# ENG1004 - Electronics & Sensors

# Part 3 – Bipolar Junction Transistors

**S. Le Goff, School of Engineering, Newcastle University**

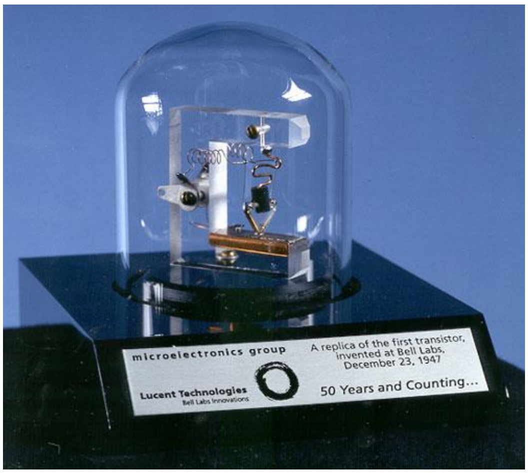
## 1. Introduction

The bipolar point-contact transistor was invented in 1947 at the Bell Telephone Laboratories (USA) by John Bardeen and Walter Brattain under the direction of William Shockley. The junction version known as the bipolar junction transistor (BJT), invented by Shockley in 1948, is the version we are going to study hereafter.

In acknowledgement of this accomplishment, Shockley, Bardeen, and Brattain were jointly awarded the 1956 Nobel Prize in Physics "for their researches on semiconductors and their discovery of the transistor effect."

John Bardeen (1908 – 1991), William Bradford Shockley (1910 – 1989) and
Walter Houser Brattain (1902 – 1987) at Bell Labs, 1948

The transistor (in its various forms, not only BJT) is the key active component in practically all modern electronics. Many consider it to be one of the greatest inventions of the 20th century. Its importance in today's society rests on its ability to be mass produced using a highly automated process (semiconductor device fabrication) that achieves astonishingly low per-transistor costs.

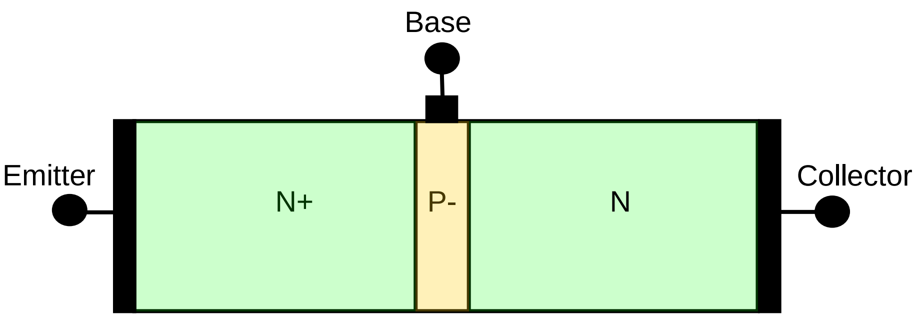


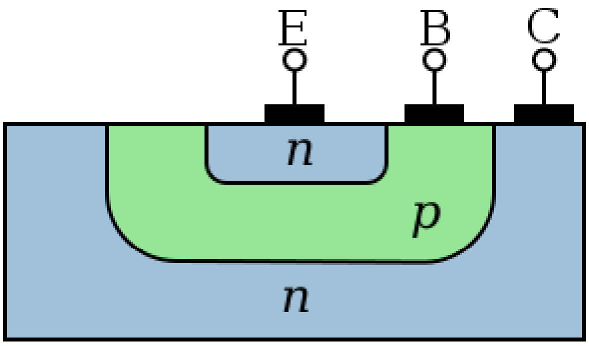

A replica of the first working transistor

Although several companies each produce over a billion individually packaged (known as discrete) transistors every year, the vast majority of transistors now are produced in integrated circuits, along with diodes, resistors, capacitors and other electronic components, to produce complete electronic circuits.

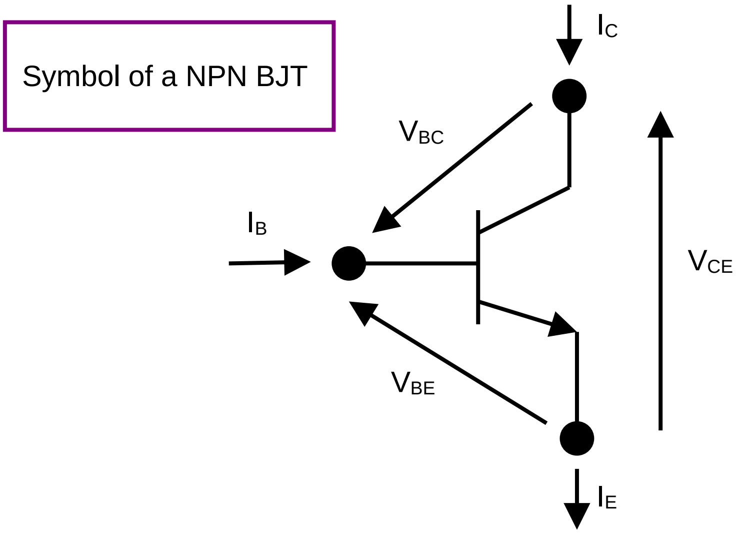
## 2. Basic Operation of a BJT

A NPN BJT is a semiconductor device consisting of a narrow P-type region between two N-type regions. The three regions are called the *emitter* (E), *base* (B), and *collector* (C), respectively. The emitter region is heavily doped (N+) with the appropriate impurity, while the base region is lightly doped (P-). The collector region has a moderate doping level (N). The whole structure is therefore not symmetrical.






Simplified cross section of a planar NPN bipolar junction transistor




We consider throughout this chapter a device consisting of N, P, and N regions in order, hereafter referred to as *NPN bipolar junction transistor*. However, we can also build equivalent devices in P, N, and P order instead (called *PNP bipolar junction transistors*). In fact, it is sometimes useful to have both types of devices available in the same circuit (see, e.g., the push-pull amplifier).

Let us see what happens when bias voltages are applied to such NPN device. To this end, let us assume the use of a silicon BJT.

Assume that a forward bias is applied to the base-emitter junction and a reverse bias is applied to the base-collector junction. These are the normal operating conditions of a bipolar junction transistor in analogue electronics (we will later say that the BJT operates in the *forward active mode* in this case).

These conditions imply that $V_{BE} = V_d$ and $V_{BC} < V_d$, where $V_d \approx 0.7$ volt denotes the threshold voltage of a silicon diode. If we take the emitter as a reference, these conditions can be re-written as $V_{BE} \approx 0.7$ volt and $V_{CE} = V_{CB} + V_{BE} = V_{BE} - V_{BC} > 0$ volt.

At first glance, we would expect to have electrons move from emitter to base and leave the device through the base at that point. With the base-collector junction reverse biased, we would expect no current to flow through that junction.

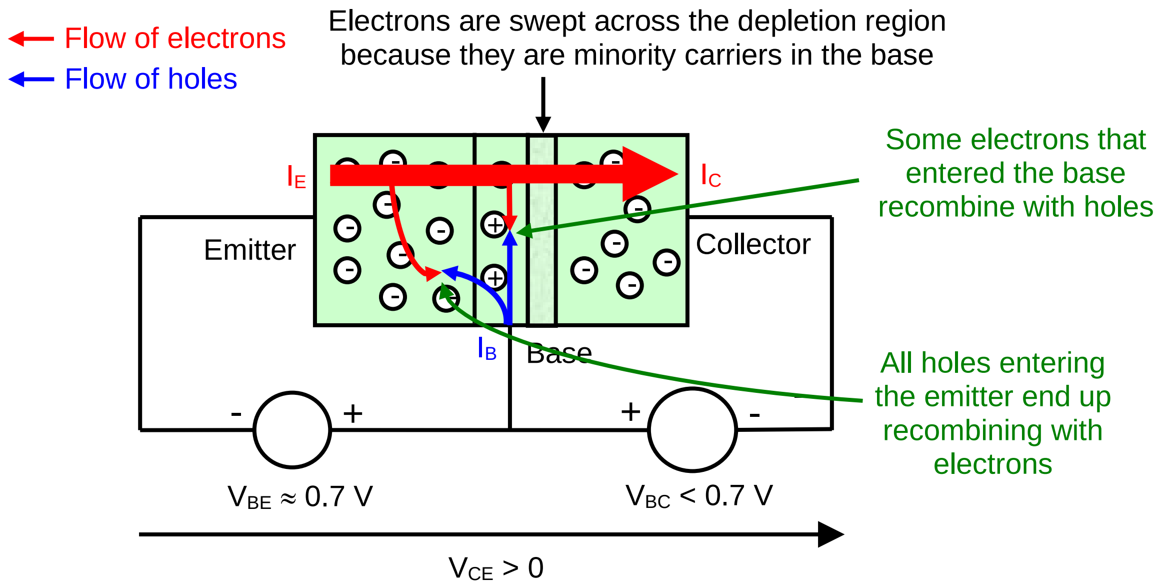
But something happens inside the base region. The forward bias on the base-emitter junction does indeed attract electrons from the emitter into the base. Once they are in the base, these electrons become minority carriers and are expected to quickly recombine with holes which are the majority carriers in the base.

