

EEE1002/EEE1010 - Electronics I

Analogue Electronics

Lecture Notes

S. Le Goff

School of Electrical and Electronic Engineering

Newcastle University

Module Organization

Lecturer for Analogue Electronics: S Le Goff (module leader)

Lecturer for Digital Electronics: N Coleman

Analogue Electronics: 24 hours of lectures and tutorials (12 weeks × 2 hours/week)

Assessment for analogue electronics:

- Mid-semester test in November, analogue electronics only, 1 hour, 8% of the final mark.
- Final examination in January, analogue & digital electronics, 2 hours, 50% of the final mark.

Recommended Books:

- Electronics – A Systems Approach, 4th Edition, by Neil Storey, Pearson Education, 2009.
- Analysis and Design of Analog Integrated Circuits, 5th Edition, by Paul Gray, Paul Hurst, Stephen Lewis, and Robert Meyer, John Wiley & Sons, 2012.
- Digital Integrated Circuits – A Design Perspective, 2th Edition, by Jan Rabaey, Ananta Chandrakasan, and Borivoje Nikolic, Pearson Education, 2003.
- Microelectronic Circuits and Devices, 2th Edition, by Mark Horenstein, Prentice Hall, 1996.
- Electronics Fundamentals – A Systems Approach, by Thomas Floyd and David Buchla, Pearson Education, 2014.
- Principles of Analog Electronics, by Giovanni Saggio, CRC Press (Taylor & Francis Group), 2014.

1. Semiconductors

Solid materials may be divided, with respect to their electrical properties, into three categories:

1. Conductors

Conductors (e.g., copper, aluminium) have a cloud of free electrons at all temperature above absolute zero. This is formed by the weakly bound “valence” electrons in the outermost orbits of their atoms. If an electric field is applied across such a material, electrons will flow, causing an electric current.

2. Insulators

In insulating materials, the valence electrons are tightly bound to the nuclei of the atoms and very few of them are able to break free to conduct electricity. The application of an electric field does not cause a current to flow as there are no mobile charge carriers.

3. Semiconductors

At very low temperatures, semiconductors have the properties of an insulator. However, at higher temperatures, some electrons are free to move and the materials take on the properties of a conductor (albeit a poor one). Nevertheless, semiconductors have some useful characteristics that make them distinct from both insulators and conductors.

To understand the operation of diodes, transistors, and other electronic devices, we need to understand the basic structure of semiconductors.

A few common semiconductor materials: silicon (Si), germanium (Ge), gallium arsenide (GaAs), indium phosphide (InP), silicon carbide (SiC), silicon-germanium (SiGe).

The first transistors were made from germanium (Ge). Silicon (Si) types currently predominate but certain advanced microwave and high performance versions employ the compound semiconductor material gallium arsenide (GaAs) and the semiconductor alloy silicon germanium (SiGe).

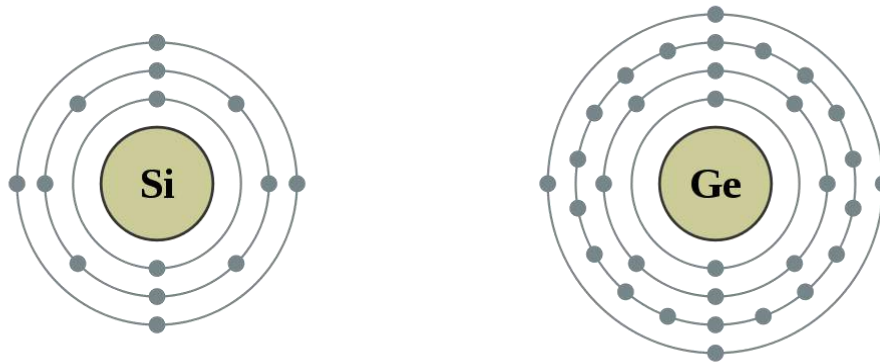
Silicon and germanium fall in column IVa of the Periodic Table. This is the carbon family of elements. The characteristic of these elements is that each atom has four electrons to share with adjacent atoms.

14: Silicon

2,8,4

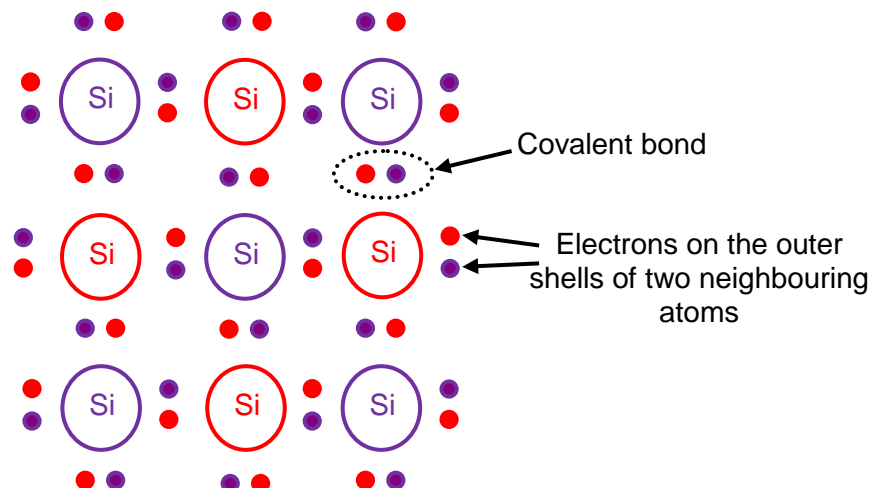
32: Germanium

2,8,18,4



| Z | Element | No. of electrons/shell |
|----|-----------|------------------------|
| 6 | carbon | 2, 4 |
| 14 | silicon | 2, 8, 4 |
| 32 | germanium | 2, 8, 18, 4 |
| 50 | tin | 2, 8, 18, 18, 4 |
| 82 | lead | 2, 8, 18, 32, 18, 4 |

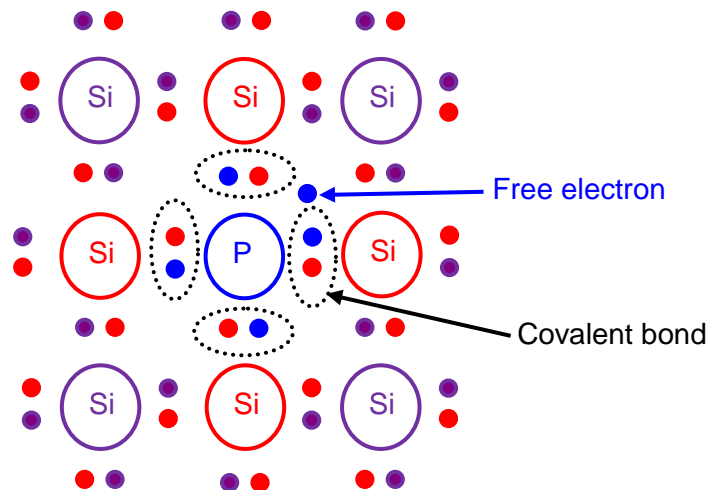
Let us have a closer look at silicon. The crystal structure of silicon is represented below.



The nature of a bond between two silicon atoms is such that each atom provides one electron to share with the other. The two electrons thus shared between atoms form a “covalent bond”. Such a bond is very stable and holds the two atoms together very tightly. It requires a lot of energy to break this bond.

All of the outer electrons of all silicon atoms are used to make covalent bonds with other atoms. There are no electrons available to move from place to place as an electrical current. Thus, a pure silicon crystal is quite a good insulator. Increasing the temperature results in some electrons breaking free from their covalent bonds and this improves the conductivity of the silicon crystal.

To allow a silicon crystal to conduct electricity without having to increase the temperature, we must find a way to allow some electrons to move from one place to the other within the crystal despite the covalent bonds between atoms. One way to accomplish this is to introduce an impurity such as arsenic or phosphorus into the crystal structure. Such process is called doping. These elements are from column Va of the Periodic Table, and have five outer valence electrons to share with other atoms.

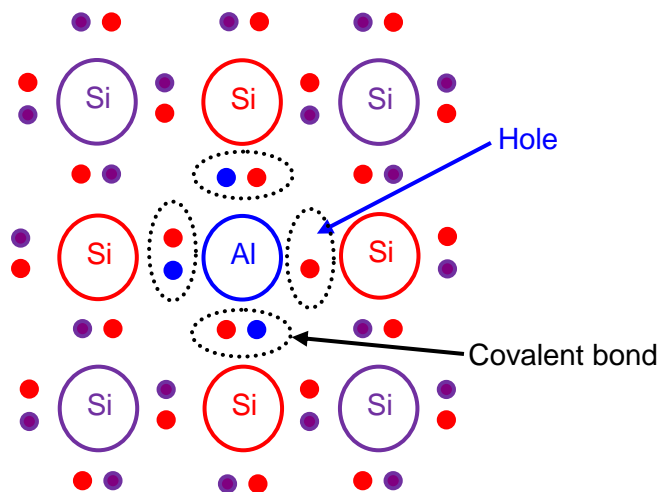


Four of these five electrons bond with adjacent silicon atoms as before, but the fifth electron cannot form a bond and is thus left “alone”. This electron can easily be moved with only a small applied electrical voltage. Because the resulting crystal has an excess of current-carrying electrons, each with a negative charge, it is known as "N-type" silicon.

Such construction does not conduct electricity as easily as, say, copper or silver since it does exhibit some resistance to the flow of electricity. It cannot properly be called a conductor, but at the same time it is no longer an insulator. Therefore, it is known as a semiconductor.

We obtained a semiconductor material by introducing a 5-electron impurity into a matrix of 4-electron atoms. We can also do the opposite and introduce a 3-electron impurity into such a crystal. Suppose we introduce some aluminium (from column IIIa in the Periodic Table) into the crystal. We could also use gallium which is also in column IIIa.

These elements only have three valence electrons available to share with other atoms. Those three electrons do indeed form covalent bonds with adjacent silicon atoms, but the expected fourth bond cannot be formed. A complete connection is impossible here, leaving a "hole" in the structure of the crystal.



There is an empty place where an electron should logically go, and often an electron will try to move into that space to fill it. However, the electron filling the hole has to leave a covalent bond behind to fill this empty space, and therefore leaves another hole behind as it moves. Yet another electron may move into that hole, leaving another hole behind, and so on. In this manner, holes appear to move as positive charges through the crystal. Therefore, this type of semiconductor material is designated "P-type" silicon.

In an N-type semiconductor, the electrons are often referred to as the majority charge carriers, whereas the holes are called the minority charge carriers since they are actually also present but

at much lower concentration. In a similar way, in a P-type semiconductor, the holes are referred to as the majority charge carriers, whereas the electrons, which are present at much lower concentration, are the minority charge carriers.

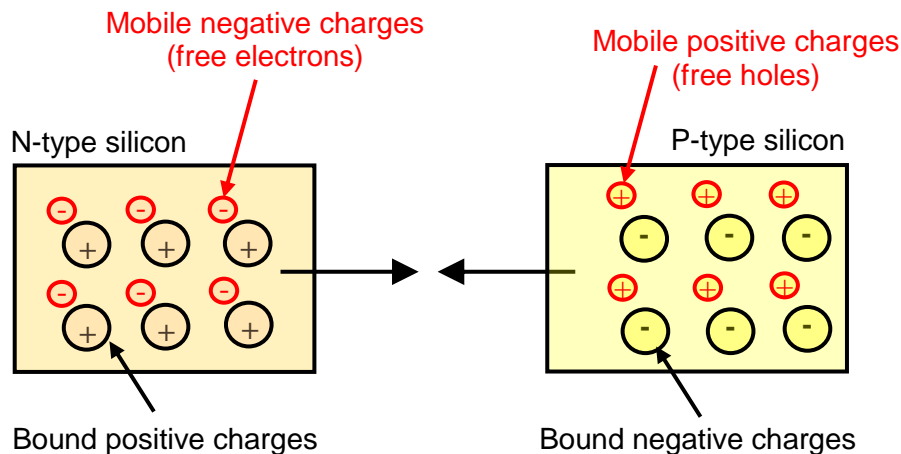
The role played by the minority charge carriers can sometimes be ignored for simplicity purposes.

2. The PN Junction (Diode)

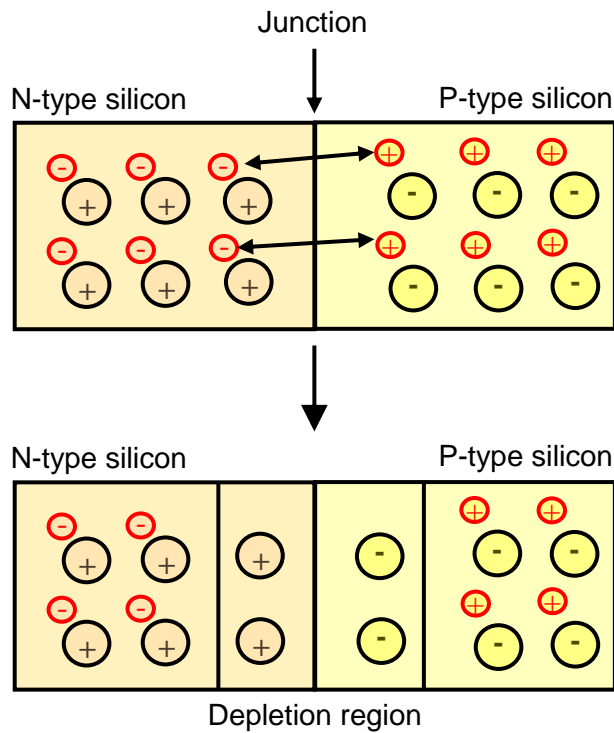
• Basic Operation

We have just seen that a crystal of pure silicon can be turned into a relatively good electrical conductor by adding an impurity such as arsenic or phosphorus (for an N-type semiconductor) or aluminium or gallium (for a P-type semiconductor). By itself, a single type of semiconductor material is not very useful. But, something interesting happens when a single semiconductor crystal contains both P-type and N-type regions.

Hereafter, we examine the properties of a single silicon crystal which is half N-type and half P-type. The two types are shown separated, as if they were two separate crystals being put in contact. In the real world, two such crystals cannot be joined together usefully. Therefore, a practical PN junction can only be created by inserting different impurities into different parts of a single crystal.

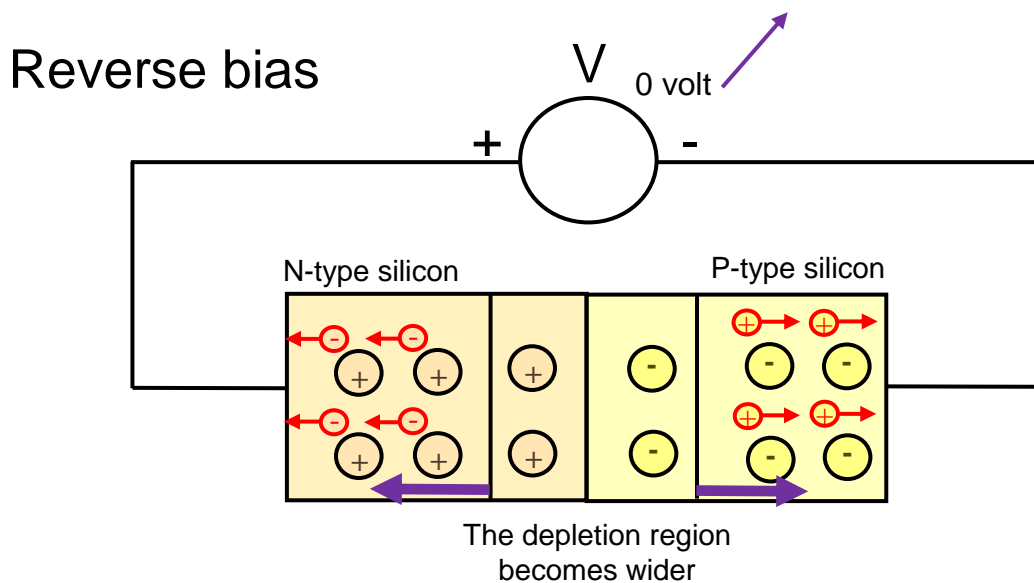


When we join the N- and P-type crystals together, an interesting interaction occurs around the junction. The extra electrons in the N region will combine with the extra holes in the P region. This leaves an area where there are no mobile charges, known as depletion region, around the junction.



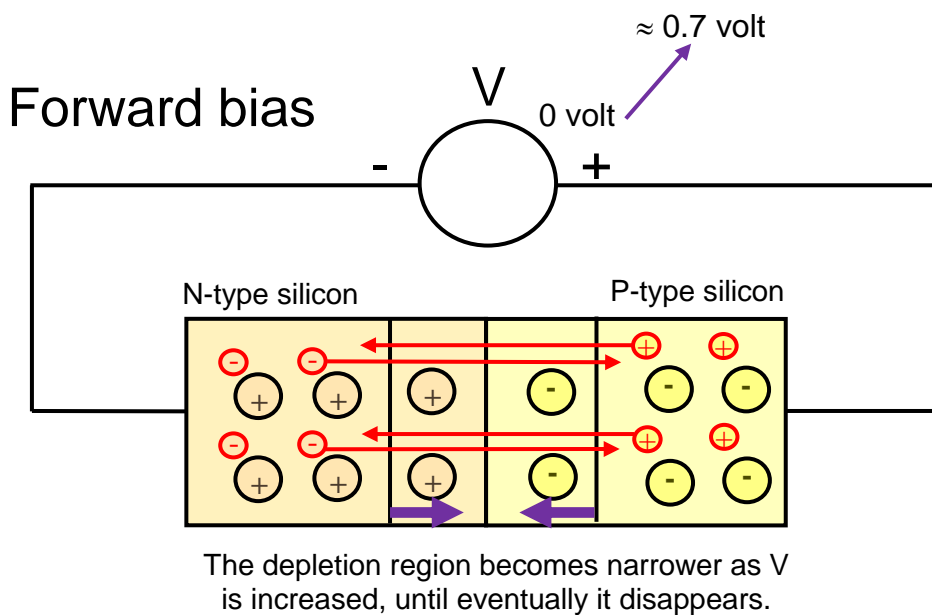
Suppose now that we apply a voltage to the outside ends of our PN crystal.

