4.1. BJT_ general operational principle

Goal;

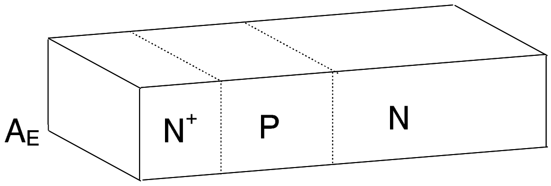
To understand basic BJT operation principles and analytical model for base, collector current as function of V_{BE} and V_{BC}.

Ref : 1. Semicoductor theory with NANOCAD,
Daeyoung Sa 2003(Korean)
2. Grove, Phys. and Tech. of Semicoductor devices.

A. General operation principle

Bipolar Junction Transistor is the "transistor" whose current (Emiiter to Collector) current is controlled by the "base current, i_B"(see fig.1 below). Unlike FET, the current is composed of two carriers,"electron current" and "hole" currents. Also it is a bulk device so that E-C current flows through bulk, so that current capability is generally larger compared with "MOSFET" whose current flows through the surface(Q_n). (Estimate the cross sectional area of the current flow for two devices) .

(a)



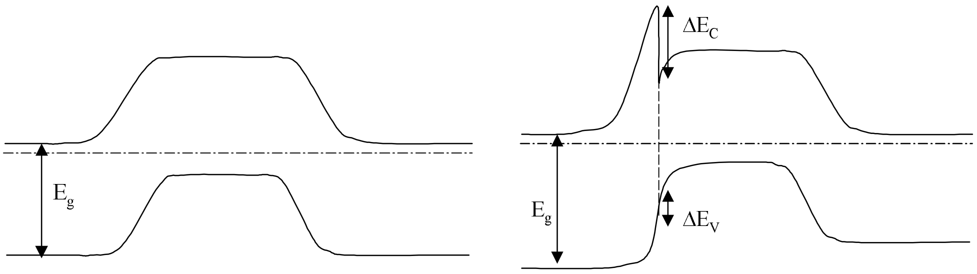
(b)

Fig. 1. (a) advanced BJT and (b) its 1 dimensional representation.

- As i_B increases, i_C increases with the "amplication factor, β so that i_C in i_C vs, V_{CE} curve shows constant value if i_B(input) is constant. If i_B is small(or nearly 0), i_C also is nearly 0 so that the case is called the "cut-off " region. If V_{CE} is reduced to some value(called the saturation voltabe , $V_{CE,sat}$), i_C begins to decrease and eventually becomes zero when V_{CE} has "some value".

Fig. 2. 5-4 of YJP(ref. 1).

B. General Band diagram



a) Homojunction
(Si)

b) Hetero junction
(AlGaAs - GaAs - GaAs)

Fig. 3.

— Considering the above two cases, BJT can be considered as a device whose potential barrier is controlled by the "BASE" voltage (V_{BE}), so that the minority carrier concentration at the base side(edge of the EB depeltion region) is controlled(this is exactly the same for the MOS case). But since "BASE" terminal is directly connected to the base region(Unlike the case for MOS where the GATE is isolated by the gate oxide) so that there exists the base current.

Two prominent differences between two devices can be noticed from the above comments:

the log(I_C) vs. V_{BE} curve has the ideal slope, $\eta=1$ whereas log(I_{DS}) vs. V_{GS} has the slope determined by $\eta = 1+ (C_{ss}+C_d)/C_{ox}$. 2nd difference is related with the range of the energy barrier modulation. In MOSFET case, only the surface is controlled by the gate so that inversion layer exists very close to the surface. However, in BJT, all the base region are controlled by the base voltage, so that IC flows along all the base cross section.

HW : From the above point of view, BJT is more ideal device than MOS device. Then why MOSFET device is so prevalent compared with BJT in modern VLSI application?

C. Regional Approach and Junction Law

- Regional approach : Consider the ideal 1D NPN BJT with the following doping profile (Fig. 4(a)). The device is assumed to be clearly divided by "neutral regions" and two "depletion regions", The assumption renders a very convenient analysis method for the ideal 1D BJT characteristics to be seen in the following.