However, this does not happen that way because the transistor's base region is so thin that minority carriers can diffuse across it in much less time than the semiconductor's minority carrier lifetime. In practice, the probability of recombination of electron-hole pairs in the base is minimised by ensuring that the thickness of the base is much less than the diffusion length of the electrons.

In addition, the light doping of the base region also ensures that the probability of recombination of electron-hole pairs is kept to a minimum in that region.

Therefore, most electrons injected into the base from the emitter are able to diffuse through the base region and reach the depletion region formed by the reverse bias of the base-collector junction.

While the reverse-bias voltage acts as a barrier to holes in the base, it actively propels electrons across it. Thus, any electrons coming close to the depletion region are swept across it into the collector and give rise to a collector current.




This phenomenon should not surprise us once we remember what happens with a simple diode under reverse bias: the depletion region cannot be crossed by the majority carriers on both sides of a PN junction, but it *can* definitely be crossed by the minority carriers. This is simply due to the fact that, under reverse bias, the external voltage applied across the PN junction does not "sufficiently attract" the majority carriers towards this junction.

As the majority and minority carriers carry, by definition, opposite electric charges, the former statement is equivalent to saying that, under reverse bias, the external voltage applied across the junction does not "sufficiently repel" the minority carriers from this junction. This is why the latter are able to cross it despite the depletion region.

In a simple diode, the current due to minority carriers under reverse bias is negligible because there are so few of them on each side of the junction. But, in a device such as the BJT, there are plenty of minority carriers (electrons in the case of an NPN BJT) injected in the base and most of them manage to reach the edge of the depletion region without recombining with majority carriers (holes in the case of an NPN BJT) in the base. The collector current thus generated by these minority carriers crossing the reverse-biased base-collector junction can certainly not be neglected.

The thin shared base and asymmetric collector and emitter doping levels are what differentiates a bipolar transistor from two separate and oppositely biased diodes connected in series.

We can write the following expression linking the emitter and collector currents: $I_C = \alpha_F \cdot I_E$, where $\alpha_F$ is a parameter called the *common-base current gain*. Its value is slightly smaller than the unit for a well-designed BJT (e.g., $\alpha_F = 0.99$).

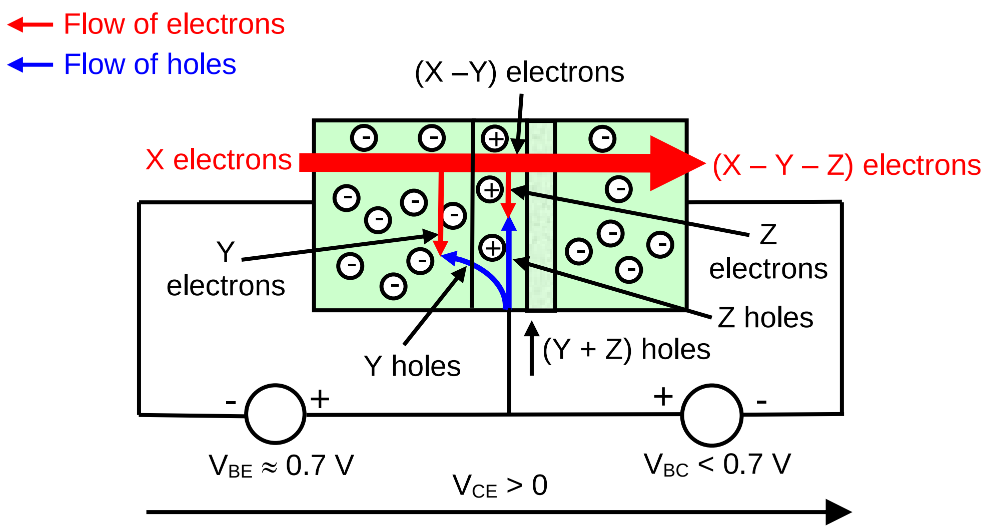
We can summarize the whole process as follows: X electrons, where X is an arbitrary number, coming from the (-) terminal of the voltage source $V_{BE}$, enter the emitter. They represent the emitter current $I_E$. Among these X electrons, (X – Y) electrons are able to enter the base, whereas the remaining Y electrons recombine with holes that have entered the emitter from the base.

Then, among the (X – Y) electrons that have entered the base, (X – Y – Z) electrons reach the collector region after being swept across the depletion region, while Z electrons end up recombining with holes within the base region.

The (X – Y – Z) electrons that have reached the collector region end up leaving the device through the collector terminal and thus represent the collector current $I_C$. The ratio $I_C/I_E$ is thus given by (X – Y – Z)/X. This ratio, denoted as $\alpha_F$ in practice, can be easily measured for any BJT and is termed the common-base current gain.

We can view the holes reaching the base as coming from the (+) terminal of the voltage source $V_{BE}$. They represent the base current $I_B$. To allow for the recombination of electron-hole pairs in both the emitter and base regions, (Y + Z) holes have to be brought to the base.

We recall that, in reality, no hole does actually come directly from the (+) terminal of the voltage source because the only current carriers in a conductor are the electrons. However, our terminology remains correct: when we refer to a hole entering the base, we imply that it is in fact an electron that leaves the base to flow towards the (+) terminal of the voltage source and what remains in the base after this electron's "departure" is a hole. So the whole thing appears as if a hole enters the base.




In addition to the equation $I_C = \alpha_F \cdot I_E$, we have another equation that links the three current entering and leaving the device: $I_E = I_B + I_C$, where $I_B$ denotes the current entering the base.

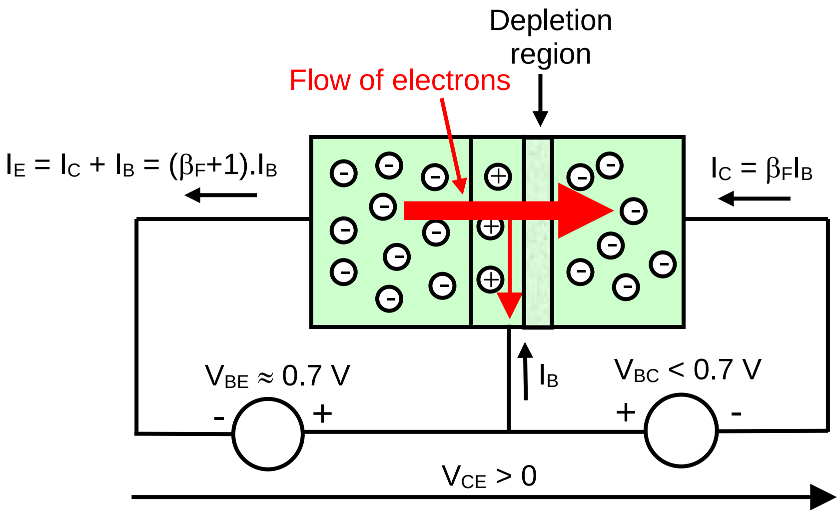
By combining these two expressions, we obtain

$$I_C = \alpha_F(I_B + I_C) \Rightarrow I_C = \frac{\alpha_F}{1-\alpha_F} I_B \Rightarrow I_C = \beta_F I_B,$$

where $\beta_F$ is a constant, called the *forward current gain*, that can take its value in the range from approximately 50 to 300 for typical bipolar technologies.

This expression is crucial because it clearly shows that the collector current $I_C$ is proportional to the base current $I_B$, and also much greater than the latter. The BJT is thus able to perform significant current amplification. This is known as the *transistor effect*.

The simplified diagram shown below summarises the working principle of the BJT.



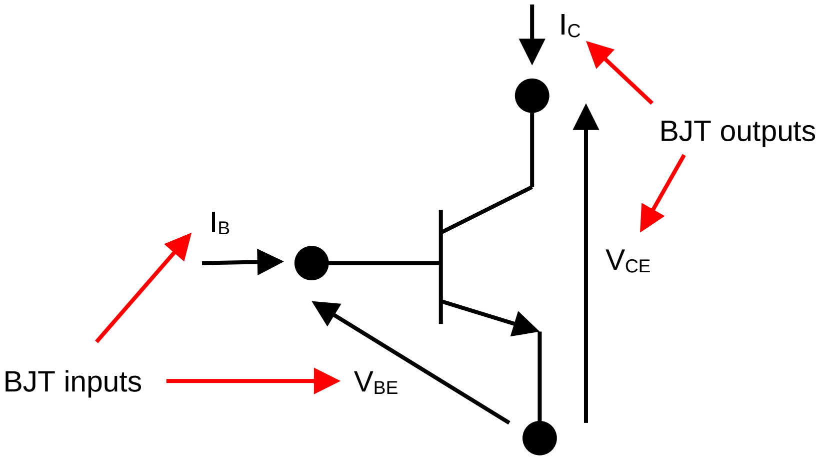

As previously mentioned, it is also possible to build a transistor by reversing the region types, thus leading to a PNP-type arrangement. With a PNP transistor, holes are drawn from the emitter into the base region by the forward bias, and are then pulled into the collector region by the higher negative bias.

