Bipolar Transistor Basics

In the **Diode** tutorials we saw that simple diodes are made up from two pieces of semiconductor material, either silicon or germanium to form a simple PN-junction and we also learnt about their properties and characteristics. If we now join together two individual signal diodes back-to-back, this will give us two PN-junctions connected together in series that share a common P or N terminal. The fusion of these two diodes produces a three layer, two junction, three terminal device forming the basis of a **Bipolar Transistor**, or **BJT** for short.

Transistors are three terminal active devices made from different semiconductor materials that can act as either an insulator or a conductor by the application of a small signal voltage. The transistor's ability to change between these two states enables it to have two basic functions: "switching" (digital electronics) or "amplification" (analogue electronics). Then bipolar transistors have the ability to operate within three different regions:

- 1. Active Region - the transistor operates as an amplifier and $Ic = \beta.Ib$

- 2. Saturation - the transistor is "fully-ON" operating as a switch and $Ic = I(saturation)$

- 3. Cut-off - the transistor is "fully-OFF" operating as a switch and $Ic = 0$

Typical Bipolar Transistor

The word Transistor is an acronym, and is a combination of the words Transfer Varistor used to describe their mode of operation way back in their early days of development. There are two basic types of bipolar transistor construction, NPN and PNP, which basically describes the physical arrangement of the P-type and N-type semiconductor materials from which they are made.

The **Bipolar Transistor** basic construction consists of two PN-junctions producing three connecting terminals with each terminal being given a name to identify it from the other two. These three terminals are known and labelled as the Emitter (E), the Base (B) and the Collector (C) respectively.

Bipolar Transistors are current regulating devices that control the amount of current flowing through them in proportion to the amount of biasing voltage applied to their base terminal acting like a current-controlled switch. The principle of operation of the two transistor types NPN and PNP, is exactly the same the only difference being in their biasing and the polarity of the power supply for each type.

Bipolar Transistor Construction

a). Construction

b). Two-diode Analogy
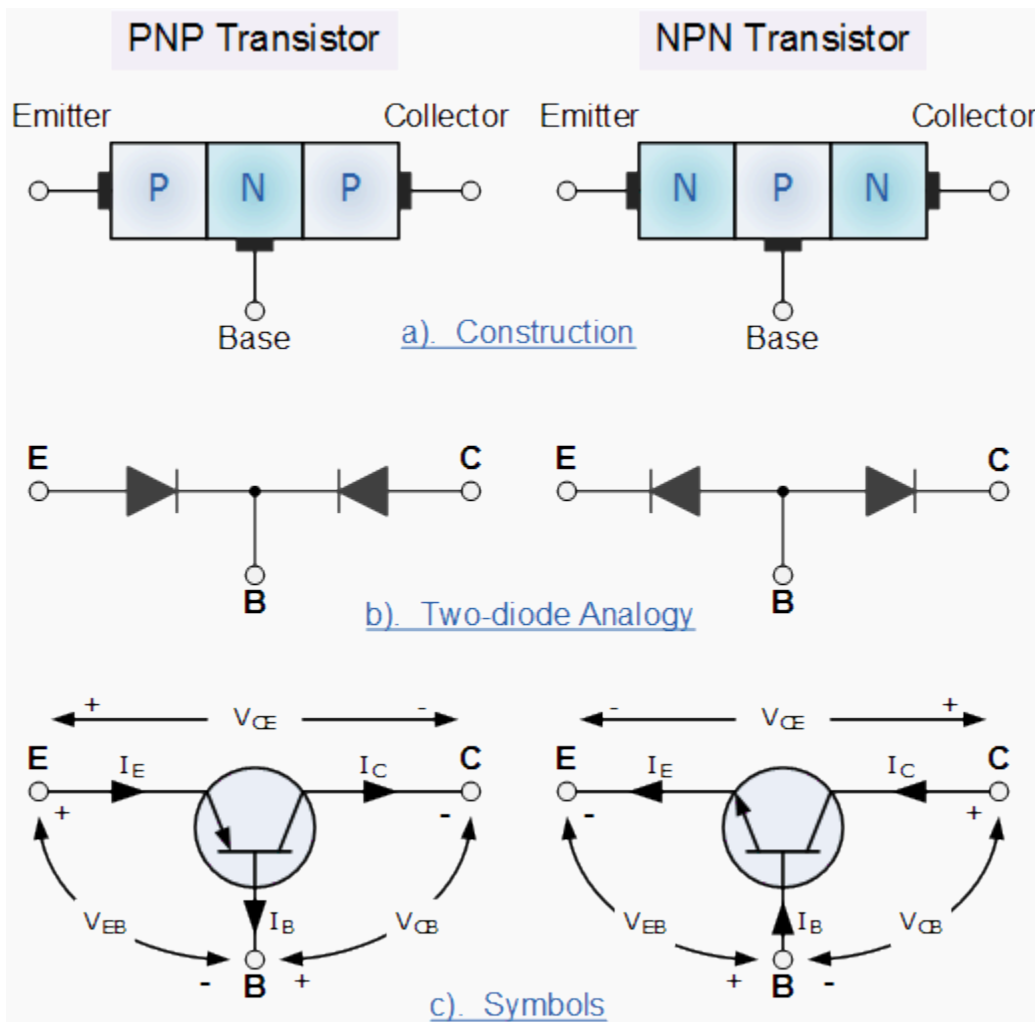
c). Symbols

The construction and circuit symbols for both the NPN and PNP bipolar transistor are given above with the arrow in the circuit symbol always showing the direction of "conventional current flow" between the base terminal and its emitter terminal. The direction of the arrow always points from the positive P-type region to the negative N-type region for both transistor types, exactly the same as for the standard diode symbol.
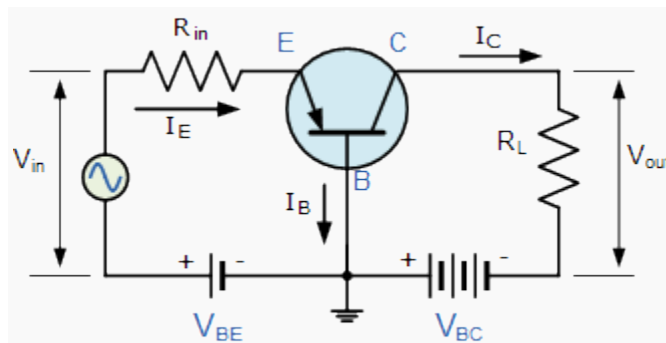
Bipolar Transistor Configurations

As the **Bipolar Transistor** is a three terminal device, there are basically three possible ways to connect it within an electronic circuit with one terminal being common to both the input and output. Each method of connection responding differently to its input signal within a circuit as the static characteristics of the transistor vary with each circuit arrangement.

- 1. Common Base Configuration   -   has Voltage Gain but no Current Gain.
-
- 2. Common Emitter Configuration   -   has both Current and Voltage Gain.
-
- 3. Common Collector Configuration   -   has Current Gain but no Voltage Gain.

The Common Base (CB) Configuration

As its name suggests, in the **Common Base** or grounded base configuration, the BASE connection is common to both the input signal AND the output signal with the input signal being applied between the base and the emitter terminals. The corresponding output signal is taken from between the base and the collector terminals as shown with the base terminal grounded or connected to a fixed reference voltage point. The input current flowing into the emitter is quite large as its the sum of both the base current and collector current respectively therefore, the collector current output is less than the emitter current input resulting in a current gain for this type of circuit of "1" (unity) or less, in other words the common base configuration "attenuates" the input signal.

The Common Base Transistor Circuit

This type of amplifier configuration is a non-inverting voltage amplifier circuit, in that the signal voltages Vin and Vout are in-phase. This type of transistor arrangement is not very common due to its unusually high voltage gain characteristics. Its output characteristics represent that of a forward biased diode while the input characteristics represent that of an illuminated photo-diode. Also this type of bipolar transistor configuration has a high ratio of output to input resistance or more importantly "load" resistance (RL) to "input" resistance (Rin) giving it a value of "Resistance Gain". Then the voltage gain (Av for a common base configuration is therefore given as:

Common Base Voltage Gain

$$A_V = \frac{Vout}{Vin} = \frac{I_C \times R_L}{I_E \times R_{IN}}$$

The common base circuit is generally only used in single stage amplifier circuits such as microphone pre-amplifier or radio frequency (Rf) amplifiers due to its very good high frequency response.

The Common Emitter (CE) Configuration

In the **Common Emitter** or grounded emitter configuration, the input signal is applied between the base, while the output is taken from between the collector and the emitter as shown. This type of configuration is the most commonly used circuit for transistor based amplifiers and which represents the "normal" method of bipolar transistor connection. The common emitter amplifier configuration produces the highest current and power gain of all the three bipolar transistor configurations. This is mainly because the input impedance is LOW as it is connected to a forward-biased PN-junction, while the output impedance is HIGH as it is taken from a reverse-biased PN-junction.

The Common Emitter Amplifier Circuit
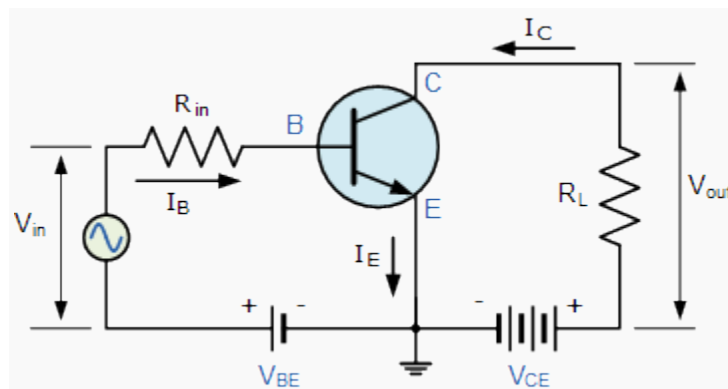
In this type of configuration, the current flowing out of the transistor must be equal to the currents flowing into the transistor as the emitter current is given as $Ie = Ic + Ib$. Also, as the load resistance (R_L) is connected in series with the collector, the current gain of the common emitter transistor configuration is quite large as it is the ratio of Ic/Ib and is given the Greek symbol of Beta, (β). As the emitter current for a common emitter configuration is defined as $Ie = Ic + Ib$, the ratio of Ic/Ie is called Alpha, given the Greek symbol of α. Note: that the value of Alpha will always be less than unity.