- The minority carrier concentration for each bias condition can be as shown in the figure (Fig. 4(b)). The governing equation for the minority carrier concentration from the equilibrium value, Δ in the neutral emitter, the base and the collector, are the carrier continuity equations as,

$$\frac{d\Delta(x)}{dt} = \nabla \cdot F + U \tag{1}$$

with the Junction law stating that

$$\Delta = C_o(e^{\frac{V_j}{V_t}} - 1) \tag{2}$$

where C_o is the equilibrium minority carrier concentration.
Since the drift current component is negligible in the neutral regions,

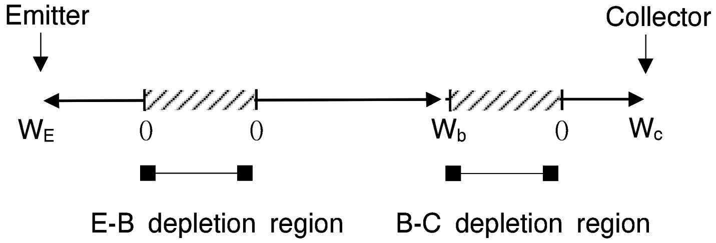
$$F_n = -\nabla \frac{d\Delta}{dx} \tag{3}$$

then eq. (1) becomes

$$0 = \frac{d^2\Delta}{dx^2} + \frac{\Delta}{\tau} \tag{4}$$

Here τ is the minority carrier life time.

(a) fig. 5-1(a) of YJP(ref.1)



(b)

Fig. 4

- The junction law in eq. (2) assumes that the majority Fermi level(of electrons in the emitter) is constant across the depletion region and becomes the minority Fermi level at the depletion region edge in the other side of the neutral region.(of electrons in the base). This assumption is only good when the current level is small enough so that the variation of the majority Fermi level does not vary in the depletion region.

- The solution for the minority carrier concentration in each neutral region can be obtained with two boundary conditions on both sides given as, BC for electrons in the base region:

$$n_p(0) = n_{po}\ e^{V_{BE}/V_t}$$

$$n_p(W_b) = n_{po}\ e^{V_{BC}/V_t} \qquad (5)$$

BC for holes in the emitter region:

$$p_n(0) = p_{no,E}\ e^{V_{BE}/V_t}$$

$$p_n(x_E) = n_{po,E} \qquad (6)$$

BC for hole in the collector region

$$p_n(0) = p_{no,C}\ e^{V_{BC}/V_t}$$

$$p_n(W_C) = p_{no,C} \qquad (7)$$

It should be noticed that $p_{no,E}$ and $p_{no,C}$ in eq. (6) and (7) are the equilibrium hole concentration in the emitter region and collector region, respectively.

D. Idear situation where the monority concentration is linear in the BJT (negligible recombination) in the base region.

- In the ideal situation where the recombination in base region is neglected,

$n_p(x) = n_p(0)(1-x/Wb)$

meaning that

$Fn(0) = Fn(Wb) = Dn^* n_p(0)/Wb$.

E. General situation in the neutral re

gion(for reference)

— Solutions for the continuity equations can be expressed for Δ(x) as,

$$\Delta(x) = A\, e^{-\frac{x}{L}} + B\, e^{\frac{x}{L}} \tag{8}$$

where L is the diffusion length($= \sqrt{D\tau}$) and

$$A = \frac{\Delta(x_L)e^{y} - \Delta(x_R)}{e^{y} - e^{-y}}$$

$$B = \frac{\Delta(x_R) - \Delta(x_L)e^{-y}}{e^{y} - e^{-y}}$$

$$y = \frac{W}{L}\,.$$

It is worthwhile to notice that the increase in the minority carrier from the equilibrium concentration, Δ, is composed of two parts; one from the left boundary and the other from the right boundary.

ex) Electron concentration in the base region can be expressed as,

$$\begin{aligned} n_p(x) - n_{po} &= \Delta \\ &= \Delta(x_L)\frac{e^{y}e^{-x/L} - e^{-y}e^{x/L}}{e^{y} - e^{-y}} \\ &\quad + \Delta(x_R)\frac{-e^{-x/L} + e^{x/L}}{e^{y} - e^{-y}} \end{aligned}$$

with

$$\Delta(x_L) = n_{po}\left(e^{\frac{V_{BE}}{V_t}} - 1\right)$$

$$\Delta(x_R) = n_{po}\left(e^{\frac{V_{BC}}{V_t}} - 1\right)$$

— With the electron and hole concentrations given in the neutral regions in eq. (8) and the total recombination(or generation) rates in the depletion regions, the steady state terminal currents can be readily found out.