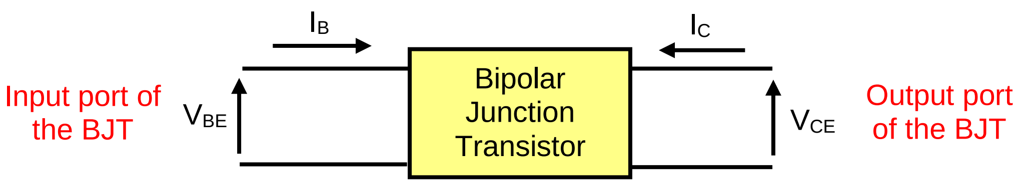
A PNP device basically works the same way and has the same general properties as the NPN transistor described above, except that all voltage polarities ($V_{BE}$, $V_{BC}$, and $V_{CE}$) and current directions ($I_B$, $I_C$, and $I_E$) are reversed.

It is also worthwhile mentioning that PNP transistors often have lower $\beta_F$ values and are slower (i.e., operate at lower frequencies) than their NPN counterparts.

### 3. Common-Emitter Configuration of a BJT

A BJT can be viewed as a semiconductor device with an input and an output. Usually, the input parameters are the current $I_B$ and the voltage $V_{BE}$, whereas the output parameters are the current $I_C$ and the voltage $V_{CE}$. This particular arrangement is referred to as common-emitter configuration because the emitter terminal is common to both input and output.




Note that common-collector and common-base configurations are also sometimes considered.

## 4. The Ebers-Moll Model of a BJT

In 1954, Jewell Ebers and John Moll introduced their model of a BJT.

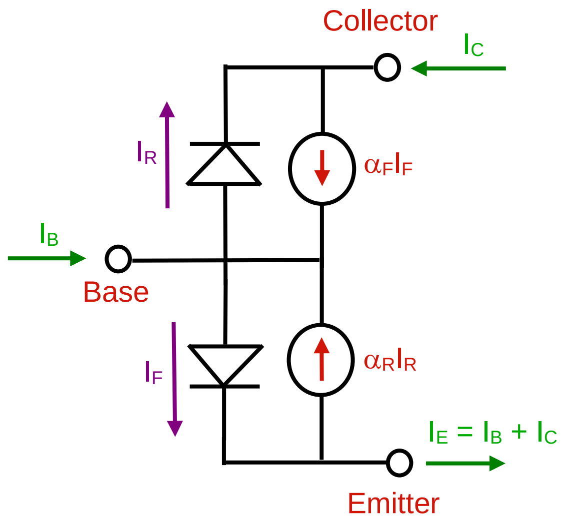
Jewell James Ebers (1921 – 1959)

No picture available for John Louis Moll (1921 - 2011)

The Ebers-Moll model for a NPN BJT is depicted below. Note that we only consider here the static version of the Ebers-Moll model in which the various parasitic capacitances of a BJT are not taken into account. The static model shown below is thus only suitable when the device operates at frequencies that remain sufficiently low.

$$I_F = I_S \left( \exp\left\{ \frac{V_{BE}}{V_T} \right\} - 1 \right)$$

$$I_R = I_S \left( \exp\left\{ \frac{V_{BC}}{V_T} \right\} - 1 \right)$$




In this model, $\alpha_F$ is the forward common-base current gain (typically ranging from 0.98 to 0.998 for most BJT technologies, i.e. $\alpha_F$ slightly smaller than the unit), and $\alpha_R$ is the reverse common-base current gain (typically, $\alpha_R \approx 0.5$).

Based on the Ebers-Moll model, we can write the general equations for the three currents entering or leaving a BJT:

- Base current: $I_B = I_F + I_R - \alpha_F I_F - \alpha_R I_R = (1 - \alpha_F) I_F + (1 - \alpha_R) I_R$
- Collector current: $I_C = \alpha_F I_F - I_R$
- Emitter current: $I_E = I_F - \alpha_R I_R = I_B + I_C$

with $I_F \approx I_S \exp\left\{ \frac{V_{BE}}{V_T} \right\}$ and $I_R \approx I_S \exp\left\{ \frac{V_{BC}}{V_T} \right\}$.

It is also worthwhile mentioning that $I_E = I_B + I_C$. This equation simply means that the current leaving a BJT ($I_E$) is equal to the sum of the two currents entering it ($I_B$ and $I_C$).

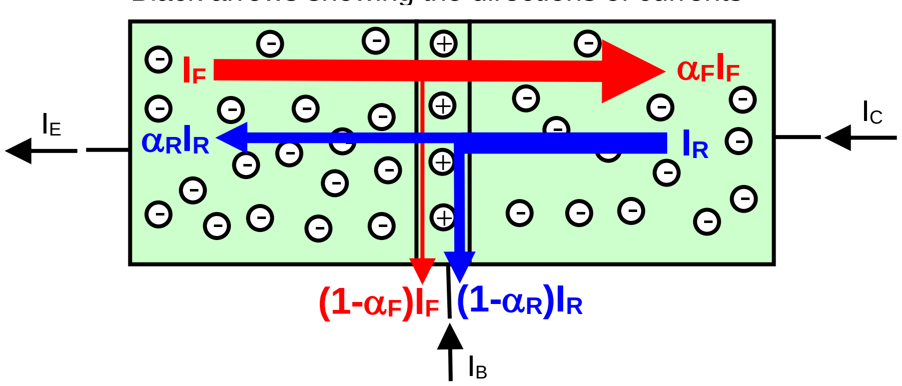
The physical phenomena behind the Ebers-Moll model are rather simple to understand:

1. Both diodes represent the base-emitter and base-collector PN junctions.

2. The parameter $\alpha_F$ represents the proportion of electrons coming from the emitter that are able to reach the collector. The fact that $\alpha_F$ is very close to the unit implies that the majority of electrons coming from the emitter do reach the collector, while the remaining electrons leave the device through the base.

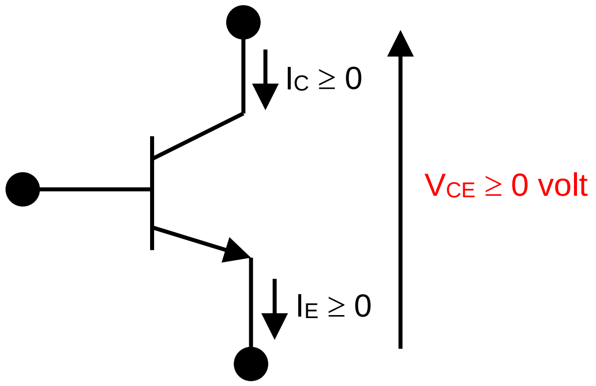
3. The parameter $\alpha_R$ represents the proportion of electrons coming from the collector that are able to reach the emitter. The fact that the value of $\alpha_R$ is (typically) approximately equal to 0.5 means that roughly half of the electrons coming from the collector end up leaving the transistor through the emitter.

The difference in values between $\alpha_F$ and $\alpha_R$ is due to the inherent non-symmetrical physical structure of a BJT (different doping levels for emitter and collector).




We are now going to derive two equations that will help us further understand the operation of a BJT. To simplify our calculations, we are hereafter going to assume that the BJT is connected to other electronic components so that the voltage $V_{CE}$ is always greater than or equal to zero ($V_{CE} \geq 0$).

In almost all circuits where NPN BJTs are employed, this assumption is valid simply because the collector is always on the side of the highest voltage whereas the emitter is on the side of the lowest voltage, thus implying that we always have $V_{CE} \geq 0$ volt.




Note that the inequality $V_{CE} \geq 0$ volt is strictly equivalent to $V_{BE} \geq V_{BC}$ because $V_{CE} = V_{CB} + V_{BE} = V_{BE} - V_{BC}$.

For simplicity sake, we are also going to assume hereafter the use of a silicon BJT.

**5. Input Characteristic of a BJT**

The input characteristic of a BJT shows the variation of the input current, i.e. the base current $I_B$, as a function of the input voltage, i.e. the voltage $V_{BE}$ between base and emitter.