Since the electrical relationship between these three currents, Ib, Ic and Ie is determined by the physical construction of the transistor itself, any small change in the base current (Ib), will result in a much larger change in the collector

current (I_C). Then, small changes in current flowing in the base will thus control the current in the emitter-collector circuit. Typically, Beta has a value between 20 and 200 for most general purpose transistors.

By combining the expressions for both Alpha, α and Beta, β the mathematical relationship between these parameters and therefore the current gain of the transistor can be given as:

$$Alpha, (\alpha) = \frac{I_C}{I_E} \quad and \quad Beta, (\beta) = \frac{I_C}{I_B}$$

$$\therefore I_C = \alpha.I_E = \beta.I_B$$

$$as: \quad \alpha = \frac{\beta}{\beta + 1} \quad \beta = \frac{\alpha}{1 - \alpha}$$

$$I_E = I_C + I_B$$

Where: "I_C" is the current flowing into the collector terminal, "I_b" is the current flowing into the base terminal and "I_e" is the current flowing out of the emitter terminal.

Then to summarise, this type of bipolar transistor configuration has a greater input impedance, current and power gain than that of the common base configuration but its voltage gain is much lower. The common emitter configuration is an inverting amplifier circuit resulting in the output signal being 180° out-of-phase with the input voltage signal.

The Common Collector (CC) Configuration

In the **Common Collector** or grounded collector configuration, the collector is now common through the supply. The input signal is connected directly to the base, while the output is taken from the emitter load as shown. This type of configuration is commonly known as a **Voltage Follower** or **Emitter Follower** circuit. The emitter follower configuration is very useful for impedance matching applications because of the very high input impedance, in the region of hundreds of thousands of Ohms while having a relatively low output impedance.

The Common Collector Transistor Circuit
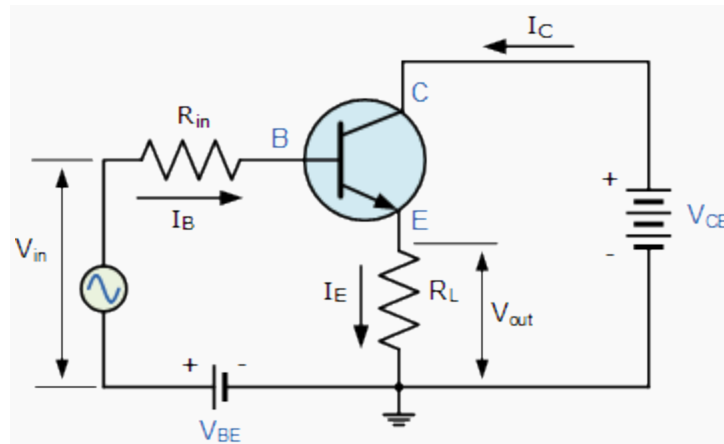
The common emitter configuration has a current gain approximately equal to the β value of the transistor itself. In the common collector configuration the load resistance is situated in series with the emitter so its current is equal to that of the emitter current. As the emitter current is the combination of the collector AND the base current combined, the load resistance in this type of transistor configuration also has both the collector current and the input current of the base flowing through it. Then the current gain of the circuit is given as:

The Common Collector Current Gain

$$I_E = I_C + I_B$$

$$A_i = \frac{I_E}{I_B} = \frac{I_C + I_B}{I_B}$$

$$A_i = \frac{I_C}{I_B} + 1$$

$$A_i = \beta + 1$$

This type of bipolar transistor configuration is a non-inverting circuit in that the signal voltages of Vin and Vout are in-phase. It has a voltage gain that is always less than "1" (unity). The load resistance of the common collector transistor receives both the base and collector currents giving a large current gain (as with the common emitter configuration) therefore, providing good current amplification with very little voltage gain.

Bipolar Transistor Summary

Then to summarise, the behaviour of the bipolar transistor in each one of the above circuit configurations is very different and produces different circuit characteristics with regards to input impedance, output impedance and gain whether this is voltage gain, current gain or power gain and this is summarised in the table below.

Bipolar Transistor Characteristics

The static characteristics for a **Bipolar Transistor** can be divided into the following three main groups.

Input Characteristics:-	Common Base -	$\Delta V_{EB} / \Delta I_E$
	Common Emitter -	$\Delta V_{BE} / \Delta I_B$
Output Characteristics:-	Common Base -	$\Delta V_C / \Delta I_C$
	Common Emitter -	$\Delta V_C / \Delta I_C$
Transfer Characteristics:-	Common Base -	$\Delta I_C / \Delta I_E$
	Common Emitter -	$\Delta I_C / \Delta I_B$

with the characteristics of the different transistor configurations given in the following table:

Characteristic	Common Base	Common Emitter	Common Collector
Input Impedance	Low	Medium	High
Output Impedance	Very High	High	Low
Phase Angle	0°	180°	0°
Voltage Gain	High	Medium	Low
Current Gain	Low	Medium	High
Power Gain	Low	Very High	Medium

In the next tutorial about **Bipolar Transistors**, we will look at the NPN Transistor in more detail when used in the common emitter configuration as an amplifier as this is the most widely used configuration due to its flexibility and high gain. We will also plot the output characteristics curves commonly associated with amplifier circuits as a function of the collector current to the base current.

The NPN Transistor

In the previous tutorial we saw that the standard **Bipolar Transistor** or BJT, comes in two basic forms. An NPN (Negative-Positive-Negative) type and a PNP (Positive-Negative-Positive) type, with the most commonly used transistor type being the **NPN Transistor**. We also learnt that the transistor junctions can be biased in one of three different ways - **Common Base**, **Common Emitter** and **Common Collector**. In this tutorial we will look more closely at the "Common Emitter" configuration using **NPN Transistors** with an example of the construction of a NPN transistor along with the transistors current flow characteristics is given below.

An NPN Transistor Configuration

Note: Conventional current flow.
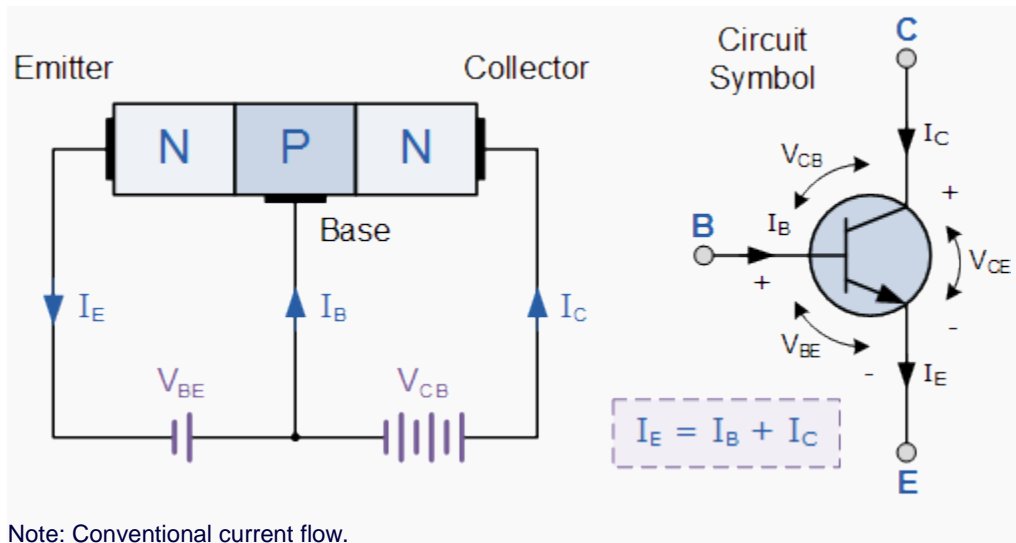
We know that the transistor is a **"current"** operated device (Beta model) and that a large current (Ic) flows freely through the device between the collector and the emitter terminals when the transistor is switched "fully-ON". However, this only happens when a small biasing current (Ib) is flowing into the base terminal of the transistor at the same time thus allowing the Base to act as a sort of current control input. The transistor current in an NPN transistor is the ratio of these two currents (Ic/Ib), called the *DC Current Gain* of the device and is given the symbol of hfe or nowadays Beta, (β). The value of β can be large up to 200 for standard transistors, and it is this large ratio between Ic and Ib that makes the NPN transistor a useful amplifying device when used in its active region as Ib provides the input and Ic provides the output. Note that Beta has no units as it is a ratio.

Also, the current gain of the transistor from the Collector terminal to the Emitter terminal, Ic/Ie, is called Alpha, (α), and is a function of the transistor itself (electrons diffusing across the junction). As the emitter current Ie is the product of a very small base current plus a very large collector current, the value of alpha α, is very close to unity, and for a typical low-power signal transistor this value ranges from about 0.950 to 0.999

α and β Relationship in a NPN Transistor

$$\text{DC Current Gain} = \frac{\text{Output Current}}{\text{Input Current}} = \frac{I_C}{I_B}$$

$$I_E = I_B + I_C \ \dots\ (\text{KCL}) \quad \text{and} \quad \frac{I_C}{I_E} = \alpha$$

$$\text{Thus}: \ I_E = I_B + I_C = \frac{I_C}{\alpha}$$

$$\text{and} \quad I_B = I_C\left(1 - \frac{1}{\alpha}\right)$$

$$\therefore \ \beta = \frac{I_C}{I_B} = \frac{1}{\left(1 - \dfrac{1}{\alpha}\right)} = \frac{\alpha}{1-\alpha}$$

By combining the two parameters α and β we can produce two mathematical expressions that gives the relationship between the different currents flowing in the transistor.

$$\beta = \frac{\alpha}{1-\alpha} \quad \text{and} \quad \alpha = \frac{\beta}{\beta+1}$$

$$\text{If}\ \ \alpha = 0.99 \quad \beta = \frac{0.99}{0.01} = 99$$

The values of Beta vary from about 20 for high current power transistors to well over 1000 for high frequency low power type bipolar transistors. The value of Beta for most standard NPN transistors can be found in the manufactures datasheets but generally range between 50 - 200.

The equation above for Beta can also be re-arranged to make Ic as the subject, and with a zero base current (Ib = 0) the resultant collector current Ic will also be zero, (β x 0). Also when the base current is high the corresponding

collector current will also be high resulting in the base current controlling the collector current. One of the most important properties of the **Bipolar Junction Transistor** is that a small base current can control a much larger collector current. Consider the following example.