F. Depletion region

In the depletion region, there is no minority and majority carrier concentrations since they may be comparable. Then we have to have a certain model for U.

Total recombination rate in the depletion region

$$= \int_0^W U dx$$

Let us consider only U due to SRH recombination mechanism which is dominant in the normal situation.

If the junction is forward biased,

$$U = \frac{CNt(np - n_i^2)}{n + p + 2n_i \frac{\cosh(E_t - E_i)}{k_B T}}$$

and considering that

$$np = n_i^2 \, e^{\frac{E_{fn} - E_{fp}}{k_B T}} = constant,$$

U_{max} occurs when $n = p = n_i \, e^{\frac{E_{fn} - E_{fp}}{k_B T}}$

Then

$$U_{max} = CN_t n_i^2 \frac{e^{\frac{E_{fn} - E_{fp}}{k_B T}} - 1}{2n_i e^{\frac{E_{fn} - E_{fp}}{V_t}} + 1}$$

$$= \frac{CN_t}{2} n_i \, e^{\frac{V_{BE}}{2V_t}}$$

if N_t located at $E_t = E_i$ in the band gap is considered.

Then

$$\int Udx = \frac{n_i\, e^{\frac{V_{BE}}{V_t}}}{2\tau}\, W_d.$$

Note that the voltage dependence is $e^{\frac{V_{BE}}{2V_t}}$.

If the junction is reverse biased,

n=p=0,

c Then $U = -CN_t\frac{n_i}{2} = -\frac{n_i}{2\tau_i}$ so that

$$\int Udx = -\frac{n_i}{2\tau_i}\, W_d.$$

D . Terminal currents and transistor α and β

With Δ given in section C, F$_n$ and F$_p$ at any point in the neural regions can be easily obtained.

Usually, F$_n$ and F$_p$ at the neutral boundaries(edge of the depletion region) are calculated since it is convenient to relate with the steady state terminal currents.

-General solution for the current component (for reference only)

For the base region,

c

$$\begin{aligned}
F_n(0;B) &= \left. -D_n \frac{d\Delta}{dx}\right)_{x=0} \\
&= \frac{D_n}{L_n}(A-B) \\
&= \frac{D_n}{L_n}\left(\coth y \Delta(0) - \frac{1}{\sinh y}\Delta(W_b)\right) \\
&= \frac{D_n}{L_n}\left(\coth y n_{p0}(e^{V_{BE}/V_t}-1) - \frac{1}{\sinh y} n_{p0}(e^{V_{BC}/V_t}-1)\right)
\end{aligned}$$

$$\begin{aligned}
F_n(W_b) &= \left. -D_n \frac{d\Delta}{dx}\right)_{x=W_b} \\
&= \frac{D_n}{L_n}\left(\frac{1}{\sinh y}\Delta(0) - \coth y \Delta(W_b)\right) \\
&= \frac{D_n}{L_n}\left(\frac{1}{\sinh y} n_{p0}(e^{V_{BE}/V_t}-1) - \coth y\ n_{p0}(e^{V_{BC}/V_t}-1)\right)
\end{aligned}$$

Also, for the emitter region,

$$\begin{aligned} F_p(0;E) &= -D_p \frac{d\Delta}{dx}\Big)_{x=0} \\ &= \frac{D_p}{L_P} \coth y \; n_{p0}(e^{V_{BE}/V_t}-1) \end{aligned}$$

$$\begin{aligned} F_p(W_e) &= -D_p \frac{d\Delta}{dx} x = W_e \\ &= \frac{D_p}{L_p} \frac{1}{\sinh y} p_{n0}(e^{V_{BE}/V_t}-1) \end{aligned}$$

and for the collector region,

$$F_p(0;C) = +\frac{D_p}{L_p} \coth y \; p_{n0}(e^{V_{BC}/V_t}-1)$$

- I_E, I_C and I_B.

