Bipolar Junction Transistor

Introduction

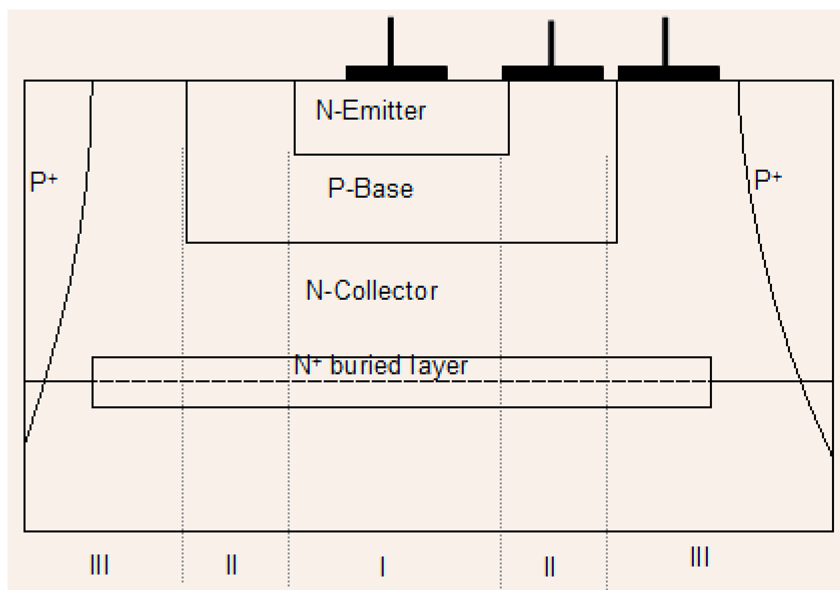
A Bipolar Junction Transistor (BJT) is a prominent semiconductor device used in a wide variety of applications

Our discussion of BJT characteristics will proceed in the following order:

(1) Basic characteristics of a BJT
(2) BJT in forward active mode of operation
(3) Dynamic characteristics of BJT in active mode
(4) High level injection effects
(5) Breakdown in BJT
(6) Circuit Models
(7) Design perspective

Most of our discussion will center around Silicon NPN transistor meant for high frequency applications.

The Figure below shows a simplified schematic of a planar npn transistor:

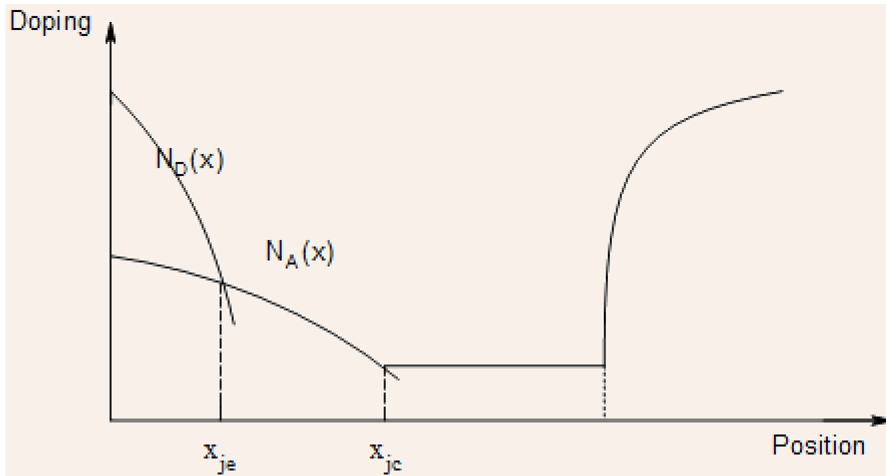


- The transistor is fabricated on an N-epilayer grown over a P-type substrate. The P-type base is diffused first. The base diffusion depth determines the thickness of the collector region.
- This is followed by + N emitter diffusion. This determines both the emitter thickness as well as the base thickness.

An + *N* buried layer is also used in the collector to reduce the collector series resistance. The contacts for all three layers are formed at the top.

- Since several transistors have to be fabricated on the same piece of Silicon, there is a need for isolating each transistor from the other. There are several ways of doing this, one of which is junction isolation. This is done by creating a reverse biased junction between two transistors through diffusion of P-type impurity right down to the substrate.

- The doping profile and the junctions are illustrated by the Figure below :



.

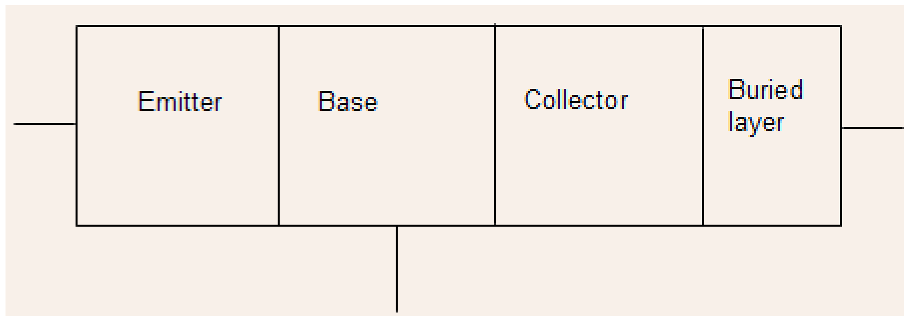
- In the schematic of the transistor, three distinct regions can be identified:

(i) Region I: this is known as the intrinsic transistor region where all the interesting things occur

(ii) Region II: this is known as the extrinsic collector-base region. This region provides a contact to the base

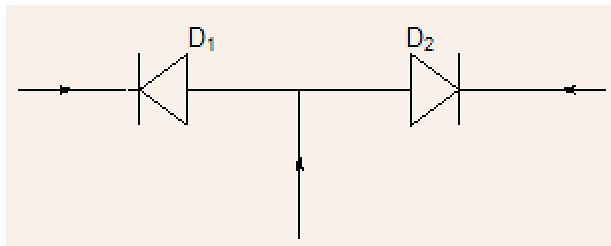
(iii) Region III: this extrinsic collector region again serves the role of providing a contact to the collector

- Since the transistor action occurs only in the intrinsic region, the other two regions are called parasitic regions. One of the goals in design of transistor is to minimize the area of the parasitic regions to a minimum through process innovation.
- In most of our discussions on BJT, we shall confine our attention to the intrinsictransistor region and model it simply as that shown below.



The structure of the transistor shows that there are two PN junctions within it; one at the emitter-base junction and the other at the collector-base junction.

- However, characteristics unique to a BJT cannot be obtained by simply taking two PN junction diodes and externally connecting them as shown below:



- In the circuit above, the two diodes do not interact with each other in the sense that their current-voltage characteristics when they are connected to each is same as that in isolation. If diode D1 is forward biased and D2 reverse biased, then all the current of D1 would flow through the base contact and D2 would be carrying only its reverse saturation current. The base current would be practically equal to the emitter current.

- Unlike the situation above, where the two diodes function in isolation of each other, the two PN junctions are closely coupled to each other as explained below:

- Consider the case where the base-emitter (EB) junction is forward biased and base collector junction (CB) junction reverse biased. This mode of operation is known as the forward active mode.

- At the EB- junction holes will be injected by the P-type base into the emitter and electrons will be injected by the N-type emitter into the base.

- The holes injected into the emitter would diffuse and get collected at the emitter contact as expected. However, the electrons injected into the base are not collected at the base contact, which is relatively far away but most of the electrons diffuse to the much nearer CB-junction from where the junction electric field sweeps them into the collector!
- As a result, a large current flows into the collector despite the reverse bias

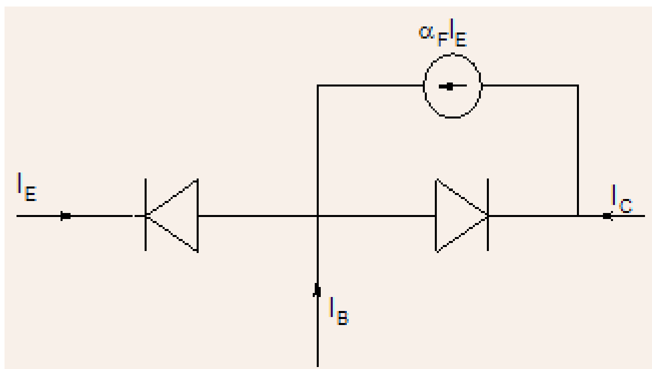
- This is not strange, if it is remembered that in reverse bias electrons are required to flow from P®N and holes from N®P and current is small not because it is difficult for carriers to move from one region to another due to any barrier but because there are so few electrons in P-region and holes in N-region.

- The forward biasing of the emitter base junction floods the P-region of the CB diode with electrons thereby increasing the current flow.

- The collector current would be determined primarily by the electrons injected into the base by the emitter, its own reverse current being extremely small.

- As expected, this current flowing through the reverse biased collector-base junction will be little influenced by the collector-base voltage but very sensitive to the emitter-base voltage.

The transistor in forward active mode of operation can therefore be represented by the following circuit:



- Apart from the expected diodes, the Figure shows a current source F F I a , which represents the fraction of emitter-base current that is coupled into the collector-base junction.

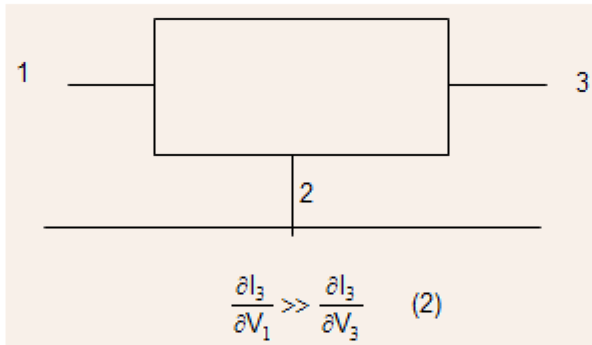
- As mentioned earlier, the collector current would be very sensitive to the emitter-base voltage and insensitive to collector-base voltage .

- We have a very interesting situation here of a current flowing between pair of terminals being insensitive to changes in voltage across it but very sensitive to voltage applied across a different pair of terminals!

$$\frac{\partial I_C}{\partial V_{BE}} >> \frac{\partial I_C}{\partial V_{BC}} \qquad (1)$$

- Because of the prominence of transconductance or its inverse transresistance, the device is appropriately called a Transistor!

One of the tests for any three terminal device being called a transistor is :

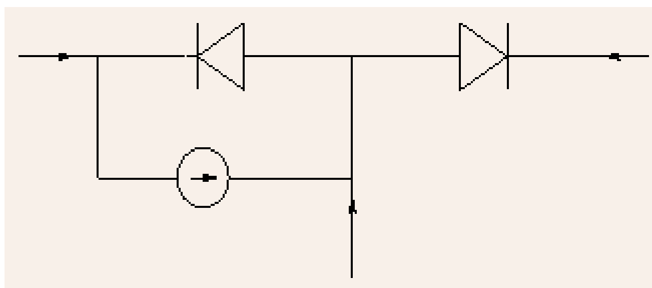


A rearrangement of this equation shows that a device which has much larger transconductance as compared to output conductance can exhibit voltage gain.

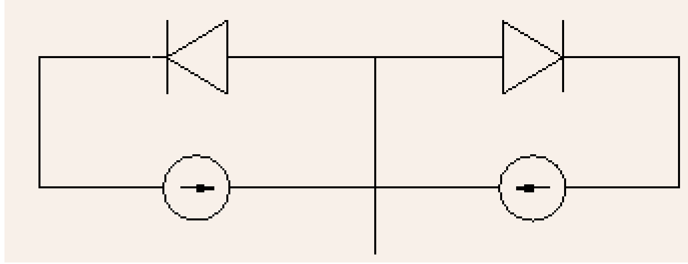
$$\frac{\partial V_3}{\partial V_1} >> 1 \qquad (3)$$

- Just as a fraction of current of forward biased emitter-base junction gets coupled into the collector-base junction, similarly we have a complimentary situation where current of forward biased CB junction will also get coupled into reverse biased emitter-base junction in what is known as reverse active mode of operation.

The circuit representation for the reverse active mode would therefore be as shown by the Figure below :



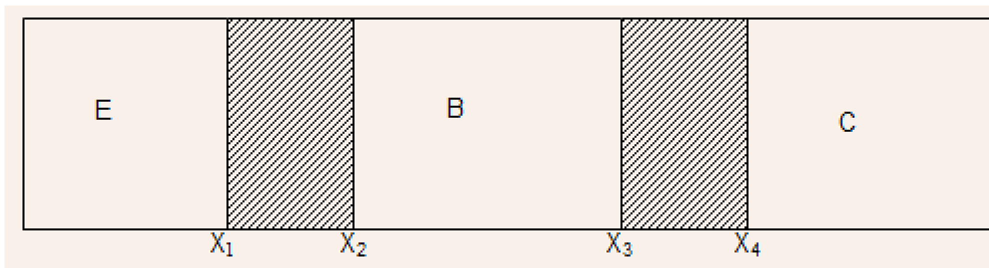
More generally, it may be expected that a transistor can be represented as :



This is the well-known Ebers-Moll[i] model of the BJT that can be used to understand transistor operation in different modes of operation.

Although the Ebers-Moll representation is a very general one, we shall for the sake of simplicity derive it under the following assumptions:

(i) 1D transistor structure with uniform doping in all the regions
(ii) Narrow base width so that carrier recombination within base can be neglected
(iii) Low level injection condition in all the regions
(iv) SCR recombination is neglected
(v) Steady state conditions



Using the results derived earlier for PN junctions, the emitter, base and collector currents can be expressed as

$$I_E = I_P(x_1) + I_n(x_2) \quad (4)$$
$$I_C = I_P(x_4) + I_n(x_3) \quad (5)$$
$$I_B = I_P(x_1) + I_P(x_4) \quad (6)$$

- The emitter and collector current are composed of electron and hole currents flowing through their respective junctions. The base current is due to holes injected into the emitter and collector regions.

- The neglect of recombination within the base (assumption ii) means that the steady state electron continuity equation simply reduces to

$$\frac{1}{qA}\frac{\partial I_n}{\partial x} \cong 0 \quad (7)$$

In other words, electron current within the base is constant. All the electrons that are injected from the emitter reach the collector and viceversa.