Example No1

An NPN Transistor has a DC current gain, (Beta) value of 200. Calculate the base current Ib required to switch a resistive load of 4mA.

$$I_B = \frac{I_C}{\beta} = \frac{4 \times 10^{-3}}{200} = 20\mu A$$

Therefore, β = 200, Ic = 4mA and Ib = 20μA.

One other point to remember about **NPN Transistors**. The collector voltage, (Vc) must be greater and positive with respect to the emitter voltage, (Ve) to allow current to flow through the transistor between the collector-emitter junctions. Also, there is a voltage drop between the Base and the Emitter terminal of about 0.7v (one diode volt drop) for silicon devices as the input characteristics of an NPN Transistor are of a forward biased diode. Then the base voltage, (Vbe) of a NPN transistor must be greater than this 0.7V otherwise the transistor will not conduct with the base current given as.

$$I_B = \frac{V_B - V_{BE}}{R_B}$$

Where: Ib is the base current, Vb is the base bias voltage, Vbe is the base-emitter volt drop (0.7v) and Rb is the base input resistor. Increasing Ib, Vbe slowly increases to 0.7V but Ic rises exponentially.

Example No2

An NPN Transistor has a DC base bias voltage, Vb of 10v and an input base resistor, Rb of 100kΩ. What will be the value of the base current into the transistor.

$$I_B = \frac{V_B - V_{BE}}{R_B} = \frac{10 - 0.7}{100k\Omega} = 93\mu A$$

Therefore, Ib = 93μA.

The Common Emitter Configuration.

As well as being used as a semiconductor switch to turn load currents "ON" or "OFF" by controlling the Base signal to the transistor in ether its saturation or cut-off regions, **NPN Transistors** can also be used in its active region to produce a circuit which will amplify any small AC signal applied to its Base terminal with the Emitter grounded. If a suitable DC "biasing" voltage is firstly applied to the transistors Base terminal thus allowing it to always operate within its linear active region, an inverting amplifier circuit called a single stage common emitter amplifier is produced.

One such *Common Emitter Amplifier* configuration of an NPN transistor is called a **Class A Amplifier**. A "Class A Amplifier" operation is one where the transistors Base terminal is biased in such a way as to forward bias the Base-emitter junction. The result is that the transistor is always operating halfway between its cut-off and saturation regions, thereby allowing the transistor amplifier to accurately reproduce the positive and negative halves of any AC input signal superimposed upon this DC biasing voltage. Without this "Bias Voltage" only one half of the input waveform would be amplified. This common emitter amplifier configuration using an NPN transistor has many applications but is commonly used in audio circuits such as pre-amplifier and power amplifier stages.

With reference to the common emitter configuration shown below, a family of curves known as the **Output Characteristics Curves**, relates the output collector current, (I_C) to the collector voltage, (V_{ce}) when different values of Base current, (I_b) are applied to the transistor for transistors with the same β value. A DC "Load Line" can also be drawn onto the output characteristics curves to show all the possible operating points when different values of base current are applied. It is necessary to set the initial value of V_{ce} correctly to allow the output voltage to vary both up and down when amplifying AC input signals and this is called setting the operating point or Quiescent Point, **Q-point** for short and this is shown below.

Single Stage Common Emitter Amplifier Circuit
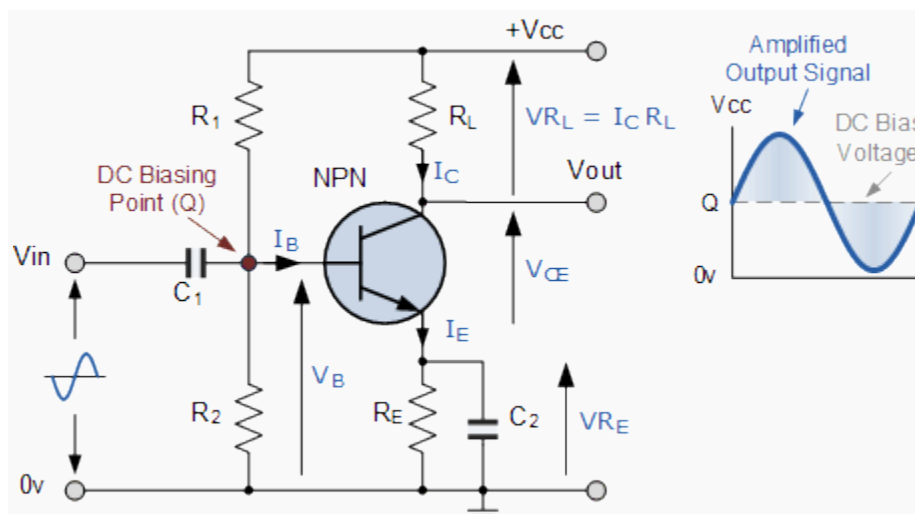
Output Characteristics Curves for a Typical Bipolar Transistor
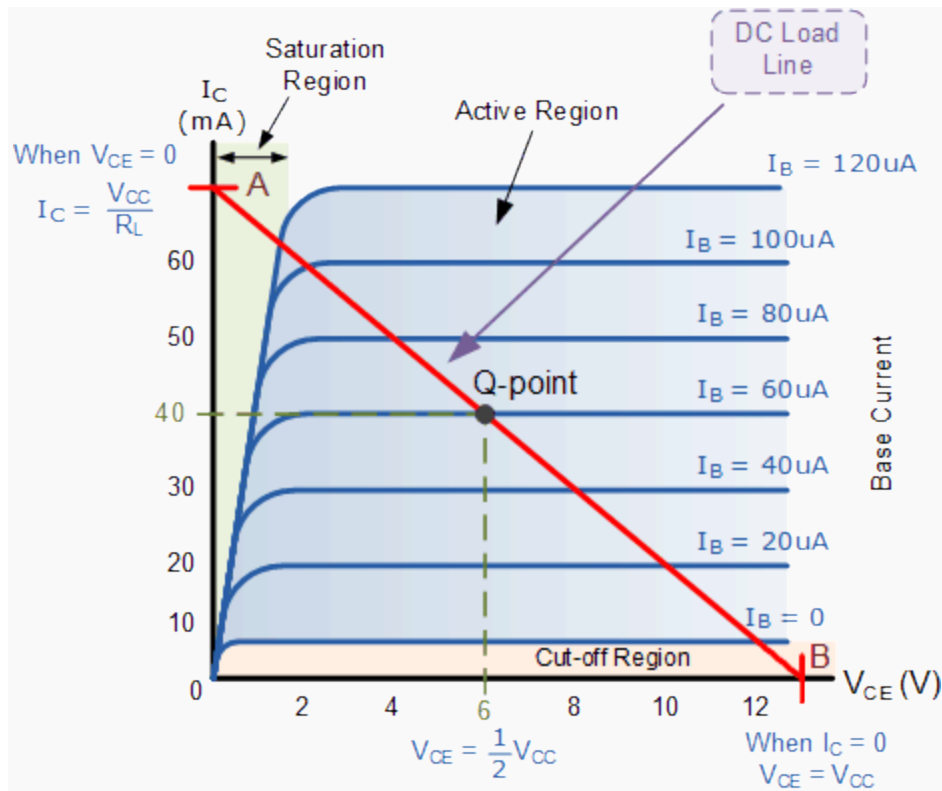
The most important factor to notice is the effect of Vce upon the collector current Ic when Vce is greater than about 1.0 volts. We can see that Ic is largely unaffected by changes in Vce above this value and instead it is almost entirely controlled by the base current, Ib. When this happens we can say then that the output circuit represents that of a "Constant Current Source". It can also be seen from the common emitter circuit above that the emitter current Ie is the sum of the collector current, Ic and the base current, Ib, added together so we can also say that " Ie = Ic + Ib " for the common emitter configuration.

By using the output characteristics curves in our example above and also Ohm´s Law, the current flowing through the load resistor, (RL), is equal to the collector current, Ic entering the transistor which inturn corresponds to the supply voltage, (Vcc) minus the voltage drop between the collector and the emitter terminals, (Vce) and is given as:

$$\text{Collector Current, } I_C = \frac{V_{CC} - V_{CE}}{R_L}$$

Also, a straight line representing the **Load Line** of the transistor can be drawn directly onto the graph of curves above from the point of "Saturation" (A) when Vce = 0 to the point of "Cut-off" (B) when Ic = 0 thus giving us the "Operating" or **Q-point** of the transistor. These two points are joined together by a straight line and any position along

this straight line represents the "Active Region" of the transistor. The actual position of the load line on the characteristics curves can be calculated as follows:

$$\text{When}: \left(V_{CE} = 0 \right) \quad I_C = \frac{V_{CC} - 0}{R_L}, \quad I_C = \frac{V_{CC}}{R_L}$$

$$\text{When}: \left(I_C = 0 \right) \quad 0 = \frac{V_{CC} - V_{CE}}{R_L}, \quad V_{CC} = V_{CE}$$

Then, the collector or output characteristics curves for **Common Emitter NPN Transistors** can be used to predict the Collector current, I_C, when given V_{ce} and the Base current, I_b. A Load Line can also be constructed onto the curves to determine a suitable Operating or **Q-point** which can be set by adjustment of the base current. The slope of this load line is equal to the reciprocal of the load resistance which is given as: $-1/R_L$

In the next tutorial about **Bipolar Transistors**, we will look at the opposite or compliment form of the NPN Transistor called the **PNP Transistor** and show that the PNP Transistor has very similar characteristics to their NPN transistor except that the polarities (or biasing) of the current and voltage directions are reversed.

The PNP Transistor

The **PNP Transistor** is the exact opposite to the **NPN Transistor** device we looked at in the previous tutorial. Basically, in this type of transistor construction the two diodes are reversed with respect to the NPN type, with the arrow, which also defines the Emitter terminal this time pointing inwards in the transistor symbol. Also, all the polarities are reversed which means that *PNP Transistors* "sink" current as opposed to the NPN transistor which "sources" current. Then, PNP Transistors use a small output base current and a negative base voltage to control a much larger emitter-collector current. The construction of a PNP transistor consists of two P-type semiconductor materials either side of the N-type material as shown below.

