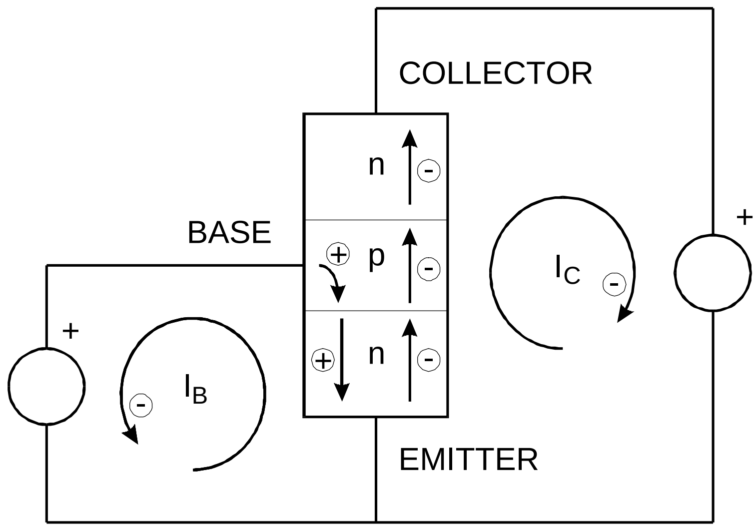
# IX.3. Bipolar Transistors

Consider the *npn* structure shown below.




The base and emitter form a diode, which is forward biased so that a base current $I_B$ flows.

> The base current injects holes into the base-emitter junction.
>
> As in a simple diode, this gives rise to a corresponding electron current through the base-emitter junction.

If the potential applied to the collector is sufficiently positive so that the electrons passing from the emitter to the base are driven towards the collector, an external current $I_C$ will flow in the collector circuit.

> The ratio of collector to base current is equal to the ratio of electron to hole currents traversing the base-emitter junction. In an ideal diode

$$\frac{I_C}{I_B} = \frac{I_{nBE}}{I_{pBE}} = \frac{D_n / N_A L_n}{D_p / N_D L_p} = \frac{N_D}{N_A} \frac{D_n L_p}{D_p L_n}$$

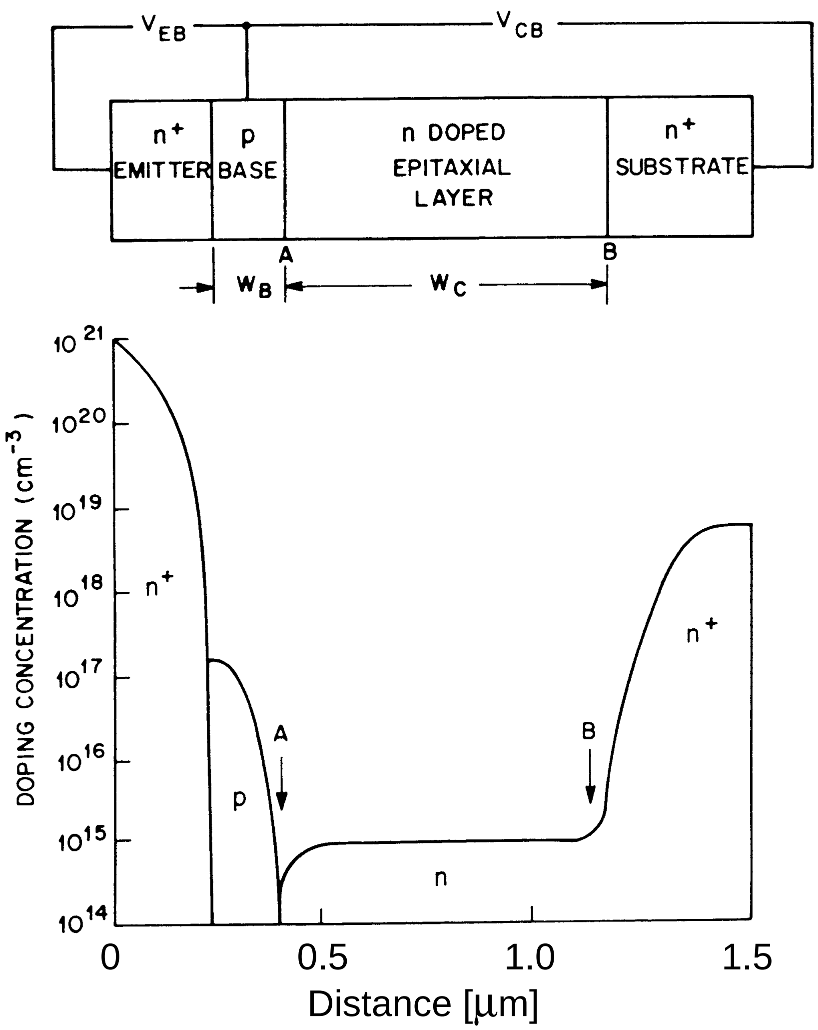
If the ratio of doping concentrations in the emitter and base regions $N_D/N_A$ is sufficiently large, the collector current will be greater than the base current.

$\Rightarrow$ DC current gain

Furthermore, we expect the collector current to saturate when the collector voltage becomes large enough to capture all of the minority carrier electrons injected into the base.