- Use of low level injection assumption means that the electron current in a uniformly doped base can be modeled as a diffusion current alone :

$$I_n = qAD_N \frac{\partial n}{\partial x} \quad (8)$$

Substitution of Eq. (8) in Eq. (7) gives

$$\frac{\partial^2 n}{\partial x^2} = 0$$

$$\frac{\partial n}{\partial x} = \text{constant} \quad (9)$$

The electron density varies linearly within the base so that :

$$I_n = qaD_N \times \left(\frac{\Delta n(x_3) - \Delta n(x_2)}{W_B}\right) \quad (10)$$

The excess electron densities at the two ends of the base can be expressed as :

$$\Delta n(x_2) = \frac{n_i^2}{N_A} \times \left(\exp\left(\frac{qV_{BE}}{kT}\right) - 1\right) \quad (11)$$

$$\Delta n(x_3) = \frac{n_i^2}{N_A} \times \left(\exp\left(\frac{qV_{BC}}{kT}\right) - 1\right) \quad (12)$$

The electron current can therefore, be written as :

$$I_n = I_S \times \left(\exp\left(\frac{qV_{BC}}{kT}\right) - 1\right) - I_S \times \left(\exp\left(\frac{qV_{BE}}{kT}\right) - 1\right) \quad (13)$$

$$I_S = Aqn_i^2 \frac{D_N}{N_A W_B} \quad (14)$$

Without going into the details, the hole current injected into the emitter being proportional to the excess hole density in the emitter, can be written as :

$$I_p(x_1) = -C_1' \times \Delta p(x_1) \quad (15)$$

$$\Delta p(x_1) = \frac{n_i^2}{N_D} \times \left(\exp\left(\frac{qV_{BE}}{kT}\right) - 1\right) \quad (16)$$

$$I_p(x_1) = -C_1 \times \frac{n_i^2}{N_D} \times \left(\exp\left(\frac{qV_{BE}}{kT}\right) - 1\right) \quad (17)$$

The negative sign indicates that the current is in negative direction.

Similarly, the hole current injected into the collector can be written as :

$$I_p(x_4) = -C_4 \times \frac{n_i^2}{N_D} \times (\exp(\frac{qV_{BC}}{kT}) - 1) \qquad (18)$$

Using Eq. (4-18), the emitter and collector currents can be expressed as :

$$I_E = -I_{ES} \times (\exp(\frac{qV_{BE}}{kT}) - 1) + I_S \times (\exp(\frac{qV_{BC}}{kT}) - 1) \qquad (19)$$

$$I_C = -I_S \times (\exp(\frac{qV_{BE}}{kT}) - 1) + I_{CS} \times (\exp(\frac{qV_{BC}}{kT}) - 1) \qquad (20)$$

Let

$$\alpha_F = \frac{I_S}{I_{ES}} \qquad (21)$$

$$\alpha_R = \frac{I_S}{I_{CS}} \qquad (22)$$

$$\Rightarrow \alpha_F I_{ES} = \alpha_R I_{CS} \qquad (23)$$

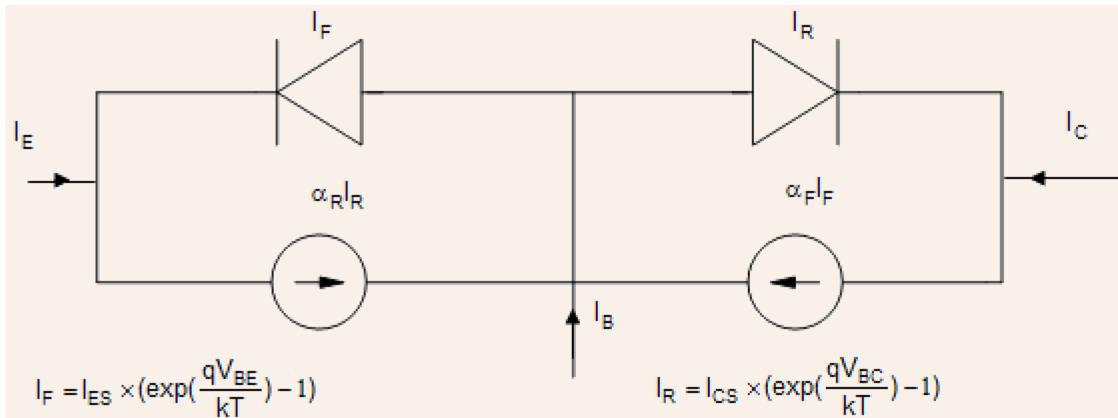
With these definitions, the Ebers-Moll can be expressed in terms of the following two equations :

$$I_E = -I_{ES} \times (\exp(\frac{qV_{BE}}{kT}) - 1) + \alpha_R I_{CS} \times (\exp(\frac{qV_{BC}}{kT}) - 1) \qquad (24)$$

$$I_C = -\alpha_F I_{ES} \times (\exp(\frac{qV_{BE}}{kT}) - 1) + I_{CS} \times (\exp(\frac{qV_{BC}}{kT}) - 1) \qquad (25)$$

$$I_B = -I_E - I_C$$
$$= (1 - \alpha_F) \times (\exp(\frac{qV_{BE}}{kT}) - 1) + (1 - \alpha_R)(\exp(\frac{qV_{BC}}{kT}) - 1) \qquad (26)$$

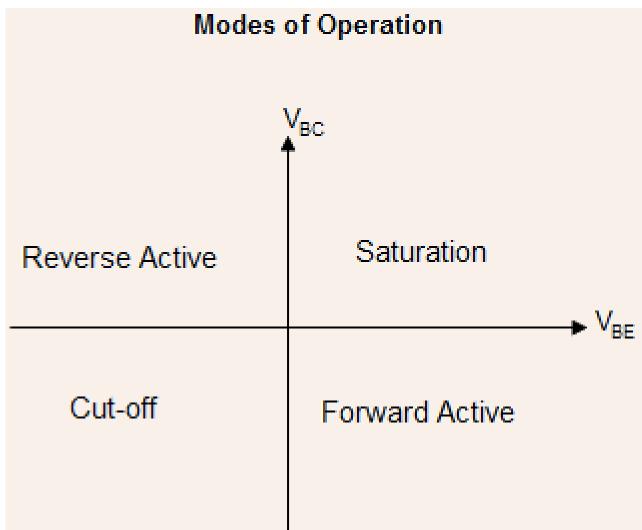
Eq. (24-26) can be represented by an equivalent circuit shown below :



The model shows that current flowing through each junction is made of two components :

(i) Normal diode current
(ii) Current due to coupling of a fraction of current flowing in the neighboring diode.

The Ebers-Moll Model can be used to understand transistor operation in all four modes of operation depicted below :



As shown by the diagram, there are four modes in which a BJT can be operated :

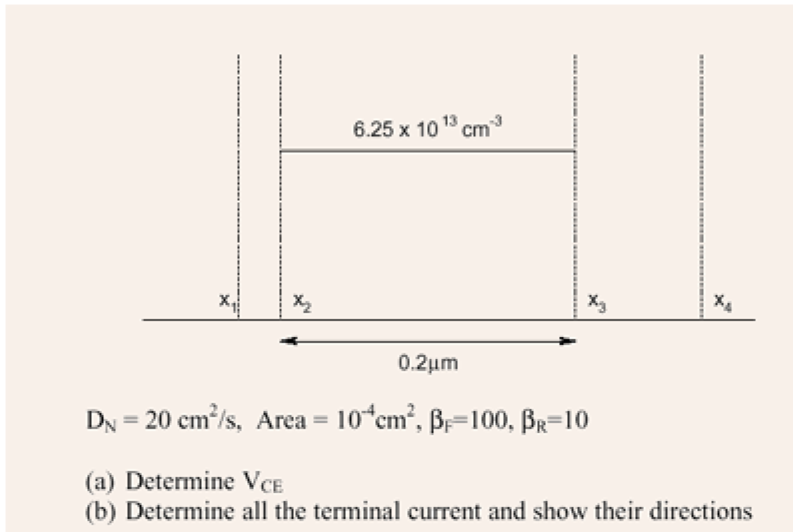
(i) Forward active Mode: The emitter-base junction is forward biased and collector-base junction is reverse biased.

(ii) Saturation Mode: Both the junctions are forward biased

(iii) Reverse active Mode: The collector-base junction is forward biased and emitter-base junction reverse biased.

- The forward active mode of operation is the most important region of transistor operation and will be discussed in detail in the next few lectures. This would be followed by a discussion of characteristics of BJT in cutoff and then saturation.

Example 1.1 : The electron density within the base of a uniformly dopde NPN BJT is constant as shown below :



Solution : (a) It is obvious that substantial minority carrier injection has taken place from both the junctions so that both are forward biased and the transistor is in saturation.

$$\Delta n(x_2) = \frac{n_i^2}{N_A} \times (\exp(\frac{qV_{BE}}{kT}) - 1) \cong \frac{n_i^2}{N_A} \times \exp(\frac{qV_{BE}}{kT})$$

$$\Delta n(x_3) = \frac{n_i^2}{N_A} \times (\exp(\frac{qV_{BC}}{kT}) - 1) \cong \frac{n_i^2}{N_A} \times \exp(\frac{qV_{BC}}{kT})$$

$$\Rightarrow V_{CE} = 0.$$

(b) Since the electron current in the base is given by the expression $I_n = qAD_N \frac{\partial n}{\partial x}$ N it is zero because the electron density is constant. Let us calculate the hole current IP(x1), injected into the emitter. Using the definition of forward current gain :

$\beta_F = \frac{I_N(x_2)}{I_P(x_1)}|$ forward active mode.

To use this expression, we have to calculate the electron current for the same emitter-base forward bias as the present case but the transistor in active mode. In that case, $\Delta n(x_2)$ remains the same but

$\Delta n(x_3) = 0$ so that

$$I_n(x_2) = qAD_N \times (\frac{\Delta n(x_3) - \Delta n(x_2)}{W_B}) = 1mA$$

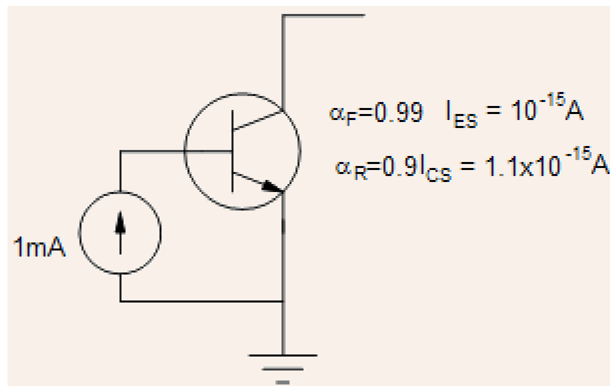
$\Rightarrow I_P(x_1) = 0.01$ mA.

Using a similar procedure, the hole current injected into the collector can be determined as $I_P(x_4) = 0.1$ mA.

$$\begin{aligned} I_E &= I_P(x_1) + I_n(x_2) &= 0.01mA \\ I_C &= I_P(x_4) + I_n(x_3) &= 0.1mA \\ I_B &= I_P(x_1) + I_P(x_4) &= 0.11mA \end{aligned}$$

n the calculations above, we have used only the magnitudes of the currents into account.The emitter current being composed of hole injection into the emitter would be negative, for the convention that current entering a terminal is taken as positive. Similarly the collector current would be negative and only the base current would be positive.

Example 1.2 : Determine the collector-emitter voltage of the transistor shown below with its collector open circuited.



Using the Ebers-Moll model we obtain

$I_C = \alpha_F I_F - I_R = 0 \Rightarrow \alpha_F I_F = I_R$ (a1)

Since 1mA current is flowing through the base and emitter terminals, the emitter-base junction is obviously forward biased so that $(\exp(\frac{qV_{BE}}{kT}) - 1) \cong \exp(\frac{qV_{BE}}{kT})$

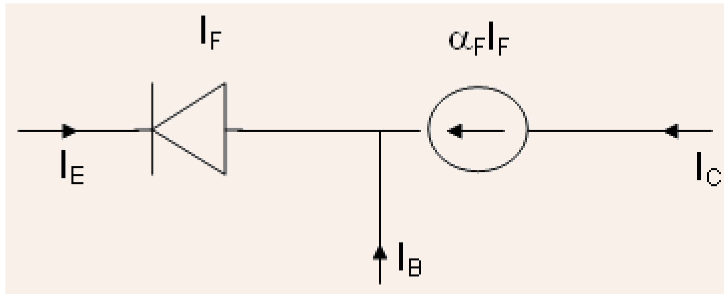
Using the expression derived above, it is also clear that collector-base junction will also be forward biased so that $(\exp(\frac{qV_{BC}}{kT}) - 1) \cong \exp(\frac{qV_{BC}}{kT})$ kT

This simplification along with Eq. (a1) allows the collector-emitter voltage to be expressed as $V_{CE} = kT/q \ln(I_{CS}/\alpha_F I_{ES})$

Using the reciprocity relationship : $\alpha_F I_{ES} = \alpha_R I_{CS}$ we obtain the final expression as

$V_{CE} = kT/q \ln(1/\alpha_R) = 2.7$ mV

The emitter-base junction is forward biased and collector-base junction is reverse biased.Ignoring the reverse saturation current of the collector-base diode, the transistor can be represented as:



$$I_C = \alpha_F I_F = I_S \times (\exp(\frac{qV_{BE}}{kT}) - 1)$$

$$= qAn_i^2 \frac{D_N}{N_A W_B} \times (\exp(\frac{qV_{BE}}{kT}) - 1) \qquad (1)$$

The base current can be written as

$$I_B = (1-\alpha_F)I_F = \frac{I_C}{\beta_F} \qquad (2)$$

$$\beta_F = \frac{\alpha_F}{1-\alpha_F} \qquad (3)$$

As we shall soon see, α_F is often very close to unity so that $\beta_F >> 1$.

A small change in the base current would appear in an amplified form at the collector current. In other words, the BJT exhibits current gain and β_F is called its forward current gain.

(Henceforth the subscript F will be dropped. Unless otherwise stated, the mode of operation by default is the forward active mode).
Let us look at current gain in more detail:

The base current supplies the holes needed to replenish those that are lost from the base due to injection/recombination. It has four distinct components :

(i) I_{B1} : this supplies the holes lost due to recombination in the neutral base region .
(ii) I_{B2} : this supplies the holes that are injected into the emitter.
(iii) I_{B3} : this supplies the holes that recombine within the EB space charge region.

(iv) I_{B4}: this represents the holes that flow from the CB space charge region into the base as part of the reverse generation current.