Assume first that the positive voltage is applied to the N-type material. In such case, the positive voltage applied to the N-type material attracts free electrons towards the end of the crystal and away from the junction, while the negative voltage applied to the P-type end attracts holes away from the junction.



The result is that all available current carriers are attracted further away from the junction, and the depletion region grows correspondingly larger. Therefore, there is no current flow through the crystal because no current carriers can cross the junction. This is known as reverse bias applied to the semiconductor crystal.

Assume now that the applied voltage polarities are reversed. The negative voltage applied to the N-type end pushes electrons towards the junction, while the positive voltage at the P-type end pushes holes towards the junction. This has the effect of shrinking the depletion region.



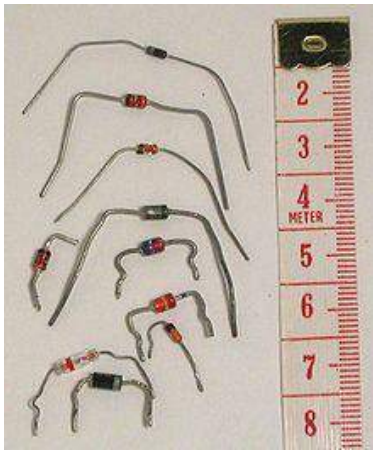
Once the applied voltage V has become large enough to make the depletion region completely disappear, i.e. once the value of V becomes equal to the threshold voltage V_d of the PN junction ($V_d \approx 0.7$ volt for silicon and $V_d \approx 0.3$ volt for germanium), current carriers of both types are finally able to cross the junction into the opposite ends of the crystal. Now, electrons in the P-type end are attracted to the positive applied voltage, while holes in the N-type end are attracted to the negative applied voltage. This is the condition of forward bias.

The conclusion is that an electrical current can flow through the junction in the forward direction, but not in the reverse direction. This is the basic property of a semiconductor diode.

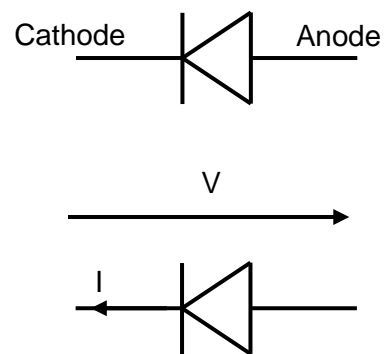
It is important to realize that holes exist only within the crystal. A hole reaching the negative

terminal of the crystal is filled by an electron from the power source and simply disappears. At the positive terminal, the power supply attracts an electron out of the crystal, leaving a hole behind to move through the crystal toward the junction again.

• **Current-Voltage Characteristic of a Diode**



Symbol of a diode

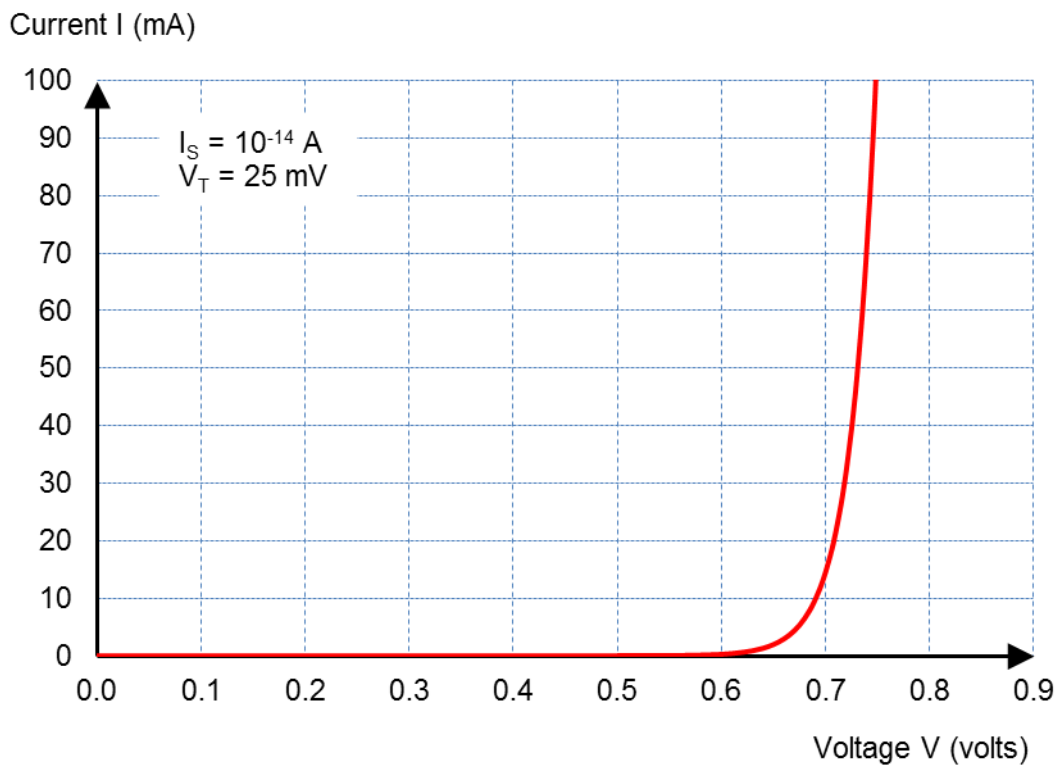
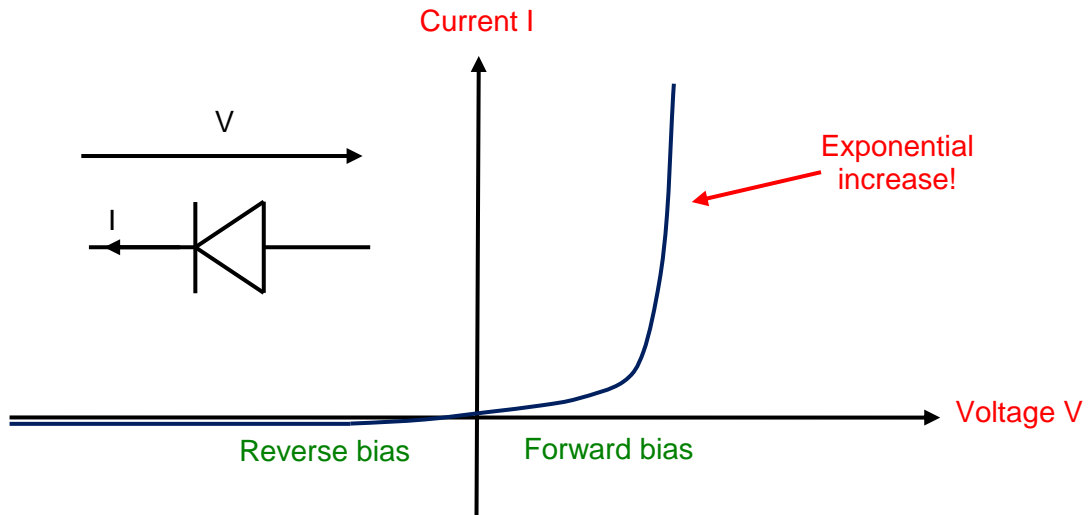


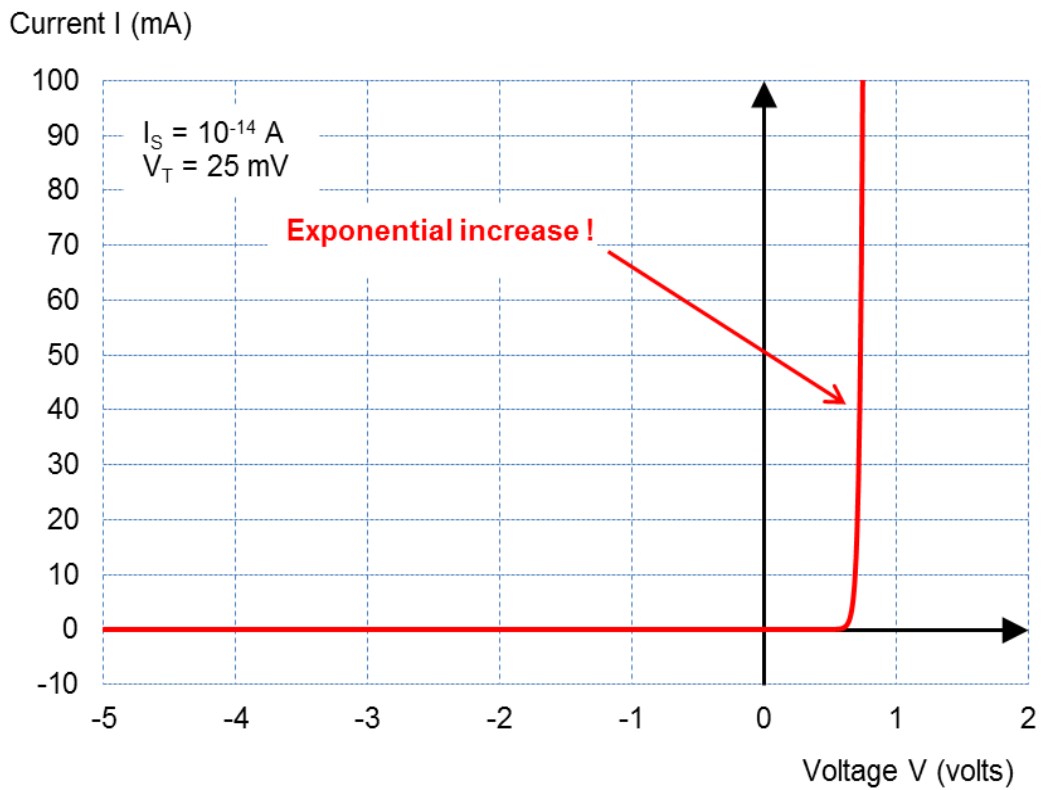
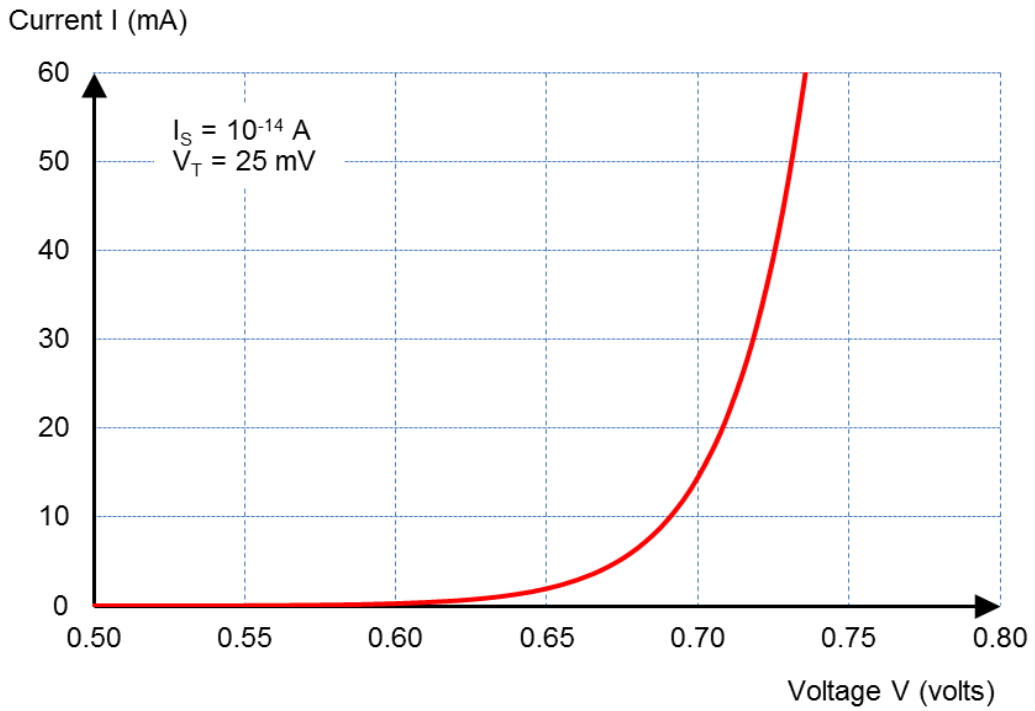
The Shockley diode equation, named after transistor co-inventor William Shockley, gives the current–voltage characteristic of a diode in either forward or reverse bias. The equation is given by

$$I = I_S \left(\exp \left\{ \frac{qV}{\eta kT} \right\} - 1 \right) = I_S \left(\exp \left\{ \frac{V}{\eta V_T} \right\} - 1 \right),$$

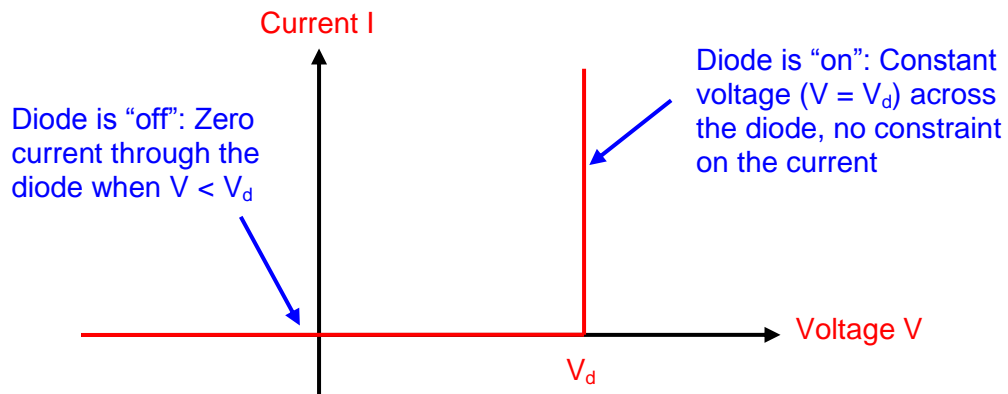
- where
- I_S : Saturation current of the diode (in the range 10^{-8} to 10^{-16} A, typically);
 - η : Emission coefficient. This is an empirical constant that varies from 1 to 2 depending on the fabrication process and semiconductor material and in many cases is assumed to be approximately equal to 1 (and thus omitted).
 - q : Electron charge ($= 1.602 \times 10^{-19}$ C);
 - T : Temperature in degrees Kelvin;
 - k : Boltzmann’s constant ($= 1.38 \times 10^{-23}$ J/K);
 - V_T : Thermal voltage (≈ 25 mV at room temperature).

This expression means that the current flowing through a diode varies exponentially with the applied voltage.





This rather complicated equation is a bit difficult to use for manual circuit analysis. Electronic engineers deal with this problem by simplifying things and using the much simpler model of the diode given below.

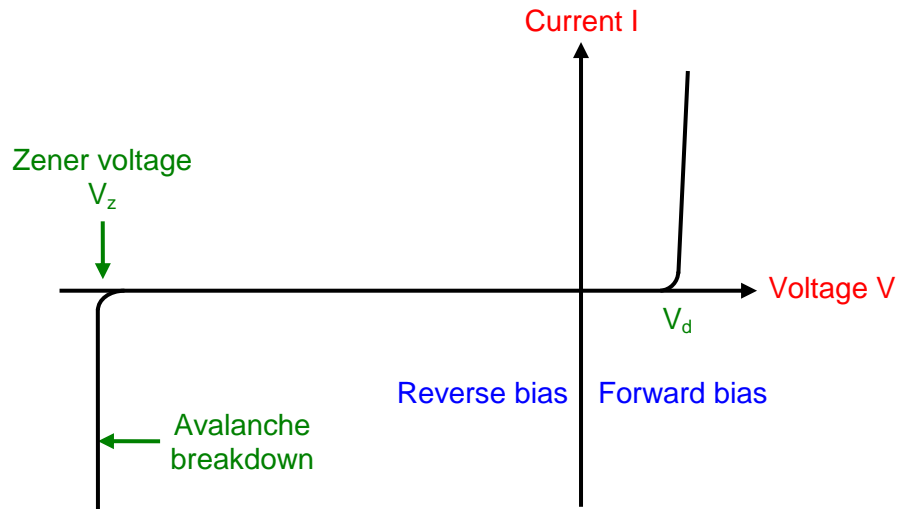


Note that V_d is called *threshold voltage* or *forward voltage drop* of the diode. We have $V_d \approx 0.7 \text{ V}$ for a silicon diode, $V_d \approx 0.3 \text{ V}$ for a germanium diode, and $V_d \approx 0.25 \text{ V}$ for a Schottky diode.

In this model, the current is zero for any voltage below the threshold voltage V_d . In effect, the diode is viewed as a switch which is open when we apply low or negative voltages across it but which closes when we apply a voltage equal to V_d across it. It is important to understand that, with this model, it is strictly impossible to get a voltage larger than V_d across the diode.

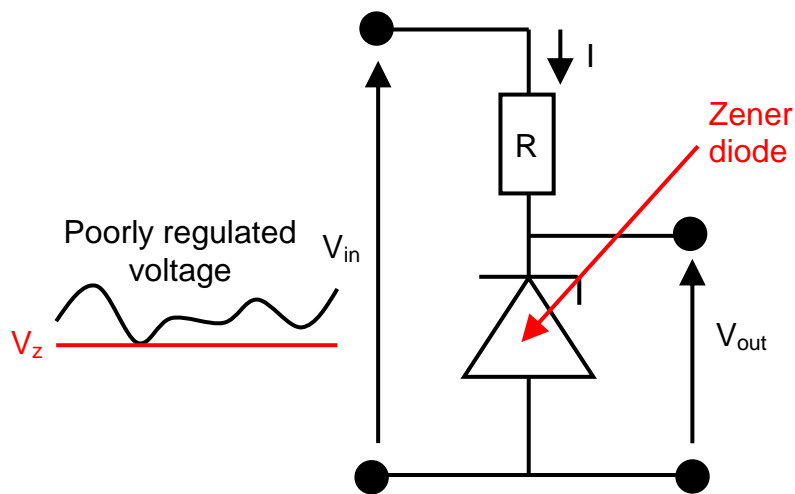
• Zener Diodes

With the application of sufficient reverse voltage, a diode experiences a breakdown and conduct current in the reverse direction. Electrons which break free under the influence of the applied electric field can be accelerated enough that they can knock loose other electrons and the subsequent collisions quickly become an avalanche.