To obtain this characteristic, we use the Ebers-Moll expression for the base current:

$$I_B = (1-\alpha_F)I_F + (1-\alpha_R)I_R \approx (1-\alpha_F)I_S \exp\left(\frac{V_{BE}}{V_T}\right) + (1-\alpha_R)I_S \exp\left(\frac{V_{BC}}{V_T}\right).$$

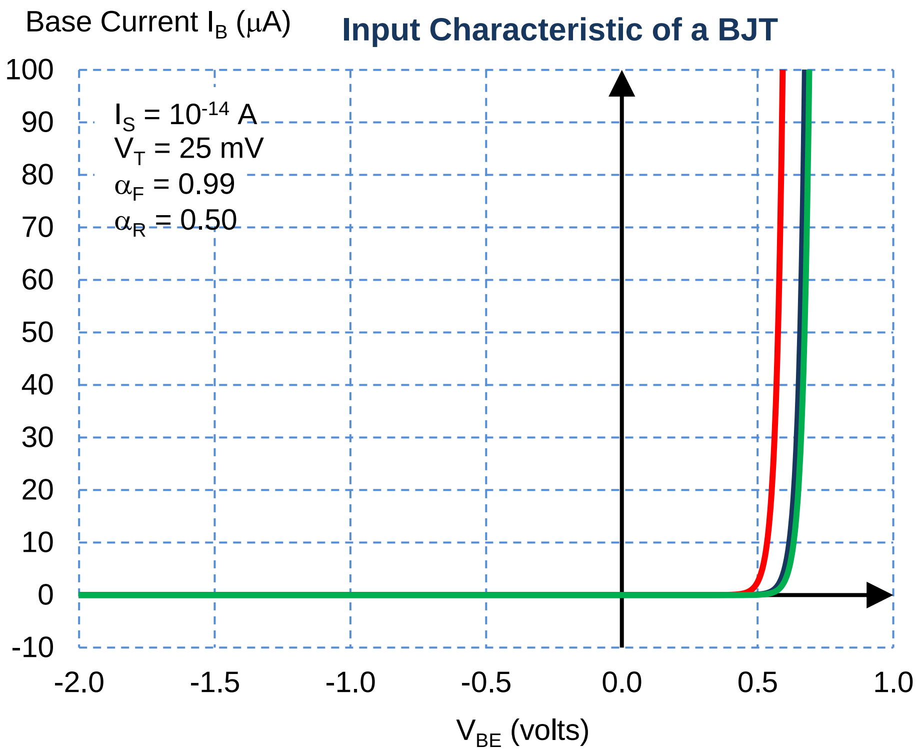
This equation can also be written as

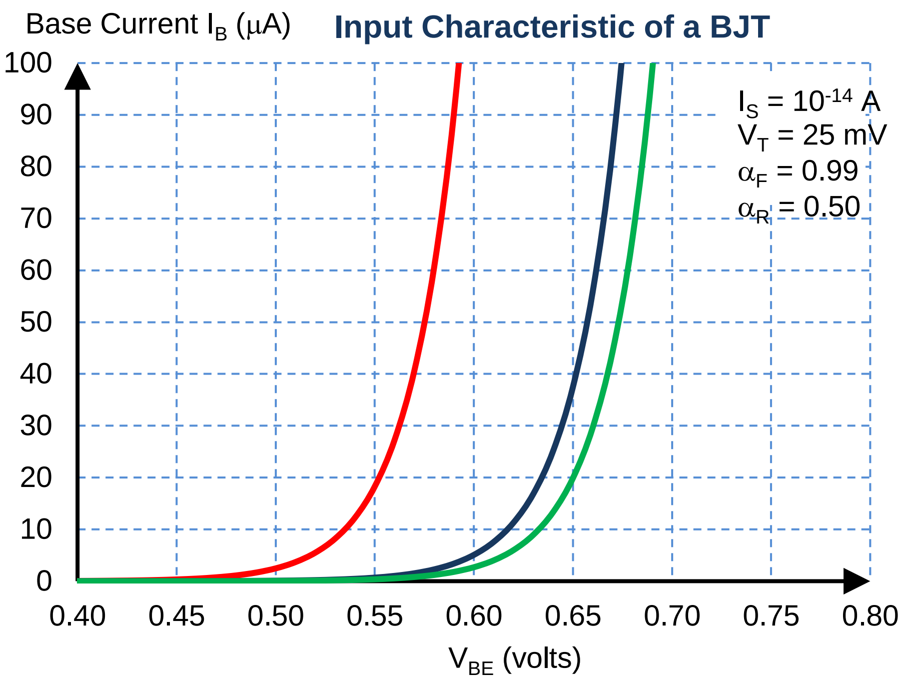
$$I_B \approx (1-\alpha_F)I_S \exp\left(\frac{V_{BE}}{V_T}\right) + (1-\alpha_R)I_S \exp\left(\frac{V_{BE}-V_{CE}}{V_T}\right)$$

=> $$I_B \approx (1-\alpha_F)I_S \exp\left(\frac{V_{BE}}{V_T}\right) + (1-\alpha_R)I_S \exp\left(\frac{V_{BE}}{V_T}\right)\exp\left(-\frac{V_{CE}}{V_T}\right)$$

=> $$I_B \approx \left[(1-\alpha_F)+(1-\alpha_R)\exp\left(-\frac{V_{CE}}{V_T}\right)\right] I_S \exp\left(\frac{V_{BE}}{V_T}\right).$$

The input characteristic corresponding to this equation is shown below for three different values of $V_{CE}$: $V_{CE}$ = 0 volt (red curve), $V_{CE}$ = 0.1 volt (blue curve), and $V_{CE}$ = 0.2 volt (green curve). For $V_{CE}$ > 0.2 volt, the corresponding characteristic is identical to that obtained with $V_{CE}$ = 0.2 volt.






These curves indicate that the base current $I_B$ varies exponentially with the voltage $V_{BE}$. In fact, the input characteristic of a BJT corresponds to that of a simple diode as if the current $I_B$ was the current flowing through the base-emitter junction.

Note that, strictly speaking, the threshold voltage of the base-emitter junction, hereafter denoted as $V_{BE,on}$, slightly depend on $V_{CE}$ if $V_{CE}$ becomes very close to zero. However, this can always be ignored in practice.

The fact that the input characteristic of a BJT corresponds to that of a diode allows us to define two possible configurations at this stage:

1. If $V_{BE} < V_{BE,on}$ ($V_{BE,on} \approx 0.7$ volt for a silicon BJT), the base current $I_B$ is equal to zero. In the next section, it will be shown that this is also the case for the collector and emitter currents ($I_C = I_E = 0$). The transistor is then said to be in the *cut-off mode of operation*.

2. If $V_{BE} = V_{BE,on}$, the base current $I_B$ is greater than zero ($I_B > 0$). In the next section, it will be demonstrated that this implies that $I_C > 0$ and $I_E > 0$. The transistor is then in a mode of operation which is not the cut-off mode.

To further study the operation of a BJT, we now need to introduce a second type of characteristic: the output characteristic.

## 6. Output Characteristic of a BJT

The output characteristic of a BJT shows the variation of the output current, i.e. the collector current $I_C$, as a function of the output voltage, i.e. the voltage $V_{CE}$ between collector and emitter.

To obtain this characteristic, we use the Ebers-Moll expression for the collector current:

$$I_C = \alpha_F I_F - I_R .$$

This equation can also be written as

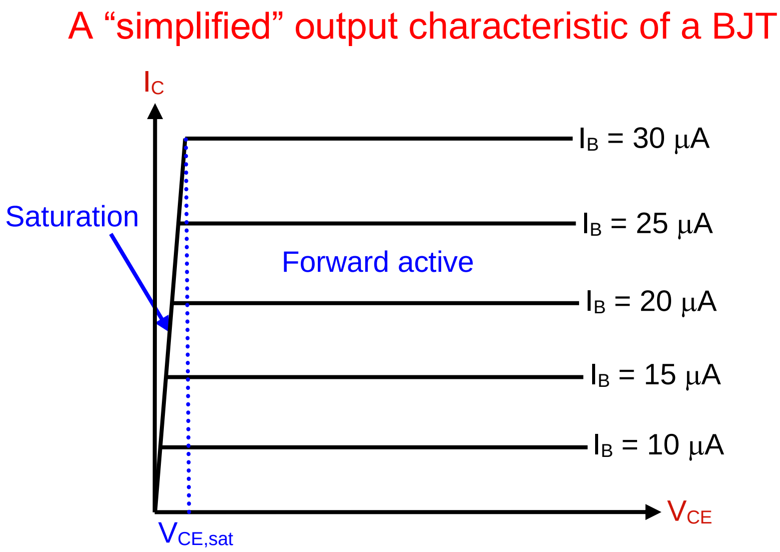
$$I_C \approx \alpha_F I_S \exp\left(\frac{V_{BE}}{V_T}\right) - I_S \exp\left(\frac{V_{BE}}{V_T}\right)\exp\left(-\frac{V_{CE}}{V_T}\right)$$