Noticing the outflow of current is positive for I_E,

$$\begin{aligned} I_E &= qA_E(F_p(0;E) + F_n(0;E)) \\ &= qA_E(F_p(0;E) + F_n(0;B) + \int U dx) \\ &= qA_E\left\{ \left(\frac{D_n}{L_n} n_{p0} \cos thy + \frac{D_p}{L_p} p_{n0} \coth y \right)(e^{V_{BE}/V_t}-1) \right. \\ &\quad \left. + W_{be}\frac{n_i}{2\tau}(e^{V_{BE}/2V_t}-1) - \frac{D_n}{L_n}\frac{1}{\sinh y} n_{p0}(e^{V_{BC}/V_t}-1) \right\} \end{aligned}$$

Also, inflow of the current is set to positive for I_C as,

$$\begin{aligned} I_C &= qA_E[F_n(0;C) - F_p(0;C)] \\ &= qA_E[F_n(W_b) + \int(-U)dx - F_p(0{:}C)] \\ &= qA_E\left\{ \left(-\frac{D_n}{L_n} \coth y \; n_{p0} - \frac{D_p}{L_p} \coth y \; p_{n0} \right)(e^{V_{BC}/Vt}-1) \right. \\ &\quad \left. - W_{bc}\frac{n_i}{2\tau}(e^{V_{BC}/2V_t}-1) + \frac{D_n}{L_n}\frac{1}{\sinh y} n_{p0}(e^{V_{BE}/V_t}-1) \right\} \end{aligned}$$

Now inflow of I_B is positive,

$$
\begin{aligned}
I_B &= I_E - I_C \\
&= qA_E\bigg\{\bigg(\frac{D_n}{L_n}\coth y\ n_{p0} + \frac{D_p}{L_p}\coth y\ p_{n0} - \frac{D_n}{L_n}\frac{1}{\sinh y}\ n_{p0}\bigg)(e^{V_{BE}/V_t}-1) \\
&\quad - qA_E\bigg(\frac{D_n}{L_n}\frac{1}{\sinh y}\ n_{p0} - \frac{D_n}{L_n}\coth y\ n_{p0} - \frac{D_p}{L_p}\coth y p_{n0}\bigg)(e^{V_{BC}/V_t}-1) \\
&\quad + W_{be}\frac{n_i}{2\tau}(e^{V_{BE}/2V_t}-1) + W_{bc}\frac{n_i}{2\tau}(e^{V_{BC}/2V_t}-1)
\end{aligned}
$$

It is interesting to notice that I_E and I_C are composed of two current terms, one which is function of V_{BE} and the other function of V_{BC}.

I_E is composed of V_{BE} term which is the "diode" current and V_{BC} dependent term which is the "transistor" current. Also I_C is composed of V_{BC} term which is the "diode" current and V_{BE} dependent term which is the " transistor" current.

The V_{BE} dependent diode current, I_F, is

$$
\begin{aligned}
I_F &= qA_E\bigg\{\bigg(\frac{D_n}{L_n}\coth y\ n_{p0} + \frac{D_p}{L_p}\coth y\ p_{n0}\bigg)(e^{V_{BE}/V_t}-1) \\
&\quad + W_{be}\frac{n_i}{2\tau}(e^{V_{BE}2V_t}-1)\bigg\} \\
I_R &= qA_E\bigg\{\bigg(\frac{D_n}{L_n}\coth y\ n_{p0} + \frac{D_p}{L_p}\coth y\ p_{n0}\bigg)(e^{V_{BC}/V_t}-1) \\
&\quad + W_{bc}\frac{n_i}{2\tau}(e^{V_{BC}/2V_t}-1)
\end{aligned}
$$

with I_E and I_R, I_C and I_E can be represented as,

$$
\begin{aligned}
&I_C = -I_R + \alpha_F I_F \\
&I_E = I_F - \alpha_R I_R
\end{aligned}
$$

where $\alpha_F I_F$ and $\alpha_R I_R$ are the transistor current and α_F, α_R will be defined in the next section.