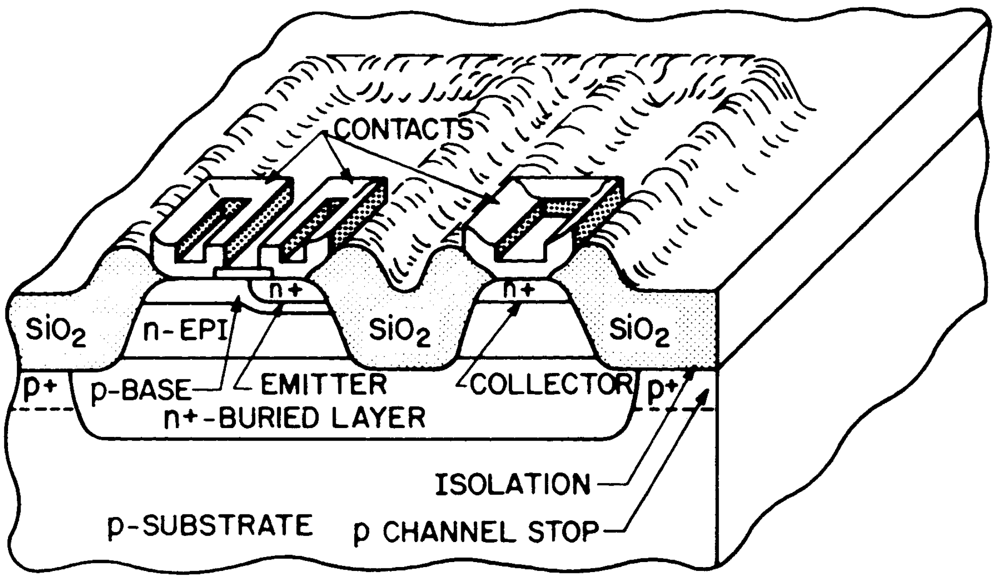
Since the current inside the transistor comprises both electrons and holes, the device is called a bipolar transistor.

Dimensions and doping levels of a modern high-frequency transistor (5 – 10 GHz bandwidth)




(adapted from Sze)

High-speed bipolar transistors are implemented as vertical structures.




(from Sze)

The base width, typically 0.2 µm or less in modern high-speed transistors, is determined by the difference in diffusion depths of the emitter and base regions.

The thin base geometry and high doping levels make the base-emitter junction sensitive to large reverse voltages.

Typically, base-emitter breakdown voltages for high-frequency transistors are but a few volts.

As shown in the preceding figure, the collector region is usually implemented as two regions: one with low doping (denoted “epitaxial layer in the figure) and the other closest to the collector contact with a high doping level. This structure improves the collector voltage breakdown characteristics.

## Quantitative analysis of the bipolar junction transistor (BJT)

1. Holes are injected into the base through the external contact.

   The potential distribution drives them towards the emitter.
   Since they are majority carriers in the base, few will recombine.

   Holes entering the base-emitter depletion region will either

   a) pass through the depletion region into the emitter

   b) be lost due to recombination

2. As shown in the discussion of the *pn*-junction, coupled with the hole current is an electron current originating in the emitter. This electron current will flow towards the collector, driven by the more positive potential.

   These electrons either

   a) enter the collector to become the collector current

   b) recombine in the base region. The holes required for the recombination are furnished by the base current.

3. Thus, the base current is the sum of the

   a) hole current entering the emitter

   b) hole losses due to recombination in the base-emitter depletion region

   c) electron losses due to recombination in the base during transport to the collector

The transport of minority carriers in the base is driven by diffusion, so

$$D_n \frac{d^2 n_p}{dx^2} - \frac{n_p - n_{p0}}{\tau_n} = 0$$

At the boundary to the base-emitter depletion region

$$n_p(0) = n_{p0} e^{q_e V_{be}/k_B T}$$

The equilibrium concentration of electrons in the base is determined by the base acceptor doping level $N_{AB}$

$$n_{p0} = \frac{n_i^2}{N_{AB}}$$

At the collector boundary all minority carriers will be immediately swept away by the reverse bias field, so that the boundary condition becomes

$$n_p(W_B) = 0$$

Then the solution of the diffusion equation is

$$n_p(x) = n_{p0}\left[1 - \frac{\sinh \frac{x}{L_n}}{\sinh \frac{W_B}{L_n}}\right] + (n_p(0) - n_{p0}) \frac{\sinh \frac{W_B - x}{L_n}}{\sinh \frac{W_B}{L_n}}$$

If $V_{BE} >> 4k_BT/q_e$ the non-equilibrium concentration will dominate

$$n_p(0) >> n_{p0}$$

so this simplifies to

$$n_p(x) = n_p(0) \frac{\sinh \frac{W_B - x}{L_n}}{\sinh \frac{W_B}{L_n}}$$

Since the base width $W_B$ in good transistors is much smaller than the diffusion length $L_n$, the concentration profile can be approximated by a linear distribution.

$$n_p(x) = n_p(0)\left(1 - \frac{x}{W_B}\right)$$

Now we can evaluate the individual current components.

In this approximation the diffusion current of electrons in the base region becomes

$$J_{nB} = -q_e D_{nB} \frac{dn_p(x)}{dx} = q_e D_{nB} \frac{n_i^2}{N_{AB} W_B} e^{q_e V_{BE}/k_B T}$$

where $D_{nB}$ is the diffusion constant of electrons in the base.

Similarly, the diffusion current of holes injected into the emitter, under the assumption that the emitter depth is much smaller than the diffusion length, is

$$J_{pE} = q_e D_{pE} \frac{n_i^2}{N_{DE} W_E} e^{q_e V_{BE}/k_B T}$$

For the moment, we'll neglect recombination of holes in the base-emitter depletion region. Under this assumption, the base current is

$$I_B = J_{pE} A_{JE}$$

where $A_{JE}$ is the area of the emitter junction.