When this process takes place, very small changes in voltage can cause very large changes in current. The breakdown process depends upon the applied electric field. Thus, by changing the thickness of the layer to which the voltage is applied, Zener diodes, named after the American physicist Clarence Zener, can be formed which break down at voltages from about a few volts to several hundred volts.

The useful feature here is that the voltage across the diode remains nearly constant even with large changes in current through the diode. Such diodes find wide use in electronic circuits as voltage regulators. To illustrate this point, let us consider the circuit shown below.



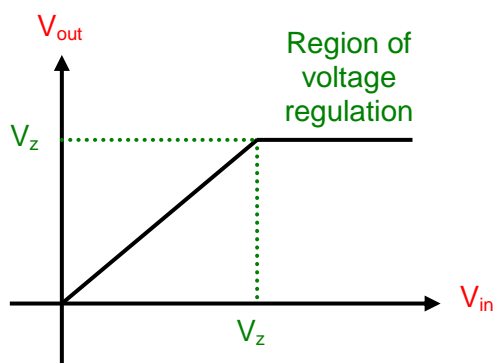
In this circuit, the Zener diode is connected so that it is reverse-biased by the input signal. We can write the following equation: $V_{in} = R \cdot I + V_{out}$.

If $V_{in} > V_z$, the diode junction will break down and conduct, drawing current from the resistance R . The diode prevents the output voltage V_{out} from going above its breakdown voltage V_z and thus generates a constant output voltage $V_{out} = V_z$, irrespective of the value of the input voltage as long as it remains higher than V_z .

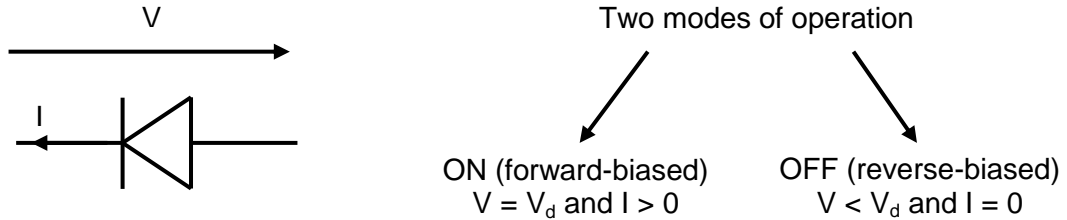
Note that, in this case, the current flowing through the resistor is given by

$$I = \frac{V_{in} - V_{out}}{R} = \frac{V_{in} - V_z}{R}.$$

If $V_{in} < V_z$, the reverse-bias voltage is not sufficient to break down the junction. As a result, the diode will conduct no current. The output voltage will be equal to the input voltage as there is no voltage drop across the resistance, i.e. $V_{out} = V_{in}$. In this situation, the Zener diode has no effect on the circuit.

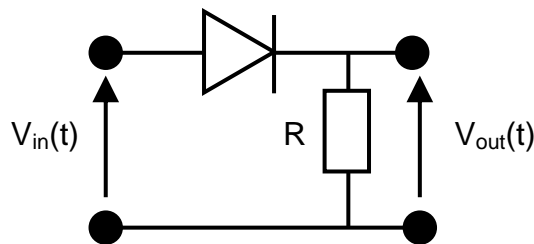


Tutorial 1 – Diode Circuits

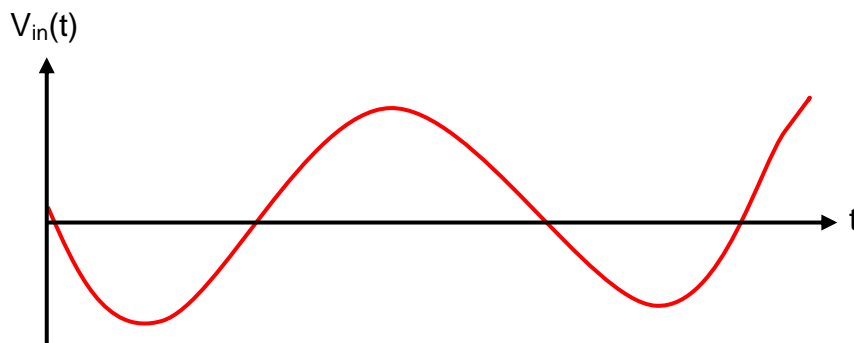


Question 1: Half-Wave Rectifier

The primary function of a rectifier circuit is to change an AC input voltage into a voltage that is only positive or only negative. In essence, a rectifier eliminates the unwanted polarity of the input waveform. As an illustration, consider the circuit below called *half-wave rectifier*.



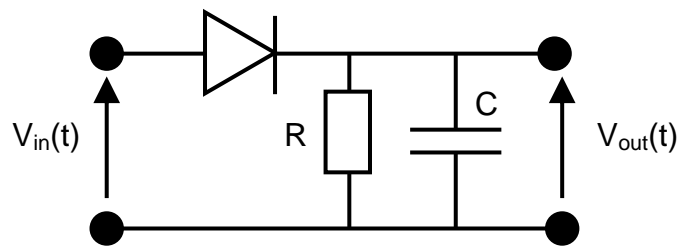
Find the output signal $V_{out}(t)$ obtained with the input signal $V_{in}(t)$ depicted below.



Question 2: Half-Wave Rectifier with Capacitor

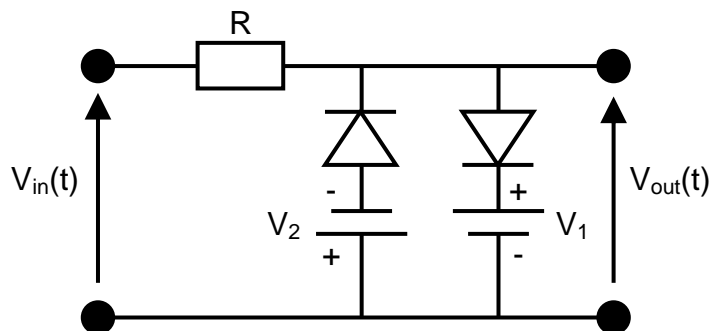
We can modify the half-wave rectifier studied in Question 1 by placing a capacitor in parallel with the resistor. We thus obtain a power supply circuit that accepts an AC voltage as its input and provides a DC voltage as its output. This circuit can also be employed as a demodulator as part of the receiver in amplitude-modulation (AM) communication systems.

Study the operation of the circuit depicted below assuming that the input signal is a sine wave.



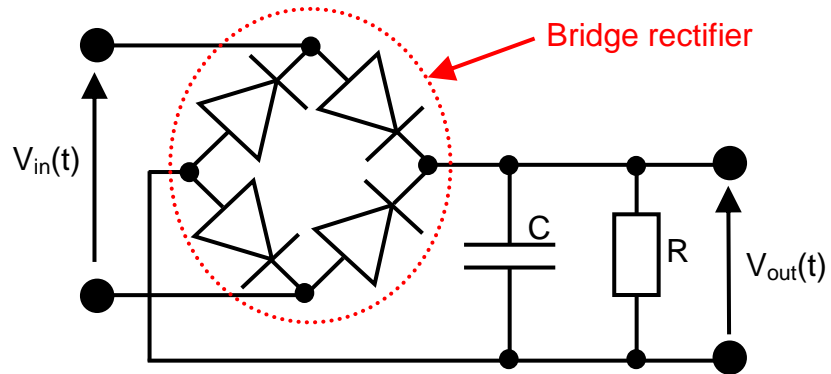
Question 3: Diode Clipping

A diode-clipping circuit can be used to limit the voltage swing of a signal. Consider the circuit depicted below and find the expression of the output voltage $V_{out}(t)$ as a function of the input voltage $V_{in}(t)$. V_1 and V_2 designate two constant voltage sources (with $V_1 > 0$ and $V_2 > 0$).



Question 4: Power-Supply Circuit using a Full-Wave Bridge Rectifier

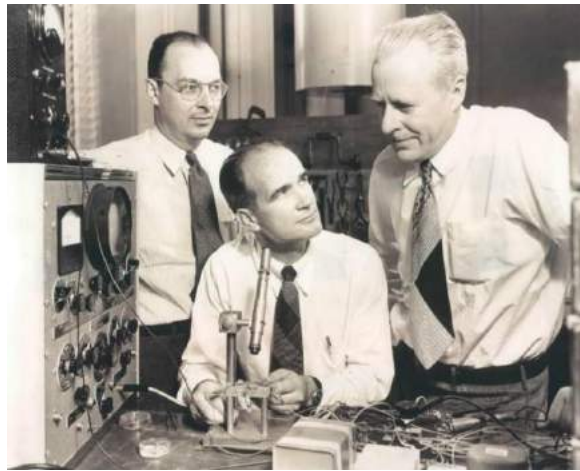
Consider the circuit shown below.



Study the operation of the circuit depicted below assuming that the input signal is a sine wave.

3. The Bipolar Junction Transistor

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, William Shockley and Walter Brattain at Bell Labs, 1948



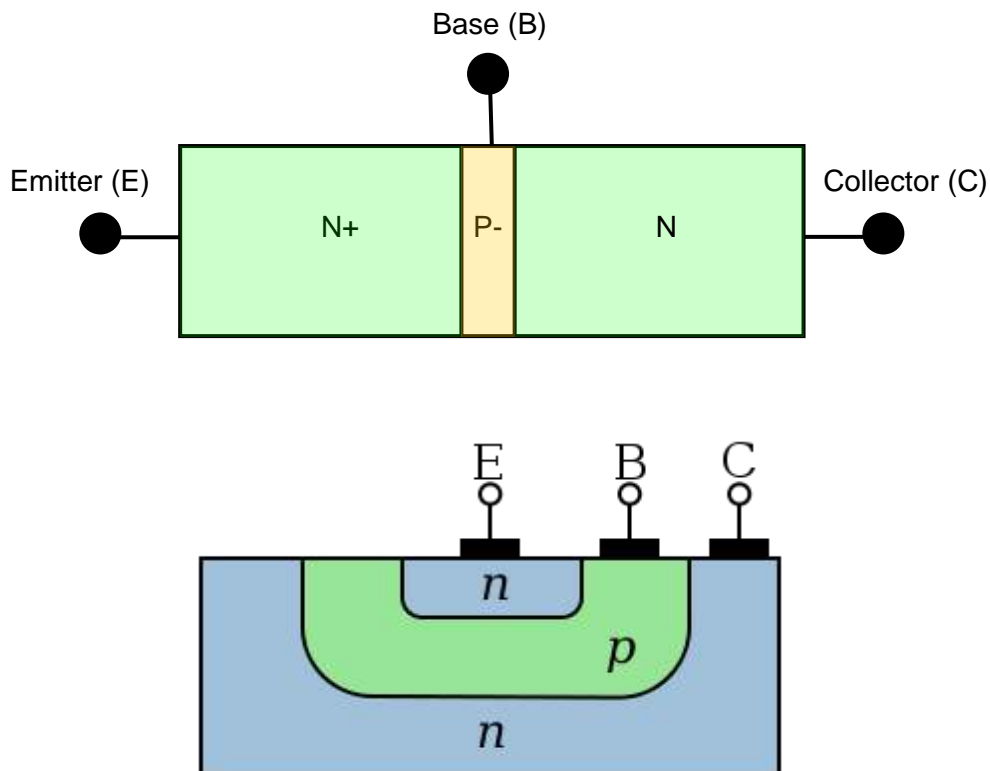
A replica of the first working transistor

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.

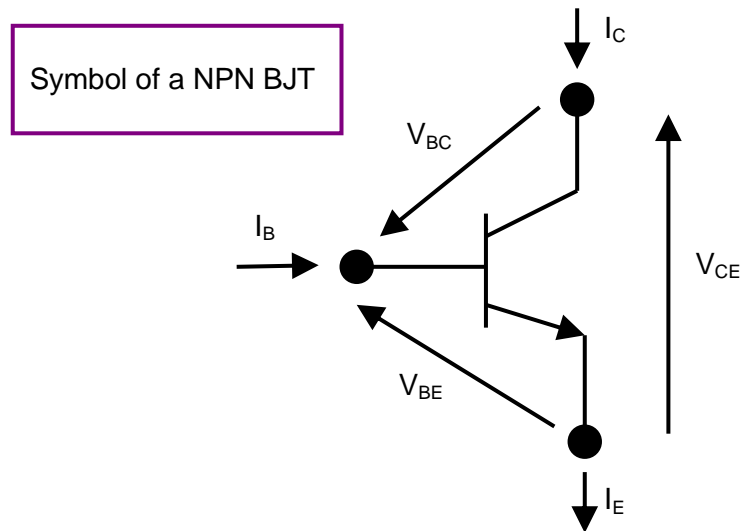
Although several companies each produce over a billion individually packaged (known as discrete) transistors every year, the vast majority of transistors now are produced in integrated circuits, along with diodes, resistors, capacitors and other electronic components, to produce complete electronic circuits.

• Basic Operation

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 with the appropriate impurity, while the base region is very lightly doped. The collector region has a moderate doping level. Note that the structure is 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, but we can also build equivalent devices in P, N, and P order instead. In fact, it is sometimes useful to have both types of devices available.

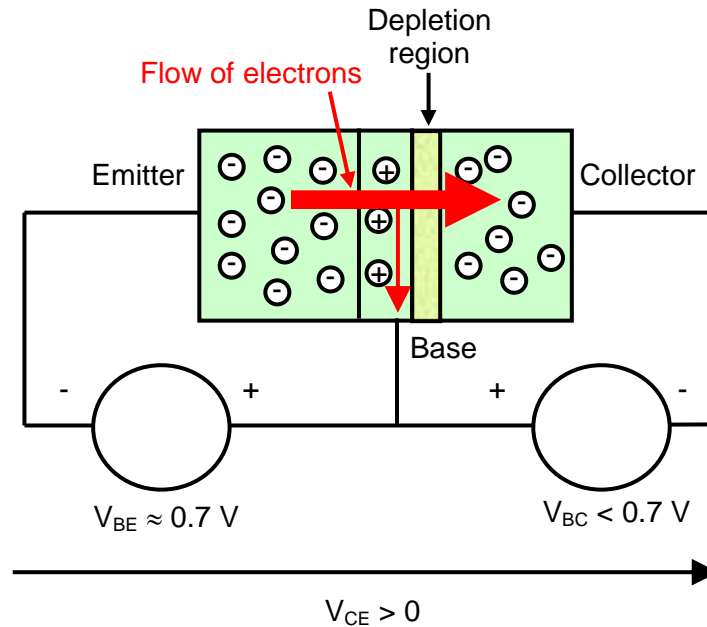
Let us see what happens when bias voltages are applied to such device. Let us assume the use of a silicon BJT.

Consider first 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. These conditions imply that $V_{BE} \approx 0.7 \text{ V}$ and $V_{BC} < 0.7 \text{ V}$. If we take the emitter as a reference, these conditions can be re-written as $V_{BE} \approx 0.7 \text{ volt}$ and $V_{CE} = V_{CB} + V_{BE} = V_{BE} - V_{BC} > 0 \text{ V}$.

Since we already know how a PN junction operates, we would expect to have electrons move from emitter to base and leave the device through the base at that point. With the 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. As the base is very thin, electrons entering the base find themselves close to 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 entering the junction area are swept across the depletion area into the collector and give rise to a collector current.



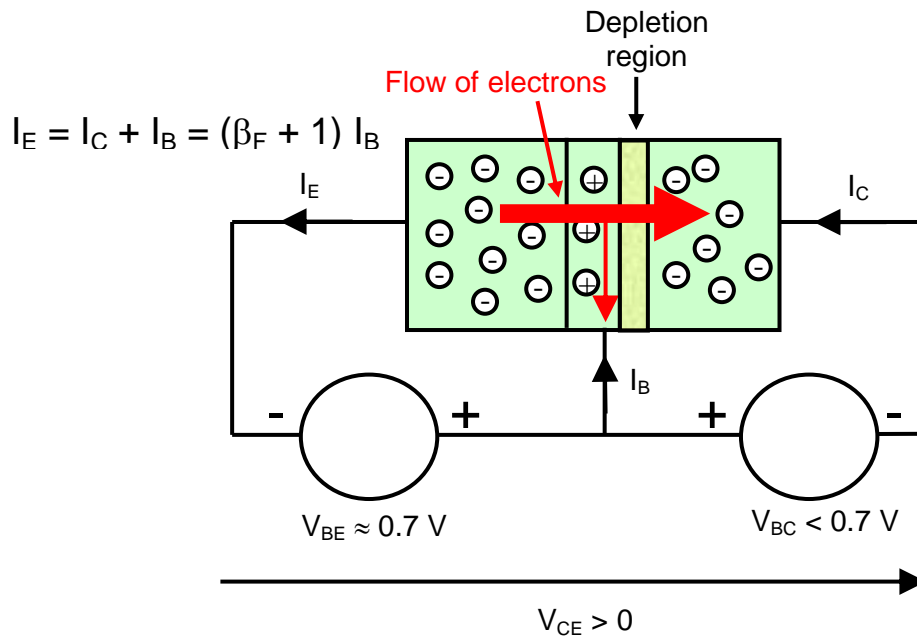
Careful design ensures that the majority of the electrons entering the base are swept across the base-collector junction into the collector.

Thus the flow of electrons from emitter to collector is many times greater than the flow from emitter to the base. In fact, the collector current I_C is proportional to the base current I_B :

$$I_C = \beta_F I_B,$$

where β_F is a constant that can take its value in the range from approximately 50 to 300 for typical bipolar technologies.

This current amplification phenomenon is known as the transistor effect.



