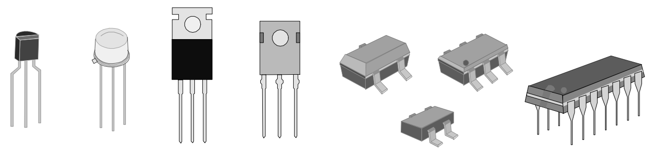
27

BIPOLAR JUNCTION TRANSISTORS



Transistor packages

27.1 Introduction

Transistors are found in virtually every electronic system.
They are the building blocks for microprocessors and most of the other integrated circuits.

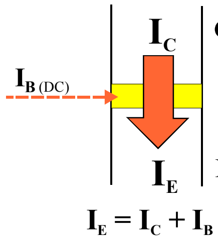
There are two main types of transistors: a bipolar junction transistor (BJT) and a field effect transistor (FET). We shall begin with the study of bipolar transistors. These are the types of transistors that electronics students have traditionally first encountered during their studies.

Transistors are used either to amplify signal or used as a switch. The term *amplification* may sometimes give a student the wrong idea that transistors are some kind of "miracle devices" that can create energy from nothing. That is of course not true. Transistors do not actually amplify current. For instance, BJTs use very small current to control much larger current that you, as a circuit designer, have made available. This "control" gives the impression of amplification.

Before we start looking at the transistor construction, it is better that we first look at a simplified view how transistors control this much higher current and how transistors operate as a switch.

Bipolar transistors have three regions: **collector**, **base**, and **emitter**.
The large main current through a NPN transistor is from the collector to the emitter and in a PNP transistor from the emitter to the collector.

NPN bipolar transistor

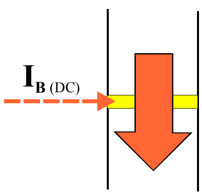


Imagine that a transistor operates like a valve. A water valve, for instance, controls the rate of water through it.
A transistor operation has the same idea.
In the illustrations, the yellow rectangular area represents a valve and the thickness of the area indicates the valve position.
A wider yellow area indicates a more closed valve position.

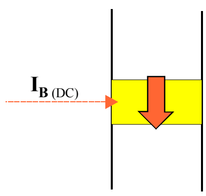
Instead of a mechanical device that controls the rate of water through the valve, in transistors, it is the transistor's base current (I_B) that controls the main current from the collector to the emitter. The higher the base current, the higher the main transistor current.

As you can see, we have labeled the collector current as I_C and the emitter current as I_E.
The reason why the two currents are not exactly the same is due to the very small base current I_B that is added to the main current flowing towards the emitter. Therefore, $I_E = I_C + I_B$.

The two illustrations below show the basic transistor operation.

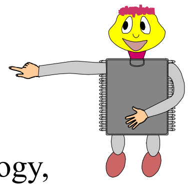


As the I_B is increased, the "valve" opens further and the main current increases.



As the I_B is decreased, the "valve" closes more and the main current decreases.

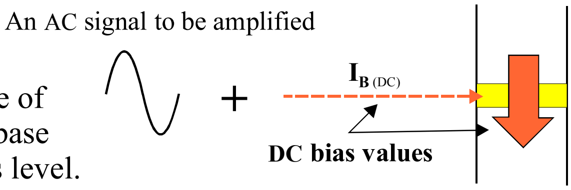
I can see how the small base current controls much larger main current but where does the signal amplification takes place?



A signal is an AC current that is also connected to the base. In our water valve analogy, an AC signal at the base is equivalent of further opening and closing a valve.

Before a transistor can amplify an AC signal, the transistor must be biased. Recall, the term biasing refers to the DC voltage that sets the operating conditions for a device. Transistor biasing sets the base and the main transistor currents at desired levels. You could think biasing is setting the valve to a certain "operating position" with desired level (rate) of water flow.

Transistor Biasing

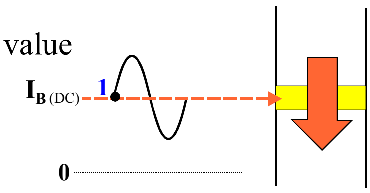


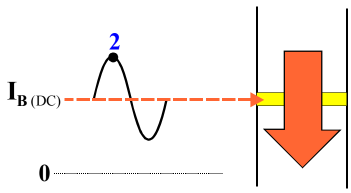
The base and the main current are set to a specific bias or oprating value.

Now, if there is an AC signal that is to be amplified, this signal is fed to the base of the transistor. The base current is now the value of an AC current about the DC bias level. So, the base current fluctuates above and below the DC bias level.

Let's see how an AC signal effects the main current.