The collector current is primarily the electron current injected into the base, minus any losses due to recombination during diffusion. The collector transport factor

$$\alpha_T = \frac{electron\ current\ reaching\ collector}{electron\ current\ injected\ \ from\ emitter} = \frac{\left.\dfrac{dn_p}{dx}\right|_{x=W_B}}{\left.\dfrac{dn_p}{dx}\right|_{x=0}}$$

Using the above result this becomes

$$\alpha_T = \frac{1}{\cosh \dfrac{W_B}{L_n}}$$

Recalling that the base width is to be much smaller than the diffusion length, this expression can be approximated as

$$\alpha_T \approx 1 - \frac{1}{2}\left(\frac{W_B}{L_n}\right)^2$$

The resulting collector current is the diffusion current of electrons in the base times the transport factor

$$I_C = J_{nB} A_{JE} \alpha_T$$

One of the most interesting parameters of a bipolar transistor is the DC current gain which is the ratio of collector current to base current

$$\beta_{DC} = \frac{I_C}{I_B}$$

Using the above results

$$\beta_{DC} = \frac{q_e D_{nB} \dfrac{n_i^2}{N_{AB} W_B} e^{q_e V_{BE}/k_B T}}{q_e D_{pE} \dfrac{n_i^2}{N_{DE} W_E} e^{q_e V_{BE}/k_B T}}$$

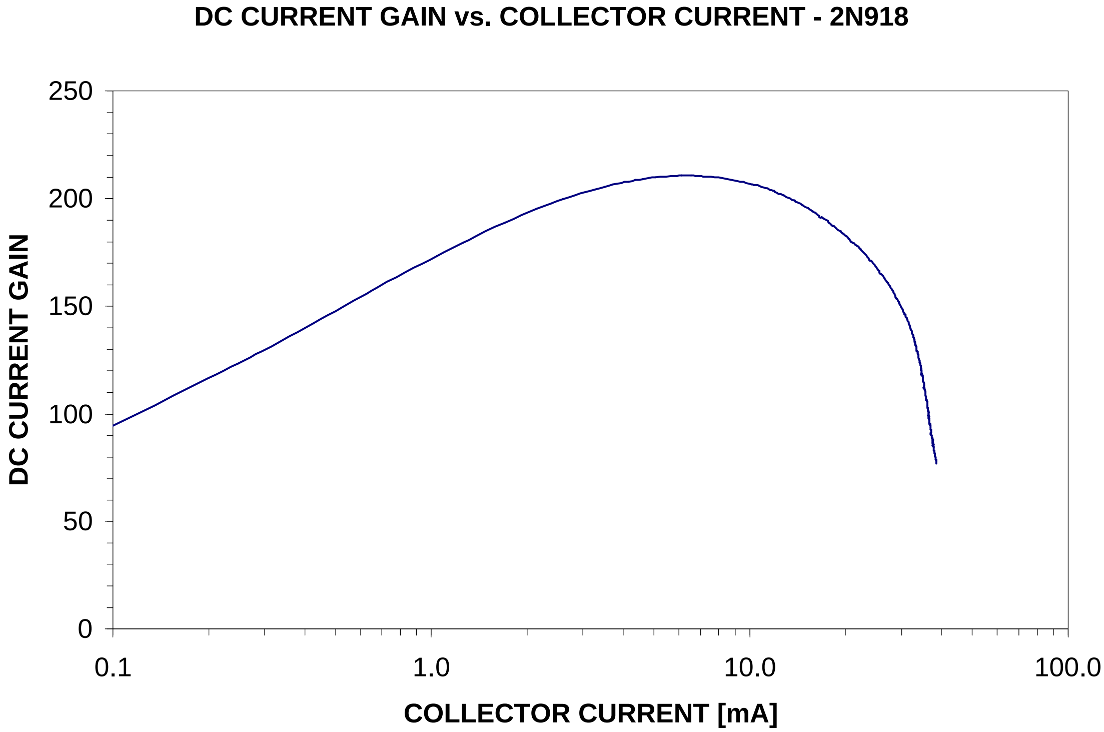
$$\beta_{DC} = \frac{N_{DE}}{N_{AB}} \cdot \frac{D_{nB} W_E}{D_{pE} W_B}$$

Primarily, the DC current gain is determined by the ratio of doping concentrations in the emitter and the base.

This simple result reflects the distribution of current in the forward biased diode between electrons and holes.

Transistors with high current gains always have much higher doping levels in the emitter than in the base.

The result of this simple analysis implies that for a given device the DC current gain should be independent of current. In reality this is not the case.




The decrease at low currents is due to recombination in the base-emitter depletion region.

Since the most efficient recombination centers are near the middle of the bandgap, we set $E_t$=$E_i$ to obtain the recombination rate

$$\frac{dN_R}{dt} = \sigma v_{th} N_t \frac{pn - n_i^2}{n + p + 2n_i}$$

Due to quasi-equilibrium in the space charge region

$$pn = n_i^2 e^{q_e V_{BE} / k_B T}$$

This gives us the product of electron concentrations, but we also need the individual concentrations.

For a given forward bias the maximum recombination rate will coincide with the minimum concentration, i.e.

$$d(p+n)=0$$

and since $pn$= const.

$$dp=-dn=\frac{pn}{p^2}dp$$

or

$$p=n$$

at the point of maximum recombination.

Hence, the carrier concentrations are

$$pn=n^2=p^2=n_i^2 e^{q_e V_{BE}/k_B T}$$

or

$$n=p=n_i e^{q_e V_{BE}/2k_B T}$$

Inserting these concentrations into the expression for the recombination rate yields

$$\frac{dN_R}{dt}=\sigma v_{th} N_t \frac{n_i^2 (e^{q_e V_{BE}/k_B T}-1)}{2n_i (e^{q_e V_{BE}/2k_B T}+1)}$$

This maximum recombination rate will not prevail throughout the depletion region, but only over a region where the potential changes no more than ~$k_B T/q_e$. If we assume that the average field is $V_{BE}/W_{BE}$, a suitable averaging distance is

$$w=\frac{k_B T}{q_e}\frac{W_{BE}}{V_{BE}}$$

Nevertheless, for simplicity assume that recombination is uniform throughout the depletion region.