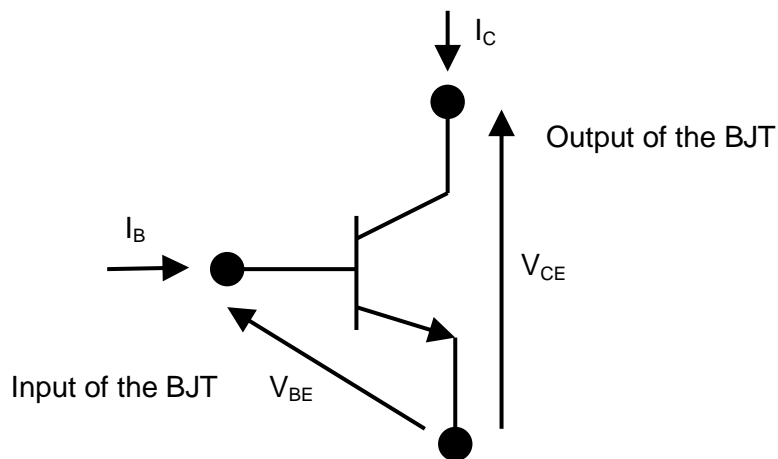
As previously mentioned, it is also possible to build a transistor with the region types reversed (PNP structure). In this case, holes will be drawn from the emitter into the base region by the forward bias, and will then be pulled into the collector region by the higher negative bias. Otherwise, this device works the same way and has the same general properties as the one described above. To distinguish between the two types of transistors, we refer to them by the order in which the different regions appear. Thus, this is a PNP transistor while the device described above is an NPN transistor.

However, PNP transistors often have lower β_F values and are slower (i.e., operate at lower frequencies) than their NPN counterparts.

• **Common-Emitter Configuration of a BJT**

The BJT is 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.

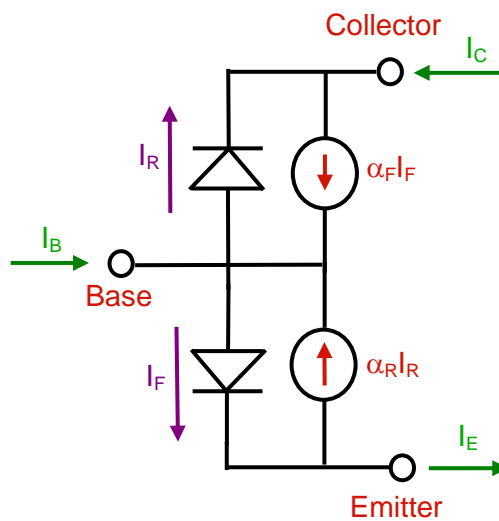


• The Four Different Modes of Operation of a Transistor

In 1954, Jewell Ebers and John Moll introduced their (static) model of a BJT. The Ebers-Moll model is depicted below.

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

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



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

We can write the general equations for the Ebers-Moll model:

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

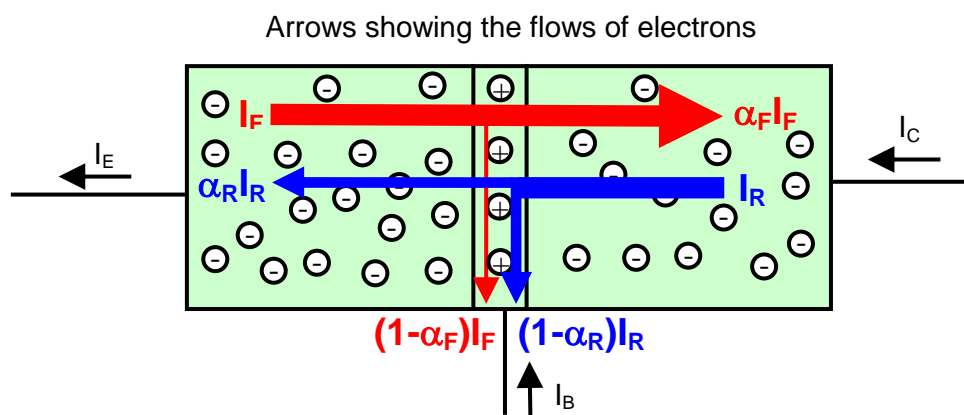
Collector current: $I_C = \alpha_F I_F - I_R$

Emitter current: $I_E = I_F - \alpha_R I_R = I_B + I_C$

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

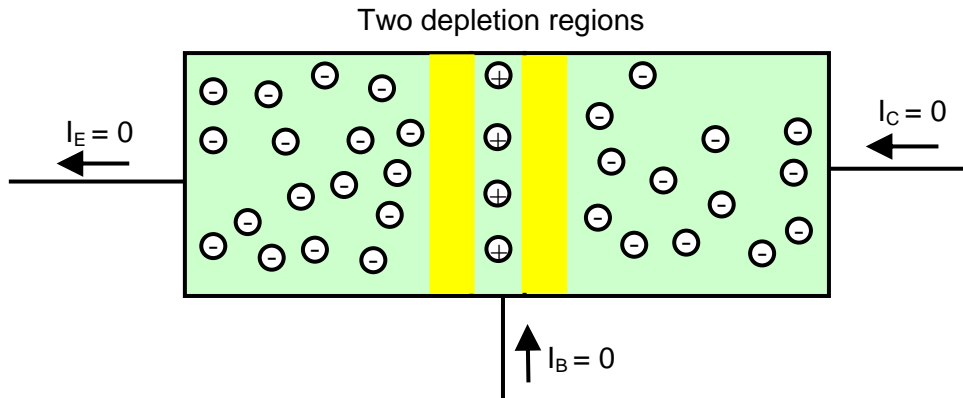
1. Both diodes represent the base-emitter and base-collector PN junctions.
2. The parameter α_F represents the proportion of electrons coming from the emitter that are able to reach the collector. The fact that α_F is very close to the unit implies that the majority of electrons coming from the emitter do reach the collector, while the remaining electrons leave the device through the base.
3. The parameter α_R represents the proportion of electrons coming from the collector that are able to reach the emitter. The fact that the value of α_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 α_F and α_R is due to the inherent non-symmetrical physical structure of a BJT.

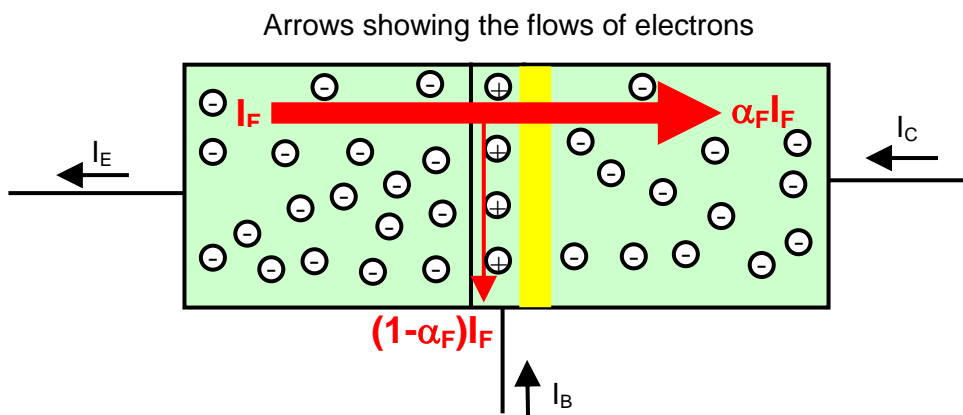
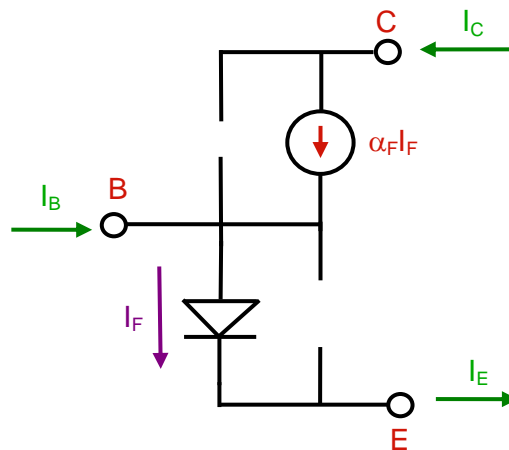


A BJT has four modes of operation.

- *First mode of operation:* The transistor is in the cut-off mode when $V_{BE} < 0.7 \text{ V}$ and $V_{BC} < 0.7 \text{ V}$. In such case, we have $I_F = I_R = 0$, which leads to $I_B = I_C = I_E = 0$.

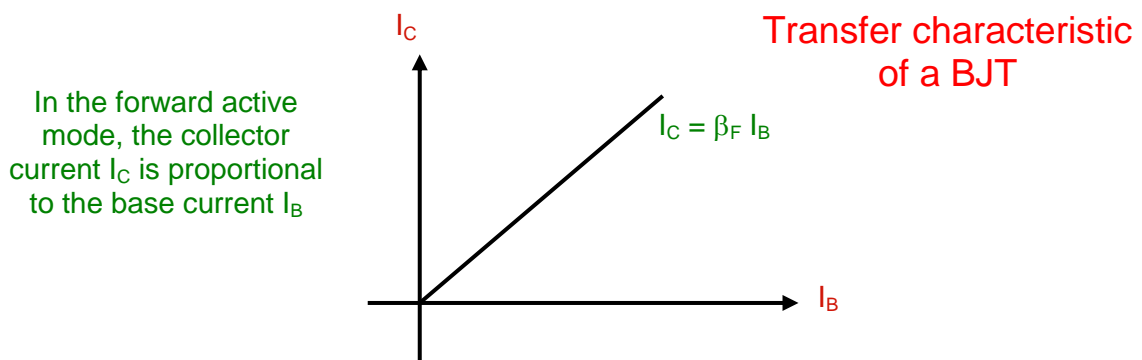


- *Second mode of operation:* The transistor is in the forward active mode when $V_{BE} \approx 0.7$ V and $V_{BC} < 0.7$ V (thus implying $V_{CE} > 0$).



In the forward active mode, we have $I_R = 0$, which yields $I_B = (1 - \alpha_F) I_F$, $I_C = \alpha_F I_F$, and $I_E = I_F$. By combining those three expressions, we obtain $I_C = \alpha_F I_F = \frac{\alpha_F}{1 - \alpha_F} I_B = \beta_F I_B$.

The parameter β_F is known as the forward current gain. If we take $0.98 < \alpha_F < 0.998$, we have $49 < \beta_F < 499$. Hereafter, we will adopt the value $\beta_F = 100$.

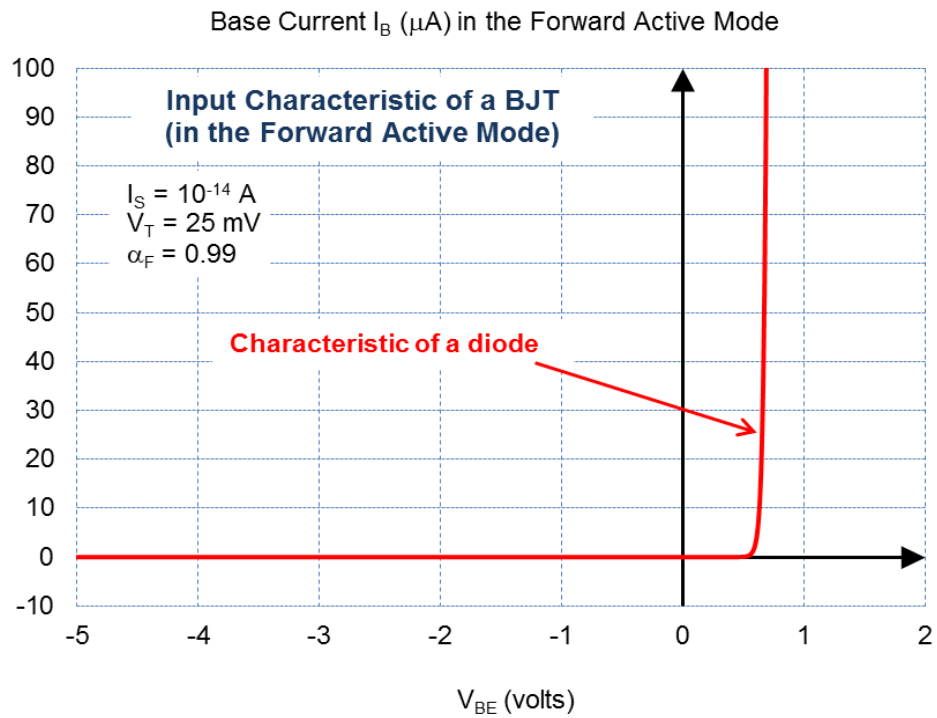


We can also notice that the emitter current is given by $I_E = I_F \approx I_S \exp\left(\frac{V_{BE}}{V_T}\right) = I_C + I_B = (\beta_F + 1) I_B$.

This result indicates that the base current varies exponentially with the voltage V_{BE} :

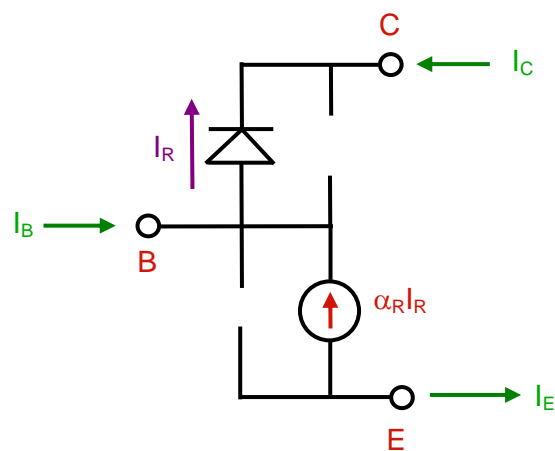
$$I_B = \frac{I_E}{\beta_F + 1} \approx \frac{I_S}{\beta_F + 1} \exp\left(\frac{V_{BE}}{V_T}\right) = I_S' \exp\left(\frac{V_{BE}}{V_T}\right).$$

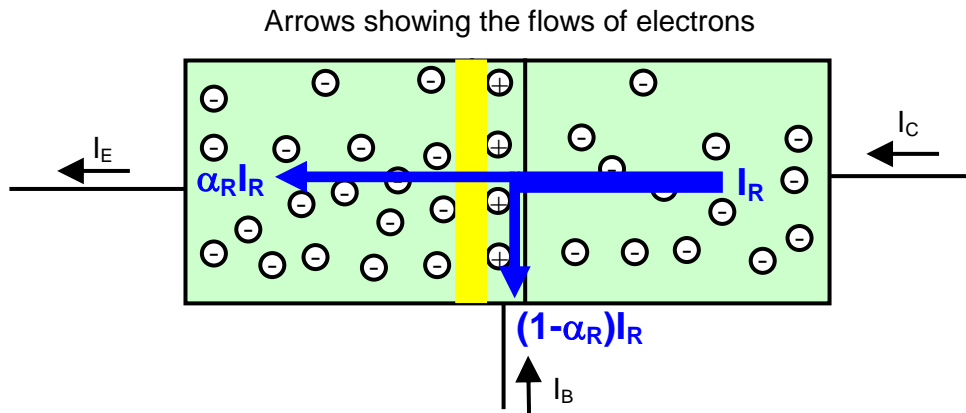
This equation linking I_B and V_{BE} provides us with the input characteristic of a BJT in the forward active mode. In fact, the equation corresponds to that of a diode as if the current I_B was the current flowing through the base-emitter junction.



The forward active mode is the mode used for designing amplifiers in analogue electronics.

- **Third mode of operation:** The transistor is in the reverse active mode when $V_{BE} < 0.7 \text{ V}$ and $V_{BC} \approx 0.7 \text{ V}$ (thus implying $V_{CE} = V_{CB} + V_{BE} = V_{BE} - V_{BC} < 0$).





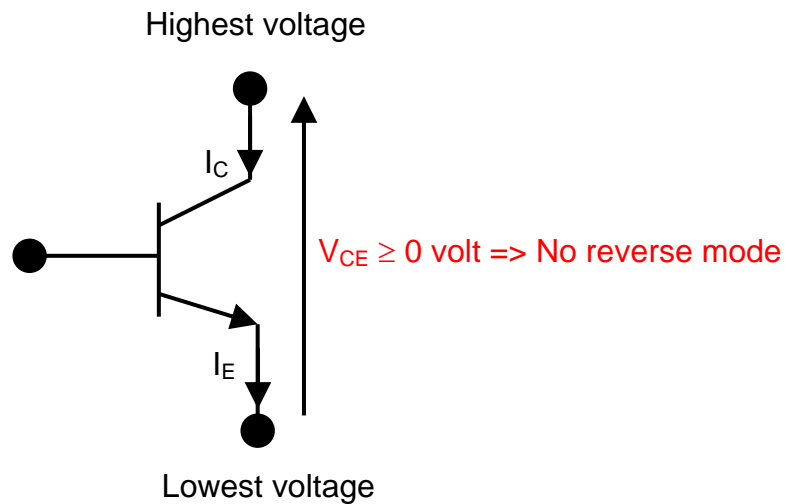
In the reverse active mode, we have $I_F = 0$, which yields $I_B = (1 - \alpha_R)I_R$, $I_C = -I_R$, and $I_E = -\alpha_R I_R = I_B + I_C$. By combining these three expressions, we obtain $I_C = -I_R = -\frac{I_B}{1 - \alpha_R}$ and

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

The parameter β_R is known as the reverse current gain. If we take $\alpha_R \approx 0.5$, we have $\beta_R \approx 1$.

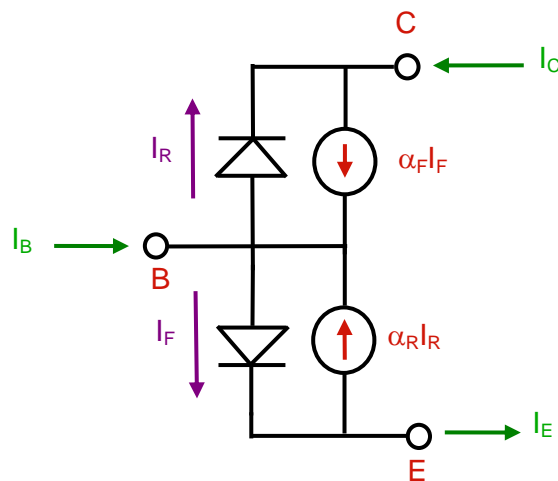
Since $\beta_R \ll \beta_F$, it is clear that the transistor effect obtained in the reverse active mode is much weaker than that achieved in the forward active mode. This is why the reverse active mode is of no particular interest in practice.