A PNP Transistor Configuration

Note: Conventional current flow.
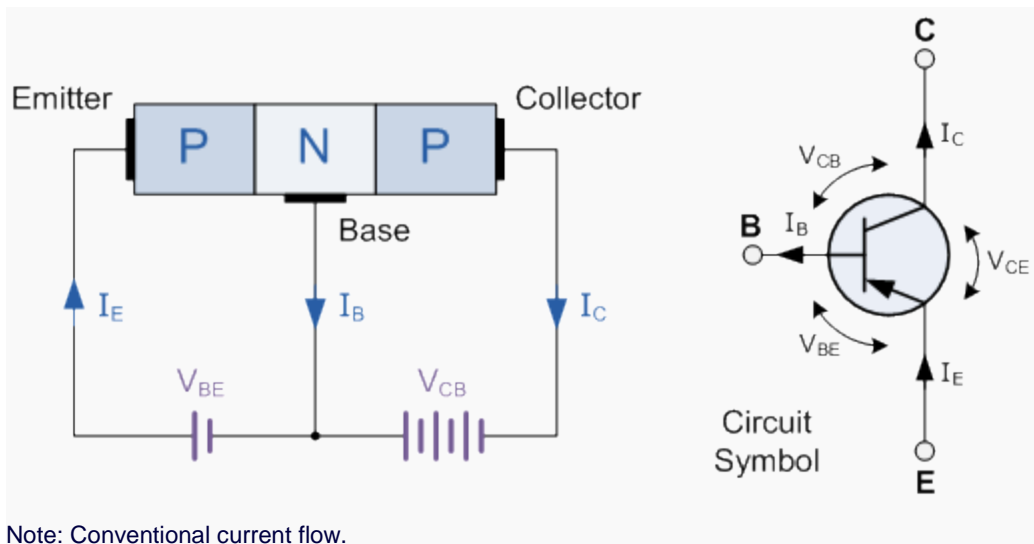
The **PNP Transistor** has very similar characteristics to their NPN bipolar cousins, except that the polarities (or biasing) of the current and voltage directions are reversed for any one of the possible three configurations looked at in the first tutorial, Common Base, Common Emitter and Common Collector. Generally, PNP Transistors require a negative (-ve) voltage at their Collector terminal with the flow of current through the emitter-collector terminals being **Holes** as opposed to **Electrons** for the NPN types. Because the movement of holes across the depletion layer tends to be slower than for electrons, PNP transistors are generally more slower than their equivalent NPN counterparts when operating.

To cause the Base current to flow in a PNP transistor the Base needs to be more negative than the Emitter (current must leave the base) by approx 0.7 volts for a silicon device or 0.3 volts for a germanium device with the formulas used to calculate the Base resistor, Base current or Collector current are the same as those used for an equivalent NPN transistor and is given as.

$$I_E = I_C + I_B$$

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

Generally, the PNP transistor can replace NPN transistors in electronic circuits, the only difference is the polarities of the voltages, and the directions of the current flow. PNP Transistors can also be used as switching devices and an example of a PNP transistor switch is shown below.

A PNP Transistor Circuit
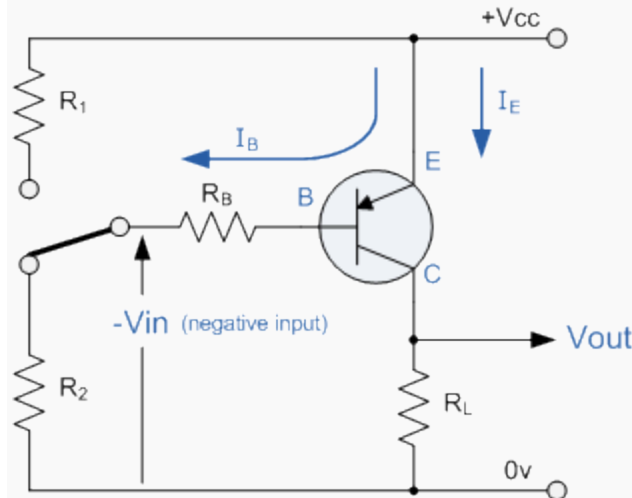
The **Output Characteristics Curves** for a PNP transistor look very similar to those for an equivalent NPN transistor except that they are rotated by 180° to take account of the reverse polarity voltages and currents, (the currents flowing out of the Base and Collector in a PNP transistor are negative).

Transistor Matching

You may think what is the point of having a **PNP Transistor**, when there are plenty of NPN Transistors available?. Well, having two different types of transistors PNP & NPN, can be an advantage when designing amplifier circuits such as **Class B Amplifiers** that use "Complementary" or "Matched Pair" transistors or for reversible **H-Bridge** motor control circuits. A pair of corresponding NPN and PNP transistors with near identical characteristics to each other are called **Complementary Transistors** for example, a TIP3055 (NPN), TIP2955 (PNP) are good examples of complementary or matched pair silicon power transistors. They have a DC current gain, Beta, (Ic / Ib) matched to within 10% and high Collector current of about 15A making them suitable for general motor control or robotic applications.

Identifying the PNP Transistor

We saw in the first tutorial of this Transistors section, that transistors are basically made up of two **Diodes** connected together back-to-back. We can use this analogy to determine whether a transistor is of the type PNP or NPN by testing its Resistance between the three different leads, Emitter, Base and Collector. By testing each pair of transistor leads in both directions will result in six tests in total with the expected resistance values in Ohm's given below.

- 1. Emitter-Base Terminals - The Emitter to Base should act like a normal diode and conduct one way only.
-
- 2. Collector-Base Terminals - The Collector-Base junction should act like a normal diode and conduct one way only.

-
- **3. Emitter-Collector Terminals** - The Emitter-Collector should not conduct in either direction.

Transistor Resistance Values for the PNP transistor and NPN transistor types

Between Transistor Terminals		PNP	NPN
Collector	Emitter	R_{HIGH}	R_{HIGH}
Collector	Base	R_{LOW}	R_{HIGH}
Emitter	Collector	R_{HIGH}	R_{HIGH}
Emitter	Base	R_{LOW}	R_{HIGH}
Base	Collector	R_{HIGH}	R_{LOW}
Base	Emitter	R_{HIGH}	R_{LOW}

The Transistor as a Switch

When used as an AC signal amplifier, the transistors Base biasing voltage is applied so that it operates within its "**Active**" region and the linear part of the output characteristics curves are used. However, both the NPN & PNP type bipolar transistors can be made to operate as an "ON/OFF" type solid state switch for controlling high power devices such as motors, solenoids or lamps. If the circuit uses the **Transistor as a Switch**, then the biasing is arranged to operate in the output characteristics curves seen previously in the areas known as the "**Saturation**" and "**Cut-off**" regions as shown below.

Transistor Curves
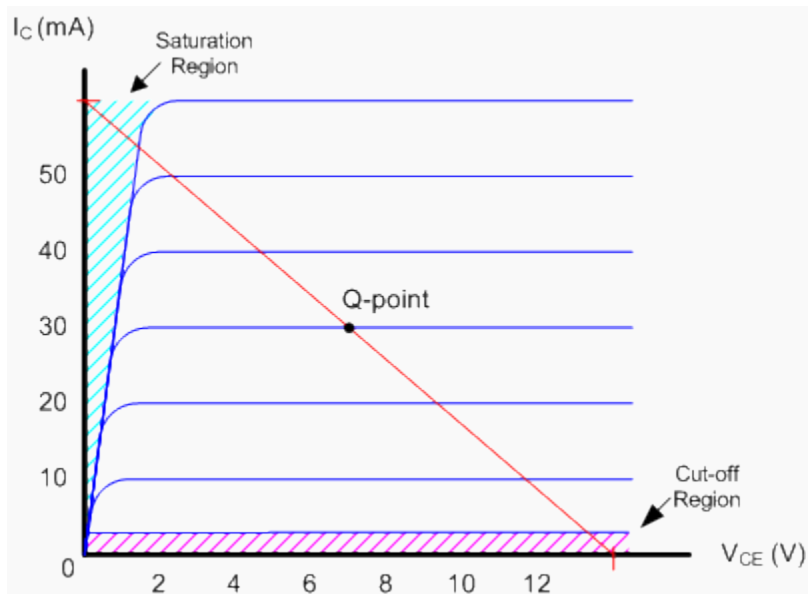
The pink shaded area at the bottom represents the "Cut-off" region. Here the operating conditions of the transistor are zero input base current (Ib), zero output collector current (Ic) and maximum collector voltage (Vce) which results in a large depletion layer and no current flows through the device. The transistor is switched "Fully-OFF". The lighter blue area to the left represents the "Saturation" region. Here the transistor will be biased so that the maximum amount of base current is applied, resulting in maximum collector current flow and minimum collector emitter voltage which results in the depletion layer being as small as possible and maximum current flows through the device. The transistor is switched "Fully-ON". Then we can summarize this as:

- 1. Cut-off Region - Both junctions are Reverse-biased, Base current is zero or very small resulting in zero Collector current flowing, the device is switched fully "OFF".
-
- 2. Saturation Region - Both junctions are Forward-biased, Base current is high enough to give a Collector-Emitter voltage of 0v resulting in maximum Collector current flowing, the device is switched fully "ON".

An example of an NPN Transistor as a switch being used to operate a relay is given below. With inductive loads such as relays or solenoids a flywheel diode is placed across the load to dissipate the back EMF generated by the inductive load when the transistor switches "OFF" and so protect the transistor from damage. If the load is of a very high current or voltage nature, such as motors, heaters etc, then the load current can be controlled via a suitable relay as shown.

Transistor Switching Circuit
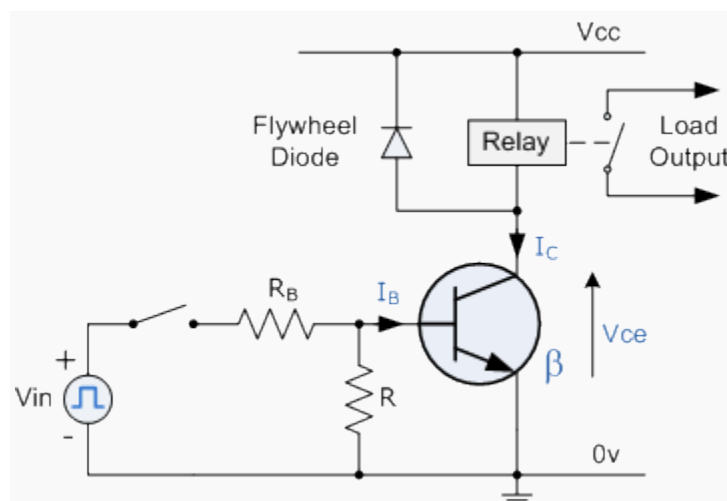
The circuit resembles that of the **Common Emitter** circuit we looked at in the previous tutorials. The difference this time is that to operate the transistor as a switch the transistor needs to be turned either fully "OFF" (Cut-off) or fully "ON" (Saturated). An ideal transistor switch would have an infinite resistance when turned "OFF" resulting in zero current flow and zero resistance when turned "ON", resulting in maximum current flow. In practice when turned "OFF", small leakage currents flow through the transistor and when fully "ON" the device has a low resistance value causing

a small saturation voltage (Vce) across it. In both the Cut-off and Saturation regions the power dissipated by the transistor is at its minimum.