For $V_{BE} >> k_B T/q_e$ this yields the recombination current to be made up by holes from the base current

$$J_{pr} = q_e \frac{dN_R}{dt} \approx \frac{1}{2} q_e \sigma v_{th} N_t n_i W_{BE} e^{q_e V_{BE}/2k_B T}$$

Now the base current becomes

$$I_B = (J_{pE} + J_R) A_{JE} \ ,$$

so the DC current gain with recombination is

$$\beta_{DC} = \frac{q_e D_{nB} \frac{n_i^2}{N_{AB} W_B} e^{q_e V_{BE}/k_B T}}{q_e D_{pE} \frac{n_i^2}{N_{DE} W_E} e^{q_e V_{BE}/k_B T} + \frac{1}{2} q_e \sigma v_{th} N_t n_i W e^{q_e V_{BE}/2k_B T}}$$

Because of the factor ½ in the exponent of the recombination term

$$\beta_{DC} = \frac{D_{nB} \frac{n_i^2}{N_{AB} W_B}}{D_{pE} \frac{n_i^2}{N_{DE} W_E} + \frac{n_i}{2\tau_0} W_{BE} e^{-q_e V_{BE}/2k_B T}} \ ,$$

which can be rewritten in a more informative form as

$$\frac{1}{\beta_{DC}} = \frac{1}{\beta_0} + \frac{N_{AB} W_{BE}}{D_{nB}} W_B \frac{N_t \sigma v_{th}}{n_i e^{q_e V_{BE}/2k_B T}}$$

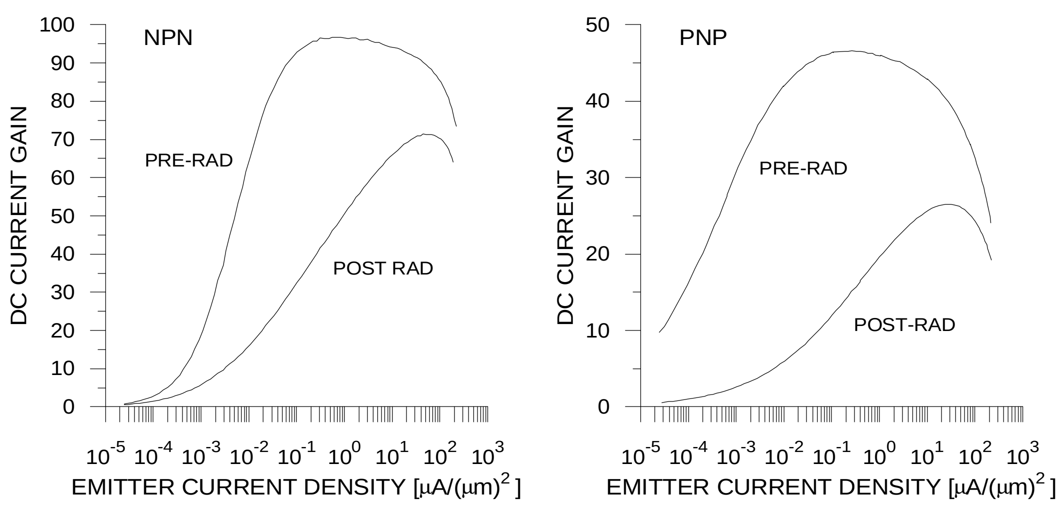
where $\beta_0$ is the DC current gain without recombination. Increasing the concentration of traps $N_t$ decreases the current gain $\beta_{DC}$, whereas decreasing the base width $W_B$ reduces the effect of traps. Since a smaller base width translates to increased speed (reduced transit time through the base), fast transistors tend to be less sensitive to trapping.

The current dependence enters through the exponential, which relates the injected carrier concentrations to the base-emitter voltage $V_{BE}$. With decreasing values of base-emitter voltage, i.e. decreasing current, recombination becomes more important. This means that the DC current gain is essentially independent of current above some current level, but will decrease at lower currents.

Furthermore note that - although not explicit in the above expression - the "ideal" DC current gain depends only on device and material constants, whereas the recombination depends on the local density of injected electrons and holes with respect to the concentration of recombination centers. Consider two transistors with different emitter areas, but operating at the same current, so the larger transistor operates at a lower current density. The larger area transistor will achieve a given current at a lower base-emitter voltage $V_{BE}$ than the smaller device, which increases the effect of trapping. Thus, the relative degradation of DC current gain due to recombination depends on the current density.

This means that within a given fabrication process, a large transistor will exhibit more recombination than a small transistor at the same current. Stated differently, for a given current, the large transistor will offer more recombination centers for the same number of carriers.

The figures below show the DC current gain of *npn* and *pnp* transistors over a wide range of current densities.



As shown in the figures, the DC current gain in modern devices is quite uniform over orders of magnitude of emitter current. The decrease in DC current gain at low current densities due to increased recombination is apparent. The figures also show the degradation of current gain after irradiation, here after exposure to the equivalent of $10^{14}$ minimum ionizing protons/cm$^2$. Radiation damage increases the concentration of trapping sites $N_t$ proportional to fluence $\Phi$, so the above equation can be rewritten to express the degradation of current gain with fluence for a given current density (fixed $V_{BE}$).

$$\frac{1}{\beta_{DC}} = \frac{1}{\beta_0} + K\Phi$$

where the constant $K$ encompasses the device constants and operating point. In a radiation-damaged transistor the reduction in current gain for a given DC current will be less for smaller devices and faster transistors (small base width) tend to be less sensitive to radiation damage.

Why does the DC current gain drop off at high currents?

a) with increasing current the high field region shifts towards the collector, effectively increasing the base width (“Kirk effect”).

b) at high current levels the injected carrier concentration becomes comparable with the bulk doping.
    a) reduction in injection efficiency
    b) at very high current densities band-gap narrowing

c) voltage drops in base and emitter resistance

d) Auger recombination

Auger effect:

In an atomic transition, instead of photon emission an electron is emitted.