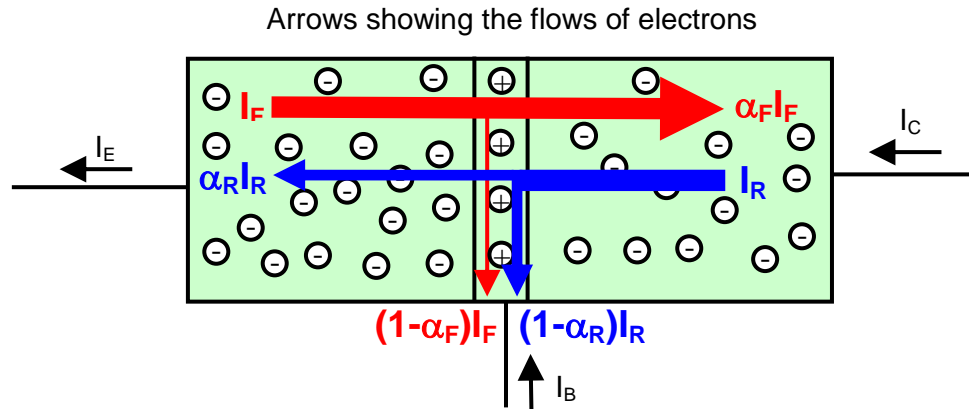
In EEE1002/EEE1010, we will never have to consider the reverse mode of operation in any of our circuits because the collector will always be on the side of the highest voltage whereas the emitter will be on the side of the lowest voltage, thus implying that we will always have $V_{CE} \geq 0$ volt (which contradicts the condition required for the reverse mode of operation).



- *Fourth mode of operation*: The transistor is in the saturation mode of operation when $V_{BE} \approx 0.7 \text{ V}$ and $V_{BC} \approx 0.7 \text{ V}$ (thus implying $V_{CE} \approx 0$).

In the saturation mode, the expressions for the base, collector, and emitter currents are complicated and, in fact, not very interesting. The most important thing to remember is that $V_{CE} \approx 0$ volt in this mode of operation.





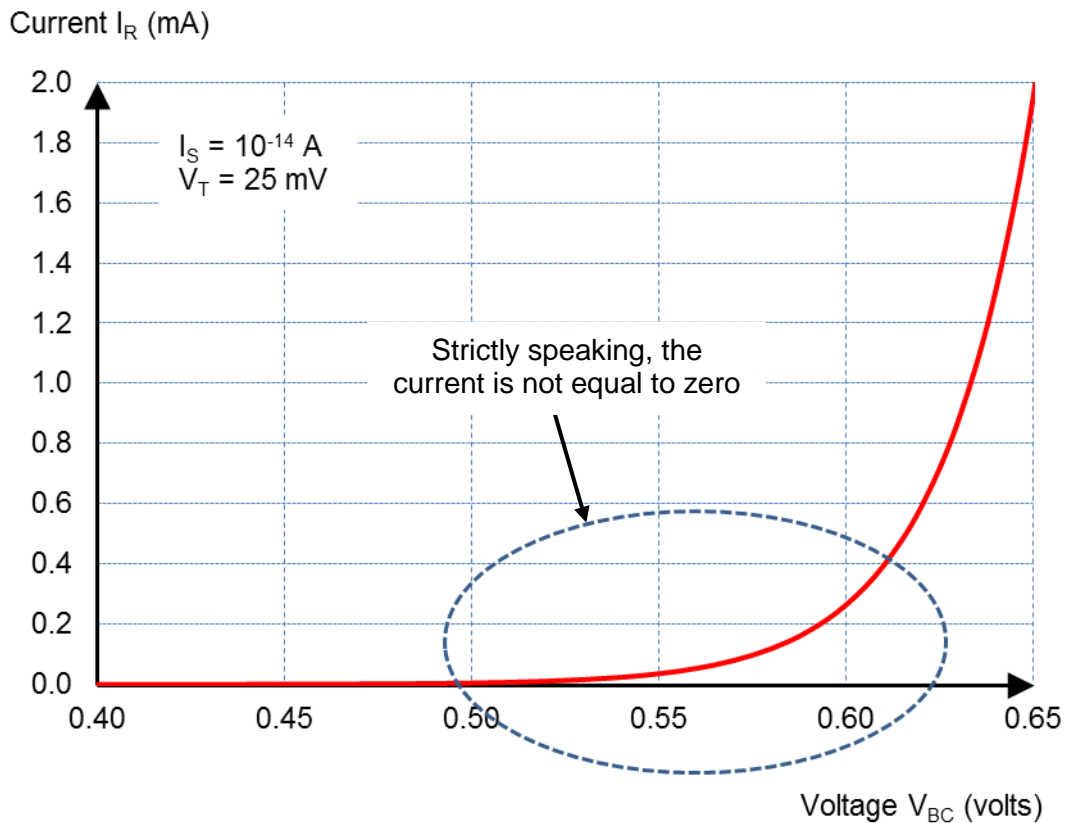
We now know enough to be able to understand the way the collector current I_C varies with the voltage V_{CE} when the BJT is either in the saturation mode or in the forward active mode, i.e. when $V_{BE} \approx 0.7$ volt.

When V_{CE} is close to zero, the BJT is in the saturation mode. In this case, we could show that a small increase in V_{CE} above 0 volt results in a very large increase in the collector current. This was not demonstrated earlier as it is of little interest to us.

In practice, the saturation mode corresponds to any value of the voltage V_{CE} ranging from 0 to roughly 0.2 volts.

This value of 0.2 volt can be explained as follows: Strictly speaking, the condition for the saturation mode is that a current does flow through the base-collector junction (i.e., $I_R \neq 0$). In a practical PN junction, a forward bias ranging from approximately 0.5 to 0.7 volt is often sufficient for the existence of a non-negligible current (see figure below).

In other words, the condition $V_{BC} > 0.5$ V can be considered sufficient for the BJT to be in the saturation mode. It is thus reasonable to say that, for $V_{CE} = V_{BE} - V_{BC} < 0.2$ V, the BJT is actually in the saturation mode.

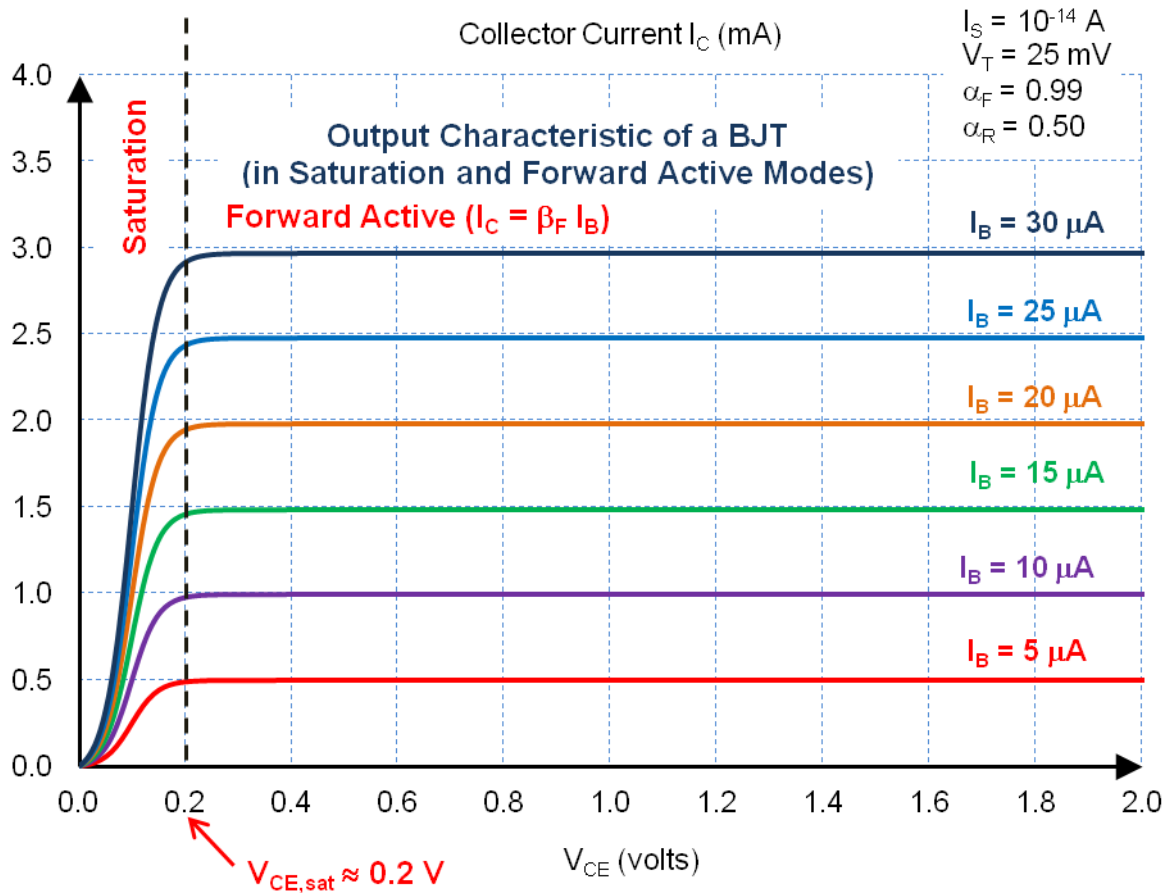


Once V_{CE} is increased beyond $V_{CE} \approx 0.2$ volt, i.e. $V_{BC} < 0.5$ volt, the current I_R is completely negligible, and the BJT is clearly in the forward active mode.

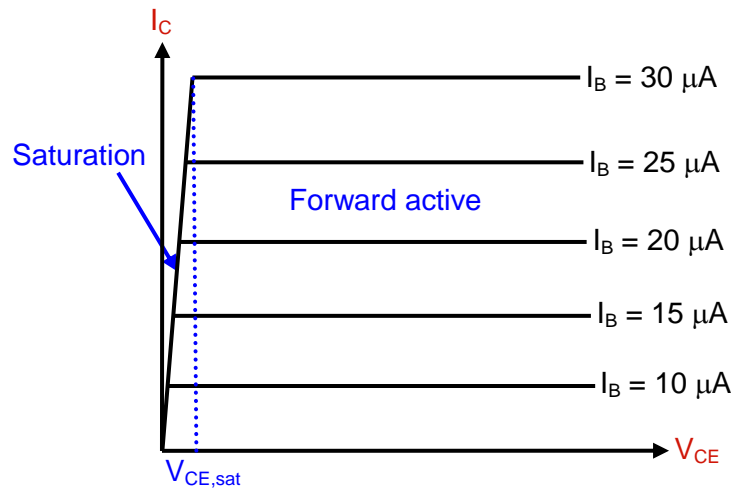
It may not always be easy to determine the exact value of the voltage V_{CE} in the saturation mode when performing the manual analysis of a circuit. However, we can make our life easier by simply assuming that, in the saturation mode, the voltage V_{CE} is a constant slightly greater than zero and called $V_{CE,sat}$.

Throughout these lecture notes, we will use the value $V_{CE,sat} \approx 0.2$ volt whenever the BJT is in the saturation mode of operation. This simplification does not result in any significant error as the actual value of V_{CE} always lies somewhere between 0 volt and $V_{CE,sat} \approx 0.2$ volt.

We finally obtain 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).

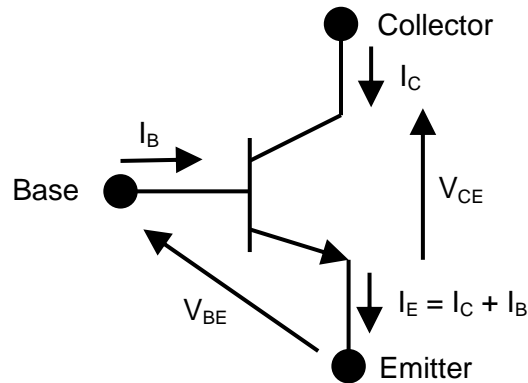


Output characteristic of a BJT

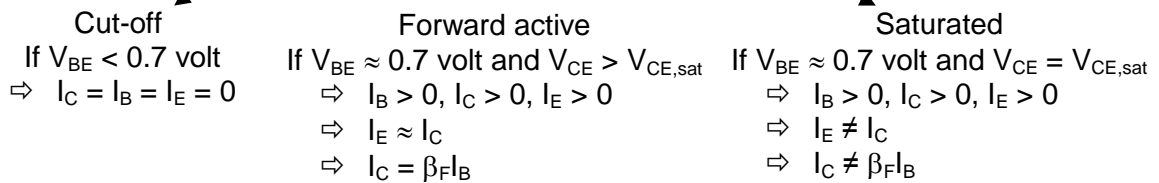


Tutorial 2 – Bipolar Junction Transistors

Throughout this Tutorial, we use silicon BJTs with $\beta_F = 100$ and $V_{CE,sat} \approx 0.2$ volt.

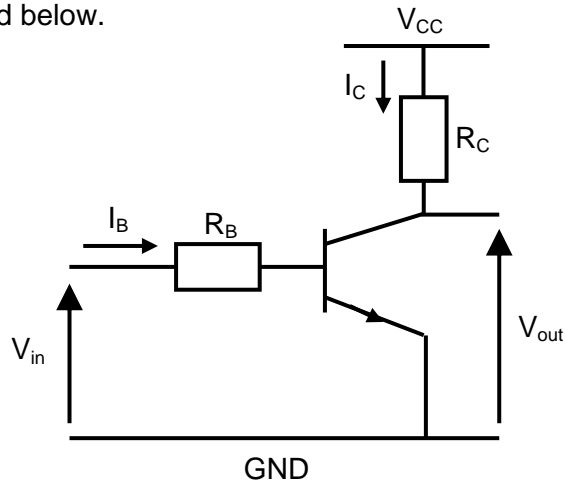


Three modes of operation



Question 1: The BJT as a Logic Inverter

Consider the circuit depicted below.



Find the DC transfer characteristic for this circuit. In other words, for each value of the input voltage V_{in} ranging from 0 to V_{CC} , determine the corresponding output voltage V_{out} .

Question 2: Design of a Linear Common-Emitter Amplifier

Modify the circuit studied in Question 1 in order to design a *linear common-emitter amplifier* with a *voltage gain* $A_v = -5$, using a power supply $V_{CC} = 10\text{ V}$.

The amplifier will be biased for maximum symmetrical voltage swing at both its input and output, and will therefore be suitable for applications where the input voltage $V_{in}(t)$ to be amplified has symmetrical positive and negative excursions.

In other words, the DC operating point will be positioned midway between $V_{CE,sat}$ and V_{CC} . This is ideal since it allows the output voltage $V_{out}(t) = A_v \cdot V_{in}(t)$ to swing by nearly $V_{CC}/2 = 5$ volts in either direction without any distortion.

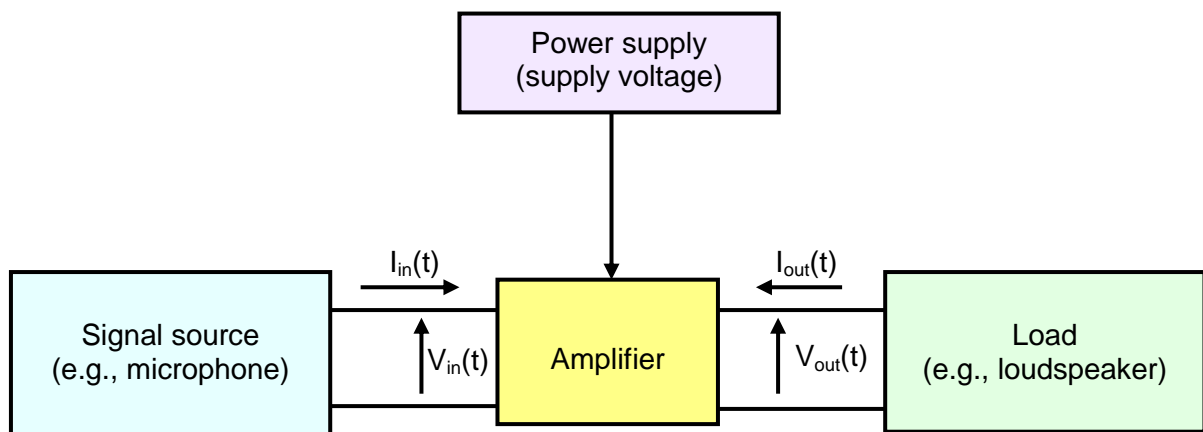
4. Linear Amplifiers

An analogue circuit that acts as an amplifier reproduces changes in its input signal as proportionately larger changes in its output signal. The term signal is used to denote the information-carrying fluctuations of a given voltage or current.

The amplification function performed by the circuit can be either linear or non-linear. If the amplification is linear, the output signal will be an amplified replica of the input signal (thus implying that no distortion will be introduced during the amplification process). If the amplification is non-linear, the output signal will be correlated to the input signal, but not an exact replica of it.

In EEE1002/EEE1010, we are only concerned with linear amplifiers for which the output signal is proportional to the input signal. Linear amplifiers are used to amplify an analogue input signal which can be either a voltage (voltage amplifier) or a current (current amplifier) or a power (power amplifier). Usually, both voltage and current amplifiers also increase the power of the signal and can thus be considered as power amplifiers.

Amplifiers are active circuits (unlike resistors, capacitors and diodes) since the output signal magnitude is greater than the input signal magnitude. Such a result can be obtained only when using active components such as transistors. The amplifier requires some form of power supply to enable it to boost the input signal.



The amplification produced by the circuit is described by its gains:

(1) Voltage gain: $A_v = \frac{V_{out}(t)}{V_{in}(t)}$,

(2) Current gain: $A_i = \frac{I_{out}(t)}{I_{in}(t)}$,

(3) Power gain: $A_p = \frac{P_{out}(t)}{P_{in}(t)} = \frac{V_{out}(t)I_{out}(t)}{V_{in}(t)I_{in}(t)} = A_v A_i$.