With this, the following equivalent circuit representation of the BJT can be obtained which is called the "Ebers-Moll model".

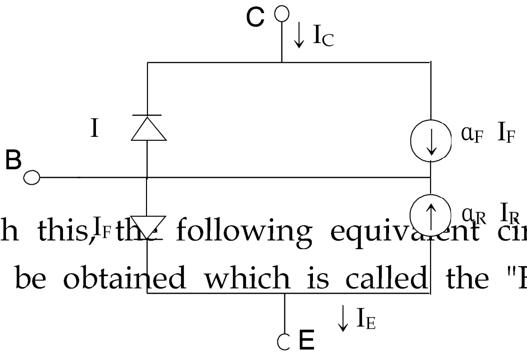


Fig. 2. Ebers-Moll model for BJT

– Important transistor DC parameters, α_F and β_F.

α_F was defined as,

$$\alpha_F = \frac{I+C+I_R}{I_F}$$

$$= \frac{qA_E \frac{D_n}{L_n} \frac{1}{\sinh y} n_{p0}(e^{V_{BE}/V_t} - 1)}{I_F}$$

$$= \frac{qA_E F_n(W_b)}{I_F} \simeq \frac{qA_E F_n(W_b)}{I_E}$$

α_F means the ratio of I_E to reach B-C depletion without being recombined during.

Emitter efficiency is γ is defined as,

$$\frac{qA_E F_n(0:B)}{I_E}$$

$$= \frac{D_n}{L_n} n_{p0} \frac{\coth y (e^{V_{BE/Vt}} - 1) - \frac{1}{\sinh y}(e^{V_{BC}/V_t} - 1)}{F_n(0) + F_p(0) + \frac{2n_i}{\tau}(e^{2V_{BE}/V_t} - 1)}$$

$$= \frac{\frac{D_n}{W_b} n_{p0}(e^{V_{BE}/V_t} - 1)}{\left(\frac{D_n}{W_b} n_{p0} + \frac{D_p}{L_p}\coth y\right)(e^{V_{BE}/V_t} - 1) + W_{be}\frac{2n_i}{\tau}(e^{V_{BE}/2V_t} - 1)}$$

for the forward active bias where $V_{BE}>0$ and $V_{BC}<0$.

The emitter efficiency indicates the ratio of electrons from the emitter to reach the neutral base without being recombined in the emitter and emitter-base junction.

- Base transport factor, α_T.

Base transport factor is defined as,

$$\frac{F_n(W_b)}{F_n(0)} = \frac{\frac{D_n}{L_n} n_{p0}\left(\coth y (e^{V_{BE}/V_t} - 1) - \frac{1}{\sinh y}(e^{V_{BC}/V_t} - 1)\right)}{\frac{D_n}{L_n} n_{p0}\left(\frac{1}{\sinh y}(e^{V_{BE}/V_t} - 1) - \coth y (e^{V_{BC}/V_t} - 1)\right)}$$

$$= 1 - \frac{1}{2}\frac{W_b^2}{L_n}$$

for $W_b<<L_n$ and $V_{BE}>0$ and $V_{BC}<<V_t$.

Base transport factor indicates the ratio of electrons injected into base to the electrons reaching the B_C junction without being recombined during excursion the neutral base region.

Now "transistor current" of I_C is $F_n(W_b)$ and α_F is defined as,

$$\alpha_F = qA_E F_n(W_b)\frac{1}{I_E} = \gamma\alpha_T.$$

In terms of alpa,F, the IC can be written as,

$$qA_E F_n(W_b) = \alpha_F I_F \approx \alpha_F I_E.$$

- Transistor β.

$$I_C = -I_R + \alpha_F I_E$$

Since

$$I_B = I_E - I_C,$$

$$I_C = -I_R + \alpha_F(I_B + I_C)$$

Or,

$$I_C = I_B\beta - I_R(1+\beta)$$

where

$$\beta = \frac{\alpha_F}{1-\alpha_F}$$