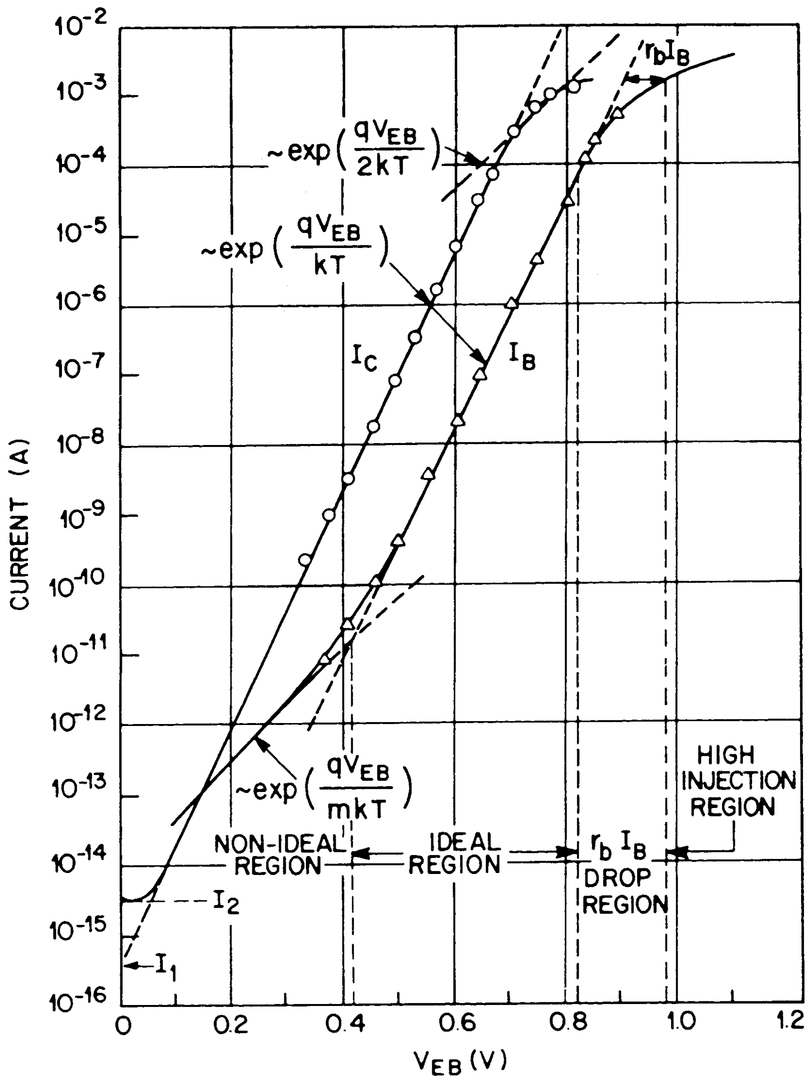
Assume a $K_\alpha$ transition (i.e. from the $L$ to the $K$ shell).

If the ionization of an $L$ electron is less than the $K_\alpha$ transition energy, instead of a photon an $L$ electron can be emitted with an energy

$$E_{el} = E_{K\alpha} - E_L = (E_K - E_L) - E_L = E_K - 2E_L$$

In semiconductors the Auger effect manifests itself as recombination of an electron-hole pair with emission of an energetic majority carrier.

Summary of effects that degrade DC current gain:



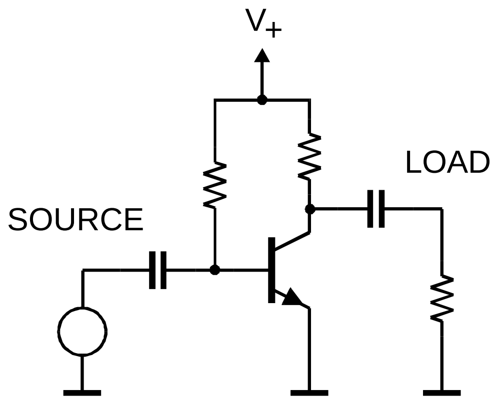

(from Sze)

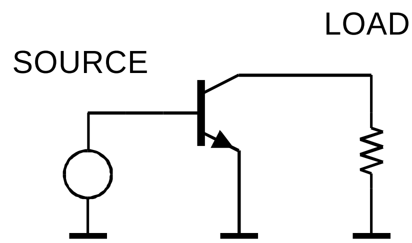
## Bipolar Transistors in Amplifiers

Three different circuit configurations are possible:

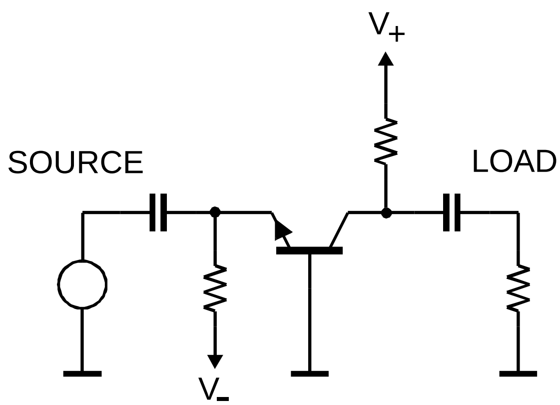
1. Common Emitter

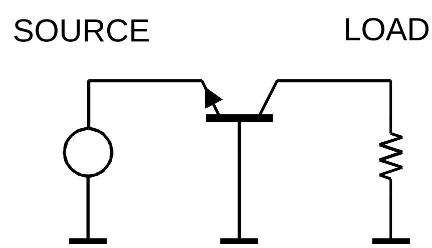
Equivalent Circuit



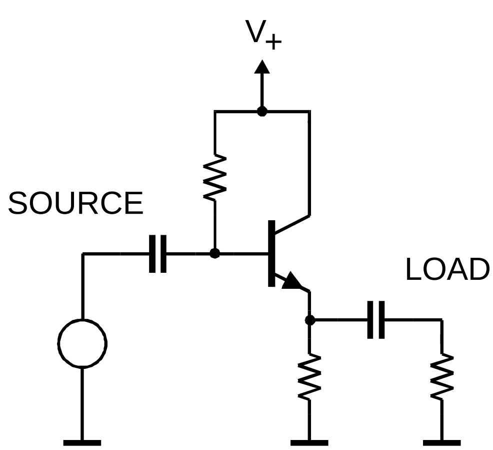



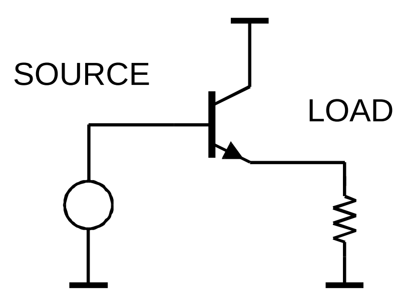

2. Common Base



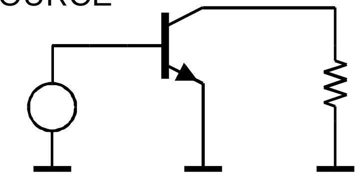




3. Common Collector (Emitter Follower)



Although the bipolar transistor is a current driven device, it is often convenient to consider its response to input voltage.




Consider a transistor in the common emitter (CE) configuration.

As discussed in a previous lecture, the voltage gain

$$A_V = \frac{dV_{out}}{dV_{in}} = \frac{dI_C}{dV_{BE}} R_L = g_m R_L$$

Since the dependence of base current on base-emitter voltage is given by the diode equation

$$I_B = I_R (e^{q_e V_{BE}/k_B T} - 1) \approx I_R e^{q_e V_{BE}/k_B T}$$

the resulting collector current is

$$I_C = \beta_{DC} I_B = \beta_{DC} I_R e^{q_e V_{BE}/k_B T}$$

and the transconductance, i.e. the change in collector current vs. base-emitter voltage

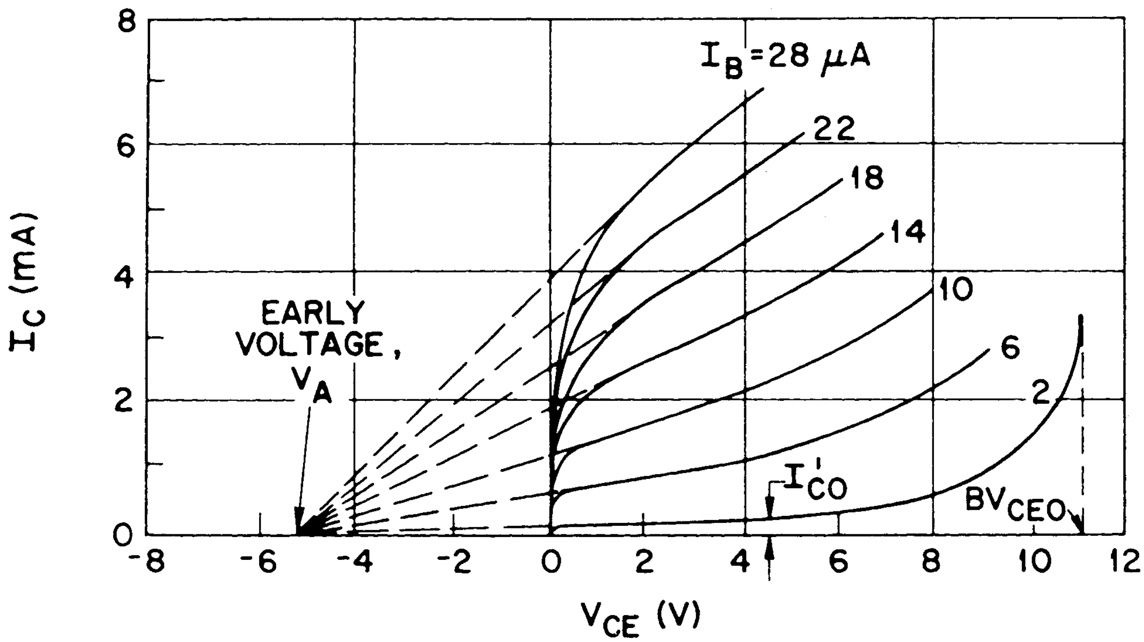
$$g_m \equiv \frac{dI_C}{dV_{BE}} = \beta_{DC} I_R \frac{q_e}{k_B T} e^{q_e V_{BE}/k_B T} = \frac{q_e}{k_B T} I_C$$

The transconductance depends only on collector current, so for any bipolar transistor – regardless of its internal design – setting the collector current determines the transconductance.

Since at room temperature $k_B T/q_e$ = 26 mV

$$g_m = \frac{I_C}{0.026} \approx 40 I_C$$

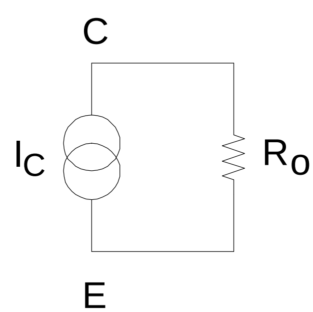
The obtainable voltage gain of an amplifier depends on the output characteristics of the transistor.




At low collector voltages the field in the collector-base region is not sufficient to transport all injected carriers to the collector without recombination. At higher voltages the output current increases gradually with voltage (saturation region), due to the change in effective base width.

An interesting feature is that the extrapolated slopes in the saturation region intersect at the same voltage $V_A$ for $I_c = 0$ ("Early voltage").

The finite slope of the output curves is equivalent to a current generator with a shunt resistance



where

$$R_o = K \frac{V_A}{I_C}$$

$K$ is a device-specific constant of order 1, so usually it's neglected.

The total load resistance is the parallel combination of the external load resistance and the output resistance of the transistor. In the limit where the external load resistance is infinite, the load resistance is the output resistance of the amplifier.