• Equivalent Model of an Amplifier

In order for the amplifier to perform some useful function, something (the source) must be connected to the input to provide an input signal and something (the load) must be connected to the output to make use of the output signal.

An ideal amplifier would always give an output signal that is determined only by the input signal and gain, irrespective of what is connected to the output. Also, an ideal amplifier would not affect the signal produced by the source. In fact, real amplifiers do not fulfil these requirements. This is why it is necessary to introduce the concepts of input and output resistances of an amplifier. We will see that the values of these resistances can strongly affect the overall gain when several circuits/amplifiers are connected to each other.

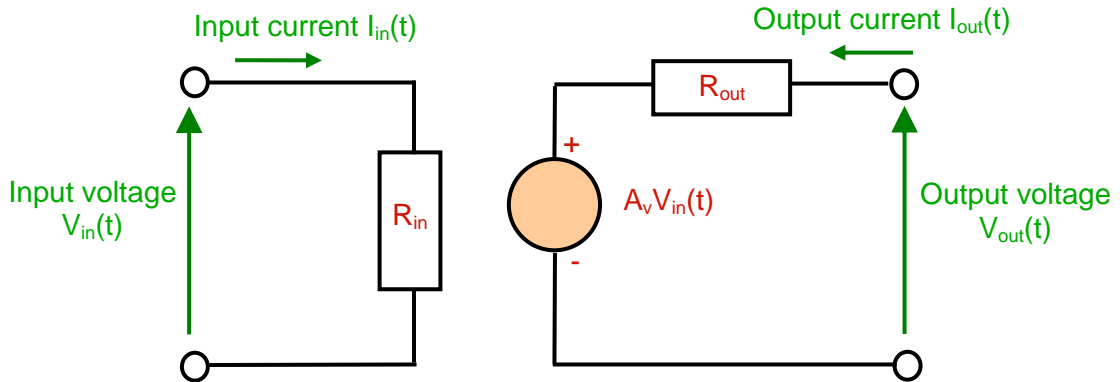
To model the characteristics of the amplifier input, we need to describe the way in which it appears to circuits that are connected to it. In other words, we need to model how it appears when it represents the load of another circuit. In most cases, the input circuitry of an amplifier can be modelled adequately by a single fixed resistance R_{in} which is termed *input resistance*.

To model the characteristics of the amplifier output, we need to describe the way in which it appears to circuits that are connected to it. In other words, we need to model how it appears when it represents a source to another circuit. In order to do so, we can use both Thévenin and Norton's theorems: Any circuit, no matter how complicated it is, can be seen as a voltage/current source in series/parallel with a resistance R_{out} which is termed *output resistance*.

• **Voltage Amplifiers**

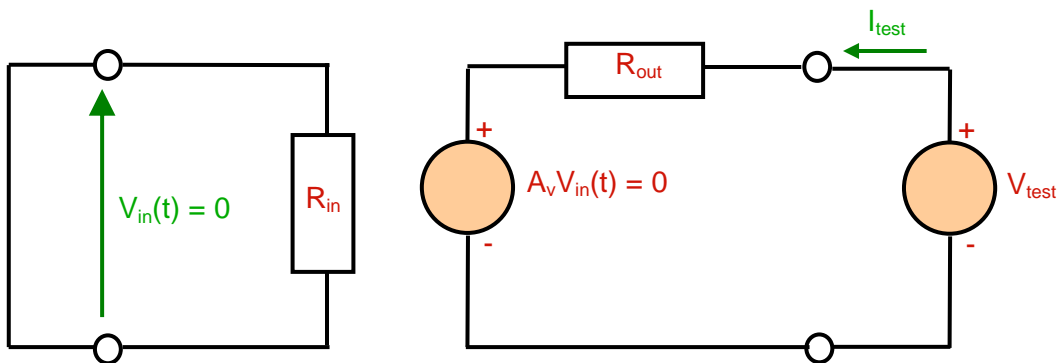
Any voltage amplifier, no matter how complicated it is, can be replaced inside a larger circuit by the following black box defined by its three parameters A_v , R_{in} , and R_{out} .

A_v is the voltage gain measured between the two output terminals when the load of the amplifier is an open circuit ($I_{out}(t) = 0$). This is, in other words, the voltage gain of the amplifier in isolation. A_v is often referred to as unloaded voltage gain because it is the ratio of the output voltage to the input voltage in the absence of any loading effects.

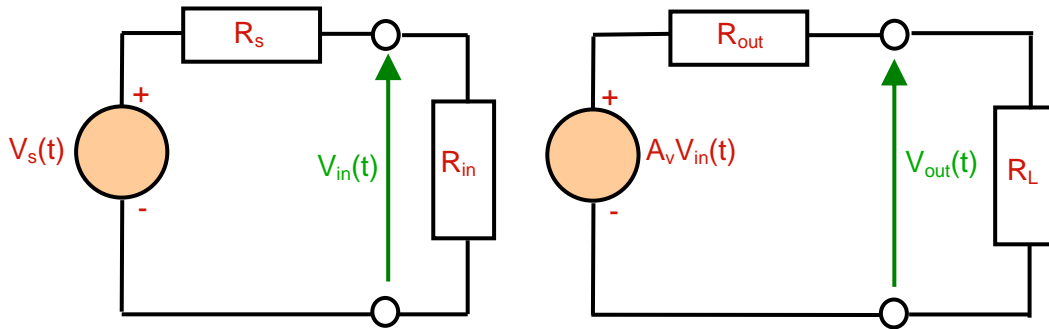


The input resistance R_{in} can be computed as $R_{in} = \frac{V_{in}(t)}{I_{in}(t)}$.

The output resistance R_{out} can be computed as $R_{out} = \frac{V_{test}}{I_{test}} \Big|_{V_{in}(t) = 0}$.



As an illustration, consider the circuit depicted below. The input of the amplifier is connected to a voltage source $V_s(t)$ with an output resistance R_s . The load resistance R_L , that represents the input resistance of the loading circuit, is connected to the output of the amplifier.



We can show that the voltage gain of the whole circuit is given by

$$\tilde{A}_V = \frac{V_{out}(t)}{V_s(t)} = A_V \frac{R_{in}}{R_{in} + R_s} \frac{R_L}{R_L + R_{out}} .$$

This expression clearly shows that, to maximise the overall gain \tilde{A}_V , one must ensure that R_{in} is as large as possible and R_{out} is as small as possible. In order to obtain $\tilde{A}_V = A_V$ (as one might have expected at first glance), we would need to have $R_{in} = +\infty$ and $R_{out} = 0$. These are the characteristics of an ideal voltage amplifier.

A “good” voltage amplifier has a large input resistance, a low output resistance, and of course a large unloaded voltage gain.

This example shows that, when an amplifier is connected to a source and a load, the resulting output voltage may be considerably less than one might have expected given the source voltage and the voltage gain of the amplifier in isolation.

The actual input voltage $V_{in}(t)$ to the amplifier is less than the source voltage $V_s(t)$ due to the effects of the voltage divider formed by R_s and R_{in} . Similarly, the output voltage of the circuit is affected by the load resistance R_L due to the effects of the voltage divider formed by R_L and R_{out} .

The lower the load resistance applied across the output of a circuit, the more heavily it is loaded as more current is drawn from it.

5. Design of Linear Amplifiers using BJTs

Linear amplifiers are analogue circuits. Proper operation of any analogue circuit requires that DC components (voltages or currents) are added to the voltages and currents inside this circuit. The DC components exist independently of any signal fluctuations and do not constitute signal information passing through the circuit.

The term signal is thus used to denote only the information-carrying fluctuations of a given voltage or current. Any fixed DC levels upon which such signals are superimposed are called *bias* components. The design and analysis of an analogue circuit generally requires that the total value of a voltage or current (signal + bias) be considered, even though only the signal component may be of interest.

Biasing is most often used in circuits that contain non-linear devices like diodes and transistors. When properly implemented, biasing causes these non-linear elements to behave as linear elements, thus greatly enhancing their usefulness.

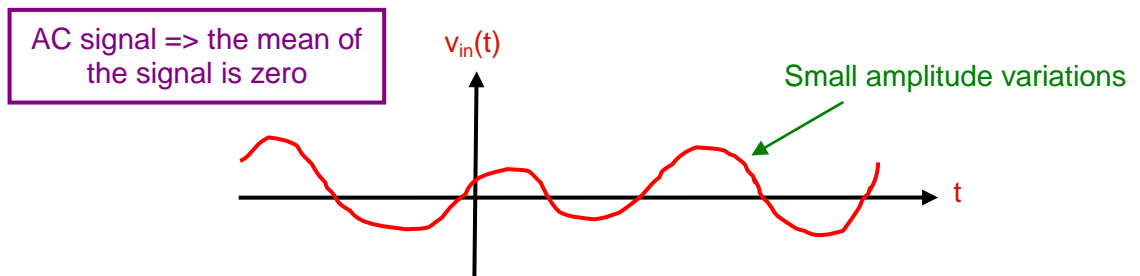
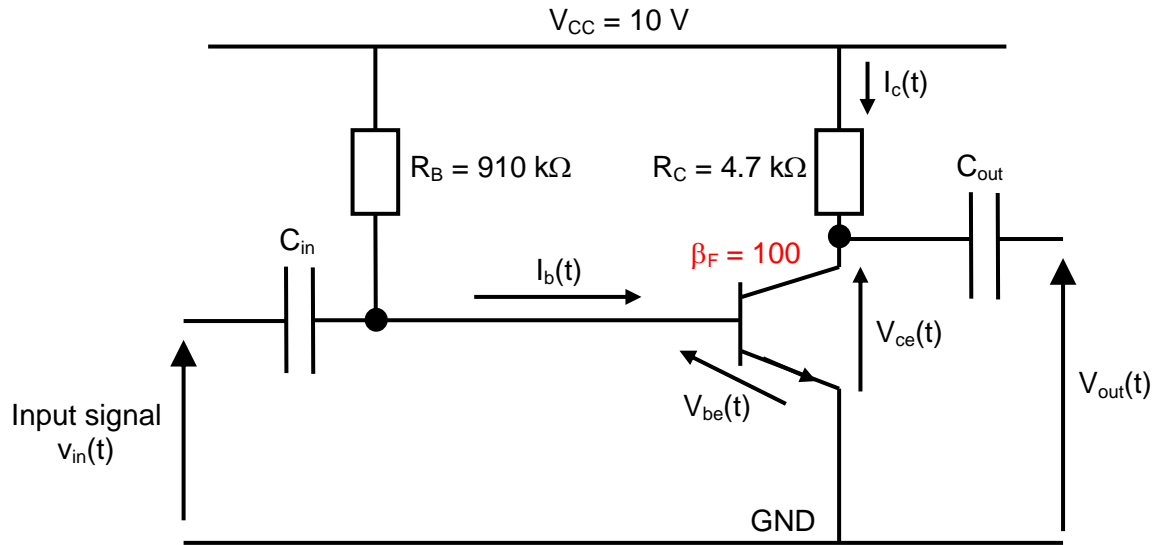
Although the various characteristics of most semi-conductor devices are non-linear, many exhibit linear behaviour over certain regions of operation. For instance, the transfer characteristic of a BJT can be described by the linear equation $I_C = \beta_F I_B$ in the forward active of operation.

The technique of biasing is employed by the circuit designer to confine a device's operating point to a region where its behaviour is linear (or at least approximately linear) while avoiding the gross non-linearities in the device's characteristics.

Throughout this Chapter, we consider the example of a common-emitter amplifier in order to show how to design and study linear amplifiers using bipolar junction transistors (BJTs).

• Study of a Common-Emitter Amplifier

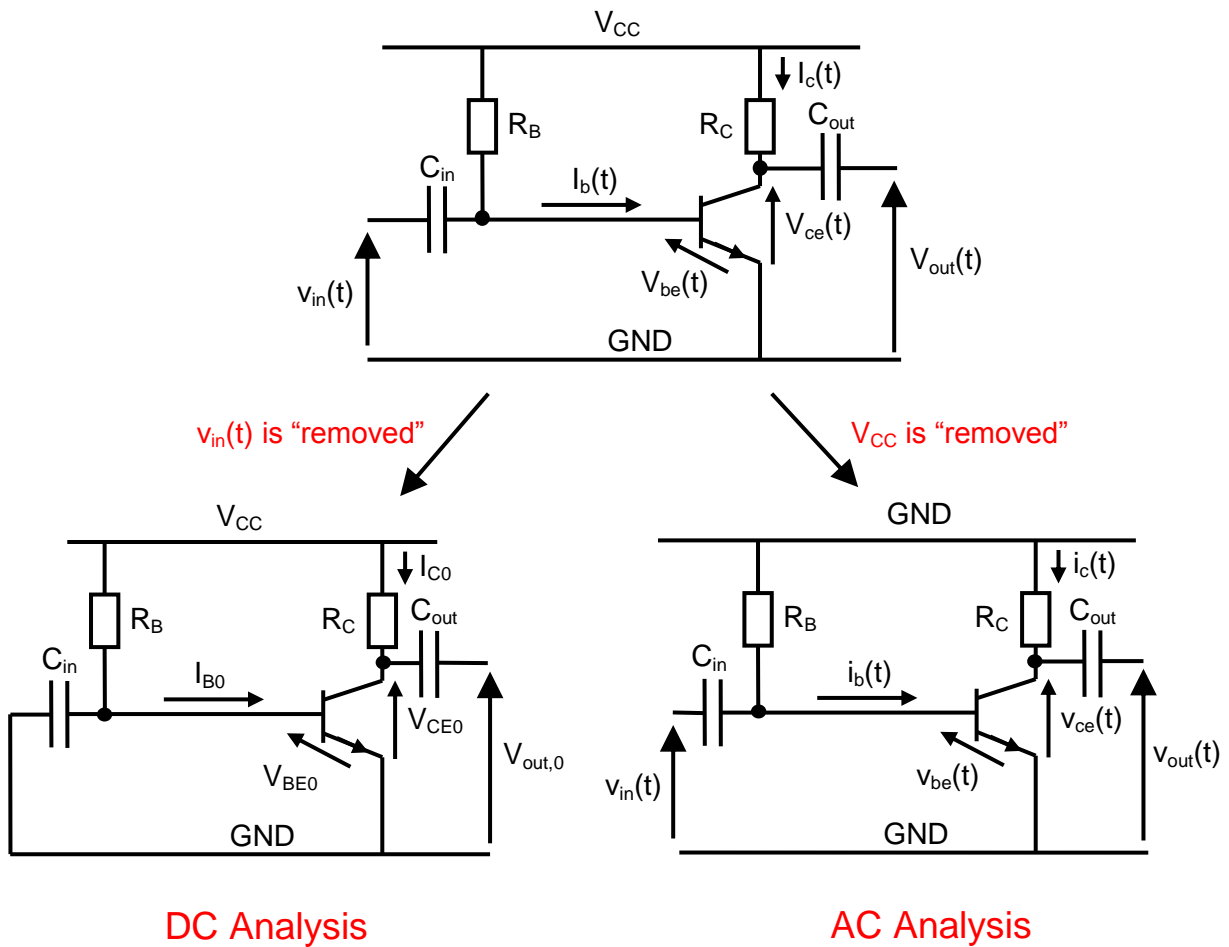
Consider the common-emitter amplifier below. As its name implies, this circuit can be used to amplify an input signal $v_{in}(t)$ that is represented by the (small) variation of an input voltage with time. For the time being, we assume that the mean of $v_{in}(t)$ is zero, i.e. $v_{in}(t)$ is an AC signal.



Note that we use two capacitors C_{in} and C_{out} , called *coupling capacitors*, in this circuit.

Apart from the AC signal $v_{in}(t)$ to be amplified, the common-emitter amplifier has another input: the supply voltage V_{CC} . It is the DC (constant) voltage source that provides the amplifier with the power necessary to boost the input voltage. It also allows the BJT to operate in the forward active mode of operation.

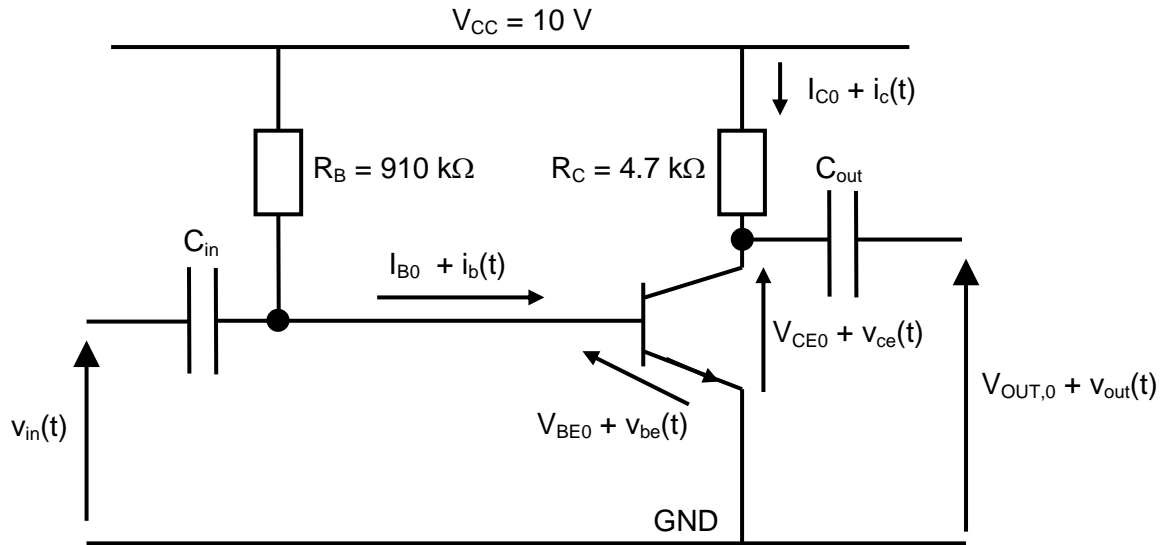
Since there are two inputs, we can use the *principle of superposition* to determine the expressions of the various currents and voltages in the circuit.