The current gain can therefore be written as:

$$\beta = \frac{I_C}{\sum I_{Bi}} \tag{4}$$

$$\Rightarrow \frac{1}{\beta} = \frac{1}{\beta_1} + \frac{1}{\beta_2} + \frac{1}{\beta_3} + \frac{1}{\beta_4} \tag{5}$$

$$\beta_i = \frac{I_C}{I_{Bi}} \tag{6}$$

Eq. (5) shows that the component of current gain, which is the least, will determine the overall current gain. Let us try to estimate each one of the current gain components:

(i) The base current due to recombination within the neutral portion of the base can be written as

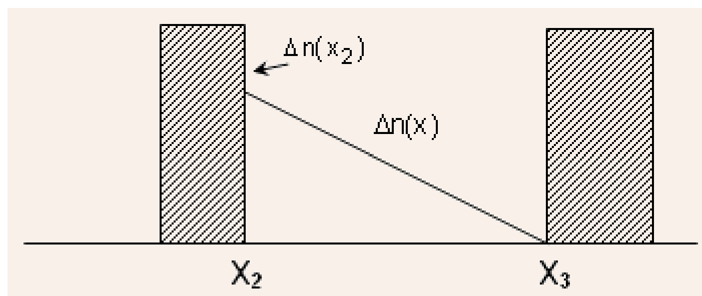
$$I_{B1} = Aq\int_{x_2}^{x_3} U dx \tag{7}$$

For low level injection:

$$U \cong \frac{\Delta n}{\tau_N} \tag{8}$$

$$I_{B1} = Aq\tau_N^{-1}\int_{x_2}^{x_3} \Delta n dx \tag{9}$$

This requires determination of electron density within the base. It was shown earlier that if base recombination is neglected, then electron density varies linearly as shown below:



- It may seem strange that we are trying to estimate recombination within the base and using neglect of recombination within the base to find the electron density!
- However, it must be remembered that an assumption is good or bad depending on the context.

- It turns out that the profile of electron density in the base with recombination taken into account differs very slightly from the case where it is assumed to be linear, so we are justified in using this assumption.

Therefore, $$\Delta n(x) = \Delta n(x_2) \times \left(1 - \frac{x - x_2}{W_B}\right) \qquad (10)$$

The base current due to recombination within the base can then be expressed as :

$$I_{B1} = 0.5Aq\tau_N^{-1}\Delta n(x_2)W_B \qquad (11)$$

The collector current given by Eq. (1) can be written in terms of $\Delta n(x_2)$:

$$I_C = Aq\frac{D_N}{W_B}\Delta n(x_2) \qquad (12)$$

$$\Rightarrow \beta_1 = \frac{2D_N\tau_N}{W_B^2} \qquad (13)$$

- Current gain is proportional to the recombination lifetime in the base and inversely proportional to the base width. The current gain is also independent of collector current.

There is a parameter called the base transport factor, which is closely related to the component of current gain just calculated. The base transport factor is defined as the fraction of electron current, injected by the emitter, that reaches the collector:

$$\alpha_T = \frac{I_n(x_3)}{I_n(x_2)} \qquad (14)$$

The fraction of electrons that do not reach the collector junction are those that get lost due to recombination so that

$$I_n(x_3) = I_n(x_2) - I_{B1} \qquad (15)$$

Eq. (13-15) can be used to obtain the expression for base transport factor:

$$\alpha_T = 1 - \frac{1}{\beta 1} = 1 - \frac{W_B^2}{2D_N\tau_N} \qquad (16)$$

Example 2.1 : Determine the current gain component due to neutral base recombination for base recombination lifetimes of 1, 0.1 and 0.01 ms. Calculate also the base transport factor .
Take $D_N = 20\,cm^2/s$, for a base of doping $2 \times 10^{17}\,cm^{-3}$ and effective thickness $W_B = 0.2\mu m$.

Solution : The calculated values are shown below :

τ (µS)	β_1	α_T
1	10^5	0.99999
0.1	10^4	0.9999
0.01	10^3	0.999

Example 2.2 : Repeat the calculations for a base width of 5 mm.

Solution: The calculated values are shown below :

τ (µS)	β_1	α_T
1	160	0.9937
0.1	16	0.937
0.01	1.6	0.375

These two examples show that for a narrow base transistor suitable for high frequency digital and analog circuit applications, the neutral base recombination does not play a role and as result the base recombination lifetime is relatively unimportant. However, transistors designed for power applications have a large base width where neutral base recombination plays an important role and thus it is crucial to obtain high recombination lifetime.

Current Gain Component due to injection of holes into the emitter

For a uniformly doped emitter, the hole current can be assumed to be a diffusion current because low-level injection condition is always valid in emitter due to its heavy doping level:

$$I_{B2} = -AqD_P \frac{dp}{dx} \qquad (17)$$

The emitter region is as thin or even thinner than the base so that despite a lower recombination lifetime, the number of holes lost due to recombination within the emitter can be assumed to be negligible.

⇒The hole density profile will be linear:

$$\Delta p(x) = \Delta p(0) + \frac{x}{W_E}\Delta p(x_1) \qquad (18)$$

$$\text{where } \Delta p(x_1) = \frac{n_{ie}^2}{N_D} \times (\exp(\frac{qV_{BE}}{kT}) - 1) \qquad (19)$$

n_{ie} is the intrinsic carrier concentration in the emitter. The intrinsic carrier concentration depends on the band gap:

$$n_i = \sqrt{N_V N_C} \times \exp(-\frac{E_g}{2kT}) \quad (20)$$

Heavy Doping Effects

For small and moderate doping levels, the bandgap is independent of doping. But at higher doping levels, the bandgap decreases with increase in doping level i . Due to different doping levels in the base and emitter, the bandgap and therefore the intrinsiccarrier concentrations will also be different.

$\Rightarrow$ There are a variety of empirical models for this bandgap narrowing effect, the simplest of which is:

$$\Delta E_g = (18mV) \times \ln(\frac{N}{N_o}) \quad \text{for} \quad N > N_o \quad (21)$$

$$\Delta E_g = 0 \quad \text{for} \quad N \leq N_o$$

For P-type Silicon: $N_o = 10^{17}\, cm^{-3}$

For N-type Silicon: $N_o = 7 \times 10^{17}\, cm^{-3}$

$\Rightarrow$ The base current due to hole injection into the emitter can then be expressed as:

$$I_{B2} = Aq\frac{D_P}{W_E}\Delta p(x_1) \quad (22)$$

Using Eq. (12) and Eq. (22) and the expressions for $\Delta p(x_1)$, $\Delta n(x_2)$, derived earlier, the current gain can be expressed as:

$$\beta_2 = \frac{n_{ib}^2}{n_{ie}^2} \times \frac{D_N W_E N_D}{D_P W_B N_A} \quad (23)$$

Using the terms ΔE_{ge} and ΔE_{gb} to denote the band gap narrowing in emitter and base regions respectively, the current gain can be re-written as :

$$\beta_2 = \exp(-\frac{(\Delta E_{ge} - \Delta E_{gb})}{kT}) \times \frac{D_N W_E N_D}{D_P W_B N_A} \quad (24)$$

Use of bandgap narrowing model described by Eq. (21) allows the above equation to be simplified to :

$$\beta_2 = 3.9 \times \frac{D_N W_E N_D^{0.3}}{D_P W_B N_A^{0.3}} \quad \text{at T=300K} \qquad (25)$$

Eq. (25) assumes that base doping is larger than $10^{17}\,cm^{-3}$.

Example 2.3 : Determine the current gain component due to hole injection into the emitter without taking band gap narrowing into account, for the following set of values:

$D_N = 17\,cm^2/s$, $N_A = 2 \times 10^{17}\,cm^{-3}$ and effective base thickness $W_B = 0.15\,\mu m$

$D_P = 2.6\,cm^2/s$, for an emitter doping $N_D = 2 \times 10^{19}\,cm^{-3}$ and thickness $W_E = 0.1\,\mu m$

Solution : The current gain without bandgap narrowing will be:

$$\beta_{2o} = \frac{D_N W_E N_D}{D_P W_B N_A} = 981$$

Example 2.4 : Calculate the bandgap narrowing in emitter and base and the resulting changes in intrinsic carrier concentrations. Re-calculate the current gain.

Using the expression given in the text $\Delta E_{ge} = 60.34\,meV$ and $\Delta E_{gb} = 12.5\,meV$

The corresponding intrinsic carrier densities are $\frac{n_{ie}}{n_{io}} = 3.2$ and $\frac{n_{ib}}{n_{io}} = 1.27$

The new expression for current gain will be $\beta_2 = \frac{n_{ib}^2}{n_{ie}^2}\beta_{2o} = 154.5$

⇒ The actual current gain may be larger because recombination velocity at the emitter contact is not infinite. As we shall soon see, this component of current gain is by far the lowest so that the overall current gain is equal to it:

$$\beta \cong \beta_2 = \exp\left(-\frac{(\Delta E_{ge} - \Delta E_{gb})}{kT}\right) \times \frac{D_N W_E N_D}{D_P W_B N_A} \qquad (26)$$

⇒ The current gain can be increased by increasing the emitter doping and reducing thebase doping level. The emitter is often doped to the maximum extent possible determined by solid solubility limits so that current gain remains a function of base doping level only.

⇒ The current gain is independent of collector current if low level injection condition prevails in the base.

Example 2.5 : In the example above, suppose the emitter doping is increased to 10^{20}. By what factor will the current gain increase?

Solution : For the new emitter doping $\Delta E_{ge} = 89.3meV \text{ and } \frac{n_{ie}}{n_{io}} = 5.5$. The new current gain will be $\beta_2 = \frac{n_{ib}^2}{n_{ie}^2}\beta_{2o} = 261$. Although the doping has increase by a factor of 5, the current gain has increased only by a factor of 1.7. The bandgap narrowing in emitter makes the dependence of current gain on emitter doping rather weak.

Emitter Injection efficiency

A parameter which is closely related to the current gain component β_2 is the emitter injection efficiency, defined as

$$\gamma_E = \frac{I_n(x_2)}{I_n(x_2) + I_p(x_1)} \quad (27)$$

The emitter injection efficiency is a measure of how efficient the emitter is in injecting electrons into the base. It is important because the entire transistor action is due to the electrons injected into the base. The holes, which are simultaneously injected into the emitter, are useless from this point of view.

Eq. (27) can be re-written as :

$$\gamma_E = \frac{I_C}{I_C + I_{B2}} = \frac{1}{1 + \beta_2^{-1}} \quad (28)$$

Using the expression for β_2 derived earlier, the emitter injection efficiency can be written as,

$$\gamma_E = \left(1 + \frac{n_{ie}^2}{n_{ib}^2} \times \frac{D_P W_B N_A}{D_N W_E N_D}\right)^{-1} \quad (29)$$

As noted earlier, the overall current gain is determined primarily by the base current resulting from injection of holes into the emitter. As a result, the current gain can be expressed in terms of emitter injection efficiency :

$$\beta \cong \frac{\gamma_E}{1 - \gamma_E} \quad (30)$$

A high emitter injection efficiency is therefore necessary to obtain a high current gain.

The component of base current due to recombination within the emitter-base space charge region can be expressed using Eq. (89) in the chapter on PN junctions as:

$$I_{B3} = Aq\frac{n_{ib}}{2\tau_N}W_{eff} \times \exp(\frac{qV_{BE}}{2kT}) \qquad (31)$$

It is assumed here that the space charge region lies entirely on the base side because the base doping is much smaller than the emitter doping. Using the expression for collector current given by Eq. (1), we obtain :

$$\beta_3 = \frac{I_C}{I_{B3}} = \sqrt{\frac{D_N}{qN_A W_B W_{eff}^2}} \times \tau_N \times \sqrt{J_C} \qquad (32)$$

This component of current gain is sensitive to recombination lifetime and varies with collector current density.

Example 2.6: Determine the current gain component due to SCR recombination for current densities of $1, 10^{-2}, 10^{-3}, 10^{-4}$ and 10^{-6} A/cm^2.
Take $D_N = 17\,cm^2/s$, $\tau_N = 1\mu s$, base doping of $2 \times 10^{17}\,cm^{-3}$, effective base thickness of $W_B = 0.15\mu m$ and effective width of SCR recombination $W_{eff} \sim 100\,A^o$.

Solution : Using Eq. (32) we obtain

$$\beta_3 = 5951 \times \sqrt{J_C\,(A/cm^2)} \qquad (33)$$

$J_C(A/cm^2)$	β_3
1	5951
10^{-2}	595
10^{-3}	188
10^{-4}	59.5
10^{-6}	5.95

The table above shows that for current densities larger than 10^{-2} A/cm^2, the SCR recombination will not play much of a role. The table also shows that current gain larger than unity can be obtained for current densities as low as 10^{-6} A/cm^2, which for a transistor area of $100\ \mu m \times 100\ \mu m$ translates into a collector current of 100pA!. This limit of course is very sensitive to recombination lifetime in the space charge region.

The last component of base current I_{B4} is due to holes which are generated in the collector-base space charge region. Unlike other base current components, this current is negative and can be described using the expression for reverse saturation current of PN junction diode:

$$I_{B4} = -Aq \frac{n_i}{2\tau_{NC}} \times W_{CB} \quad (34)$$

Since collector is lightly doped as compared to the base, most of the space charge region would be in collector. As a result recombination lifetime for collector has been used. W_{CB} is the collector-base depletion width.

- A quick calculation shows that this base current component is much smaller than the rest and can be ignored.
- There are however, situations where this current becomes the most important base current component. For example, there are devices called phototransistors, which are basically BJTs whose excitation is optical instead of electrical.

As a result of optical excitation, electron-hole pairs are generated, mostly in the collector-base depletion region. The phototransistor is operated with its base floating so that

$$I_B = 0 \quad (35)$$

However, internally what happens is that:

$$-I_{B4} = I_{B1} + I_{B2} + I_{B3} \quad (36)$$

The optically excited holes supply the rest of the base current components so that they are non-zero. As we noted earlier, the second component will be the largest so that

$$I_{B2} = -I_{B4}$$
$$I_C = \beta I_{B2} \quad (37)$$

The optically excited current gets multiplied by the current gain and appears in an amplified form at the collector.