The maximum obtainable voltage gain is

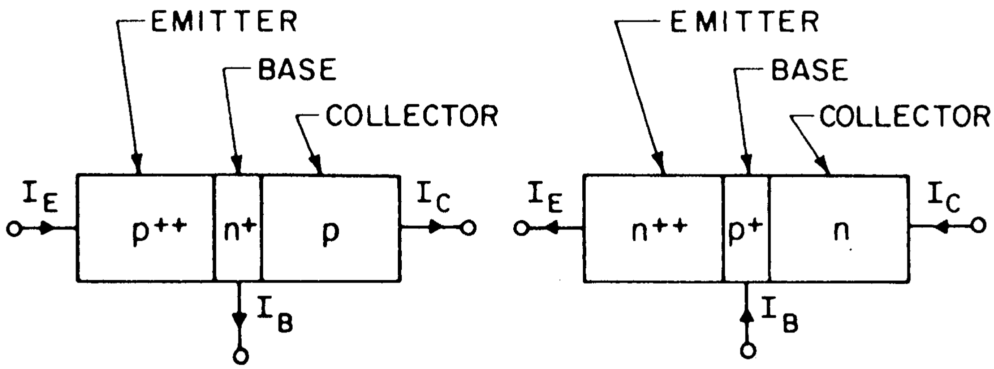
$$A_{v,\max} = \frac{dI_C}{dV_{BE}} R_o = g_m R_o \approx \frac{I_C}{k_B T / q_e} \frac{V_A}{I_C} = \frac{V_A}{k_B T / q_e}$$

which at room temperature is about ($40 V_A$).

- Note that to first order the maximum obtainable voltage gain is independent of current.

- Transistors with large Early voltages allow higher voltage gain.

Bipolar transistors can be implemented as *pnp* or *npn* structures.




The polarities of the applied voltages are the opposite:

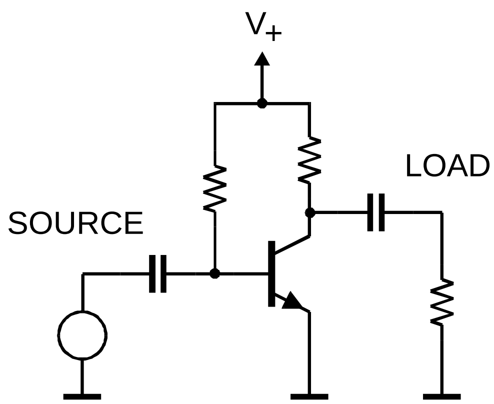
*npn*:  positive collector-emitter and base emitter voltages

*pnp*:  negative collector-emitter and base emitter voltages

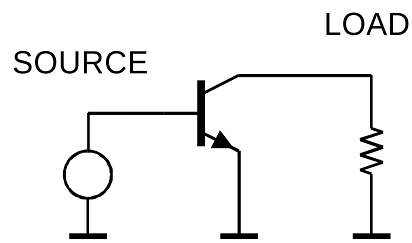
The basic amplifier equations are the same for both transistor types.

The availablility of complementary transistors offers great flexibility in circuit design.

a) Common Emitter configuration




Equivalent Circuit



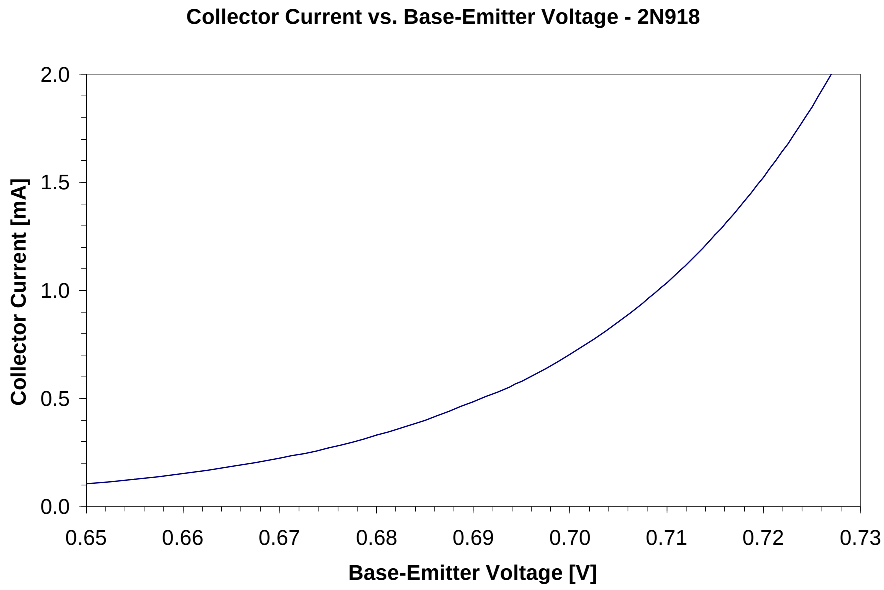

The input signal is applied to the base, the output taken from the collector.

The input resistance is proportional to the DC current gain and inversely proportional to the collector current.

$$R_i = \frac{dV_{BE}}{dI_B} \approx \beta_{DC} \frac{dV_{BE}}{dI_C} = \frac{\beta_{DC}}{g_m} = \frac{k_B T}{q_e} \frac{\beta_{DC}}{I_C}$$