According to the principle of superposition, any signal $X(t)$ in the circuit is therefore the sum of a DC component X_0 (mean of $X(t)$) and an AC component $x(t)$, i.e. $X(t) = X_0 + x(t)$.

The quantity X_0 is determined by the DC analysis of the circuit, whereas the expression of $x(t)$ is determined by the AC analysis of the circuit. For instance, we can write:

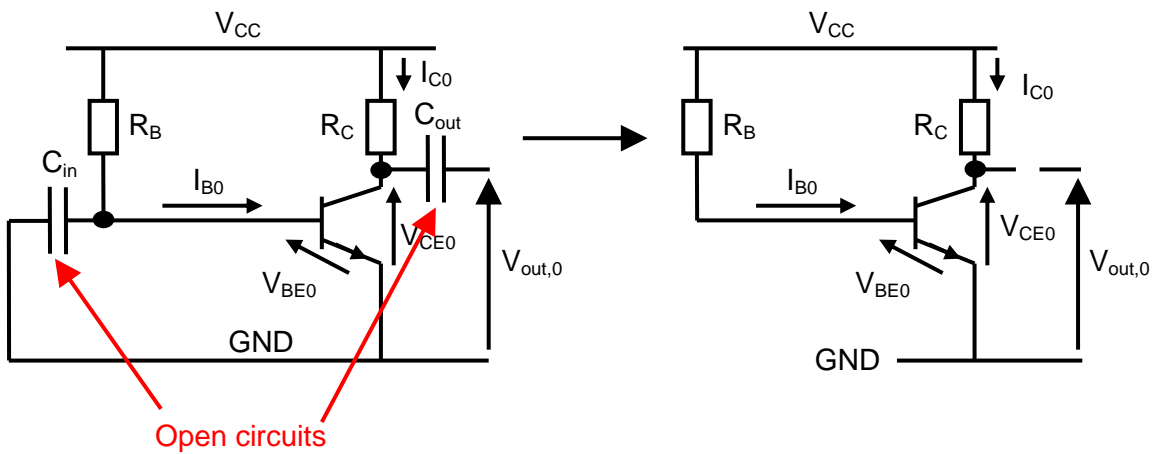
- $I_b(t) = I_{B0} + i_b(t)$ and $I_c(t) = I_{C0} + i_c(t)$;
- $V_{be}(t) = V_{BE0} + v_{be}(t)$ and $V_{ce}(t) = V_{CE0} + v_{ce}(t)$;
- $V_{out}(t) = V_{OUT,0} + v_{out}(t)$.



• DC analysis of our Common-Emitter Amplifier

We are now going to perform the DC analysis of our common-emitter amplifier. The main purpose of the DC analysis is to determine whether or not the BJT is properly biased.

Bipolar transistors have four different modes of operation. A BJT must be and always remain in the forward active mode when used to design an amplifier, whatever the value of the input signal to be amplified. Biasing an amplifier circuit consists of connecting the BJT so that it operates in the forward active mode.



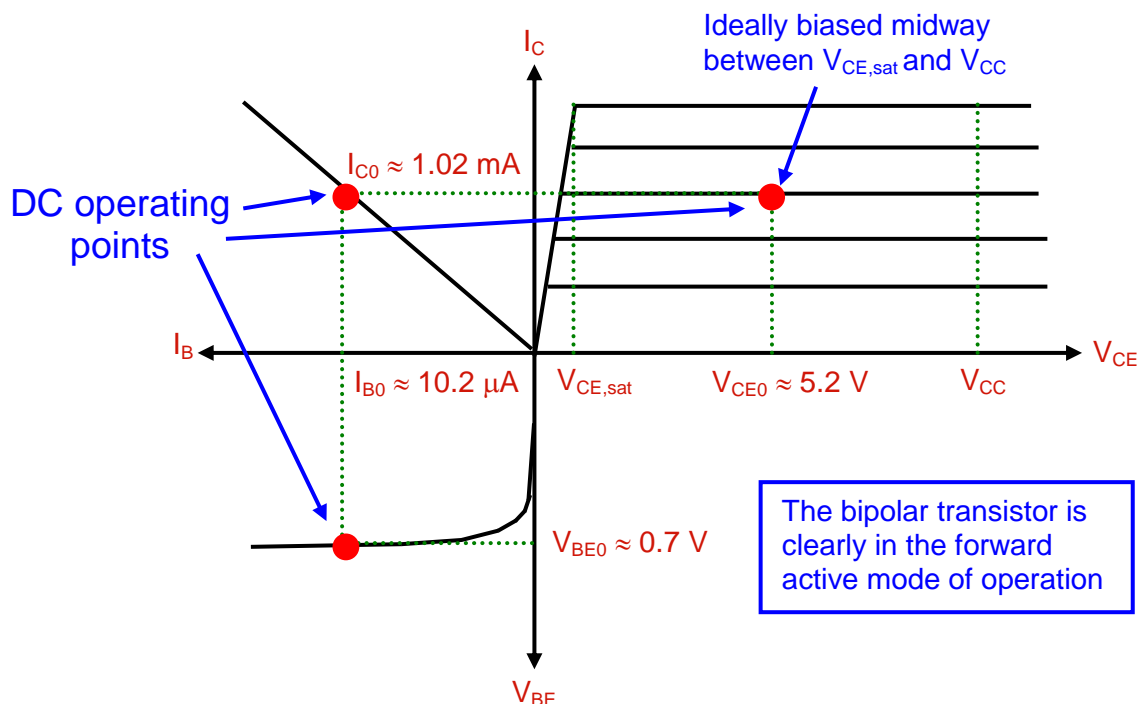
The base-emitter junction is on. We have $V_{BE0} \approx 0.7 \text{ V}$. Therefore, the base current is given by

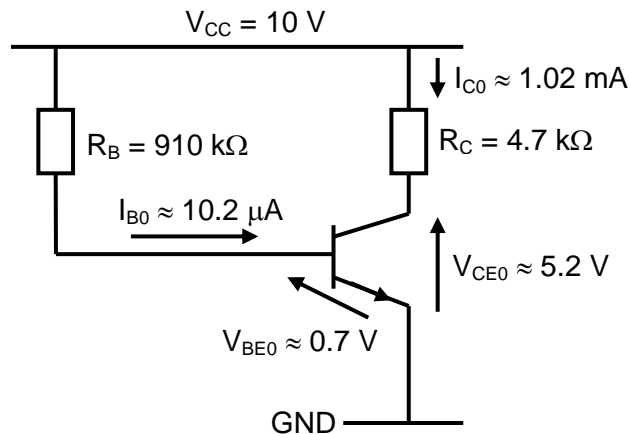
$$I_{B0} \approx \frac{V_{CC} - 0.7}{R_B} \approx 10.2 \mu\text{A}.$$

Assume now that the BJT is in the forward active mode (we actually do not know it for the time being, we will check later whether we were right or not). We also assume that the forward current gain of the BJT is given by $\beta_F = 100$.

We can thus write $I_{C0} = \beta_F I_{B0} \approx 1.02 \text{ mA}$. Then, we can compute the value of the voltage V_{CE0} as $V_{CE0} = V_{CC} - R_C I_{C0} \approx 5.2 \text{ V}$.

Since $V_{CE0} > V_{CE,sat}$, we conclude that our assumption was right and the BJT is indeed in the forward active mode.

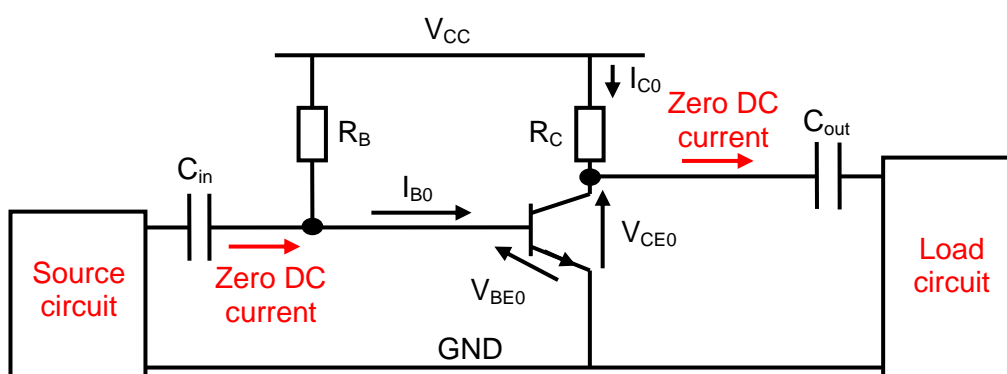




• **Role Played by the Coupling Capacitors in DC Operation**

The presence of both capacitances C_{in} and C_{out} protects the biasing circuit from any external influence. In other words, any change in the external environment of the amplifier circuit does not have any effect on the DC operating points. This is simply due to the fact that no DC current can flow through a capacitance.

To clarify this point, let us assume that our common-emitter amplifier circuit is connected to a source circuit and a load circuit, as shown below. Note that this will always be the case in practice.



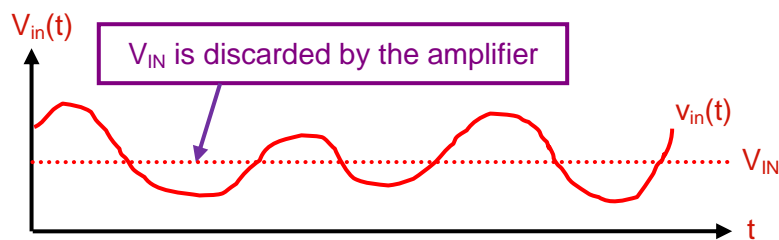
It clearly appears that, thanks to the presence of both coupling capacitors, the DC currents I_{B0} and I_{C0} as well as the DC voltages V_{BE0} and V_{CE0} are independent of the parameters of the source and

load circuits. This means that the DC operating points of our amplifier are not affected by those parameters.

If no coupling capacitors were used, the DC currents I_{B0} and I_{C0} as well as the DC voltages V_{BE0} and V_{CE0} would depend on the parameters of the source and load circuits, which is unacceptable in most cases.

Unfortunately, there is a price to pay for the use of coupling capacitors. The presence of both coupling capacitances C_{in} and C_{out} puts a lower limit on the frequency of the signals that can be amplified. This implies, in particular, that DC signals, which are zero-frequency signals, cannot be amplified.

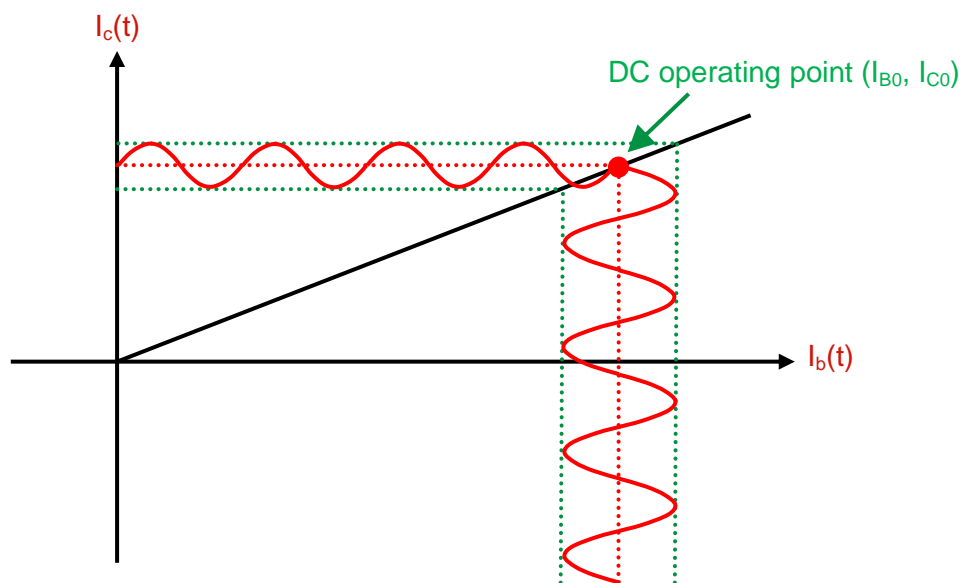
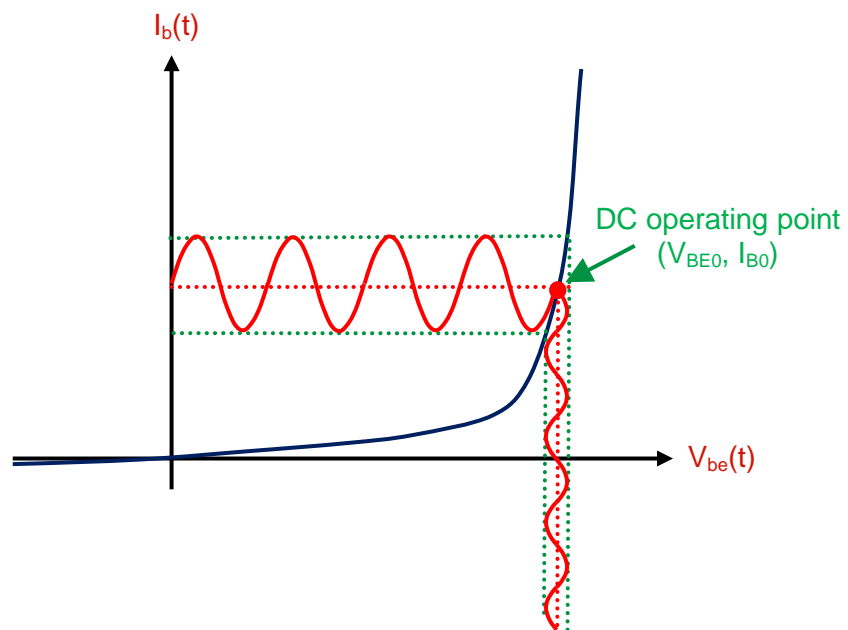
If there was a DC component in the input signal, it would be filtered out by C_{in} and ignored by the amplifier circuit as a result of this.

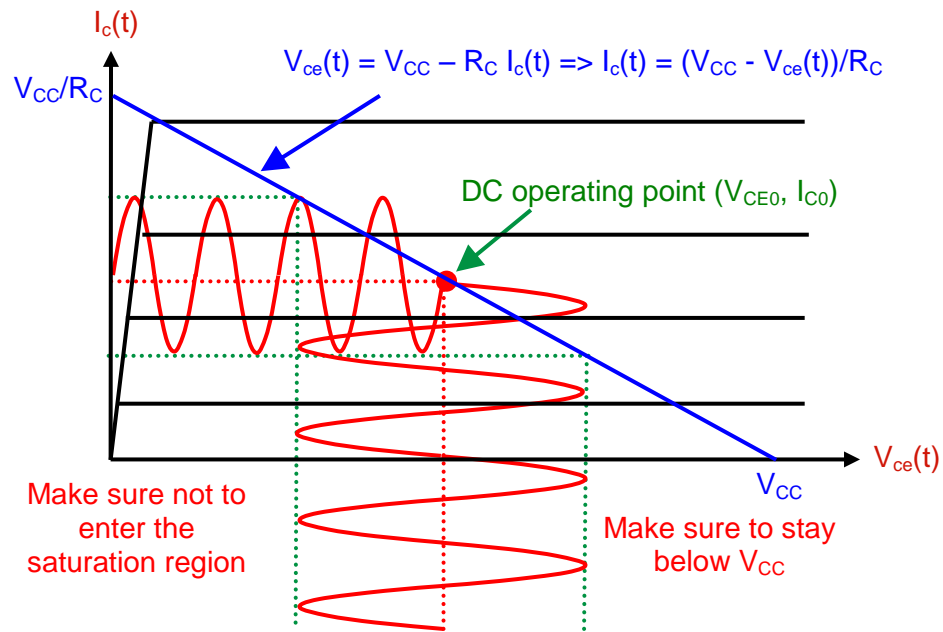


• AC analysis of our Common-Emitter Amplifier

Our bipolar transistor is properly biased and can therefore be used to amplify an input voltage $v_{in}(t)$. We assume that $v_{in}(t)$ is an AC signal with small amplitude variations. As we have just seen, if the input signal also contained a DC component, the latter would be filtered out by the coupling capacitance C_{in} and would therefore be ignored by the amplifier circuit.

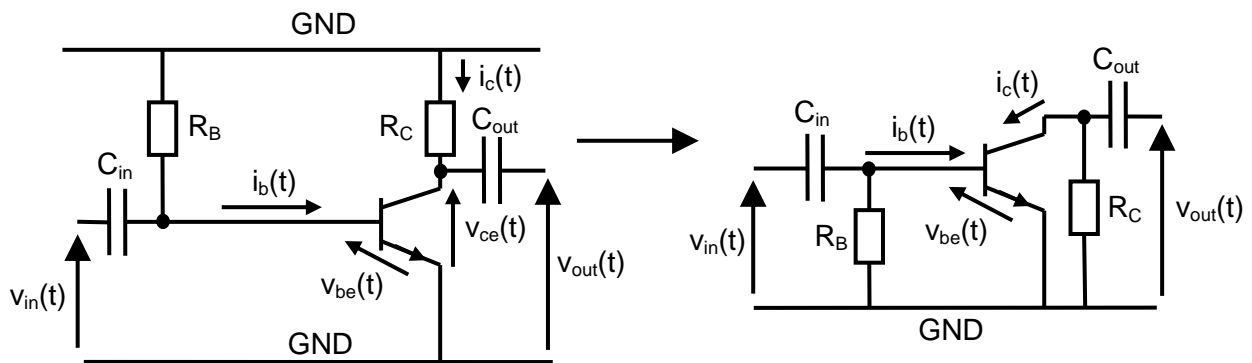
Any fluctuations of the AC signal $v_{in}(t)$ will induce corresponding AC variations of the various voltages and currents inside the circuit around their DC values, as illustrated by the three examples shown below.