To make the Base current flow, the Base input terminal must be made more positive than the Emitter by increasing it above the 0.7 volts needed for a silicon device. By varying the Base-Emitter voltage Vbe, the Base current is altered and which in turn controls the amount of Collector current flowing through the transistor as previously discussed. When maximum Collector current flows the transistor is said to be **Saturated**. The value of the Base resistor determines how much input voltage is required and corresponding Base current to switch the transistor fully "ON".

Example No1.

For example, using the transistor values from the previous tutorials of: $\beta = 200$, Ic = 4mA and Ib = 20uA, find the value of the Base resistor (Rb) required to switch the load "ON" when the input terminal voltage exceeds 2.5v.

$$R_B = \frac{V_{in} - V_{BE}}{I_B} = \frac{2.5v - 0.7v}{20x10^{-6}} = 90k\Omega$$

Example No2.

Again using the same values, find the minimum Base current required to turn the transistor fully "ON" (Saturated) for a load that requires 200mA of current.

$$I_B = \frac{I_C}{\beta} = \frac{200mA}{200} = 1mA$$

Transistor switches are used for a wide variety of applications such as interfacing large current or high voltage devices like motors, relays or lamps to low voltage digital logic IC's or gates like AND Gates or OR Gates. Here, the output from a digital logic gate is only +5v but the device to be controlled may require a 12 or even 24 volts supply. Or the load such as a DC Motor may need to have its speed controlled using a series of pulses (Pulse Width Modulation) and transistor switches will allow us to do this faster and more easily than with conventional mechanical switches.

Digital Logic Transistor Switch
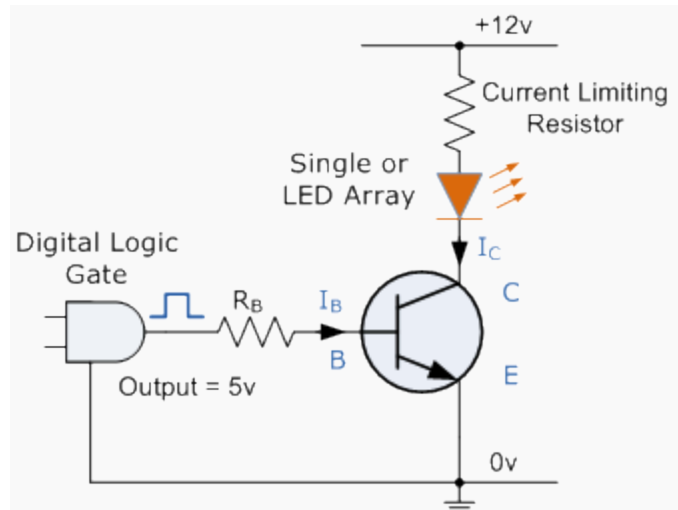
The base resistor, Rb is required to limit the output current of the logic gate.
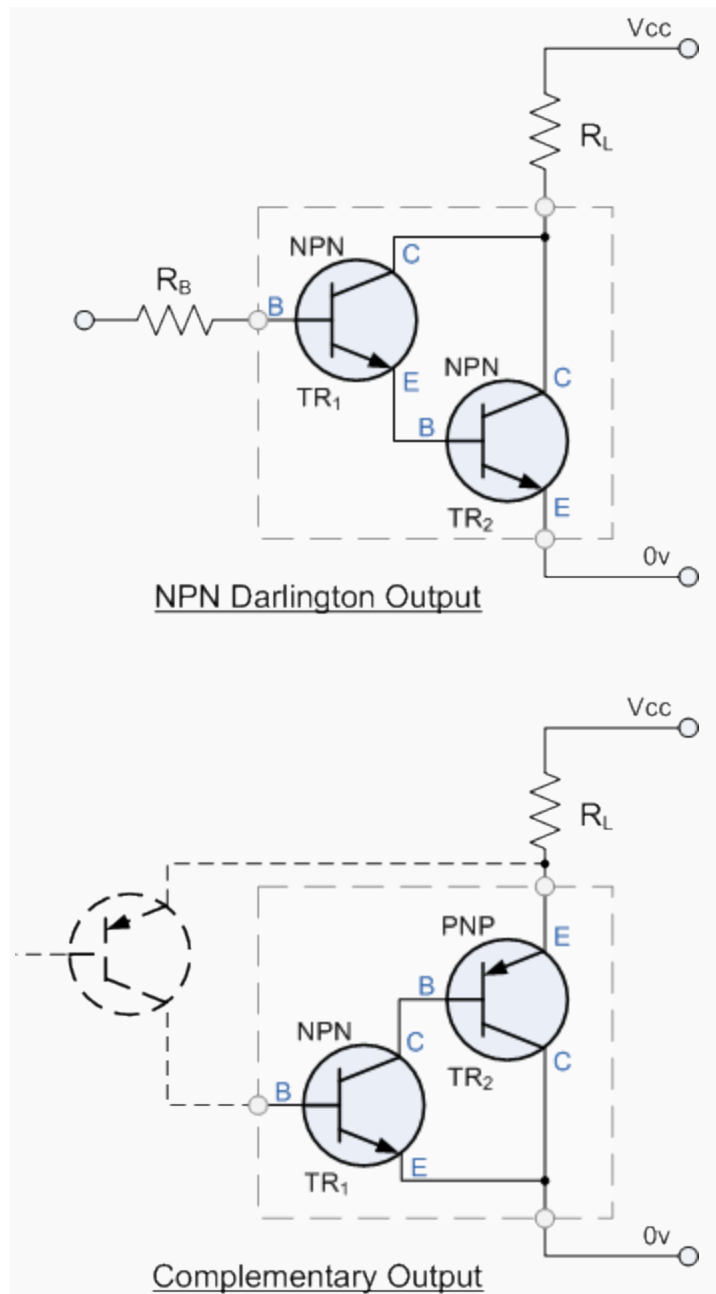
Darlington Transistors

Sometimes the DC current gain of the bipolar transistor is too low to directly switch the load current or voltage, so multiple switching transistors are used. Here, one small input transistor is used to switch "ON" or "OFF" a much larger current handling output transistor. To maximise the signal gain the two transistors are connected in a "Complementary Gain Compounding Configuration" or what is generally called a "**Darlington Configuration**" where the amplification factor is the product of the two individual transistors.

Darlington Transistors simply contain two individual bipolar NPN or PNP type transistors connected together so that the current gain of the first transistor is multiplied with that of the current gain of the second transistor to produce a device which acts like a single transistor with a very high current gain. The overall current gain Beta (β) or Hfe value of a Darlington device is the product of the two individual gains of the transistors and is given as:

$$\beta_{TOTAL} = \beta_1 \times \beta_2$$

So Darlington Transistors with very high β values and high Collector currents are possible compared to a single transistor. An example of the two basic types of Darlington transistor are given below.

Darlington Transistor Configurations

NPN Darlington Output

Complementary Output

The above NPN Darlington transistor configuration shows the Collectors of the two transistors connected together with the Emitter of the first transistor connected to the Base of the second transistor therefore, the Emitter current of the first transistor becomes the Base current of the second transistor. The first or "input" transistor receives an input signal, amplifies it and uses it to drive the second or "output" transistors which amplifies it again resulting in a very high current gain. As well as its high increased current and voltage switching capabilities, another advantage of a Darlington transistor is in its high switching speeds making them ideal for use in Inverter circuits and DC motor or stepper motor control applications.

One difference to consider when using Darlington transistors over the conventional single bipolar transistor type is that the Base-Emitter input voltage Vbe needs to be higher at approx 1.4v for silicon devices, due to the series connection of the two PN junctions.

Then to summarise when using a **Transistor as a Switch**.

- Transistor switches can be used to switch and control lamps, relays or even motors.

- When using bipolar transistors as switches they must be fully "OFF" or fully "ON".

- Transistors that are fully "ON" are said to be in their **Saturation** region.

- Transistors that are fully "OFF" are said to be in their **Cut-off** region.

- In a transistor switch a small Base current controls a much larger Collector current.

- When using transistors to switch inductive relay loads a "Flywheel Diode" is required.

- When large currents or voltages need to be controlled, **Darlington Transistors** are used.

The Field Effect Transistor

In the **Bipolar Junction Transistor** tutorials, we saw that the output Collector current is determined by the amount of current flowing into the Base terminal of the device and thereby making the Bipolar Transistor a **CURRENT** operated device. The **Field Effect Transistor**, or simply **FET** however, use the voltage that is applied to their input terminal to control the output current, since their operation relies on the electric field (hence the name field effect) generated by the input voltage. This then makes the **Field Effect Transistor** a **VOLTAGE** operated device.

The **Field Effect Transistor** is a unipolar device that has very similar properties to those of the *Bipolar Transistor* ie, high efficiency, instant operation, robust and cheap, and they can be used in most circuit applications that use the equivalent Bipolar Junction Transistors, (BJT). They can be made much smaller than an equivalent BJT transistor and along with their low power consumption and dissipation make them ideal for use in integrated circuits such as the CMOS range of chips.

We remember from the previous tutorials that there are two basic types of Bipolar Transistor construction, NPN and PNP, which basically describes the physical arrangement of the P-type and N-type semiconductor materials from which they are made. There are also two basic types of Field Effect Transistor, N-channel and P-channel. As their name implies, Bipolar Transistors are "Bipolar" devices because they operate with both types of charge carriers, Holes and Electrons. The Field Effect Transistor on the other hand is a "Unipolar" device that depends only on the conduction of Electrons (N-channel) or Holes (P-channel).

The **Field Effect Transistor** has one major advantage over its standard bipolar transistor cousins, in that their input impedance is very high, (Thousands of Ohms) making them very sensitive to input signals, but this high sensitivity also means that they can be easily damaged by static electricity. There are two main types of field effect transistor, the **Junction Field Effect Transistor** or **JFET** and the **Insulated-gate Field Effect Transistor** or **IGFET)**, which is more commonly known as the standard **Metal Oxide Semiconductor Field Effect Transistor** or **MOSFET** for short.