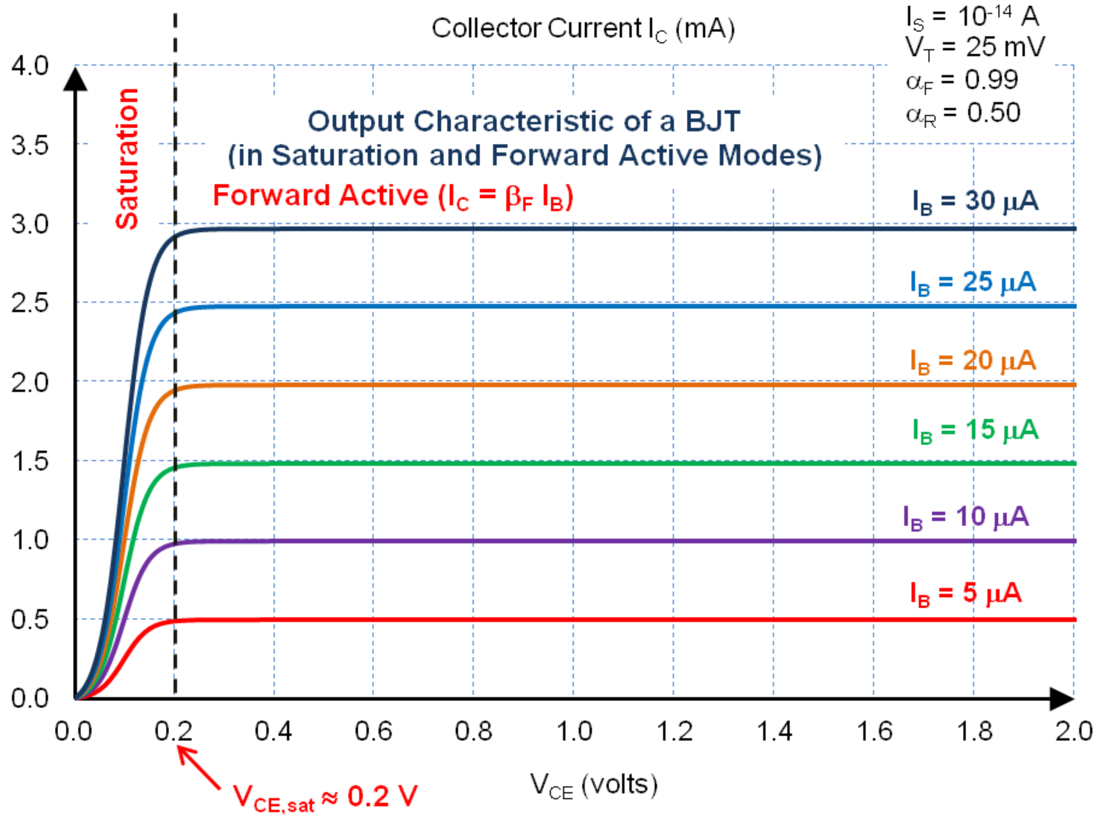
=> $$I_C \approx \left(\alpha_F - \exp\left(-\frac{V_{CE}}{V_T}\right)\right) I_S \exp\left(\frac{V_{BE}}{V_T}\right)$$

=> $$I_C \approx I_B \frac{\alpha_F - \exp\left(-\frac{V_{CE}}{V_T}\right)}{(1-\alpha_F) + (1-\alpha_R)\exp\left(-\frac{V_{CE}}{V_T}\right)} .$$

This expression clearly shows that, if $I_B = 0$, we have $I_C = 0$, and then $I_E = I_C + I_B = 0$. It also indicates that, if $I_B > 0$, we have $I_C > 0$ (because $\alpha_F > \exp\left(-\frac{V_{CE}}{V_T}\right)$ when $V_{CE} > 0$), and then $I_E = I_C + I_B > 0$.

We can see below the output characteristic of a BJT which shows the variation of the collector current $I_C$ (output current) as a function of the voltage $V_{CE}$ (output voltage), for different values of the base current $I_B$.




It is observed that, depending on the value of $V_{CE}$, the BJT exhibits two very different types of behaviour:

1. If $V_{CE}$ is smaller than a constant called $V_{CE,sat}$ (≈ 0.2 volt), the collector current $I_C$ depends on both $V_{CE}$ and $I_B$. In particular, we notice that a small increase in $V_{CE}$ above 0 volt results

in a very large increase in the collector current $I_C$. The BJT is then said to be in the *saturation mode of operation*.

2. Once $V_{CE}$ becomes greater than $V_{CE,sat}$, the collector current $I_C$ does no longer depend on $V_{CE}$. Instead, it is only a function of the base current $I_B$. The BJT is then said to operate in the *forward active mode*.

The fact that, in the forward active mode, $I_C$ only depends on $I_B$ can be confirmed by simplifying the general equation previously obtained:

$$I_C \approx I_B \frac{\alpha_F - \exp\left(-\frac{V_{CE}}{V_T}\right)}{(1-\alpha_F) + (1-\alpha_R)\exp\left(-\frac{V_{CE}}{V_T}\right)}$$

=> $I_C \approx \frac{\alpha_F}{1-\alpha_F} I_B = \beta_F I_B$, for large values of $V_{CE}$ (in fact, $V_{CE} > V_{CE,sat}$).

This is due to the fact that $\exp(-x) \to 0$ as $x \to +\infty$.

The parameter $\beta_F = \frac{\alpha_F}{1-\alpha_F}$ is known as the *forward current gain*.

For instance, with $0.98 < \alpha_F < 0.998$, we have $49 < \beta_F < 499$. For simplicity sake, the value $\beta_F = 100$ will be adopted in all examples throughout these lecture notes.

In other words, in the forward active mode, the BJT behaves as a current amplifier in the sense that the output (collector) current $I_C$ is an amplified version of the input (base) current $I_B$. This particular and fascinating property is what has made the BJT so useful for many applications in the field of analogue electronics.

The forward active mode is by far the most important mode of operation in the context of EEE1004. In fact, the design of amplifiers in analogue electronics requires the use of BJTs that operate in the forward active mode. This is explained by the fact that the only mode of operation in which a BJT acts as a current amplifier is the forward active mode.

## 7. Summary – What do we need to remember?

We have only considered NPN bipolar junction transistors throughout this chapter. We have also assumed that $V_{CE} \geq 0$, which is (almost) always the case in practice.

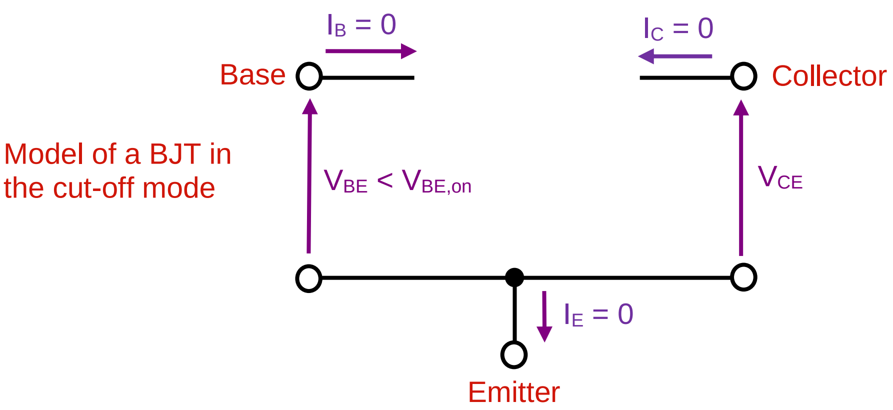
We will need to remember the following points about an NPN BJT.

(1) A BJT can operate in three possible modes: the cut-off mode, the forward active mode, and the saturation mode.

(2) The following equations are always valid: $I_E = I_C + I_B$, $I_B \geq 0$, $I_C \geq 0$, and $I_E \geq 0$.

(3) The base current $I_B$ can always be viewed as flowing through a diode with a voltage $V_{BE}$ across it. The threshold voltage of this diode is traditionally denoted as $V_{BE,on}$. Throughout these notes, we will always consider the use of silicon BJTs for simplicity sake. For a silicon BJT, we have $V_{BE,on} \approx 0.7$ volt.

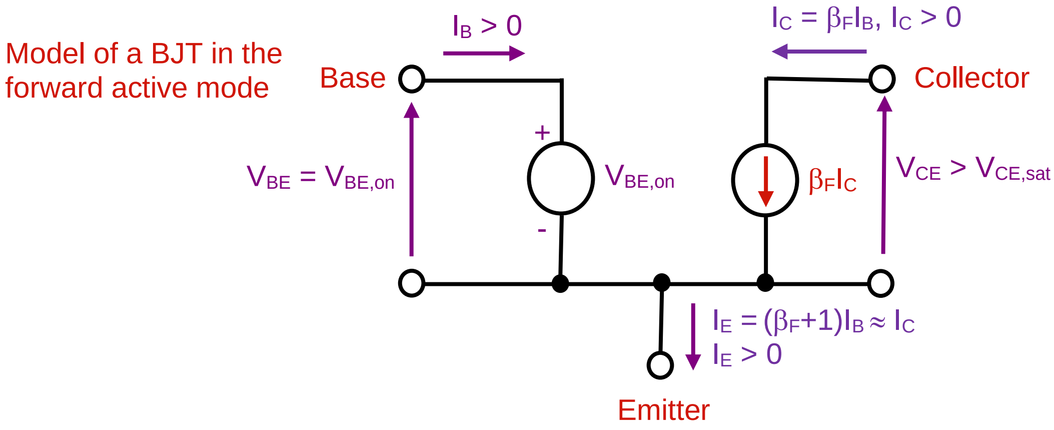
(4) If $V_{BE} < V_{BE,on}$, the BJT is in the cut-off mode of operation, and we thus have $I_B = I_C = I_E = 0$.




The cut-off mode is not suitable for designing linear analogue electronic circuits such as linear voltage or current amplifiers. However, the cut-off mode can be used to design logic gates employing BJTs (see, e.g., the RTL, DTL, TTL, and ECL logic families that were very popular from the 1960s to the 1980s) and non-linear amplifiers (class-B and class-C amplifiers that are more power-efficient than linear amplifiers).