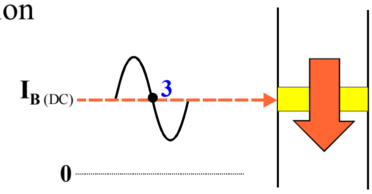
At point 1, the AC current has the same value as the bias DC current. The base current and the collector currents do not change from their bias values.

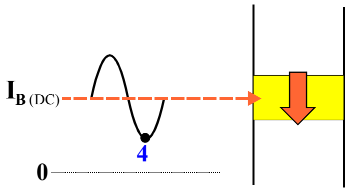




At point 2, the base current has increased by the peak AC current value. This causes "the valve" to open further and this increases the main transistor current.

At the point 3, we have the same situation as at the point 1. The total base current equals the bias DC current ($I_{B\,(DC)}$).





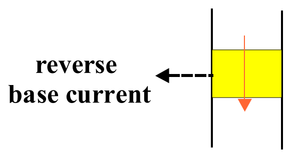
Finally, at point 4, the AC peak current value is less than the DC current ($I_{B\,(DC)}$). Therefore, the base current is less than $I_{B(DC)}$. The "valve" closes beyond its bias position. The main transistor current is less than the current at its bias value.

Do you see now where the impression of amplification comes from? The main transistor current increases and decreases according to a fluctuations of a small AC signal at the base. Therefore, the transistor current is a copy of the base AC current but is much larger (up to 200 times larger).

Transistor as a switch

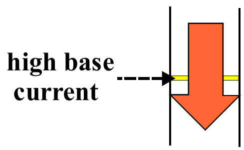
In digital electronics, transistors are used as switches. There is no need to bias a transistor. If we reversed the current to the base, there is virtually no current through the transistor. (The valve is in fully closed position.)

Transistor switch is closed



If we start increasing the base current, at certain point, the main transistor current cannot increase any further. Transistor current has reached its maximum value. (Valve is in fully open position.)

Transistor switch is open

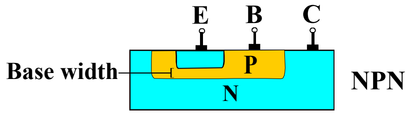


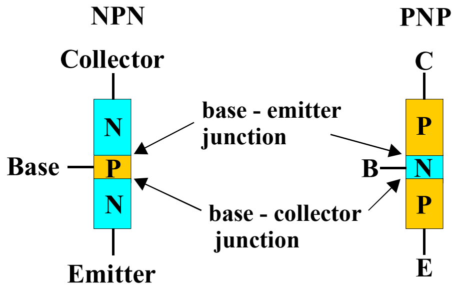
27.2 Transistor Construction

A bipolar transistor consists of three doped semiconductor regions: collector, emitter, and base which are separated by two P-N junctions: *base - collector* and *base - emitter*. A wire lead is connected to each region and are labeled C, B, and E for collector, base, and emitter.

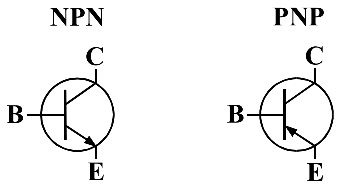
The emitter region is smaller than the collector region and is more heavily doped. The base region is very thin and only lightly doped.

Bipolar transistor construction





On the right are schematic symbols of both NPN and PNP bipolar junction transistors.
The bipolar refers to the carriers in the transistor.
Both holes and electrons are carriers, hence the term bipolar.



Bipolar transistor symbols

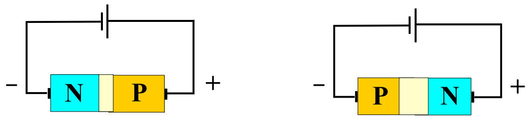
27.3 Transistor Operation

In the introduction section, the basic transistor operation was already discussed.
In this section, we shall look at the transistor operation by using an energy band diagram.

Since transistors have two back-to-back P-N junctions, you can think that a transistor consists of two diodes.

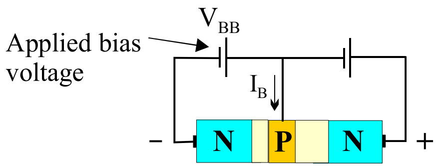
One diode is forward-biased and the other one is reversed-biased as shown on the right.
For clarity, we have also drawn the depletions layers as yellow.