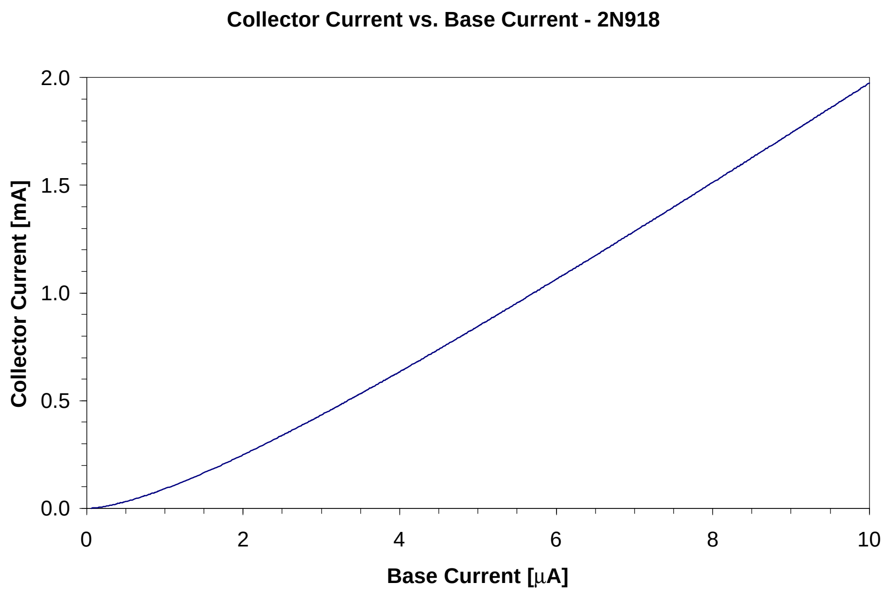
For $\beta_{DC}$ = 100 and $I_c$= 1 mA, $R_i$= 2600 Ω.

Although the bipolar transistor is often treated as a voltage-driven device, the exponential dependence of base current on input voltage means that as an amplifier the response is very non-linear.

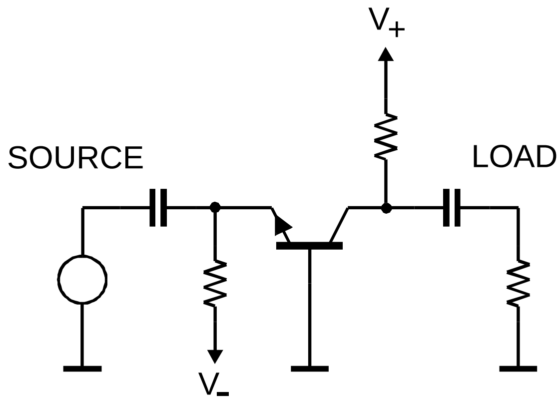



In audio amplifiers, for example, this causes distortion.
Distortion may be limited by restricting the voltage swing, which to some degree is feasible because of the high transconductance.

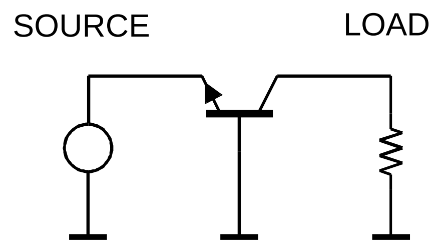
With current drive the linearity is much better.



b) Common Base configuration




Equivalent Circuit




The input signal is applied to the emitter, the output taken from the collector.

This configuration is used where a low input impedance is required.

$$R_i = \frac{dV_{EB}}{dI_E} \approx \frac{dV_{EB}}{dI_C} = \frac{1}{g_m} = \frac{k_B T}{q_e} \frac{1}{I_C}$$

Since at room temperature $k_B T/q_e$= 26 mV

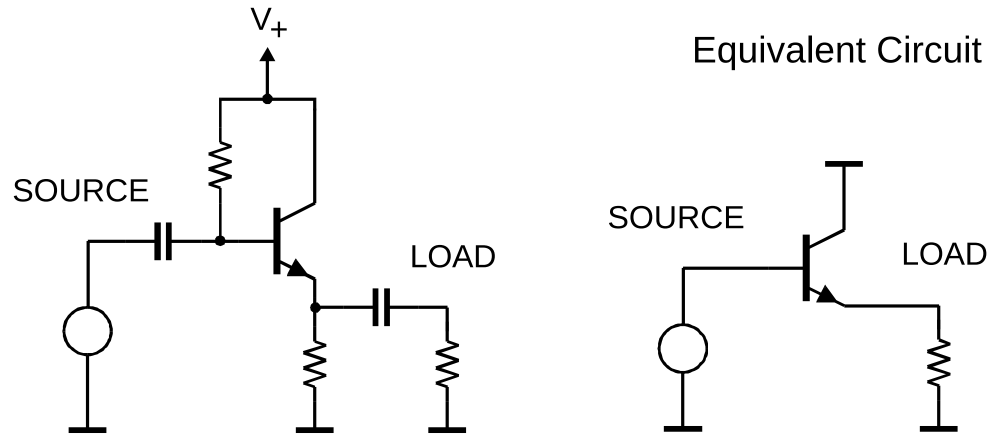
$$R_i = \frac{0.026}{I_C}$$

i.e. $R_i$= 26 $\Omega$ at $I_c$= 1 mA.

The input resistance is about $1/\beta$ times smaller than in the common emitter configuration.

c) Common Collector configuration

The signal is applied to the base and the output taken from the emitter (“emitter follower”).




The load resistance $R_L$ introduces local negative feedback,

$$V_i = V_{BE} + I_E R_L \approx V_{BE} + \beta I_B R_L$$

since $V_{BE}$ varies only logarithmically with $I_B$, it can be considered to be constant ( $\approx$ 0.6 V for small signal transistors), so

$$\frac{dV_i}{dI_i} = \frac{dV_i}{dI_B} \approx \beta R_L$$

Thus, the input resistance depends on the load: $R_i \approx \beta R_L$

Since $dV_{BE}/dI_B \approx$ const, the emitter voltage follows the input voltage, so the voltage gain cannot exceed 1.

The output resistance of the emitter follower is

$$R_o = -\frac{dV_{out}}{dI_{out}} = -\frac{d(V_{in} - V_{BE})}{dI_E} \approx \frac{dV_{BE}}{dI_E} \approx \frac{1}{g_m}$$

(since the applied input voltage is independent of emitter current, $dV_{in}/dI_E = 0$). At 1 mA current $R_o$= 26 $\Omega$.

Although the stage only has unity voltage gain, it does have current gain, so emitter followers are often used as output drivers