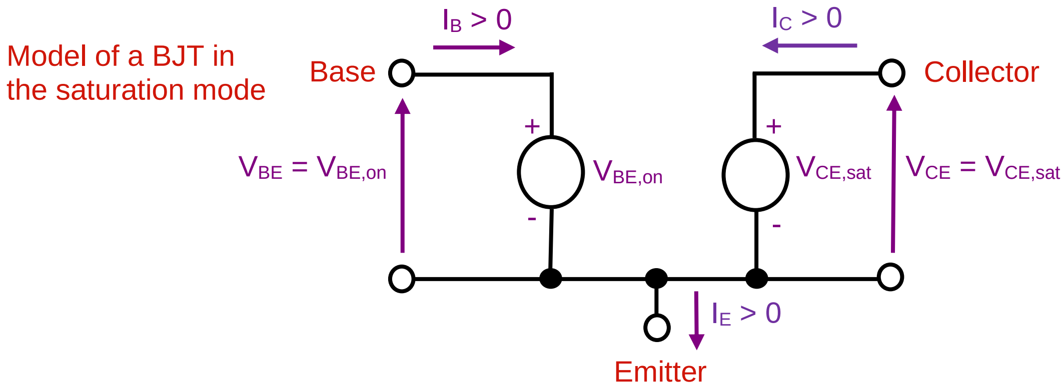
(5) If $V_{BE} = V_{BE,on}$, the BJT is either in the forward active mode or in the saturation mode, meaning that $I_B > 0$, $I_C > 0$, and $I_E > 0$.

(6) If $V_{BE} = V_{BE,on}$ and $V_{CE} > V_{CE,sat}$ (≈ 0.2 volt), the BJT is in the forward active mode, meaning that $I_C = \beta_F I_B$ and $I_E = I_C + I_B = \beta_F I_B + I_B = (\beta_F+1)I_B \approx \beta_F I_B$ (since $\beta_F >> 1$) $= I_C$.




The forward active mode is particularly suitable for all analogue electronic applications in which linear functions are to be implemented. This mode can also be used in conjunction with the cut-off mode in order to design logic gates that can switch very fast (ECL logic family which used to be the fastest logic family on silicon until the 1990s).

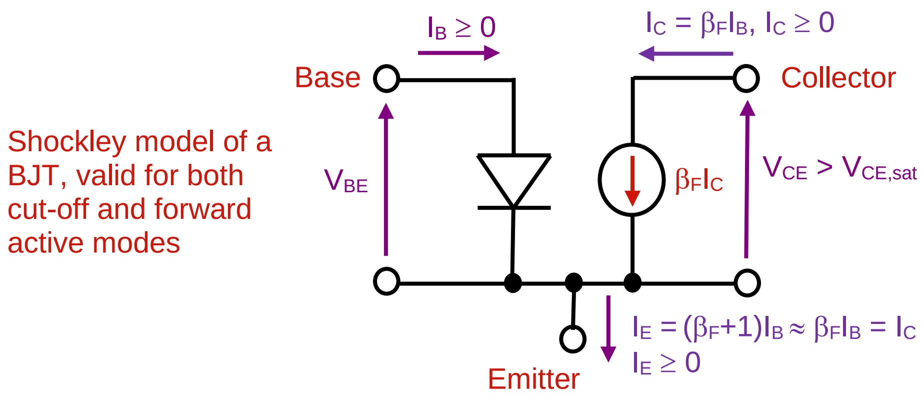
(7) If $V_{BE} = V_{BE,on}$ and $V_{CE} < V_{CE,sat}$, the BJT is in the saturation mode, meaning that $I_C$ is no longer equal to $\beta_F I_B$. For simplicity sake, we often assume that $V_{CE} = V_{CE,sat}$ in this mode. This simplification does not result in any significant error as the actual value of $V_{CE}$ lies somewhere between 0 volt and $V_{CE,sat} \approx 0.2$ volt.




The saturation mode is not suitable for designing analogue electronic circuits that perform linear functions. However, the saturation mode can be used in conjunction with the cut-off mode to design logic gates employing BJTs (saturated bipolar logic families of the 1960s and 1970s, e.g. RTL, DTL, and TTL).

(8) There is unfortunately no single model that is valid for all three modes of operation. However, there is a model that can be used for both cut-off and forward-active modes. Hence, as long as the BJT does not enter saturation, this model can be employed without having to make a distinction between cut-off and forward active modes, which can be quite convenient when analysing some particular circuits.

Hereafter, this model shown below will be referred to as *Shockley model*.




With the Shockley model, the base current is given by

$$I_B \approx I_S \exp\left\{\frac{V_{BE}}{V_T}\right\},$$

whereas the collector current can be expressed as

$$I_C = \beta_F I_B \approx \beta_F I_S \exp\left\{\frac{V_{BE}}{V_T}\right\}.$$

One can see that these equations are valid for both cut-off and forward active modes as no assumption needs to be made as to whether the base-emitter junction is on or off. The Shockley model can sometimes be used if greater accuracy is desired (and achievable) when performing the manual analysis of a circuit.

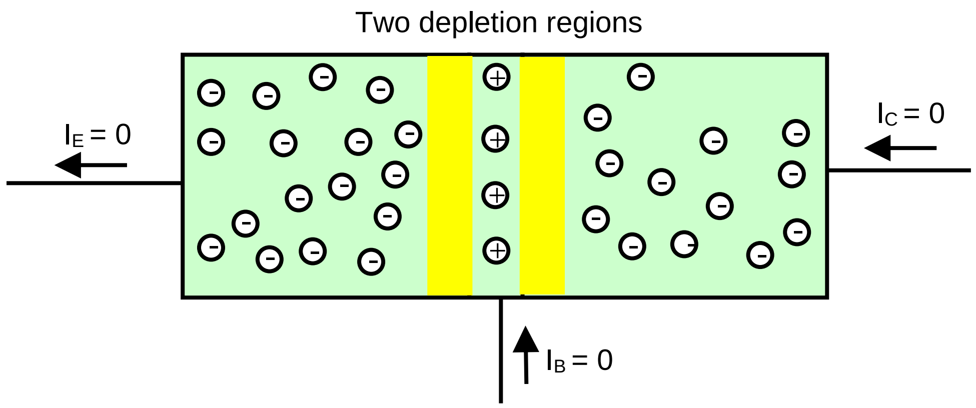
**8. Physical Interpretations of the Three Modes of Operation**

We have identified three modes of operation for a BJT using equations derived from the Ebers-Moll model. But, beyond mathematical expressions, what is the physical meaning of those modes? To answer this question, let us consider each possible mode one at a time.

- Cut-off mode: A BJT operates in the cut-off mode when $V_{BE} < V_{BE,on}$, i.e. the base-emitter junction is off.

What about the base-collector junction in this case? Combining the equation $V_{CE} = V_{CB} + V_{BE} = V_{BE} - V_{BC}$ with the assumption that $V_{CE} \geq 0$ volt, we obtain the inequality $V_{BE} \geq V_{BC}$. Since $V_{BE} < V_{BE,on}$, this means that $V_{BC} < V_{BE,on}$ in the cut-off mode.

It can be concluded that the cut-off mode corresponds to the case where both base-emitter and base-collector junctions are off at the same time. In other words, the base is isolated from both emitter and collector due to the existence of two depletion regions (shown in yellow in the drawing below). No current can therefore flow through the device and we have $I_B = I_C = I_E = 0$.

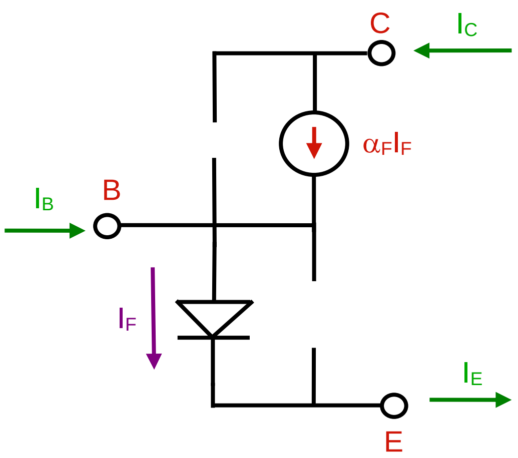


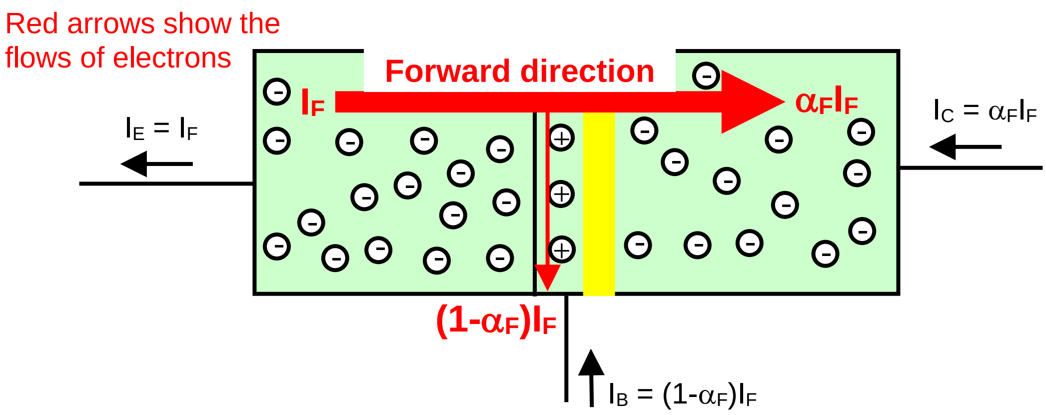

- Forward active mode: A BJT operates in the forward-active mode when $V_{BE} = V_{BE,on}$, i.e. the base-emitter junction is on, and $V_{CE} > V_{CE,sat}$.