Forward biased diode **Reverse biased diode**



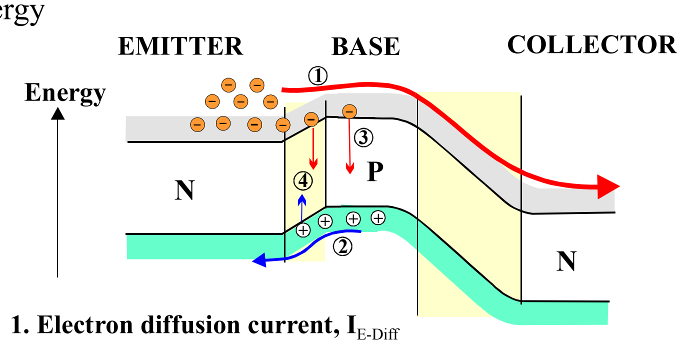
The illustration on the right shows the two diodes together. The forward biased voltage is also the applied bias voltage that controls the base current (I_B).
The other voltage source does not play part in transistor biasing. This is more clear by looking at the energy band diagram.

NPN transistor with bias voltage



If you do not understand this energy band diagram, you should review the section 25.7.

To simplify the diagram, only the most important carrier currents are shown.



The forward biasing allows electrons in the emitter region to overcome the potential hill (the built-in potential).
As the applied voltage increases, more electrons are injected across the base - emitter junction.

1. Electron diffusion current, $I_{E\text{-}Diff}$
2. Hole diffusion current, $I_{H\text{-}Diff}$
3. Base recombination current, $I_{B\text{-}R}$
4. Base-emitter depletion layer recombination current, $I_{BE\text{-}R}$

$$I_C = I_{E\text{-}Diff} - I_{B\text{-}R}$$
$$I_E = I_{E\text{-}Diff} + I_{H\text{-}Diff} + I_{B\text{-}R}$$
$$I_B = I_{H\text{-}Diff} + I_{B\text{-}R} + I_{BE\text{-}R}$$

When the applied bias voltage equals or exceeds the built-in potential (around 0.7V for silicon), the B-E diode is "turned on" and electrons flow freely across the base-emitter junction.

If we look at the electron diffusion current, the main difference between a transistor diffusion current and a diode diffusion current is that in a transistor, most of the electrons that have been injected across the base-emitter junction, now as minority carriers, do not recombine but rather diffuse across the base region. Once at the base - collector depletion layer, it is "downhill" for the electrons to the collector. The downhill slope in the energy band diagram means that the electric field in the depletion layer is pulling electrons towards the collector.

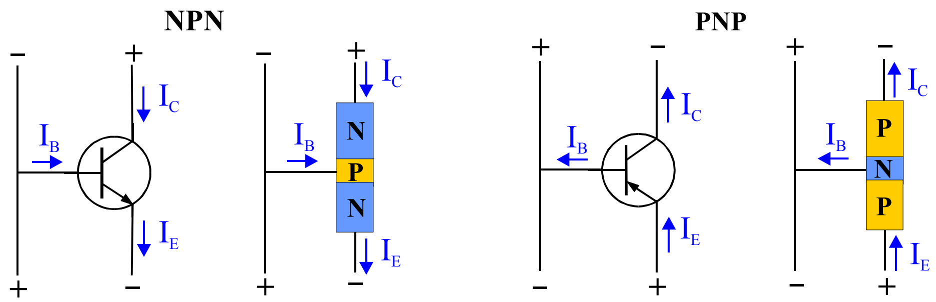
As you can see, the reverse bias voltage does not control the electron current from the emitter to the collector. Only the forward bias voltage controls it.

There are four main internal transistor carrier currents: the electron diffusion current, the hole diffusion current, the base recombination current, and the base - emitter depletion layer recombination current.
As electrons diffuse across the base - emitter junction and across the base region, some recombination takes place in those regions.

The collector current is the difference between the electron diffusion current and the base recombination current.
The emitter current consists of the electron diffusion current, the hole diffusion current, and the base recombination current.
The base current consists of the hole diffusion current, base recombination current, and base - emitter depletion layer recombination current.

Note: In the energy band diagram, (internal) current directions shown are for electron and hole currents. The directions of conventional electric currents of transistors are shown below.

Transistor currents and directions



$$I_E = I_C + I_B$$

The ratio of the collector current (I_C) to the emitter current (I_E) is called the **current gain factor** (α).

$$\alpha = \frac{I_C}{I_E}$$

Another important ratio is the collector current to the base current.
We have already discussed the relationship between the applied bias voltage (between the base and the emitter) and the base current. Instead of saying that the applied bias voltage controls the large transistor current (collector current), it is better to state that the base current controls the collector current.
The ratio of the collector current (I_C) to the base current (I_B) is called the **current gain** or **beta** (β).

$$\beta = \frac{I_C}{I_B}$$