Overall current Gain

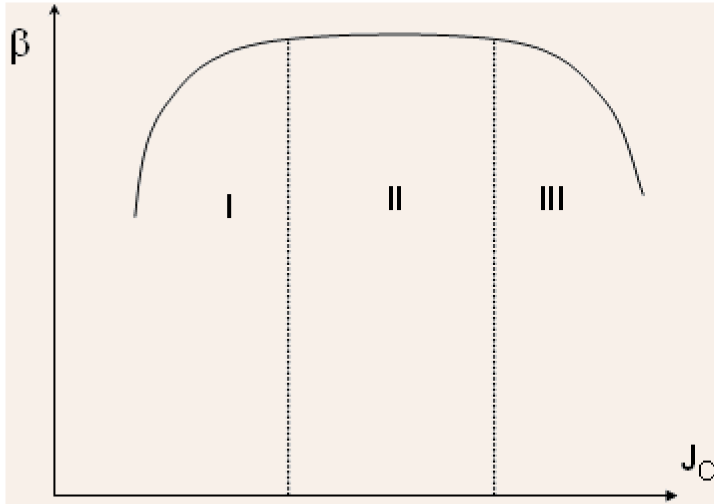
Taking all the components of the base current into account, we can draw the following conclusions:

i. at very low collector current densities, the current gain is determined primarily by recombination within the emitter-base space charge region. In this regime the current gain increases with the collector-current density and is sensitive to the recombination lifetime in the base.

ii. At higher collector current densities, the current gain is determined primarily by injection of holes into the emitter or in other words emitter injection efficiency.

iii. The current gain limited by emitter injection efficiency is independent of collector current as long as low level injection condition prevails in the base. As we shall see later, at very high collector current densities, the current gain begins to decrease.

The variation of current gain with collector current density can be represented by the following Figure:



I - Low current density
II - Medium current density
III - High current density

Example 2.7 (a) By using the values calculated in earlier examples, determine the overall current gain at a collector current density of $1\ A/cm^2$

(b) At what collector current density, does the current gain fall below unity?

(c) If recombination lifetime reduces to $0.1\ \mu s$, determine the collector current density beyond which the current gain will be practically constant.

Solution : (a) The different current gain components are $\beta_1 = 10^5$, $\beta_2 = 154.5$, $\beta_3 = 5951$ $\Rightarrow \beta = 150.36$. $\beta/\beta_2 = 0.97$ so that the current gain is primarily determined by emitter injection efficiency .

(b) This requires that the SCR limited current gain reduce almost to unity so that $J_C \sim 1/\beta_3^2 = 2.8 \times 10^{-8}\ A/cm^2$. This limit is also inversely proportional to $\sim 1/\tau_N^2$ and thus is very sensitive to material purity.

(c) If recombination lifetime is reduced to $0.1\ \mu s$, then for a collector current density of $1\ A/cm^2$, the different components become $\beta_1=10^5$, $\beta_2=154.5$, $\beta_3=595.1$ and $\beta/\beta_2 = 0.79$

If collector current density is increased to $100\ A/cm^2$, then $\beta/\beta_2 = 0.96$. Beyond this the current gain will remain constant until high level injection effects set in.

Most of the time, the BJT is operated in region II where the current gain is constant. In this region, the current gain is given by Eq. (26) (reproduced here again):

$$\beta \cong \exp\left(-\frac{(\Delta E_{ge} - \Delta E_{gb})}{kT}\right) \times \frac{D_N W_B N_D}{D_P W_E N_A} \qquad (38)$$

- As mentioned earlier, the emitter doping is often kept at the maximum allowable limit so as to obtain the highest possible gain.
- A thin, lightly doped base region is needed to obtain a high current gain.
- Since emitter is much more heavily doped as compared to the base, the band gap narrowing is stronger in emitter so that $\Delta E_{ge} > \Delta E_{gb}$. Eq. (38) then shows that the current gain will decrease with decrease in temperature due to the exponential term involving band gap-narrowing terms.

Example 2.8 : Determine the injection efficiency limited current gain for the following sets of temperatures:
$100°C$, $27°C$, $0°C$, $-50°C$ and $77°K$.

The table below shows the variation of current gain with temperature predicted by Eq.(38)

Temperature	Current gain
100 °C	221
27°C	153
0 °C	128
-50 °C	80
77 °K	<1

- The current gain varies significantly with temperature and drops to less than unity at 77 K, which is the temperature of liquid nitrogen.
- Since liquid nitrogen is readily available at a very low cost, it is one of the most widely used methods of cooling a device. It is of interest to cool a BJT so as to remove excess heat from it. Circuits such as emitter coupled logic (ECL) consume lot of power and their scale of integration is limited by the ability to extract heat generated in the bipolar

transistors. A ver simple method of overcoming this would have been to refrigerate the circuit in liquid nitrogen, but as we have seen, the current gain falls to less than unity, making the transistor useless.

- All the discussion regarding current gain so far has been for the dc current gain. As we shall see later, the small signal current gain is also of interest in many cases

$$\beta_{ss} = \frac{\partial I_C}{\partial I_B} \quad (39)$$

The small signal gain is sometimes also called the ac gain because, the small signal is often a sinusoidal signal in many analysis.

The small signal current gain can be expressed in terms of the dc current gain using the following general expressions for collector and base currents:

$$I_C = I_S \times \exp(\frac{qV_{BE}}{n_C kT}) \quad (40)$$

$$I_B = I_{BO} \times \exp(\frac{qV_{BE}}{n_B kT}) \quad (41)$$

- The collector and base currents have ideality factors n_C and n_B respectively. The ideality factor may depart from unity at low or very high current densities.

Using Eq. (39-41), we obtain

$$\beta_{ss} = \frac{n_b}{n_c} \times \beta \quad (42)$$

- The small signal current gain will be different from the dc gain if either the base or the collector ideality factor is different from unity.

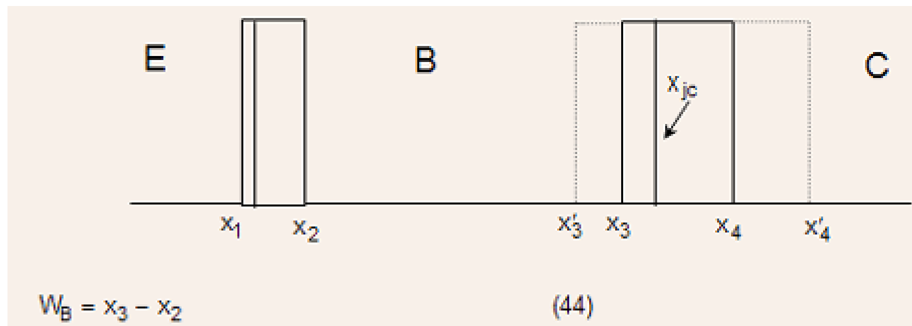
- For example, at very low collector current densities, the base current is dominated by SCR recombination current, which has an ideality factor of 2. Since the collector current has ideality factor of unity at low or moderate currents, the small signal current gain will be twice that of the dc gain.
- At very high collector current densities, we shall see later, collector current has ideality factor greater than unity, while the base current has ideality factor of unity. In this case the small signal gain will be smaller than the dc gain.
- For intermediate collector current densities, the dc gain equals the small signal current gain.

Early Effect[ii]

The collector current in active mode of operation for a uniformly doped base is given by the expression:

$$I_C = qAn_i^2 \frac{D_N}{N_A W_B} \times (\exp(\frac{qV_{BE}}{kT}) - 1) \quad (43)$$

It appears from this equation as if the collector current is independent of the collector- base voltage. However, this is not so because the effective base width W_B depends on the collector bias as illustrated by the Figure below:



$$W_B = x_3 - x_2 \quad (44)$$

- As the collector-base voltage V_{CB} increases, the depletion region associated with the collector base junction widens as illustrated by the Figure above. The base width narrows to $W_B = x'_3 - x_2$. As a result of the decrease in base width, the collector current would increase. This base width modulation effect described first by James Early, is known as the Early Effect.

The change in collector current as a result of change in collector-base voltage can be evaluated using Eq. (43)

$$\frac{\partial I_C}{\partial V_{CB}} = -I_C \times \frac{1}{W_B} \frac{\partial W_B}{\partial V_{CB}} \quad (45)$$

Using Eq. (44), we obtain

$$\frac{\partial W_B}{\partial V_{CB}} = -\frac{\partial x_3}{\partial V_{CB}} \quad (46)$$

Using the expression for the magnitude of collector-base depletion charge, we can obtain the following results:

$$Q_{jc} = qAN_A(x_{jc} - x_3) \quad (47)$$

$$C_{jc} = \frac{\partial Q_{jc}}{\partial V_{CB}} = -qAN_A \frac{\partial x_3}{\partial V_{CB}} \quad (48)$$

$$\frac{\partial I_C}{\partial V_{CB}} = I_C \times \frac{C_{jc}}{qAN_A W_B} \quad (49)$$

Defining a parameter called the Early voltage:

$$V_A = \frac{qAN_A W_B}{C_{jc}} \quad (50)$$

$$\frac{\partial I_C}{\partial V_{CB}} = \frac{I_C}{V_A} \quad (51)$$

- The expression for early voltage shows that its value depends on the collector-base bias but this dependence is weak so that for collector-base voltages not far from zero:

$$I_C(V_{CB}) = I_C(0) + I_C(0)\frac{V_{CB}}{V_A} \quad (52)$$

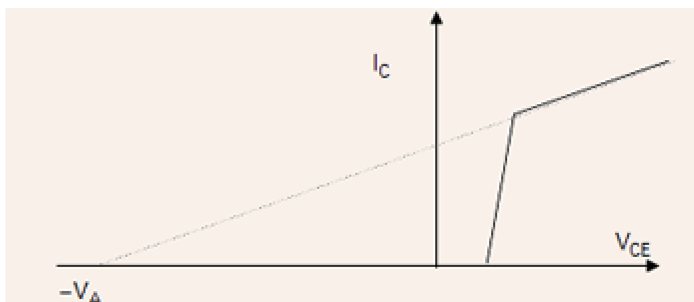
Representing $I_C(0) = I_S \times \exp(\frac{qV_{BE}}{kT})$, we obtain

$$I_C(V_{CB}) = I_S \times \exp(\frac{qV_{BE}}{kT}) \times (1 + \frac{V_{CB}}{V_A}) \quad (53)$$

- Noting that $V_{CB} = V_{CE} - V_{BE}$ and base emitter voltage in active mode being less than unity is often much smaller than the early voltage, we obtain

$$I_C(V_{CE}) = I_S \times \exp(\frac{qV_{BE}}{kT}) \times (1 + \frac{V_{CE}}{V_A}) \quad (54)$$

Eq. (54) shows that the collector current would slowly increase with the collector-emitter voltage as illustrated by the Figure below:



Through extrapolation of the characteristics in active mode, the early voltage can be measured!

Impact of Early effect on Base current

Ignoring the small reverse current of the collector-junction, the base current can be written as

$$I_B = I_{B1} + I_{B2} + I_{B3} \qquad (55)$$

Eq. (22) for I_{B2} and Eq. (31) for I_{B3} show that these currents have no dependence on the collector-base voltage so that

$$\frac{\partial I_B}{\partial V_{CB}} = \frac{\partial I_{B1}}{\partial V_{CB}} = \frac{1}{\beta_1}\frac{\partial I_C}{\partial V_{CB}} \qquad (56)$$

Use of Eq. (51) allows Eq. (56) to be re-written as:

$$\frac{\partial I_B}{\partial V_{CB}} = \frac{I_B}{V_\eta}$$

$$V_\eta = \frac{\beta_1 V_A}{\beta} \qquad (57)$$

Eq. (57) can be re-written in terms of the base transport factor

$$V_\eta = \frac{V_A}{(1-\alpha_T)\beta} \qquad (58)$$

- Because of very small value of base transport factor, the voltage $V_\eta >> V_A$ so that the base current is much less affected by the applied collector bias as compared to the collector current.

Example 2.9 Determine the Early voltage for an NPN transistor having the following characteristics at $V_{CB} = 0$ and -5 Volts.

Take, $D_N = 17\,cm^2/s$, $\tau_N = 1\mu s$, base doping of $2\times10^{17}\,cm^{-3}$, effective base thickness of $W_B = 0.15\mu m$, collector doping of $10^{16} cm^{-3}$.

Solution : The collector depletion width and junction capacitance/area can be calculated to be :

$0.32\ \mu m$ and $3.2 \times 10^{-8}\ F/cm^2$. The early voltage then can be calculated to be: $V_A = 7.4$ Volts.

$V_{bc} = -5$Volts, the collector depletion width and junction capacitance become $0.86\ \mu m$ and $1.1 \times 10^{-8}\ F/cm^2$. The new early voltage ~20Volts.

Non-Uniform Doping

- In all the analysis carried out so far, the doping in emitter, base and collector were all assumed to be uniform. However, in most devices, due to the nature of fabrication, doping in emitter and base is nonuniform, approximating gaussian or error function most of the time. We will next look at the model for collector current when base doping is nonuniform.

- As a result of nonuniform base doping, an electric field is present in the base so that the electron cannot be assumed to be simply diffusive. The magnitude of the electric field can be obtained through analysis of the hole current:

$$J_p = qp\mu_p\varepsilon - qD_p\frac{\partial p}{\partial x} \qquad (59)$$

This hole current supplies the holes that are injected into the emitter or those that recombine in the base. If the hole current is assumed to be very small so that:

$$qp\mu_p\varepsilon - qD_p\frac{\partial p}{\partial x} \cong 0 \qquad (60)$$

$$\Rightarrow \varepsilon = \frac{kT}{q}\frac{1}{p}\frac{dp}{dx} \qquad (61)$$

This expression for electric field can be used to write the expression for electron current or equivalently collector current as:

$$I_N = A(qn\mu_N\varepsilon + qD_N\frac{dn}{dx})$$

$$= AqD_N\frac{n\frac{dp}{dx} + p\frac{dn}{dx}}{p} \qquad (62)$$

Eq. (62) can be re-written as

$$I_n p = AqD_N \frac{d(pn)}{dx} \qquad (63)$$

Integration of this equation across the base gives:

$$I_n = Aq\overline{D}_N \frac{p(x_3)n(x_3) - p(x_2)n(x_2)}{\int_{x_2}^{x_3} p\,dx} \qquad (64)$$

Since doping is nonuniform in the base, the mobility and therefore the diffusion constant would also vary in the base. This variation, however, is small and $\overline{D}_N$ represents its average value in the base.

