F_{\min} \approx 1 + b f^2 \left[1 + \sqrt{1 + \frac{2}{b f^2}} \right] \tag{8.40}$$

where $b = 40 I_C r_{bb} / f_T$. In the InGaP/GaAs and AlGaAs/GaAs HBTs investigated by the author, this was not the trend. In fact the minimum noise figure had a rather linear dependence to the frequency as shown in the measured and modeled data of Fig. 8.17.

Another aspect of the noise is the location of optimum source impedance for minimum noise. This was investigated for InGaP/GaAs and AlGaAs/GaAs HBTs and the results are shown in Fig. 8.18. As frequency increases the optimum location moves counter clockwise.

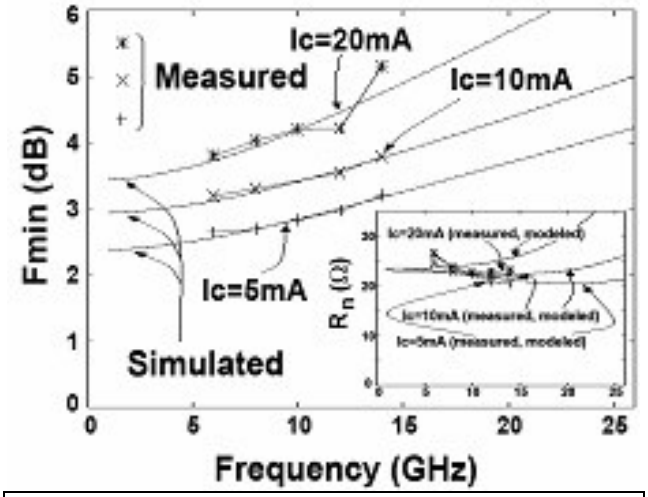
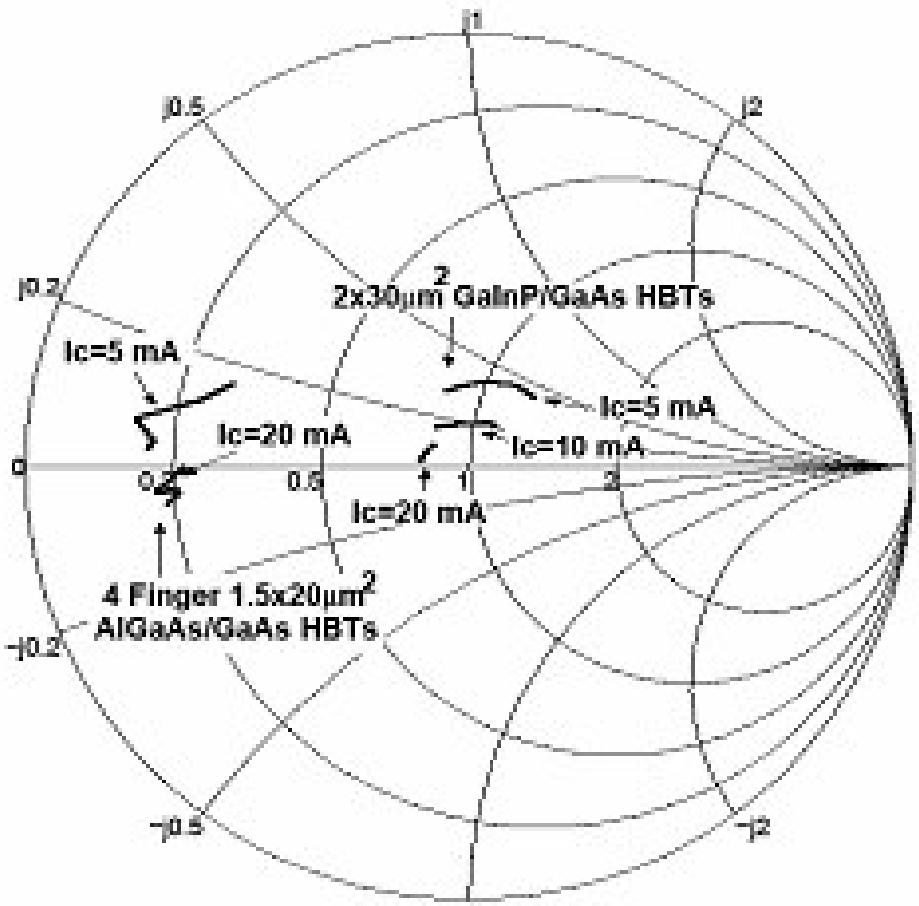


Fig. 8.17. Measured and modeled minimum noise figure and noise resistance of InGaP/GaAs HBTs as a function of frequency



.

Fig. 8.18. Optimum source impedance for minimum noise of AlGaAs/GaAs and InGaP/GaAs HBTs as a function of device collector current.