The Junction Field Effect Transistor

We saw previously that a bipolar junction transistor is constructed using two PN junctions in the main current path between the Emitter and the Collector terminals. The Field Effect Transistor has no junctions but instead has a narrow "Channel" of N-type or P-type silicon with electrical connections at either end commonly called the DRAIN and the SOURCE respectively. Both P-channel and N-channel FET's are available. Within this channel there is a third connection which is called the GATE and this can also be a P or N-type material forming a PN junction and these connections are compared below.

Bipolar Transistor	Field Effect Transistor
Emitter - (E)	Source - (S)
Base - (B)	Gate - (G)
Collector - (C)	Drain - (D)

The semiconductor "Channel" of the Junction Field Effect Transistor is a resistive path through which a voltage V_{ds} causes a current I_d to flow. A voltage gradient is thus formed down the length of the channel with this voltage becoming less positive as we go from the drain terminal to the source terminal. The PN junction therefore has a high reverse bias at the drain terminal and a lower reverse bias at the source terminal. This bias causes a "depletion layer" to be formed within the channel and whose width increases with the bias. FET's control the current flow through them between the drain and source terminals by controlling the voltage applied to the gate terminal. In an N-channel JFET this gate voltage is negative while for a P-channel JFET the gate voltage is positive.

Bias arrangement for an N-channel JFET and corresponding circuit symbols.
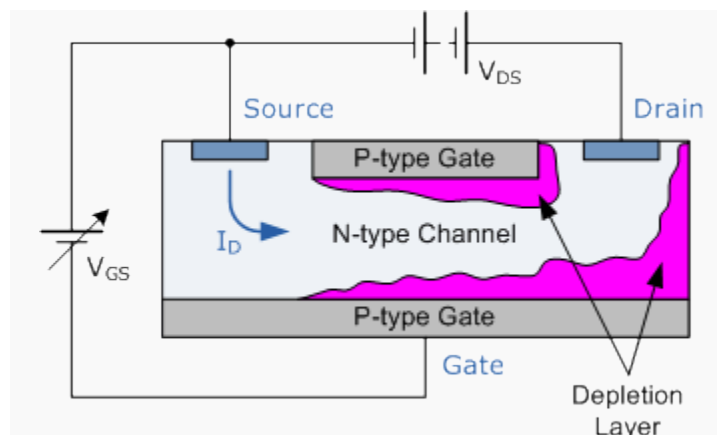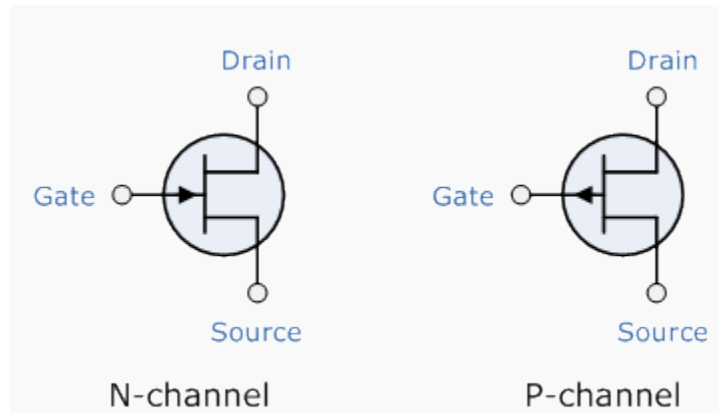
The cross sectional diagram above shows an N-type semiconductor channel with a P-type region called the gate diffused into the N-type channel forming a reverse biased PN junction and its this junction which forms the **depletion layer** around the gate area. This depletion layer restricts the current flow through the channel by reducing its effective width and thus increasing the overall resistance of the channel.

When the gate voltage V_g is equal to 0V and a small external voltage (V_{ds}) is applied between the drain and the source maximum current (I_d) will flow through the channel slightly restricted by the small depletion layer. If a negative voltage (V_{gs}) is now applied to the gate the size of the depletion layer begins to increase reducing the overall effective area of the channel and thus reducing the current flowing through it, a sort of "squeezing" effect. As the gate voltage (V_{gs}) is made more negative, the width of the channel decreases until no more current flows between the drain and the source and the FET is said to be "**pinched-off**". In this pinch-off region the gate voltage, V_{gs} controls the channel current and V_{ds} has little or no effect. The result is that the FET acts more like a voltage controlled resistor which has zero resistance when $V_{gs} = 0$ and maximum "ON" resistance (R_{ds}) when the gate voltage is very negative.

Output characteristic voltage-current curves of a typical junction FET.
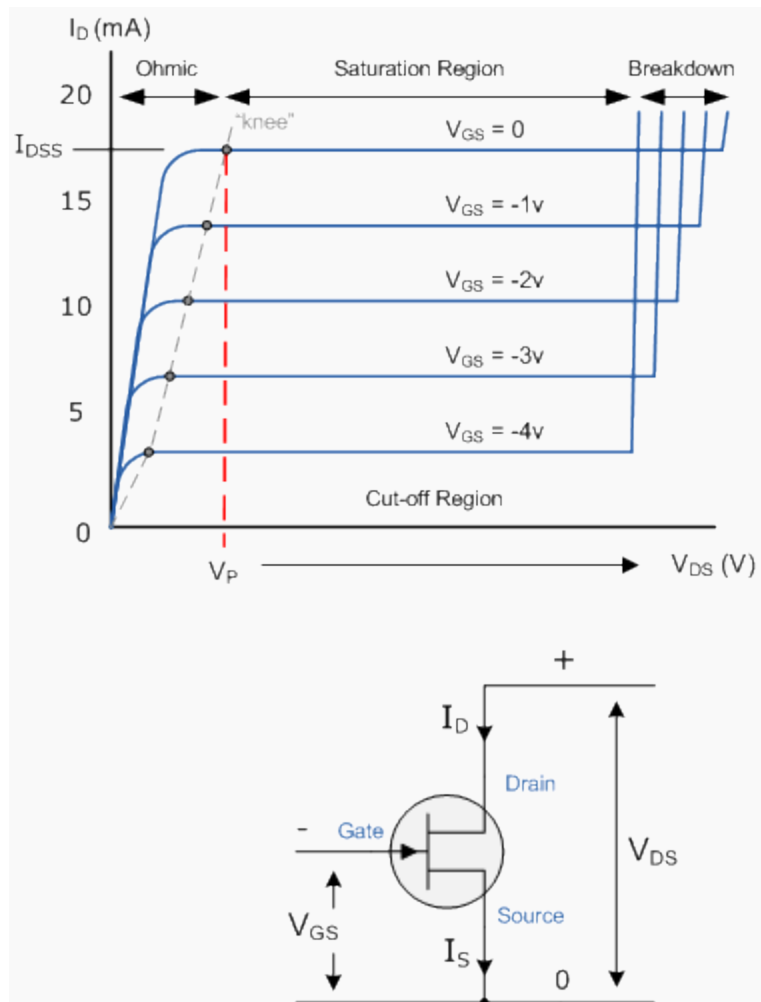
The voltage V_{gs} applied to the gate controls the current flowing between the drain and the source terminals. V_{gs} refers to the voltage applied between the gate and the source while V_{ds} refers to the voltage applied between the drain and the source. Because a **Field Effect Transistor** is a VOLTAGE controlled device, "NO current flows into the gate!" then the source current (I_s) flowing out of the device equals the drain current flowing into it and therefore ($I_d = I_s$).

The characteristics curves example shown above, shows the four different regions of operation for a JFET and these are given as:

- Ohmic Region - The depletion layer of the channel is very small and the JFET acts like a variable resistor.
-
- Cut-off Region - The gate voltage is sufficient to cause the JFET to act as an open circuit as the channel resistance is at maximum.
-
- Saturation or Active Region - The JFET becomes a good conductor and is controlled by the gate-source voltage, (V_{gs}) while the drain-source voltage, (V_{ds}) has little or no effect.

-
- Breakdown Region - The voltage between the drain and source, (V_{ds}) is high enough to causes the JFET's resistive channel to break down and pass current.

The control of the drain current by a negative gate potential makes the **Junction Field Effect Transistor** useful as a switch and it is essential that the gate voltage is never positive for an N-channel JFET as the channel current will flow to the gate and not the drain resulting in damage to the JFET. The principals of operation for a P-channel JFET are the same as for the N-channel JFET, except that the polarity of the voltages need to be reversed.

The MOSFET

As well as the Junction Field Effect Transistor, there is another type of Field Effect Transistor available whose Gate input is electrically insulated from the main current carrying channel and is therefore called an **Insulated Gate Field Effect Transistor**. The most common type of insulated gate FET or IGFET as it is sometimes called, is the **Metal Oxide Semiconductor Field Effect Transistor** or **MOSFET** for short.

The **MOSFET** type of field effect transistor has a "Metal Oxide" gate (usually silicon dioxide commonly known as glass), which is electrically insulated from the main semiconductor N-channel or P-channel. This isolation of the controlling gate makes the input resistance of the **MOSFET** extremely high in the Mega-ohms region and almost infinite. As the gate terminal is isolated from the main current carrying channel ""NO current flows into the gate"" and like the JFET, the **MOSFET** also acts like a voltage controlled resistor. Also like the JFET, this very high input resistance can easily accumulate large static charges resulting in the **MOSFET** becoming easily damaged unless carefully handled or protected.