·The electron density at the edge of reverse biased collector-base junction will be negligible so that $n(x_2) \cong 0$, and the electron-hole product at the edge of emitter-base junction can be written as :

$$n(x_2)p(x_2) = n_i^2 \times \exp(\frac{qV_{BE}}{kT}) \qquad (65)$$

Eq. (64) can now be simplified to

$$I_C = Aq\overline{D}_N \frac{n_i^2 \exp(\frac{qV_{BE}}{kT})}{Q_B} \qquad (66)$$

$$Q_B = q\int_{x_2}^{x_3} p\,dx \qquad (67)$$

·Eq. (66) is known as the Gummel-Poon[iii] representation of the collector current. A model forBJT will be derived using this equation later on.

Q_B is known as the Gummel number for the base. Using quasi-neutrality condition for the base:

$$p(x) = N_A + n(x) \qquad (68)$$

For low level injection: $n(x) << N_A$ so that

$$Q_B \cong q \int_{x_2}^{x_3} N_A \, dx \qquad (69)$$

For uniformly doped base :

$$Q_B = qN_A W_B \qquad (70)$$

Substitution of Eq. (70) in Eq. (66) gives the same expression for collector current as Eq. (1) that we have used throughout the previous analysis.

Example 2.10 (a) An NPN transistor has a nonuniform doping in the base described by the expression : $N_A(x) = N_0 \times \exp(-ax)$ with $N_0 = 3.16 \times 10^{17}$ cm^{-3} and $a = 5 \times 10^4$ cm^{-1}.

Assume that average diffusion constant in the base is 17 cm^2/s and that effective base thickness is 0.2 µm. Determine the magnitude and direction of electric field in the base
(b) Obtain an expression for electron density profile within the base
(c) Determine the injection efficiency limited current gain at room temperature. Assume that

$D_p = 2.6$ cm^2/s , emitter doping $N_D = 2 \times 10^{19}$ cm^{-3} and emitter thickness $W_E = 0.1$ µm.

(a) Under low level injection conditions, $p(x) \cong N_A(x)$ so that

$$\varepsilon = \frac{kT}{q}\frac{1}{N_A}\frac{dN_A}{dx} = -\frac{kT}{q}a = -1.3 \times 10^3 \text{ V/cm}$$

The negative sign indicates that the field is directed towards emitter. It therefore aids in drift of electrons towards the base.

(b) The electron current density within the base can be written as

$$J_N = (qn\mu_N\varepsilon + qD_N \frac{dn}{dx})$$

Using the expression for electric field obtained earlier, we can re-write the above expression as

$$\frac{dn}{dx} + na = \frac{J_N}{qD_N}$$

This equation can be expressed as

$$\exp(-ax)\times\frac{d(n\exp(ax))}{dx}=\frac{J_N}{qD_N}$$

Integration of the above expression from x to WB with the boundary condition that electron density at the collector end of the base is negligible, we obtain

$$n(x)=\frac{J_N}{qaD_N}(\exp(a(W_B-x))-1)$$

(c) The expression for current gain can be modified to

$$\beta_2=\frac{n_{ib}^2}{n_{ie}^2}\times\frac{D_N N_D W_E}{D_P\times Q_B/q}$$

For calculating the band gap narrowing in the base, we find the average doping in the base and use it in the expression given in the text. With that we obtain a current gain of 116.

High Level Injection Effects

- There are two high-level injection effects that we shall discuss here:

  (i) High Level injection in Base
  (ii) High level injection in Collector

  We shall consider high level injection in the base, also termed as Webster effect[i], effecti, first:

  For simplicity, we shall consider a uniformly doped base. The expression for collector current that has been used in the past discussion was based on low level injection approximation. This approximation allowed the collector current to be expressed as a purely electron diffusion current:

$$I_C=-AqD_N\frac{dn}{dx}\qquad(1)$$

  Neglect of recombination results in linear variation of electron density across the base so that

$$I_C=AqD_N\frac{\Delta n(x_2)}{W_B}\qquad(2)$$

The second simplification that low level injection resulted in was that the expression:

$$n_i^2 \times \exp(\frac{qV_{BE}}{kT}) = n(x_2)p(x_2) = n(x_2)^2 + n(x_2)N_A$$

$$\cong n(x_2)N_A \qquad (3)$$

With these three simplifications, the collector current can be written as:

$$I_C = AqD_N \frac{n_i^2}{N_A} \times \exp(\frac{qV_{BE}}{kT}) \qquad (4)$$

When the injected electron density begins to become comparable to the base doping, the electron current begins to have an appreciable drift component as well. Assuming that the hole current in the base and therefore the base current is small, we can obtain the magnitude of the electric field in the base:.

$$J_P = qp\mu_P\varepsilon - qD_P \frac{dp}{dx} \cong 0 \qquad (5)$$

$$\Rightarrow \varepsilon = \frac{kT}{q}\frac{1}{p}\frac{dp}{dx} \qquad (6)$$

With this expression, the collector current can be expressed as sum of drift and diffusion terms:

$$I_C = -AqD_N(\frac{dn}{dx} + \frac{n}{p}\frac{dp}{dx}) \qquad (7)$$

The base is quasi-neutral and uniformly doped as that:

$$\frac{dn}{dx} = \frac{dp}{dx} \qquad (8)$$

This allows Eq. (7) to be written as

$$_C = -AqD_N \frac{dn}{dx}(1 + \frac{n}{p}) \qquad (9)$$

Eq. (9) shows that when injected electron density is much smaller, than the hole density in the base, the current is same as before and diffusive.

At very high current densities, $n \sim p$ so that

$$I_C = -Aq2D_N \frac{dn}{dx} \qquad (10)$$

In general, the collector current can be expressed as

$$I_C = -Aq\eta D_N \frac{dn}{dx} \qquad (11)$$

where $1 \leq \eta \leq 2$

Under high-level injection conditions, the approximation made in Eq. (3) is no longer

valid so that a full expression for electron density has to be used:

$$n_i^2 \times \exp(\frac{qV_{BE}}{kT}) = n(x_2)p(x_2) = n(x_2)^2 + n(x_2)N_A$$

$$\Rightarrow n(x_2) = \frac{-N_A + \sqrt{N_A^2 + 4n_i^2 \times \exp(\frac{qV_{BE}}{kT})}}{2} \qquad (12)$$

The electron density and therefore the collector current no longer increase as $\sim \exp(qV_{BE}/kT)$ but as $\sim \exp(qV_{BE}/nkT)$,where the ideality factor is unity for small base-emitter voltages but tends to approach 2 at high biases.

The collector current under high level injection can therefore be modeled as

$$I_C = Aq\eta D_N \frac{n_i^2}{N_A} \times \exp(\frac{qV_{BE}}{nkT}) \qquad (13)$$

- The major impact of high level injection is that the ideality factor of the collector current departs from unity and approaches a value of 2.

- The base current, being due to injection of holes into the emitter, continues to have an ideality factor of unity because low level injection conditions continue to prevail in emitter due to its heavy doping level.

- The increase in ideality factor of collector current and a constant ideality factor of unity for base current means that the collector current does not increase as rapidly as the base current with increase in bias. As a result current gain will decrease!

$$\beta = \frac{J_C}{J_B} \propto \frac{\exp(\frac{qV_{BE}}{nkT})}{\exp(\frac{qV_{BE}}{kT})} \qquad (14)$$

$$\Rightarrow \beta \propto (\exp(\frac{qV_{BE}}{nkT}))^{1-n}$$

$$\Rightarrow \beta \propto J_C^{1-n} \qquad (15)$$

Eq. (15) shows that current gain will decrease with increase in collector current density. For very high collector current densities:

$$\beta \propto \frac{1}{J_C} \qquad (16)$$

- This rapid decrease in current with increase in collector current density means that the device quickly becomes useless since almost all applications require at least a moderate value of current gain.

- As a measure of the collector current density at which high level injection effects begin to become appreciable, we take
  $n(x_2) = N_A$ so that
  $$J_C^{HL} \sim qAD_N \frac{N_A}{W_B} \quad (17)$$

- When the injected electron density becomes comparable to the background doping in the base, it begins to affect the collector-base depletion region on the base side also. The charge density given by:
  $$\rho = q(p - N_A - n) \quad (18)$$

  would increase because electron density adds to the charge due to acceptor atoms. This increase in charge density would result in decrease in depletion width on the base side thereby increasing the effective base width.

- This base pushout effect, besides decreasing collector current and therefore the current gain, will also tend to increase the base transit time.

- There is a counteracting effect due to the presence of electric field in the base but in general the onset of high level injection results in degradation of unity gain frequency.

  Conductivity Modulation

  The high-level injection phenomenon is also called conductivity modulation. The conductivity of base can be expressed as
  $$\sigma = q(n\mu_N + p\mu_P) \quad (19)$$
  $$p = n + N_A \quad (20)$$
  $$\sigma = q\mu_P N_A + qn(\mu_N + \mu_P) \quad (21)$$

  Eq. (21) shows that as the injected electron density begins to become comparable to the base doping, the base conductivity will begin to increase with increase in collector density.

- ---

  Example 4.1 (a)For an NPN transistor whose description is given below, determine the collector current after which the current gain will begin to fall.
  Base : N_A= 2 x 10^{17}cm^{-3}, W_B=0.15 µm, D_N = 17 cm^2/s
  Area = 10^{-4}cm^2
  (b)Discuss how can you double the value of maximum collector current.

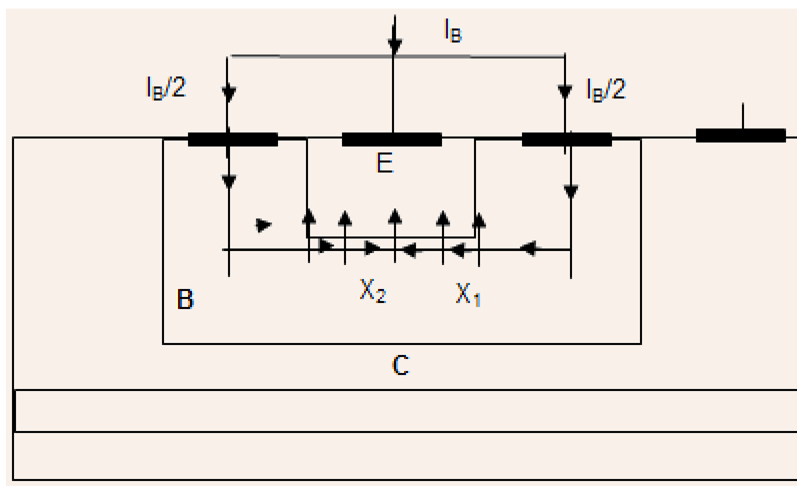
  Solution : The collector current density and collector current can be found using Eq. (17) to be 3.6 x 10^4 A/cm^2 and 3.62 mA respectively.

  (b) An obvious way of doubling the collector current would be to double the transistor

area. Besides occupying more area on Silicon wafer, this approach would also double all the junction capacitances thereby adversely affecting the frequency response of the device. Another approach could be by doubling the base doping. This would however decrease the current gain also. A third method could be by reducing the base thickness. This would decrease the Early voltage and as we shall see in the next section also lower what is called the punchthrough voltage of the transistor.

Current Crowding

The Figure below shows a simplified schematic of a Bipolar Transistor :



The structure has two base contacts, one on each side of the emitter. As a result of the lateral resistance of intrinsic base region
(R_{Bi}), there will be a voltage drop across it.

This means that the potential at the edge of the emitter (point x_1) will be larger than the potential in the middle (point x_2). Since the
emitter voltage is the same everywhere,

$$V_{BE}(\text{edge}) > V_{BE}(\text{middle}) \quad (22)$$

This means that a larger fraction of collector current would flow at the edges than at the center.

- This phenomenon will increase as the current increases so that at very high currents, the entire collector current may flow only at the edges. This is known as current crowding. There are several adverse effects of current crowding at the emitter including

  (i) Since all the current is confined at the emitter periphery, all the power will also be dissipated in a very narrow region resulting in localized heating which may damage the transistor

(ii) Since the current flows in a very narrow area close to the emitter edge, the collector current density will become very high resulting in early onset of high-level injection effects.

The onset of current crowding effect may be taken as the base current at which the lateral voltage drop becomes comparable to the thermal voltage V_T:

$$\frac{I_B(CC)}{2} \times \frac{R_{Bi}}{2} \sim V_T \qquad (23)$$

$$\Rightarrow I_C(CC) = \beta \frac{4V_T}{R_{Bi}} \qquad (24)$$

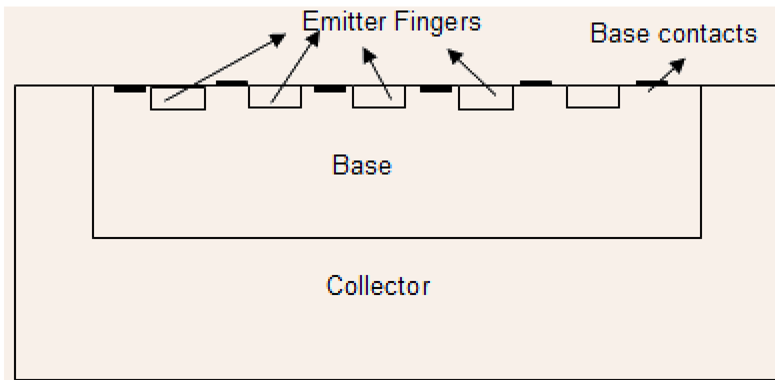
The lateral base resistance can be expressed as

$$R_{Bi} = \frac{1}{q\mu_P N_A} \frac{L_e}{W_e W_B} \qquad (25)$$

where L_e, W_e are emitter length and width respectively.
Using Eq. (25) and (24) , the collector current density for the onset of current crowding can be expressed as:

$$J_C(CC) = 4V_T\beta \frac{q\mu_P N_A W_B}{L_e^2} \qquad (26)$$

Eq. (26) shows that current crowding decreases as the emitter length is reduced. It is for this reason that high power bipolar transistors are fabricated using a finger like structure with multiple emitter regions sandwiched between base contact regions as shown in the Figure below:



- Eq. (26) also seems to suggest that an increase in base doping and thickness will also reduce current crowding due to decrease in base resistance. However, an increase in base doping or thickness will also reduce the current gain. If the expression for current gain derived earlier is taken, then it can be shown that $J_C(CC)$ is independent of base thickness and increases as $N_A^{0.7}$.