Since $V_{CE} = V_{BE} – V_{BC}$, the latter condition implies that $V_{BC} < V_{BE,on} – V_{CE,sat}$, which is approximately equivalent to $V_{BC} < 0.5$ volt. This result indicates that the base-collector junction is clearly off.

We conclude that the forward-active mode corresponds to the case where the base-emitter junction is on, whereas the base-collector junction is off.

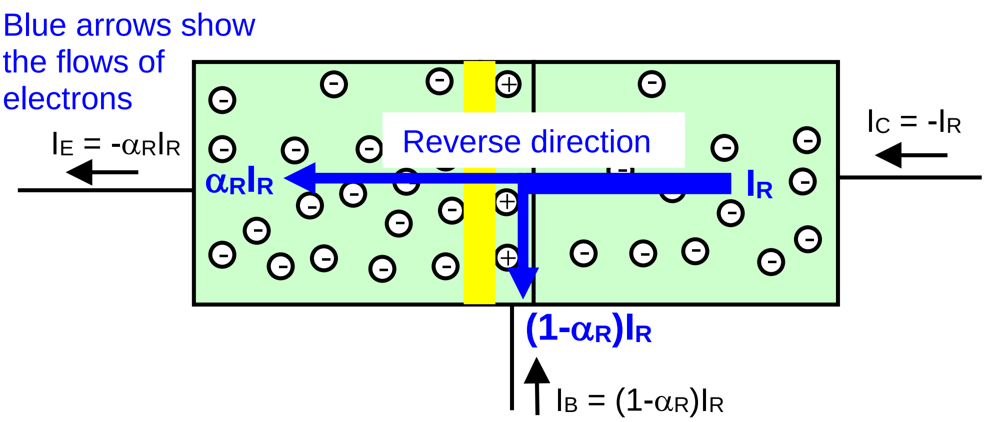
The corresponding Ebers-Moll model and physical structure of a BJT are depicted below.






The electrons flow in the "forward direction", i.e. from the heavily-doped region (emitter) towards the moderately-doped region (collector). The difference in doping levels between the emitter and collector regions is the reason why the current amplification effect ($I_C = \beta_F I_B$) works so well in the forward direction.

Note that we could also reverse the polarities of a BJT (base-emitter junction off, base-collector junction on, $V_{CE} < 0$) in order to make electrons flow from the collector towards the emitter.




The BJT would then operate in a mode called the *reverse active mode* (that is actually a fourth possible mode of operation) and the emitter current $I_E$ would be given by

$$I_E = -\frac{\alpha_R}{1-\alpha_R} I_B = -\beta_R I_B ,$$

where the parameter $\beta_R$ is known as the reverse current gain.

At first glance, the reverse active mode seems to provide a current amplification effect similar to that obtained in the forward active mode as the emitter current is clearly proportional to the base current. However, the current amplification is non-existent simply because $\beta_R << \beta_F$ (for instance, a typical $\alpha_R$ of 0.5 yields a $\beta_R$ value approximately equal to the unit only). Clearly, the current amplification effect does not work at all when electrons flow in the "reverse direction", i.e. from the moderately-doped collector region towards the heavily-doped emitter region.

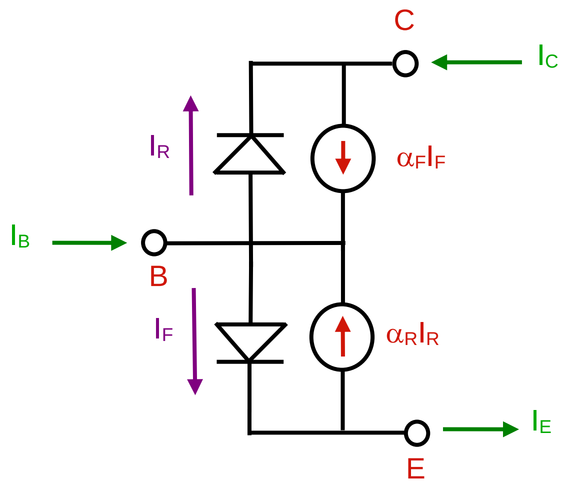
In any case, in EEE1004, we do not have to worry about the reverse mode as the assumption that $V_{CE} \geq 0$ volt makes it impossible for a BJT to operate in this mode.

- Saturation mode: A BJT operates in the saturation mode when $V_{BE} = V_{BE,on}$, i.e. the base-emitter junction is on, and $V_{CE} < V_{CE,sat}$.

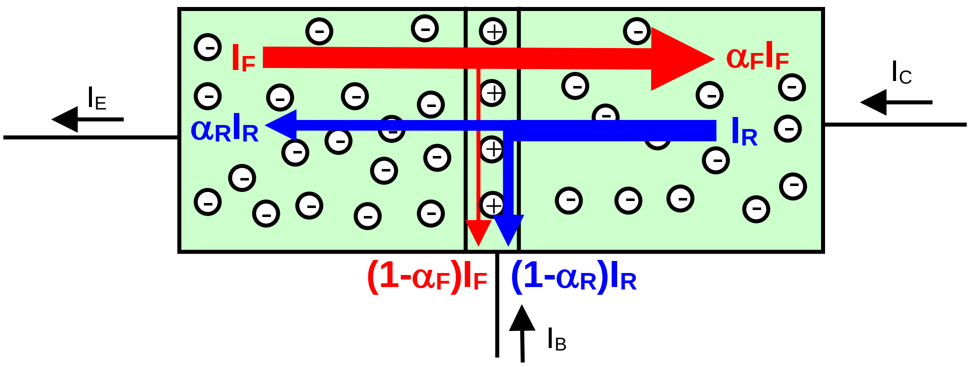
Since $V_{CE} = V_{BE} - V_{BC}$, the latter condition implies that $V_{BC} > V_{BE,on} - V_{CE,sat}$, which is approximately equivalent to $V_{BC} > 0.5$ volt. This result indicates that the base-collector junction is either on or "almost on" (in a silicon transistor, when $V_{BC} \approx 0.5 - 0.6$ volt, one can consider the base-collector junction as being not completely off, but not fully on either). As $V_{BC}$ becomes closer to the "official" threshold voltage of 0.7 volt for a silicon diode, the current flowing through the base-collector junction increases drastically and the BJT goes deeper into saturation.

It thus clearly appears that the saturation mode corresponds to the case where both base-emitter and base-collector junctions are on at the same time.

The corresponding Ebers-Moll model and physical structure of a BJT are depicted below.