Basic MOSFET Structure and Symbol
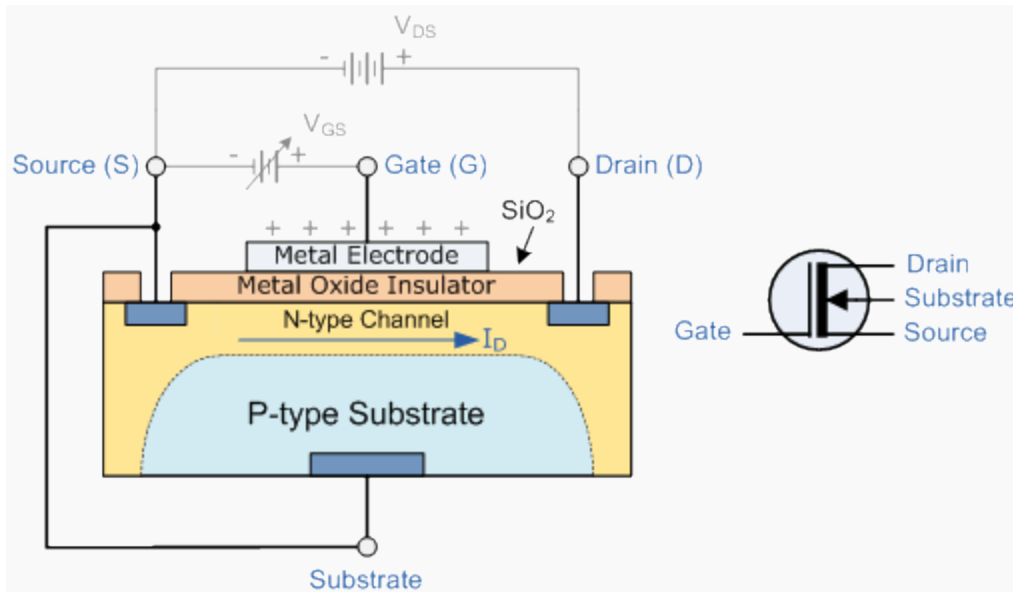
We also saw previously that the gate of a JFET must be biased in such a way as to forward-bias the PN junction but in a MOSFET device no such limitations applies so it is possible to bias the gate in either polarity. This makes MOSFET's specially valuable as electronic switches or to make logic gates because with no bias they are normally non-conducting and the high gate resistance means that very little control current is needed. Both the P-channel and the N-channel MOSFET is available in two basic forms, the **Enhancement** type and the **Depletion** type.

Depletion-mode MOSFET

The **Depletion-mode MOSFET**, which is less common than the enhancement types is normally switched "ON" without a gate bias voltage but requires a gate to source voltage (V_{gs}) to switch the device "OFF". Similar to the JFET types. For N-channel MOSFET's a "Positive" gate voltage widens the channel, increasing the flow of the drain current and decreasing the drain current as the gate voltage goes more negative. The opposite is also true for the P-channel types. The depletion mode MOSFET is equivalent to a "Normally Closed" switch.

Depletion-mode N-Channel MOSFET and circuit Symbols
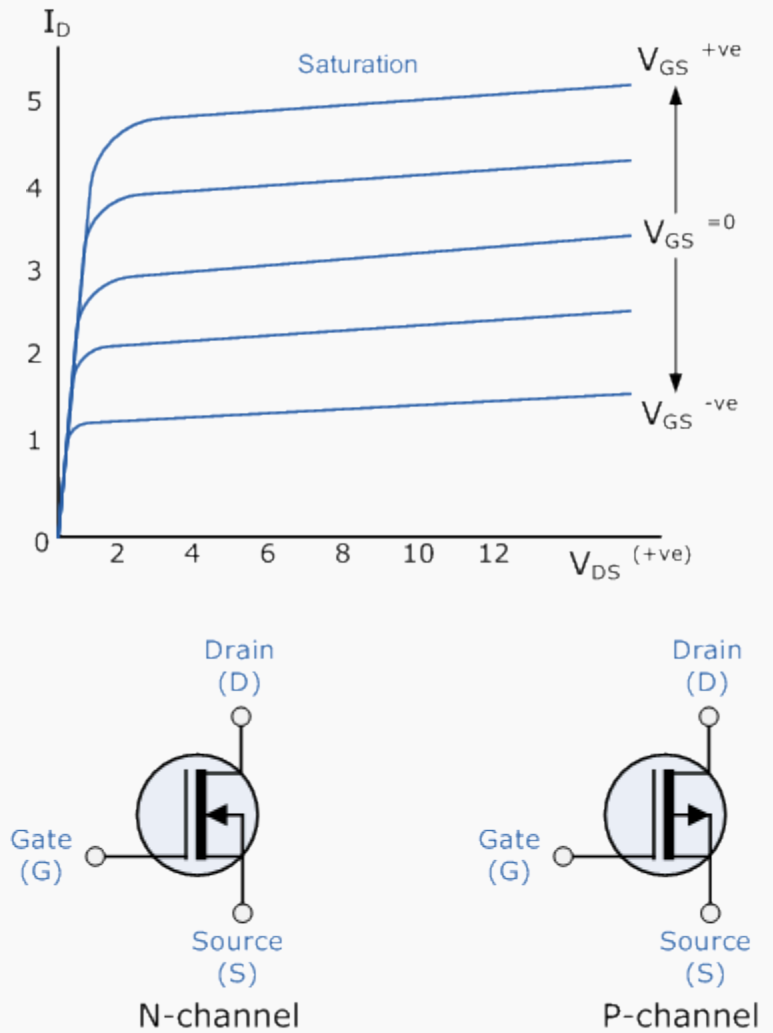
Depletion-mode MOSFET's are constructed similar to their JFET transistor counterparts where the drain-source channel is inherently conductive with electrons and holes already present within the N-type or P-type channel. This doping of the channel produces a conducting path of low resistance between the drain and source with zero gate bias.

Enhancement-mode MOSFET

The more common **Enhancement-mode MOSFET** is the reverse of the depletion-mode type. Here the conducting channel is lightly doped or even undoped making it non-conductive. This results in the device being normally "OFF" when the gate bias voltage is equal to zero.

A drain current will only flow when a gate voltage (V_{gs}) is applied to the gate terminal. This positive voltage creates an electrical field within the channel attracting electrons towards the oxide layer and thereby reducing the overall resistance of the channel allowing current to flow. Increasing this positive gate voltage will cause an increase in the drain current, I_d through the channel. Then, the Enhancement-mode device is equivalent to a "Normally Open" switch.

Enhancement-mode N-Channel MOSFET and circuit Symbols
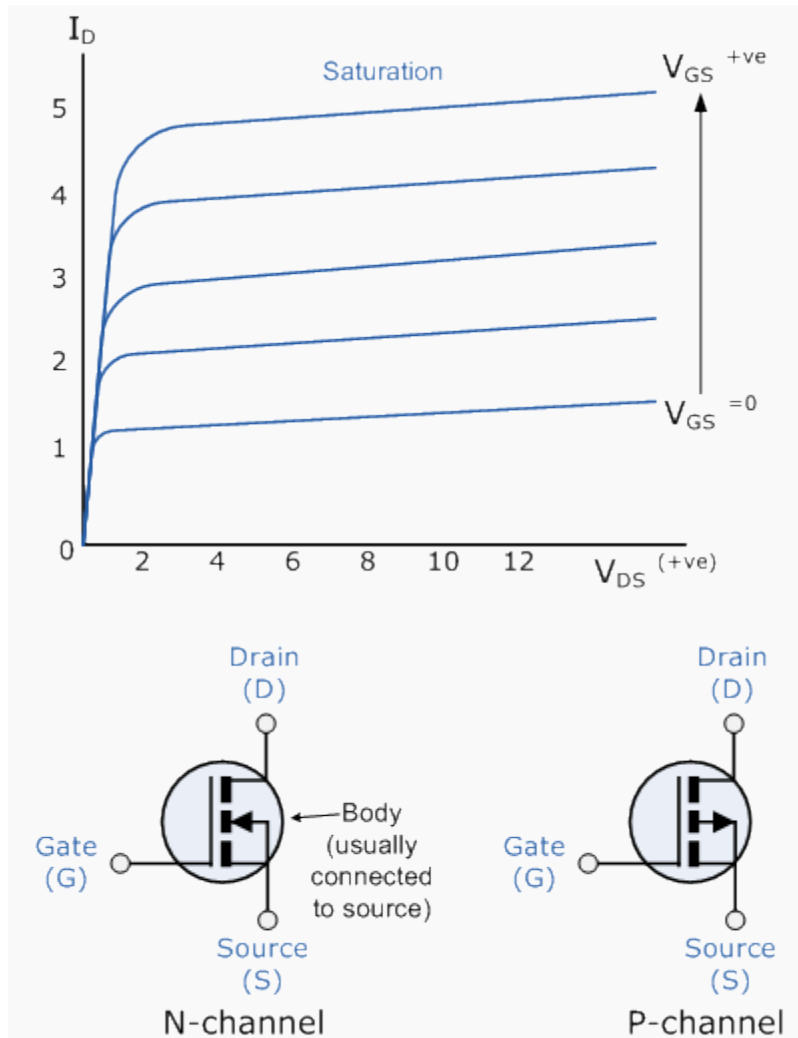
Enhancement-mode MOSFET's make excellent electronics switches due to their low "ON" resistance and extremely high "OFF" resistance and extremely high gate resistance. Enhancement-mode MOSFET's are used in integrated circuits to produce CMOS type **Logic Gates** and power switching circuits as they can be driven by digital logic levels.

MOSFET Summary

The **MOSFET** has an extremely high input gate resistance and as such a easily damaged by static electricity if not carefully protected. MOSFET's are ideal for use as electronic switches or common-source amplifiers as their power consumption is very small. Typical applications for MOSFET's are in Microprocessors, Memories, Calculators and Logic Gates etc. Also, notice that the broken lines within the symbol indicates a normally "OFF" Enhancement type showing that "NO" current can flow through the channel when zero gate voltage is applied and a continuous line within the symbol indicates a normally "ON" Depletion type showing that current "CAN" flow through the channel with

zero gate voltage. For P-Channel types the symbols are exactly the same for both types except that the arrow points outwards.

This can be summarised in the following switching table.

| MOSFET type	Vgs = +ve	Vgs = 0	Vgs = -ve
N-Channel Depletion	ON	ON	OFF
N-Channel Enhancement	ON	OFF	OFF
P-Channel Depletion	OFF	ON	ON
P-Channel Enhancement	OFF	OFF	ON

The MOSFET as a Switch

We saw previously, that the N-channel, Enhancement-mode MOSFET operates using a positive input voltage and has an extremely high input resistance (almost infinite) making it possible to interface with nearly any logic gate or driver capable of producing a positive output. Also, due to this very high input (Gate) resistance we can parallel together many different MOSFET's until we achieve the current handling limit required. While connecting together various MOSFET's may enable us to switch high current or high voltage loads, doing so becomes expensive and impractical in both components and circuit board space. To overcome this problem **Power Field Effect Transistors** or **Power FET's** where developed.

We now know that there are two main differences between FET's, Depletion-mode for JFET's and Enhancement-mode for MOSFET's and on this page we will look at using the Enhancement-mode MOSFET as a Switch.

By applying a suitable drive voltage to the Gate of an FET the resistance of the Drain-Source channel can be varied from an "OFF-resistance" of many hundreds of kΩ's, effectively an open circuit, to an "ON-resistance" of less than 1Ω, effectively a short circuit. We can also drive the MOSFET to turn "ON" fast or slow, or to pass high currents or low currents. This ability to turn the power MOSFET "ON" and "OFF" allows the device to be used as a very efficient switch with switching speeds much faster than standard bipolar junction transistors.