Example 4.2 For an NPN transistor whose description is given below, determine
(a) the maximum collector current limited by current crowding considerations
(b) Discuss different ways by which the maximum collector current can be increased by a factor of 9.

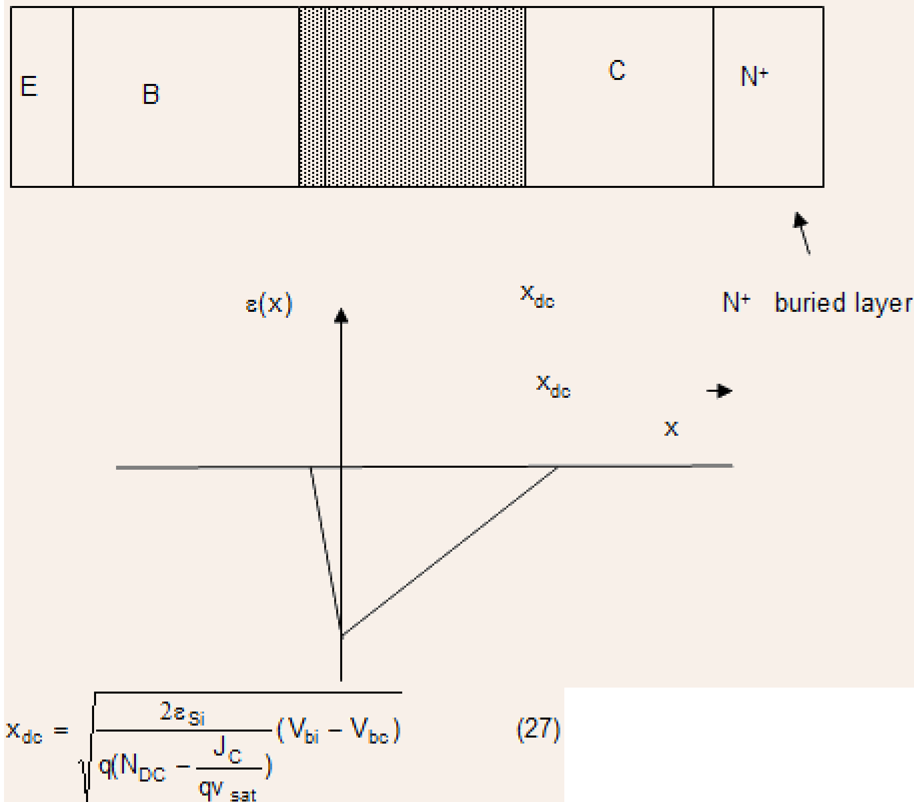
Base : $N_A = 2 \times 10^{17} cm^{-3}$, $W_B = 0.15\ \mu m$, $D_N = 17\ cm^2/s$, $\mu_P = 280\ cm^2/V/s$
Emitter Area = 5µm x 5µm.

Solution : $\beta = 154.5$ can be calculated for this case. Using Eq. (26) we obtain $J_C(CC) = 8.6 \times 10^3\ A/cm^2$ and $I_C(CC) = 2.16$ mA.

(b)To increase collector current by a factor of 9, let us increase the transistor area by same factor to 15µm x 15µm. However, now $J_C(CC) = 9.55 \times 10^2\ A/cm^2$ so that maximum collector current remains $I_C(CC) = 2.16$ mA. So a simple scaling up of transistor area will not work. Suppose we decrease the emitter length while keeping the transistor area the same so that now the transistor has a rectangular (1.66µm x 15µm) rather than square geometry. Now the current crowding limited current density increases to $J_C(CC) = 7.74 \times 10^4\ A/cm^2$ and $I_C(CC) = 19.44$ mA, which is the desired value. However, this approach requires advanced lithography to obtain a dimension of 1.6µm, which may pose problems. A third approach could be to use 9 emitter fingers, each of dimension 5µm x 5µm in a manner shown in the Figure on the previous page. Now $J_C(CC)$ remains same at $8.6 \times 10^3\ A/cm^2$ but total collector current increases nine times. There are other ways of increasing the collector current by increasing base doping etc but each comes at a price.

High Level Injection in Collector

We shall assume for the most part that collector doping is considerably smaller than the base doping. As a result, the collector depletion region would lie mostly on the collector side as illustrated by the electric field profile shown below:



$$x_{dc} = \sqrt{\frac{2\varepsilon_{Si}}{q(N_{DC} - \frac{J_C}{qv_{sat}})}(V_{bi} - V_{bc})} \qquad (27)$$

The effective doping in the collector is less than the background doping due to the storage of electrons in the depletion region required for flow of collector current.

Kirk Effect[ii]

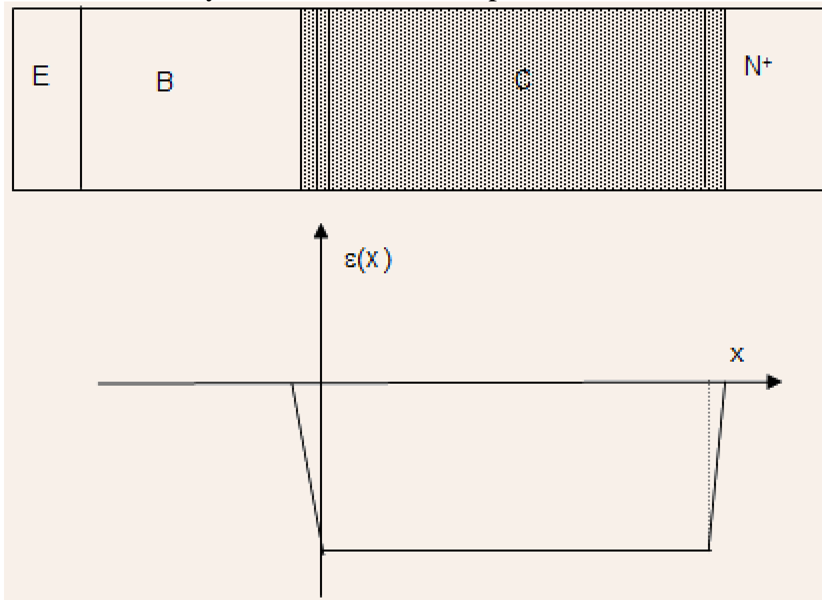
- As the collector current density increases, the effective collector doping decreases resulting in an increase of collector depletion width. This would result in increase of collector transit time τ_C and a decrease in unity gain frequency.

Eq. (27) predicts an infinite depletion width for a collector current density of

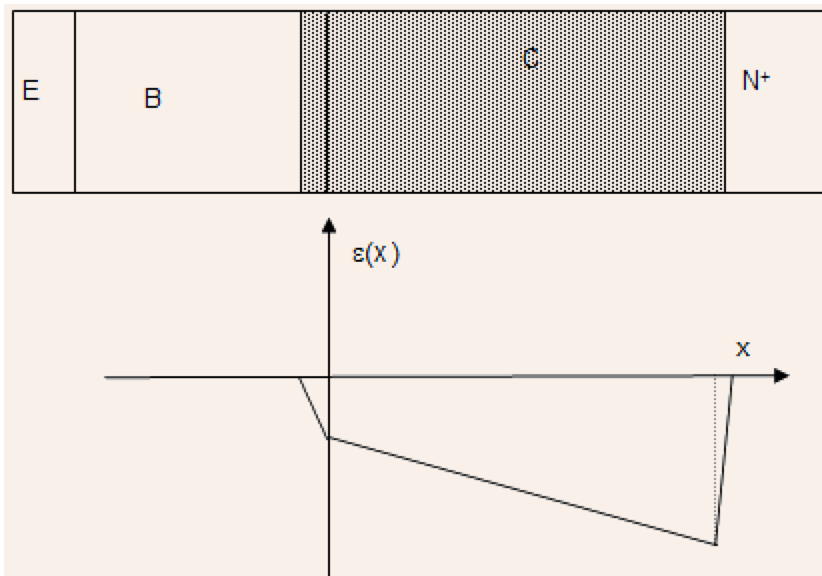
$$J_K = qN_{DC}v_{sat} \qquad (28)$$

However, the depletion width does not become infinite but simply touches the N^+ buried layer. The effective collector doping being zero for this case, the collector can be though

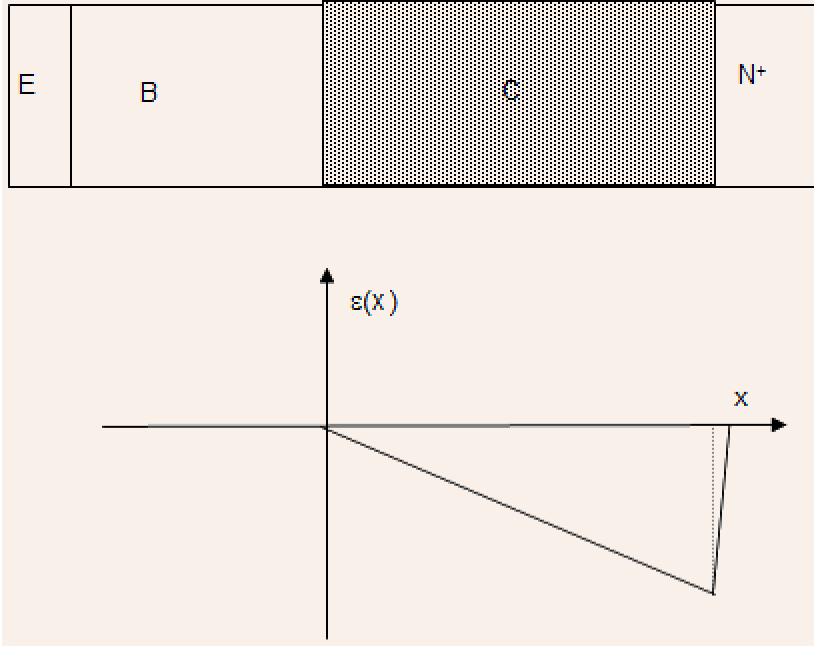
of as intrinsic layer. The electric field profile for this case will be



- The collector-base junction is like a PIN junction now!
- With further increase in collector current density, the effective collector doping becomes negative, implying a P-like semiconductor layer. The junction now shifts to the buried layer as illustrated by the Figure below:

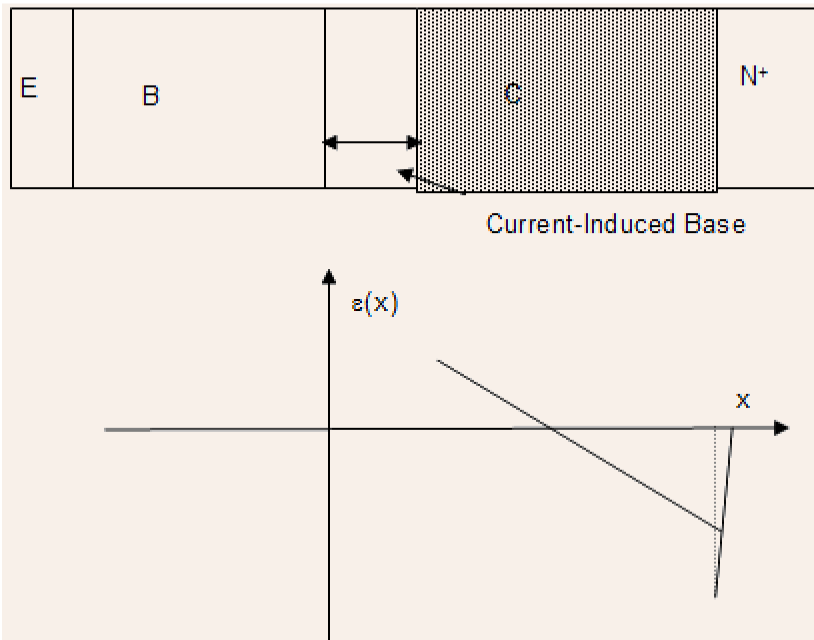


With continued increase in collector current density, the depletion region withdraws from the base completely as illustrated by the next Figure.



·With even further increase in collector current density, the effective P-like doping increases further so that the space charge region begins to withdraw even from the collector region leaving behind a quasi-neutral collector region called a current induced base!

This is illustrated in the next Figure.



The formation of current-induced base further degrades the base transit time and unity

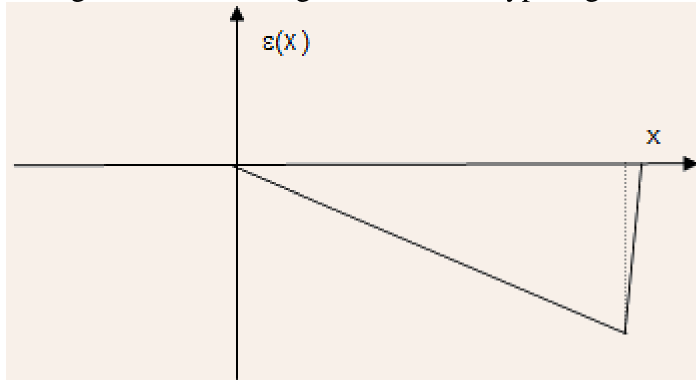
gain frequency!

The onset of high level injection or the Kirk effect can be taken as the collector current density given by Eq. (28). At this collector current density, the collector becomes like an intrinsic layer!

Example 4.3 For an NPN transistor whose description is given below, determine the collector current after which the high level injection will begin to occur in the collector:
Base : $N_A = 2 \times 10^{17} cm^{-3}$, $W_B = 0.15$ µm, $D_N = 17\ cm^2/s$
Collector: The collector consists of a lightly doped region of doping $N_{DC} = 2 \times 10^{16} cm^{-3}$ and thickness 0.5µm followed by a thick heavily doped N-layer.
Assume that V_{cb}=2 Volts and V_{bi}=0.8 Volts
Area = $10^{-7} cm^2$
(b) At what current will the current induced base begin to form.