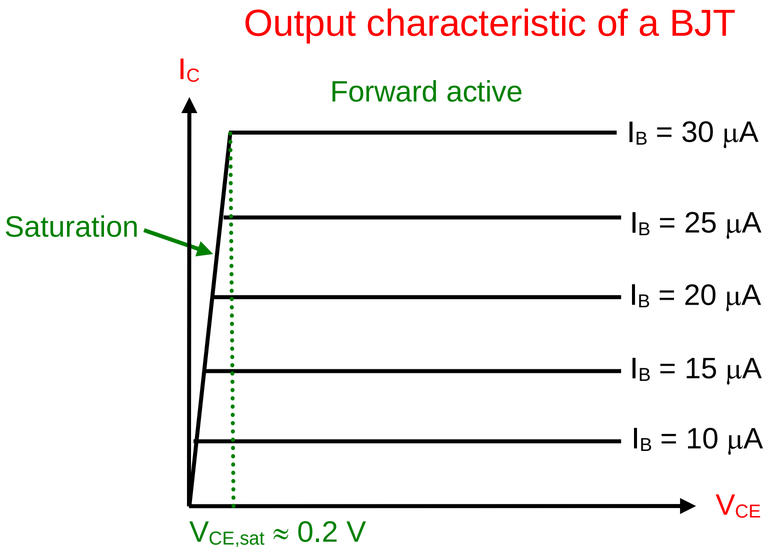



Arrows showing the flows of electrons



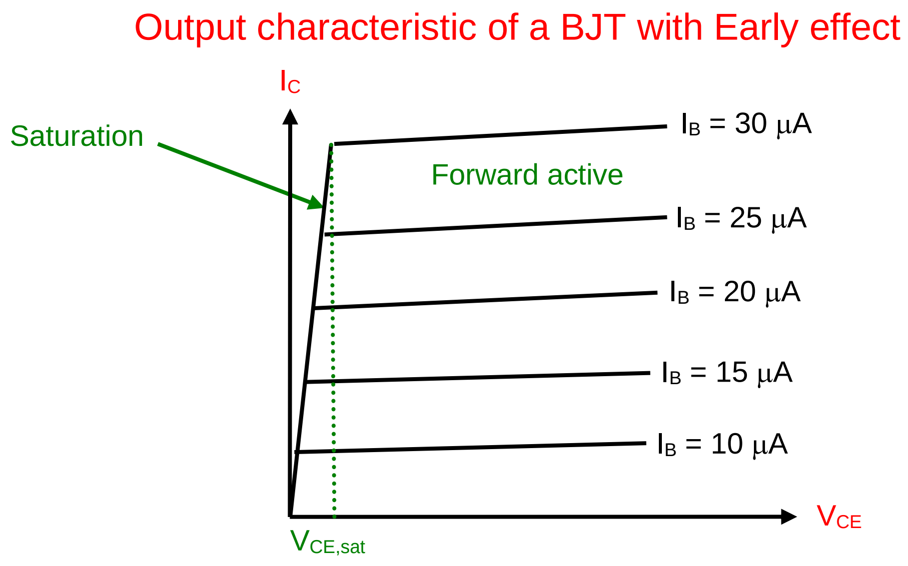

**9. The Early Effect**

In the forward active mode, we have seen that the collector current $I_C$ is proportional to the base current $I_B$, i.e. $I_C = \beta_F I_B$ . This means that $I_C$ is not expected to depend on the voltage $V_{CE}$. In other words, the output characteristic is expected to be flat in the forward mode.




However, in practice, the collector current also slightly depends on the voltage $V_{CE}$. This slight dependence of $I_C$ on $V_{CE}$, which cannot be accounted for by the Ebers-Moll model considered throughout this chapter, is due to the Early effect, named after the American engineer James Early.

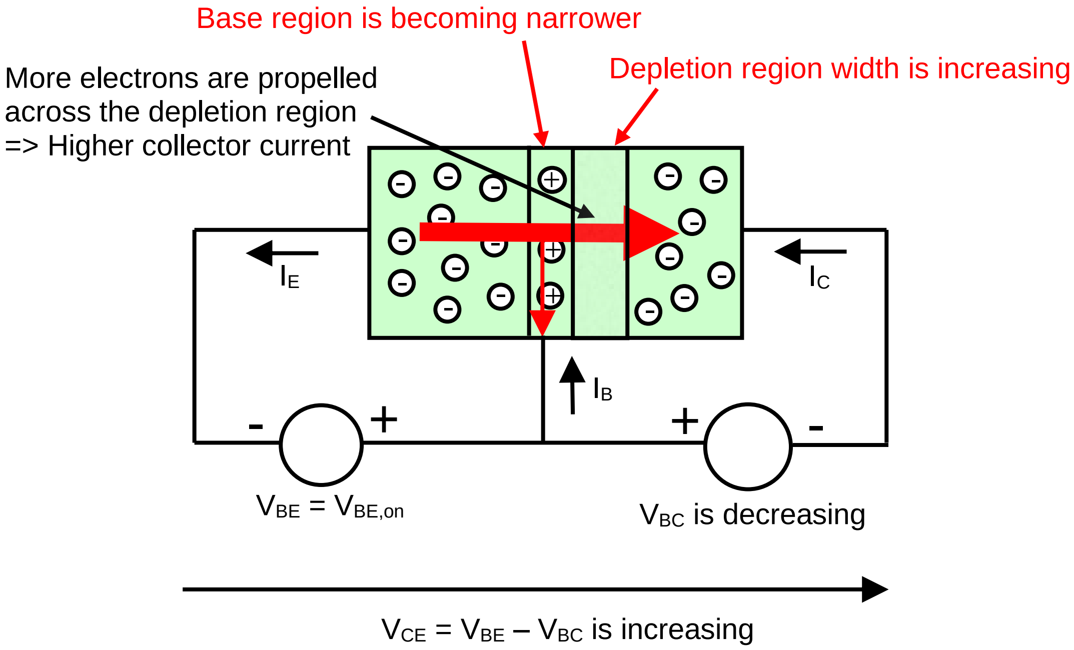
James M Early (1922 - 2004)




The Early effect is in fact the variation in the width of the base due to a variation in $V_{BC}$. Remember that $V_{BC} = V_{BE} + V_{EC} = V_{BE} - V_{CE}$. A greater reverse bias across the base-collector (due to an increasing voltage $V_{CE}$), for example, increases the base-collector depletion width, thus decreasing the width of the narrow P-type semiconductor layer forming the base.

The narrowing of the base has two consequences that affect the collector current:
(1) There is a lesser chance for electron-hole recombination within the "smaller" base region.
(2) The potential gradient is increased across the base-collector depletion region and, consequently, more electrons are propelled across it.

Both these factors increase the collector current $I_C$ with a decrease in $V_{BC}$, i.e. an increase in $V_{CE}$.



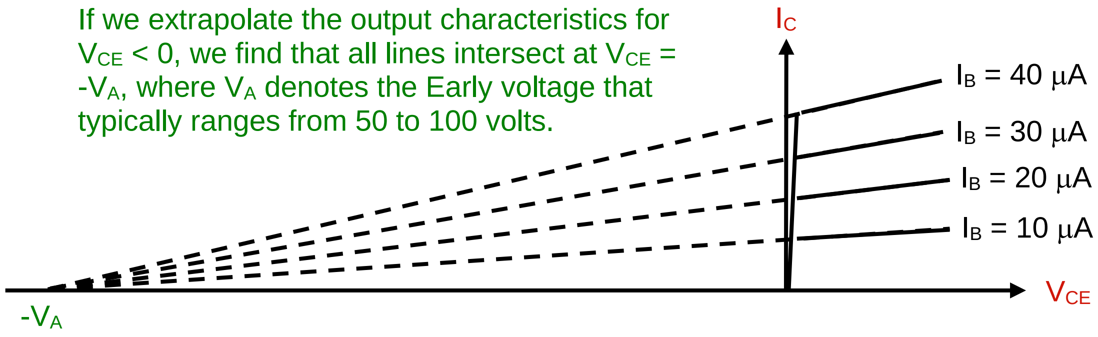

If we decide to take the Early effect into account, the expression of the collector current in the forward active mode of operation is no longer $I_C = \beta_F I_B$ . Instead, it is replaced with

$$I_C = \beta_F I_B \left(1 + \frac{V_{CE}}{V_A}\right),$$

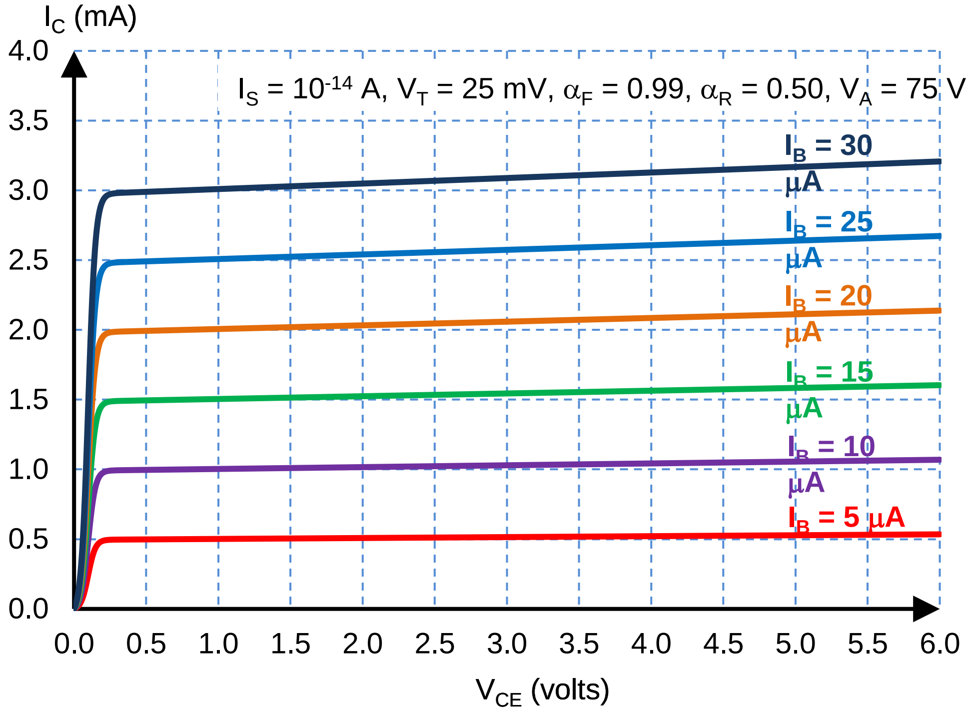
where $V_A$ denotes the Early voltage that typically ranges from 50 to 100 volts.

The slope of the output characteristic for a given base current $I_B$ is thus given by

$$\frac{dI_C}{dV_{CE}} = \frac{\beta_F I_B}{V_A}.$$



As a practical illustration, we show below the output characteristic obtained when $V_A$ = 75 volts.




**- END -**