An example of using the MOSFET as a switch
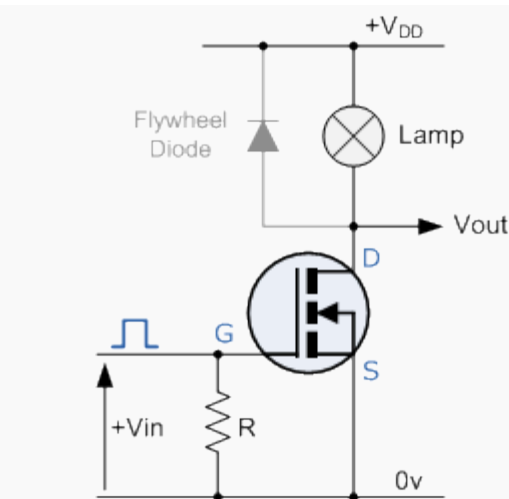
In this circuit arrangement an Enhancement-mode N-channel MOSFET is being used to switch a simple lamp "ON" and "OFF" (could also be an LED). The gate input voltage V_{GS} is taken to an appropriate positive voltage level to turn the device and the lamp either fully "ON", (V_{GS} = +ve) or a zero voltage level to turn the device fully "OFF", (V_{GS} = 0).

If the resistive load of the lamp was to be replaced by an inductive load such as a coil or solenoid, a "Flywheel" diode would be required in parallel with the load to protect the MOSFET from any back-emf.

Above shows a very simple circuit for switching a resistive load such as a lamp or LED. But when using power MOSFET's to switch either inductive or capacitive loads some form of protection is required to prevent the MOSFET device from becoming damaged. Driving an inductive load has the opposite effect from driving a capacitive load. For example, a capacitor without an electrical charge is a short circuit, resulting in a high "inrush" of current and when we remove the voltage from an inductive load we have a large reverse voltage build up as the magnetic field collapses, resulting in an induced back-emf in the windings of the inductor.

For the power MOSFET to operate as an analogue switching device, it needs to be switched between its "Cut-off Region" where V_{GS} = 0 and its "Saturation Region" where $V_{GS(on)}$ = +ve. The power dissipated in the MOSFET (P_D) depends upon the current flowing through the channel I_D at saturation and also the "ON-resistance" of the channel given as $R_{DS(on)}$. For example.

Example No1

Lets assume that the lamp is rated at 6v, 24W and is fully "ON" and the standard MOSFET has a channel "ON-resistance" ($R_{DS(on)}$) value of 0.1ohms. Calculate the power dissipated in the MOSFET switch.

The current flowing through the lamp is calculated as:

$$P = V \times I, \text{ therefore, } I_D = \frac{P}{V} = \frac{24}{6} = 4 \, amps$$

Then the power dissipated in the MOSFET will be given as:

$$P_D = I_D^2 \times R_{DS}, \text{ therefore, } P_D = 4^2 \times 0.1 = 1.6 \, watts$$

You may think, well so what!, but when using the MOSFET as a switch to control DC motors or high inrush current devices the "ON" channel resistance ($R_{DS(on)}$) is very important. For example, MOSFET's that control DC motors, are subjected to a high in-rush current as the motor first begins to rotate. Then a high $R_{DS(on)}$ channel resistance value would simply result in large amounts of power being dissipated within the MOSFET itself resulting in an excessive temperature rise, and which in turn could result in the MOSFET becoming very hot and damaged due to a thermal overload. But a low $R_{DS(on)}$ value on the other hand is also desirable to help reduce the effective saturation voltage ($V_{DS(sat)} = I_D \times R_{DS(on)}$) across the MOSFET. When using MOSFET´s or any type of Field Effect Transistor for that matter as a switching device, it is always advisable to select ones that have a very low $R_{DS(on)}$ value or at least mount them onto a suitable heatsink to help reduce any thermal runaway and damage.

Power MOSFET Motor Control

Because of the extremely high input or Gate resistance that the MOSFET has, its very fast switching speeds and the ease at which they can be driven makes them ideal to interface with op-amps or standard logic gates. However, care must be taken to ensure that the gate-source input voltage is correctly chosen because when using the **MOSFET as a switch** the device must obtain a low $R_{DS(on)}$ channel resistance in proportion to this input gate voltage. For example, do not apply a 12v signal if a 5v signal voltage is required. Power MOSFET´s can be used to control the movement of DC motors or brushless stepper motors directly from computer logic or Pulse-width Modulation (PWM) type controllers. As a DC motor offers high starting torque and which is also proportional to the armature current, MOSFET switches along with a PWM can be used as a very good speed controller that would provide smooth and quiet motor operation.

Simple Power MOSFET Motor Controller
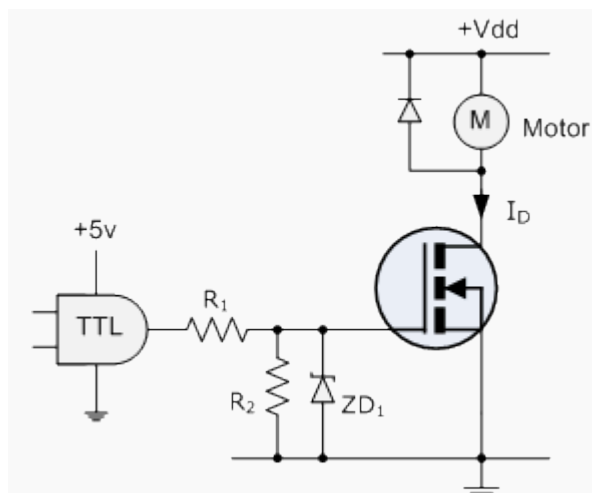
As the motor load is inductive, a simple "Free-wheeling" diode is connected across the load to dissipate any back emf generated by the motor when the MOSFET turns it "OFF".
The Zener diode is used to prevent excessive gate-source input voltages.

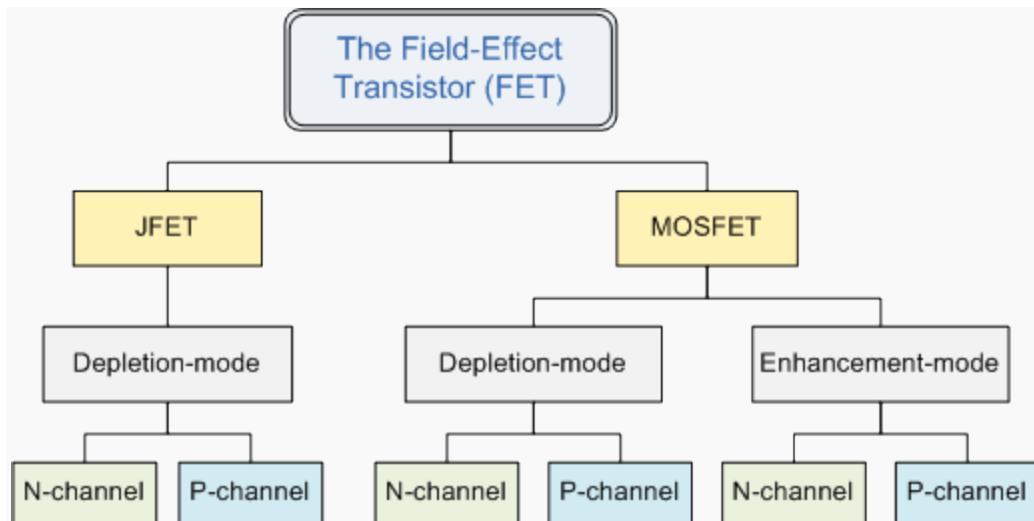## Summary of Bipolar Junction Transistors

- The **Bipolar Junction Transistor** (BJT) is a three layer device constructed form two semiconductor diode junctions joined together, one forward biased and one reverse biased.

- There are two main types of bipolar junction transistors, the NPN and the PNP transistor.

- Transistors are "**Current Operated Devices**" where a much smaller Base current causes a larger Emitter to Collector current, which themselves are nearly equal, to flow.

- The most common transistor connection is the Common-emitter configuration.

- Requires a Biasing voltage for AC amplifier operation.

- The Collector or output characteristics curves can be used to find either Ib, Ic or β to which a load line can be constructed to determine a suitable operating point, Q with variations in base current determining the operating range.

- A transistor can also be used as an electronic switch to control devices such as lamps, motors and solenoids etc.

- Inductive loads such as DC motors, relays and solenoids require a reverse biased "Flywheel" diode placed across the load. This helps prevent any induced back emf's generated when the load is switched "OFF" from damaging the transistor.

- The NPN transistor requires the Base to be more positive than the Emitter while the PNP type requires that the Emitter is more positive than the Base.

Summary of Field Effect Transistors

- **Field Effect Transistors**, or FET's are "**Voltage Operated Devices**" and can be divided into two main types: Junction-gate devices called JFET's and Insulated-gate devices called IGFET´s or more commonly known as MOSFET's.

- Insulated-gate devices can also be sub-divided into Enhancement types and Depletion types. All forms are available in both N-channel and P-channel versions.

- FET's have very high input resistances so very little or no current (MOSFET types) flows into the input terminal making them ideal for use as electronic switches.

- The input impedance of the MOSFET is even higher than that of the JFET due to the insulating oxide layer and therefore static electricity can easily damage MOSFET devices so care needs to be taken when handling them.

- FET's have very large current gain compared to junction transistors.

- They can be used as ideal switches due to their very high channel "OFF" resistance, low "ON" resistance.

The Field Effect Transistor Family-tree

Field Effect Transistors can be used to replace normal Bipolar Junction Transistors in electronic circuits and a simple comparison between FET's and transistors stating both their advantages and their disadvantages is given below.

	Field Effect Transistor (FET)	Bipolar Junction Transistor (BJT)
1	Low voltage gain	High voltage gain
2	High current gain	Low current gain
3	Very input impedance	Low input impedance
4	High output impedance	Low output impedance
5	Low noise generation	Medium noise generation
6	Fast switching time	Medium switching time
7	Easily damaged by static	Robust
8	Some require an input to turn it "OFF"	Requires zero input to turn it "OFF"
9	Voltage controlled device	Current controlled device
10	Exhibits the properties of a Resistor	
11	More expensive than bipolar	Cheap
12	Difficult to bias	Easy to bias