Solution : The Kirk current density is given by $J_K = qN_{DC}v_{sat} = 3.2 \times 10^4 A/cm^2$. The collector current can then be calculated to be 3.2 mA

(b) When the current induced is about to form, the collector junction has shifted to the N^+ region and the N-region is like a P-type region as shown below



The effective P-type doping of $N_{eff} = \frac{J_C}{qv_{sat}} - N_{DC}$ The total depletion width, when current induced just begins to form now is W_C so that

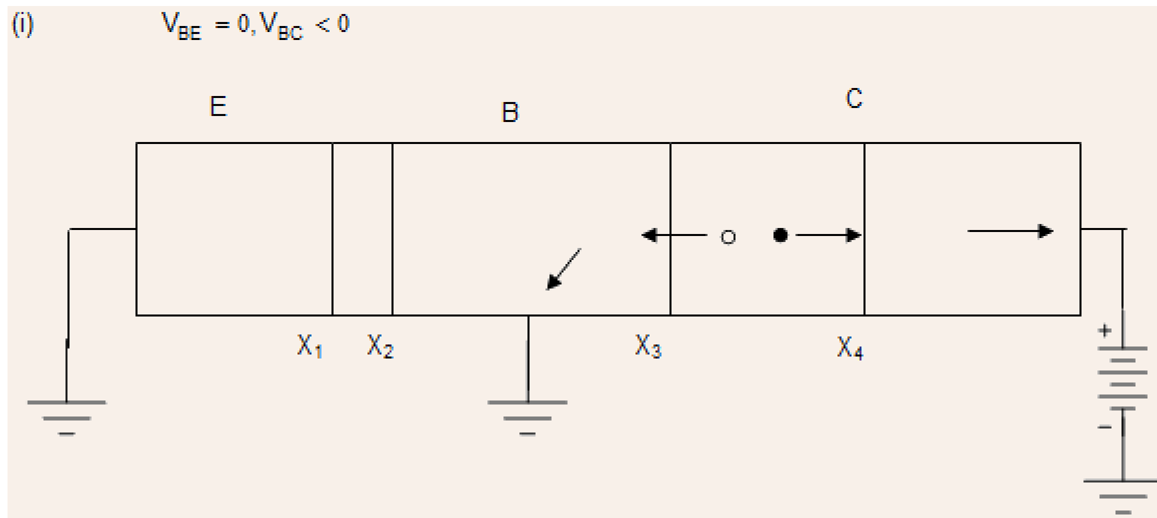
$$\sqrt{\frac{2\varepsilon_S}{qN_{eff}} \times (V_{bi} - V_{bc})} = W_C$$

This gives $N_{eff} = 1.44 \times 10^{16} cm^{-3}$ which then gives a collector current density of 5.5 x 10^4 A/cm^2 This translates into a collector current of 5.5 mA.

Cutoff Mode

- In cutoff mode, none of the junctions are forward biased so that normally the current is very small and relatively constant with respect to the applied bias.

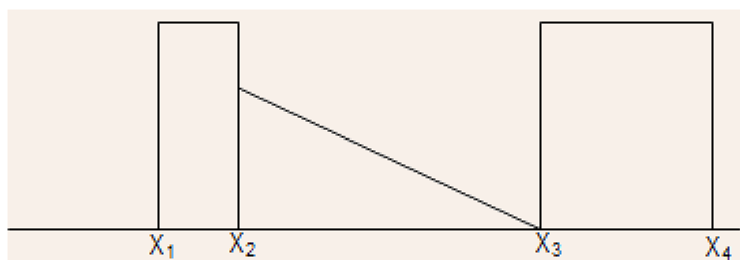
We will consider several bias conditions and determine the magnitude of leakage current for each case.



- The collector current is primarily due to thermal generation in the collector base depletion region $(x_3 \rightarrow x_4)$. The electrons generated in this region are swept by the electric field into the neutral collector region from where they flow to the contact. The holes generated are swept into the base from where they flow out of the contact. Let us call this current I_{CO}.

- The current due to generation and diffusion of carriers outside the collector depletion region also contributes to current flow but is usually much smaller than the generation current in the collector junction.

- However, this small current results in a non-zero current flowing through the emitter even though emitter-base junction is at zero bias..



Due to reverse bias, the electron density $n(x_3) \cong 0$, but electron density at the other end is

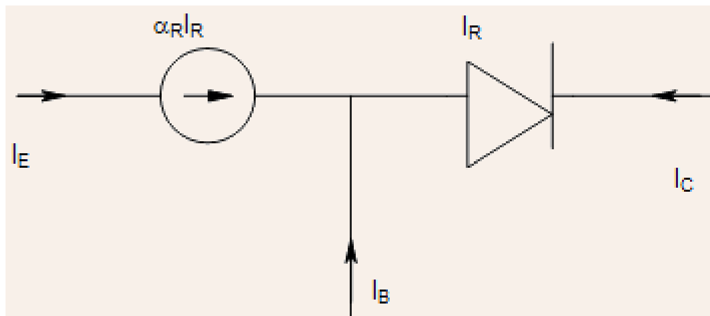
$n(x_2) = n_i^2 / N_A$.

As a result an electron would flow:

$$I_N = I_E = -qAD_N \frac{n_i^2 / N_A}{W_B} \quad (1)$$

This result can also be obtained using the Ebers-Moll model derived earlier.

For $V_{BE} = 0, V_{BC} < 0$, the Ebers-Moll Model can be simplified to

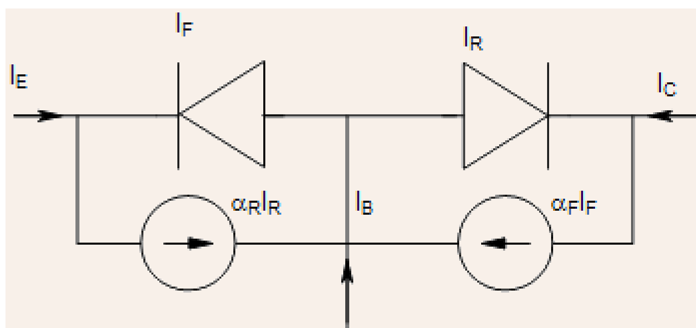


Since collector-base diode is reverse biased:

$$-I_C = I_R = -I_{CS} = -\frac{I_S}{\alpha_R} \quad (2)$$

$$I_E = \alpha_R I_R = -I_S \quad (3)$$

- The expression for emitter current agrees with that estimated earlier (Eq. (1)) but the collector current is underestimated by Eq. (2) because of neglect of generation/recombination in the space charge regions during the derivation of Ebers-Moll Model.

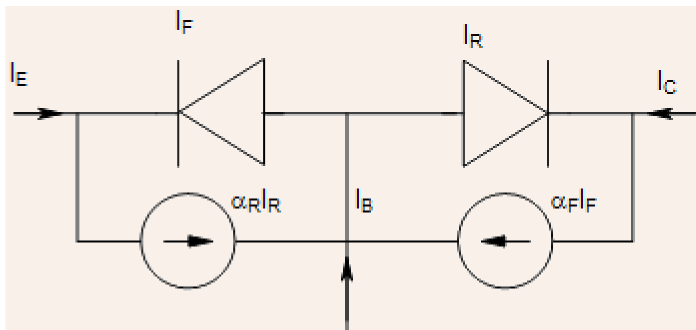


(ii) Let us consider a situation where emitter is open $(I_E = 0)$ and collector base junction reverse biased as usual

Like the first case, the collector current is primarily due to thermal generation in the collector base depletion region. The generated electrons and holes flow to collector and base contacts respectively.

- The emitter junction is slightly reverse biased even though emitter current is zero. This can be understood by noting that if emitter junction were zero biased, an electron current given by Eq. (1) would tend to flow, which is impossible here. To counteract this current the junction gets slightly reverse biased so that the reverse current of the junction cancels this diffusion current.

This can also be understood using Ebers-Moll Model :

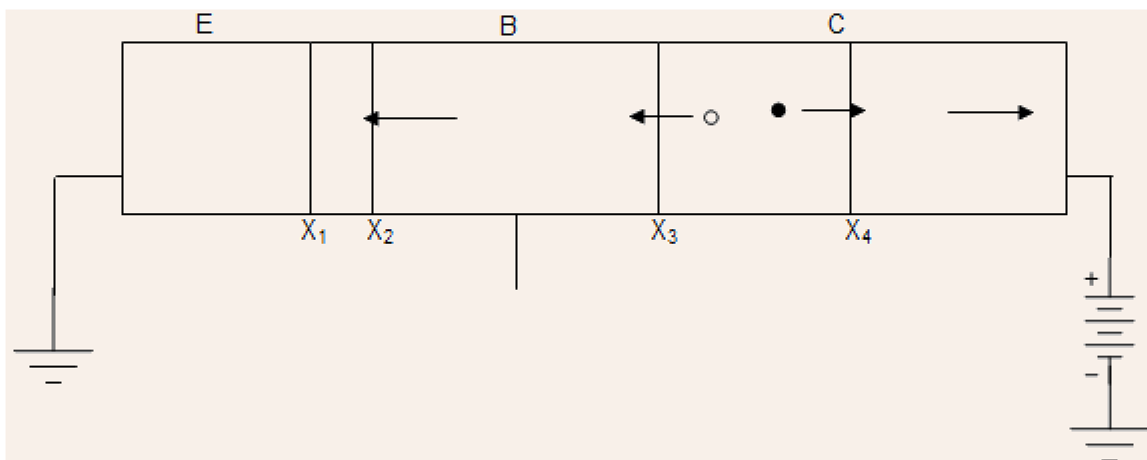


As usual $I_R = -I_{CS}$. Because the emitter current is zero:

$$I_F = \alpha_R I_R = -\alpha_R I_{CS} \quad (4)$$

The negative direction of I_F denotes that emitter junction is reverse biased.

(iii) Base is open, emitter grounded and collector reverse biased



- Since the collector-base junction is reverse biased, electron-hole pairs will be generated. While the electrons flow to the collector contact as usual, the holes cannot flow out of the base contact because it is floating.

- As indicated by the Figure, the only place where the excess holes can go is into the emitter. The injection of holes into the emitter implies forward biasing of the emitter-base junction. For every hole injected into the emitter, several electrons will be injected by the emitter into the base. These electrons would travel across the base and contribute to collector current.

The current flowing through the collector will therefore be the sum of the normal generation current I_{CO} and a current βI_{CO} due to the injected electrons. The net current can then be written as

$$I_{CEO} = (\beta + 1) I_{CO} \quad (5)$$

The multiplication of collector-base leakage current can also be derived using the Ebers-Moll Model. A zero value for base current implies:

$$I_B = (1 - \alpha_F) I_F + (1 - \alpha_R I_R) = 0 \quad (6)$$

$$\Rightarrow I_F = \frac{1 - \alpha_F}{1 - \alpha_R} I_{CS} \quad (7)$$

$$\Rightarrow I_C = I_{CS} + \beta_F (1 - \alpha_R) I_{CS} \quad (8)$$

- The expression for current gain that is appropriate for estimating the leakage current is the current gain at very low current densities limited primarily by space charge recombination in the emitter junction. The magnitude of this current gain can be significantly smaller than the current gain at moderate collector current densities.

Example 5.1 Determine the reverse leakage currents under open emitter and open base conditions for an NPN transistor whose description is given below:

Base : N_A=2x 10^{17}cm^{-3}, W_B=0.15 µm, τ_N=1µs
Emitter : N_D=2x 10^{19}cm^{-3}, W_E=0.1 µm
Collector : N_{DC}=2x 10^{16}cm^{-3}, τ_N=τ_P= 1µs
Area = 10^{-4}cm^2. Assume a collector-base bias of V_{CB}=2 Volts .

Solution : The collector depletion width (W_C)can be calculated to be 0.42 µm

$$I_{CBO} = q \frac{n_i}{\tau_N} \times W_C \times \text{Area} = 9.7\text{pA}$$

To calculate the current under open base conditions, we need a value of current gain. Let us initially assume that $\beta \cong \beta_2 = 154.5$. For this value $I_{CEO} = 1.5nA$. This corresponds to a collector current density of 1.5×10^{-5} A/cm^2. Since this is quite small, so current gain component β_3 will not be very high. This can be calculated using expressions developed in an earlier chapter as

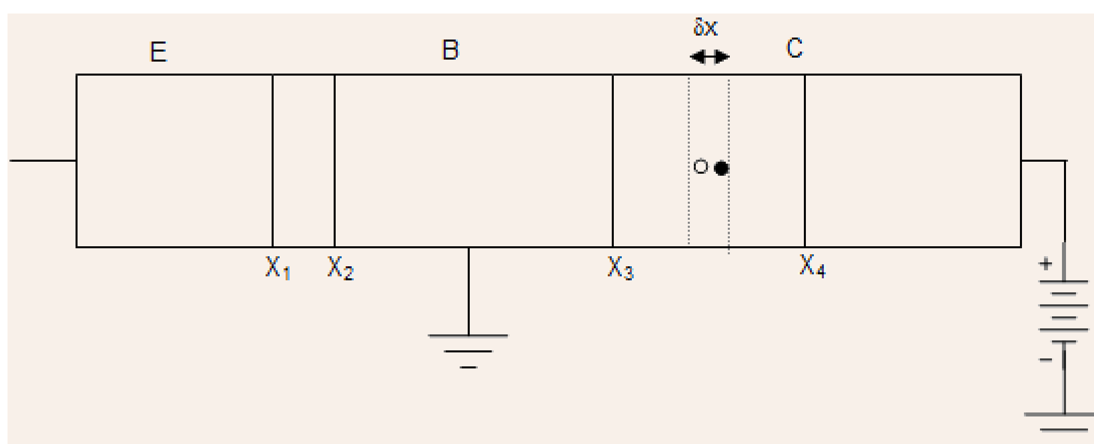
$$\beta_3 = \frac{I_C}{I_{B3}} = \sqrt{\frac{D_N}{qN_A W_B W_{eff}^2}} \times \tau_N \times \sqrt{J_C} = 23.089$$

The new value of current gain turns out to be 20, and I_{CEO} = 0.195nA.

- The current flowing in the cutoff mode remains small, till impact ionization induced breakdown occurs. We shall consider two cases of breakdown:

(i) Breakdown with emitter open circuited and base grounded.
(ii) Breakdown with base open circuited and emitter grounded.
The breakdown voltage obtained in the first case will be represented as BV_{CBO} and in the second case BV_{CBO}

- **Breakdown voltage with Emitter open**



In the absence of impact ionization, the collector leakage current is due to thermal generation within the space charge region. The number of electron-hole pairs generated within a narrow region δx is given by the expression $G_{th}\delta x$, where

$$G_{th} = \frac{qn_i}{2\tau} \quad (9)$$

The net current is given by the expression

$$I_{CBO} = qA\int_{x_3}^{x_4} G_{th}dx \quad (10)$$

With the onset of impact ionization, there is an additional source of carrier generation G_{im} which will also contribute to current flow, so that

$$I'_{CBO} = qA\int_{x_3}^{x_4} G_{th}dx + qA\int_{x_3}^{x_4} G_{im}dx \quad (11)$$

The generation rate due to impact ionization can be written as

$$qG_{im}dx = I_P\alpha_P dx + I_N\alpha_N dx \quad (12)$$

For equal electron and hole ionization coefficients, Eq. (12) simplifies to :

$$qG_{im} = \alpha\ (I_P + I_N) \quad (13)$$
$$= \alpha I'_{CBO} \quad (14)$$

Eq. (14) and Eq. (10) allow Eq. (11) to be written as:

$$I'_{CBO} = \frac{I_{CBO}}{1-G} \quad (15)$$

$$G = \int_{x_3}^{x_4} \alpha dx \quad (16)$$

Eq. (15) can also be written as:

$$I'_{CBO} = M \times I_{CBO} \quad (17)$$
$$M = \frac{1}{1-\int_{x_3}^{x_4} \alpha dx} \quad (18)$$

- Eq. (17-18) show that the normal reverse leakage current of the transistor gets multiplied by the multiplication factor M.

When $M \to \infty$, avalanche breakdown would occur. The condition for this is:

$$\int_{x_3}^{x_4} \alpha dx = 1 \quad (19)$$

- This is identical to the condition for breakdown of a diode in isolation, so that breakdown of the transistor in this configuration is simply the breakdown of the collector-base diode.

The multiplication factor M can be empirically modeled as

$$M = \frac{1}{1 - \left(\frac{V_{CB}}{BV_{CBO}}\right)^n} \quad (20)$$

where n varies with the structure of the PN junction, whether it is P^+N or N^+P etc.

Let us consider the next case of breakdown.

Breakdown voltage with Base open

As discussed earlier, due to the requirement of zero base current, the holes generated in the collector depletion region have nowhere to flow but the emitter. This results in injection of electrons by the emitter into the base. These electrons flow into the collector. The net collector current I_{CEO}, would therefore have three components:

(i) Normal thermal generation current I_{CBO}

(ii) Current due to generation of carriers as a result of impact ionization.

(iii) Current due to electrons injected by the collector

Eq. (13) shows that the second current can be expressed as:

$$I_{impact} = (I_N + I_P)\int \alpha dx \quad (21)$$

$$= I_{CEO} \times G \quad (22)$$

where G is given by Eq. (16)

- The current injected by the emitter would be β times the hole current, which is simply
the sum of the first two generation current components. Therefore, the net current can be written as:

$$I_{CEO} = I_{CBO} + G \times I_{CEO} + (I_{CBO} + G \times I_{CEO}) \times \beta \quad (23)$$

Eq. (23) can be simplified to:

$$I_{CEO} = \frac{(\beta + 1)I_{CBO}}{1 - (\beta + 1)\frac{M - 1}{M}} \quad (24)$$

where the relation M = 1/1 - G has been used. When avalanche multiplication is small, the collector current
is β times the current under emitter-open conditions as derived earlier also.

In this case breakdown would occur when

$$(\beta+1)\frac{M-1}{M}=1$$

$$\Rightarrow M=\frac{\beta+1}{\beta} \qquad (25)$$

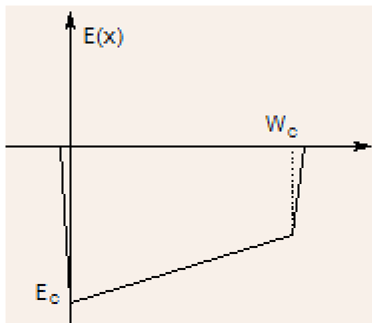
Using Eq. (20) for multiplication coefficient, we obtain

$$BV_{CEO}=BV_{CBO}\times\frac{1}{(\beta+1)^{1/n}} \qquad (26)$$

Eq. (26) shows that breakdown would occur earlier with base open due to extra injection of electrons by the emitter.

Example 5.2 Determine the breakdown voltage under open emitter conditions for an NPN transistor which has a collector consisting of a lightly doped region of doping $10^{16}cm^{-3}$ followed by a very heavily doped N-type region. Determine the breakdown voltage for collector thicknesses of 1 μm, 0.5 μm and 0.2 μm.

Solution : Considering base doping to be much higher than the collector doping, the electric field profile in the collector can be sketched as



The breakdown voltage can then be written as

$$BV_{CBO} = 0.5 \times (4x10^5 + (4x10^5 - \frac{q}{\varepsilon_S} N_{DC} \times W_C) \times W_C$$

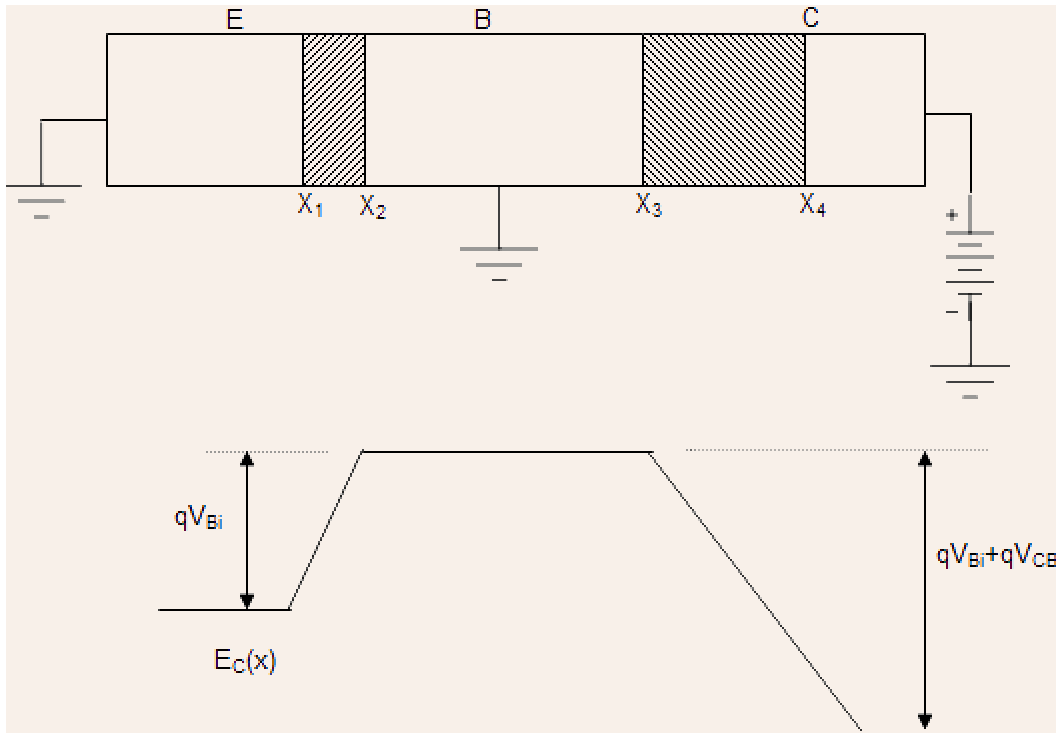
The table below shows the breakdown voltages:

$W_C(\mu m)$	BV_{CBO}
1	32.2
0.5	18
0.2	7.7

It can be seen that the breakdown voltage of cylindrical junctions is lower than that for planar junctions especially for small radii of curvature.

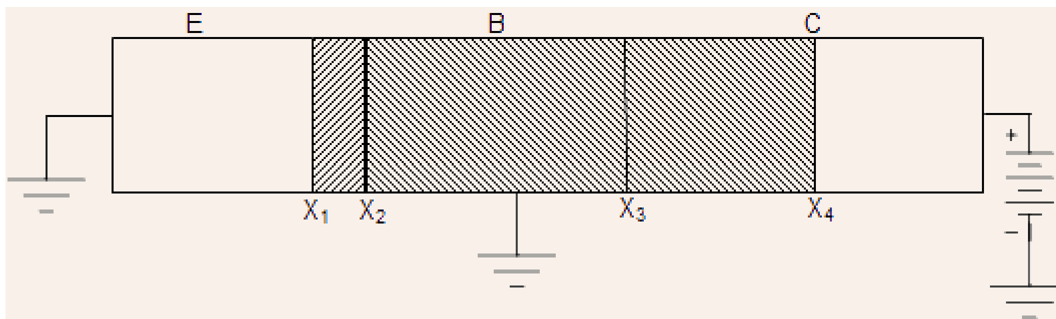
Punchthrough

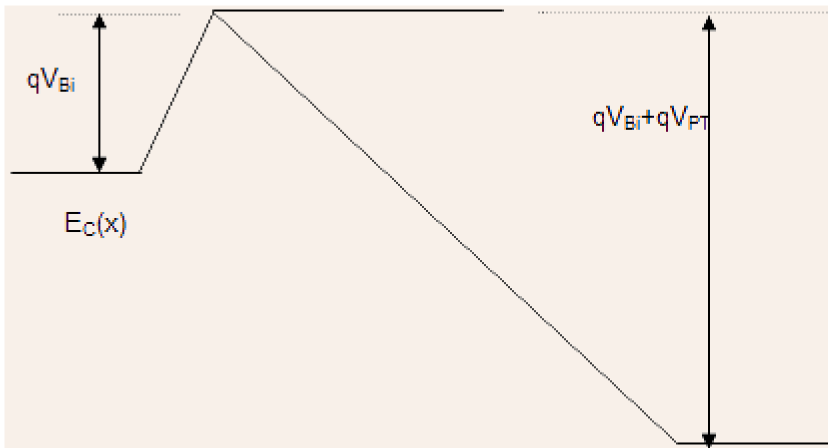
- Punchthrough is another mechanism by which a large collector current can flow in the cutoff mode. Let us consider a case where emitter-base junction is biased at zero volts and collector-base junction is reverse biased. For small collector bias, the depletion regions associated with emitter and collector junctions will be far apart as illustrated below:



- The variation of conduction band across the device is also shown. There is a barrier of height V_{bi} which prevents the numerous electrons present in the emitter from flowing into the base and collector.

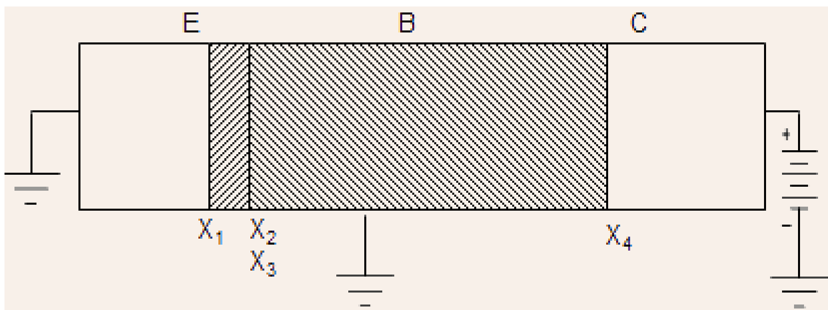
- As collector voltage increases, the depletion region extends in the base and approaches closer to the emitter depletion region. Eventually, at a sufficiently higher collector voltage, the two depletion regions would just touch each other as illustrated below:

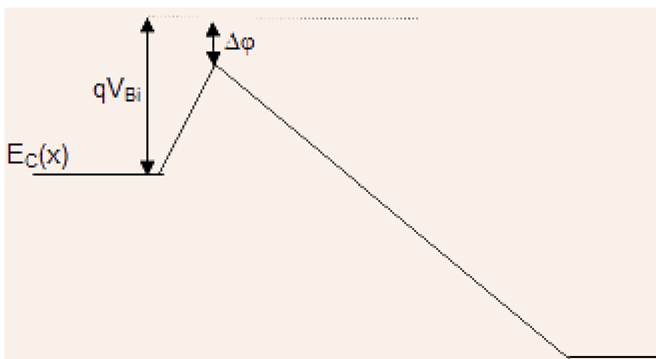




• The collector depletion region has punched through the base and merged with the emitter depletion region. The collector-base voltage at which this phenomenon occurs is known as the punchthrough voltage V_{PT}.

• The barrier seen by the electrons is still the built in voltage. However, if the collector voltage is further increased, then the collector depletion region will encroach over the emitter depletion region. The decrease in emitter depletion width will also be accompanied with a decrease in the barrier height.





• The barrier height is reduced by an amount $\Delta\varphi$. This amounts to a forward biasing of the base-emitter junction by the same amount.

As a result a large collector current begins to flow. The voltage at which punchthrough occurs can be determined using the relation:

$$x_{de} + \frac{N_{DC}}{N_A + N_{DC}} x_{dc} = W_{BO} \qquad (27)$$

where x_{de} is the width of emitter depletion region in the base and W_{BO} is the metallurgical base width. The second term in the expression above represents the width of collector depletion region in the base. Using standard expressions for depletion widths, we obtain:

$$\sqrt{\frac{2\varepsilon_{Si}}{qN_A} V_{Bi}} + \sqrt{\frac{2\varepsilon_{Si}(N_{DC}/N_A)}{q(N_A + N_{DC})}(V_{Bi} + V_{PT})} = W_{BO} \qquad (28)$$

Eq. (28) can be re-written as:

$$V_{PT} = 0.5 \frac{q}{\varepsilon_{Si}} \frac{N_A(N_{DC} + N_A)}{N_{DC}} \times \left(W_{BO} - \sqrt{\frac{2\varepsilon_{Si}}{qN_A} V_{Bi}}\right)^2 - V_{Bi} \qquad (29)$$

- Punchthrough is of importance only if the punchthrough voltage turns out to be less that the breakdown voltage. As expected, this begins to happen when base thickness is reduced.

Example 5.3 : Determine the punchthrough voltage for an NPN transistor having the following characteristics:

Base : $N_A = 2 \times 10^{17}$ cm^{-3}, $W_{BO} = 0.2$ µm, $\tau_N = 1$ µs
Emitter : $N_D = 2 \times 10^{19}$ cm^{-3}, $W_E = 0.1$ µm
Collector : $N_{DC} = 1 \times 10^{16}$ cm^{-3}, $\tau_N = \tau_P = 1$ µs. Take $V_{bi} = 0.9$ Volts

(b) How can you increase the punch through voltage without compromising the current gain?

Solution :

(a) Using Eq. (29), we obtain V_{PT} = 7.6 Volts

(b) Increase in base width or base doping will result in decrease in current gain, so it cannot be used. The only other factor is the collector doping, whose reduction will improve the punchthrough voltage. The drawback of this approach is that it would lower the maximum collector current density due to early onset of Kirk effect and also increase the collector transit time.