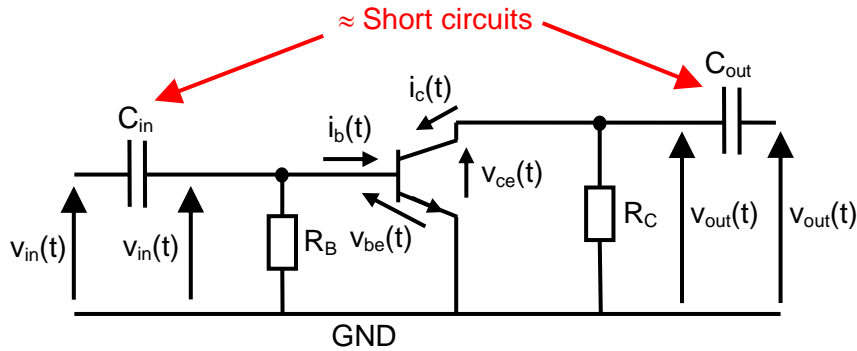
The magnitude of $v_{in}(t)$ must be sufficiently small so that the variation of $V_{ce}(t)$ around V_{CE0} allows the transistor to remain in the forward active mode of operation.

For the AC analysis of our common-emitter amplifier, we need to study the circuit depicted below.



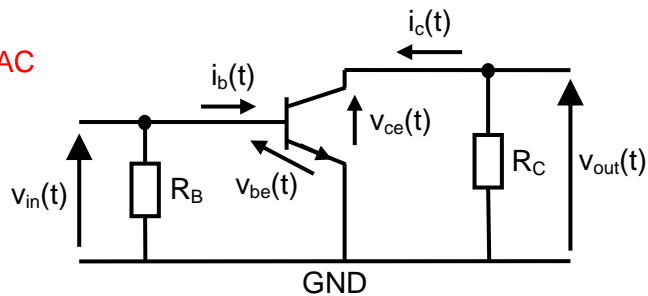
It is possible to further simplify the circuit by getting rid of the coupling capacitances C_{in} and C_{out} . To do so, we can assume that each coupling capacitance has a sufficiently high value to be considered as a short circuit (perfect wire) at the frequency of the input signal.

In other words, we can assume that both AC input and output signals $v_{in}(t)$ and $v_{out}(t)$ are passed through the capacitances without being affected at all.



Impedance of a capacitance C at a given frequency ω :
 $Z = 1/jC\omega$
 If $C\omega$ is "large enough", $Z \approx 0$

We need to perform the AC analysis of this circuit:

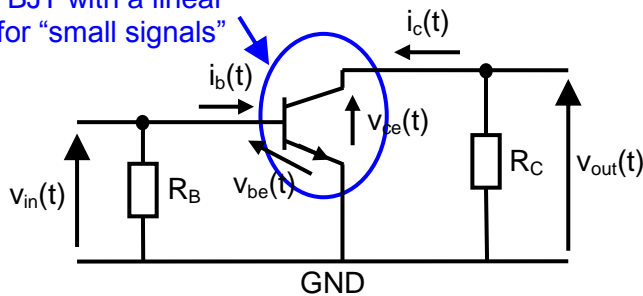


The voltage gain of our common-emitter amplifier is defined as $A_v = \frac{v_{out}(t)}{v_{in}(t)}$.

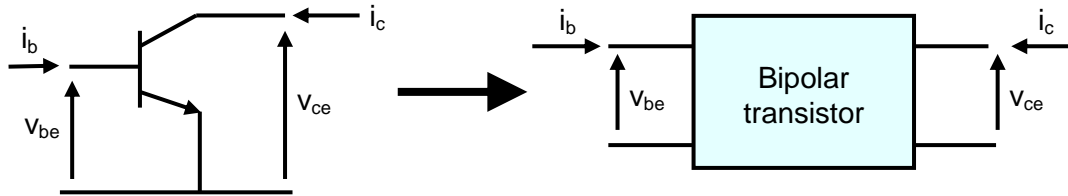
• **Small-Signal Model of a BJT**

In order to perform the AC analysis of the amplifier, we need to replace the BJT with an equivalent model valid for a small-amplitude input signal ("small signal"). Obviously, this model is derived assuming that the BJT is in the forward active mode of operation.

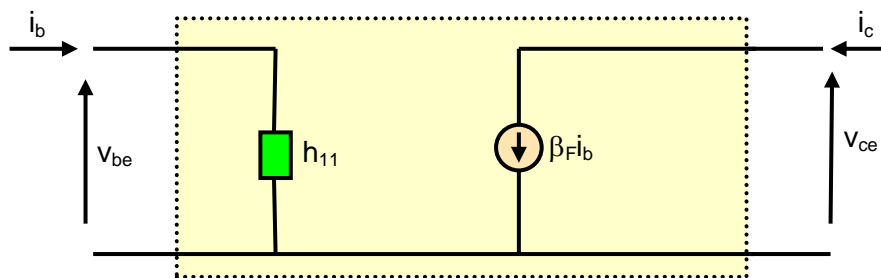
Replace the BJT with a linear model valid for "small signals"



The bipolar transistor is now seen as a “black box” with two inputs and two outputs.



We can easily show that the (generic) small-signal model of a BJT is given below:



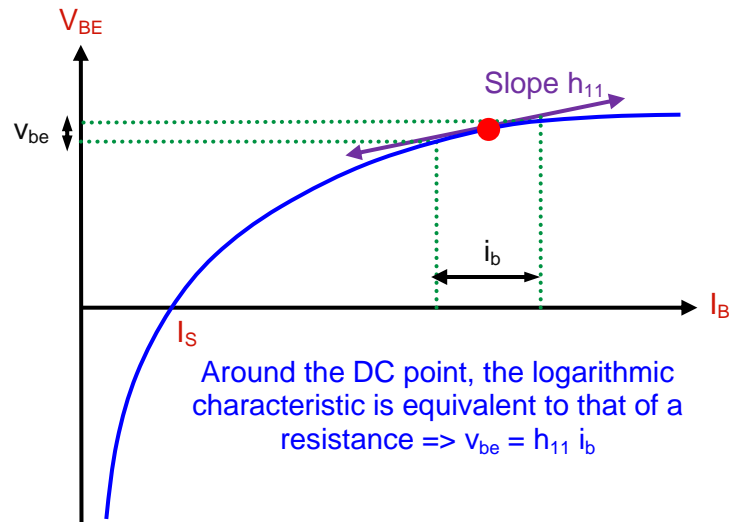
The presence of a resistance h_{11} implies that the relation between v_{be} and i_b is linear, which is not the case (the base current is in fact an exponential function of the base-emitter voltage). However, remember that the quantities v_{be} , v_{ce} , i_b , and i_c represent AC signals, i.e. fluctuations of the corresponding voltages and currents around the DC operating points. The assumption that these fluctuations are “small” allows us to approximate the expressions linking those quantities using linear functions. After all, any mathematical function can be replaced by a straight line over a sufficiently small interval.

The expression of the resistance h_{11} is thus obtained as follows:

$$h_{11} = \frac{v_{be}}{i_b} = \left. \frac{dV_{BE}}{dI_B} \right|_{V_{BE0}, I_{B0}}$$

The equation of the base-emitter junction characteristic is given by $I_B \approx I_S \cdot \exp\left\{\frac{V_{BE}}{V_T}\right\}$ (remember the input characteristic of a BJT).

This expression is equivalent to $V_{BE} \approx V_T \cdot \ln\left(\frac{I_B}{I_S}\right)$.



Therefore, we can write $h_{11} = \left. \frac{dV_{BE}}{dI_B} \right|_{V_{BE0}, I_{B0}} \approx \left. \frac{V_T}{I_B} \right|_{V_{BE0}, I_{B0}} = \frac{V_T}{I_{B0}}$.

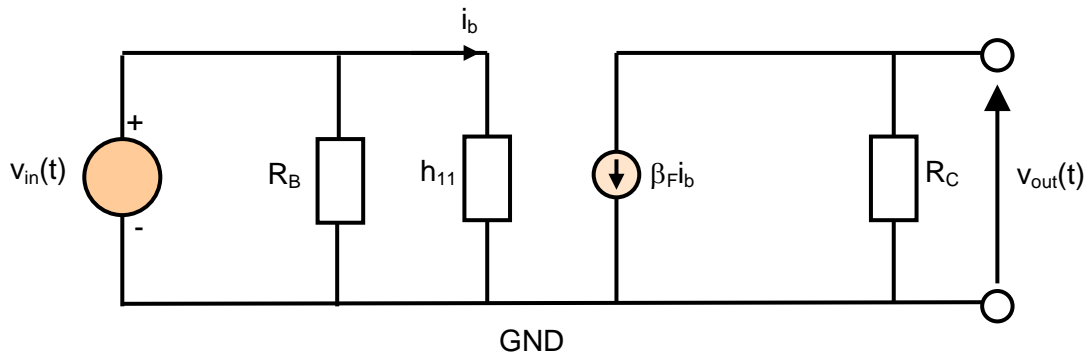
At room temperature, the expression of the parameter h_{11} is thus given by $h_{11} \approx \frac{25 \text{ mV}}{I_{B0}}$.

As for the common-emitter amplifier considered throughout this chapter, we have $h_{11} \approx 2.45 \text{ k}\Omega$ (at room temperature) since $I_{B0} \approx 10.2 \mu\text{A}$.

• Small-Signal AC Analysis of our Common-Emitter Amplifier

We can now replace the bipolar transistor with its small-signal model in order to determine the expressions of the voltage gain, input resistance, and output resistance of our common-emitter amplifier.

We recall the values of the circuit parameters: $R_B = 910 \text{ k}\Omega$, $h_{11} \approx 2.45 \text{ k}\Omega$, $\beta_F = 100$, $R_C = 4.7 \text{ k}\Omega$.



• Voltage Gain

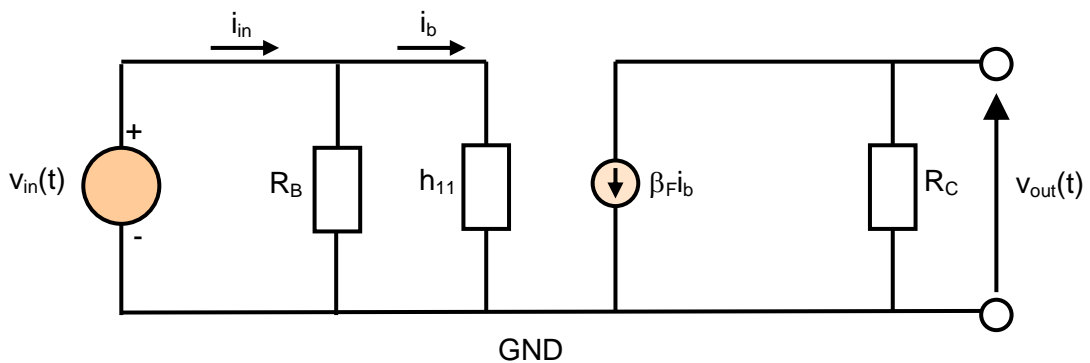
We can show that $v_{out}(t) = -R_C \beta_F i_b$. Therefore, we obtain $v_{out}(t) = -R_C \beta_F \frac{v_{in}(t)}{h_{11}}$.

The small-signal voltage gain of our common-emitter amplifier is thus given by

$$A_V = \frac{v_{out}(t)}{v_{in}(t)} = -\beta_F \frac{R_C}{h_{11}} \approx -192.$$

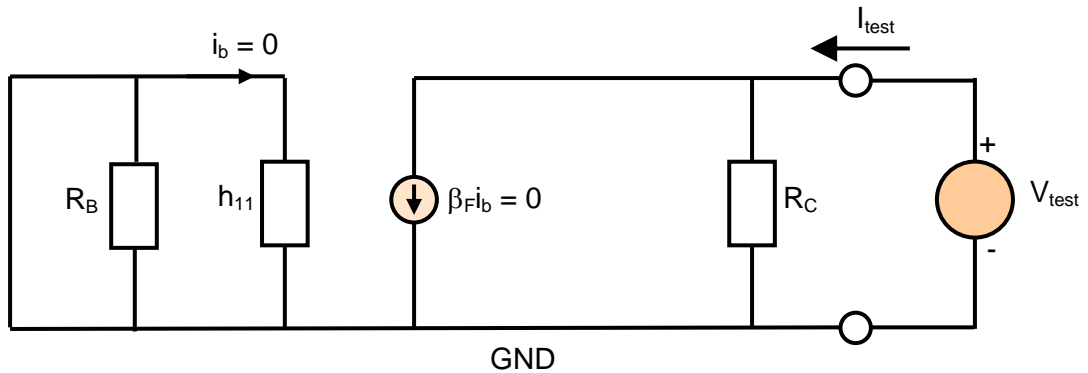
• Input Resistance

The input resistance is given by $r_{in} = \frac{v_{in}(t)}{i_{in}(t)} = \frac{R_B h_{11}}{R_B + h_{11}} \approx h_{11}$ since $h_{11} \ll R_B$. Its value is thus $r_{in} \approx 2.45 \text{ k}\Omega$.



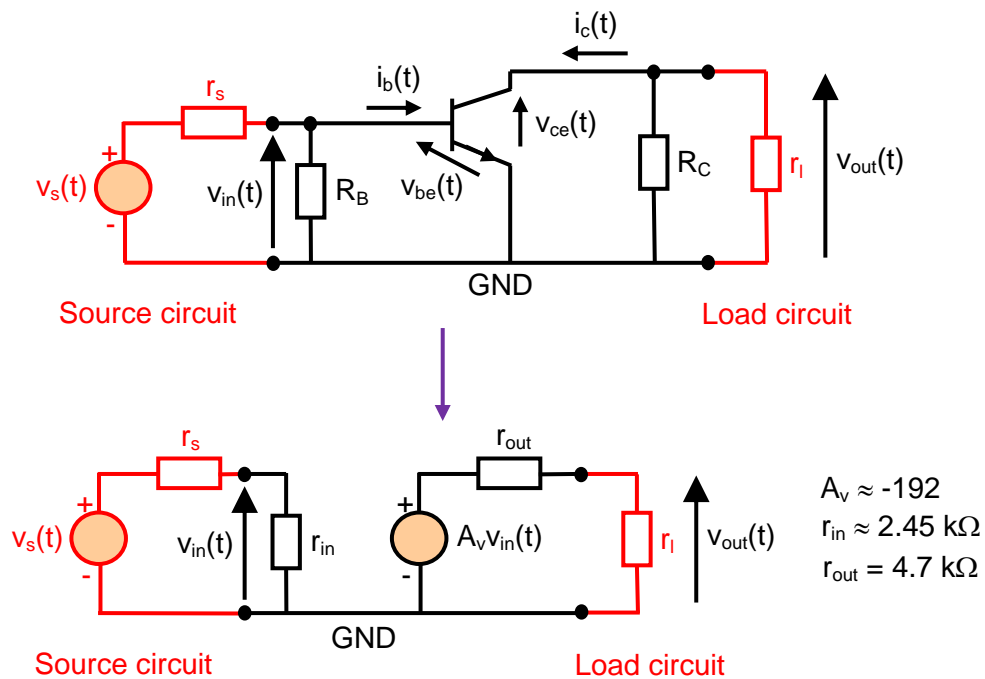
• Output Resistance

The output resistance is given by $r_{out} = \frac{V_{test}}{I_{test}} \Big|_{v_{in}(t)=0} = R_C = 4.7 \text{ k}\Omega$.



- What is the actual voltage gain when the amplifier is connected to other circuits?

In practice, the input of our common-emitter amplifier is connected to a voltage source $v_s(t)$ with an output resistance r_s , whereas its output is connected to a loading circuit that is represented by an input resistance r_l .



We can show that the small-signal voltage gain of the whole circuit is given by

$$\tilde{A}_v = \frac{v_{out}(t)}{v_s(t)} = A_v \frac{r_{in}}{r_{in} + r_s} \frac{r_l}{r_l + r_{out}}$$

Note that, if $r_{in} \ll r_s$ and/or $r_l \ll r_{out}$, we have $\tilde{A}_V \ll A_V$. This result indicates that, in order to avoid a drastic reduction in voltage gain when the amplifier is connected to other circuits, one should maximise its input resistance r_{in} and minimise its output resistance r_{out} .

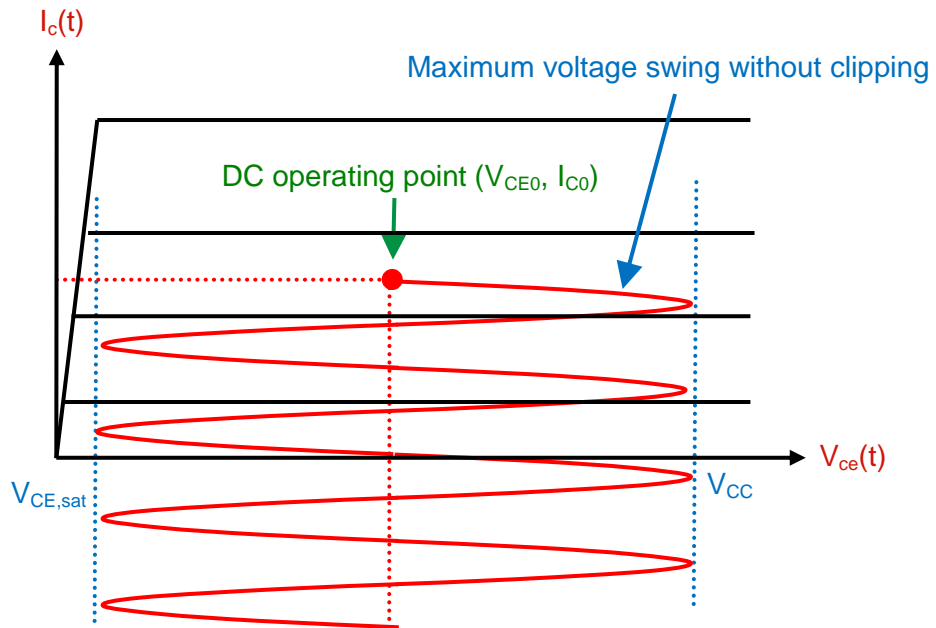
• Device Variability

The typical value of the forward current gain β_F for a general-purpose bipolar transistor is in the range 50 to 300. However, this current gain varies considerably with temperature, component aging, and operating conditions.

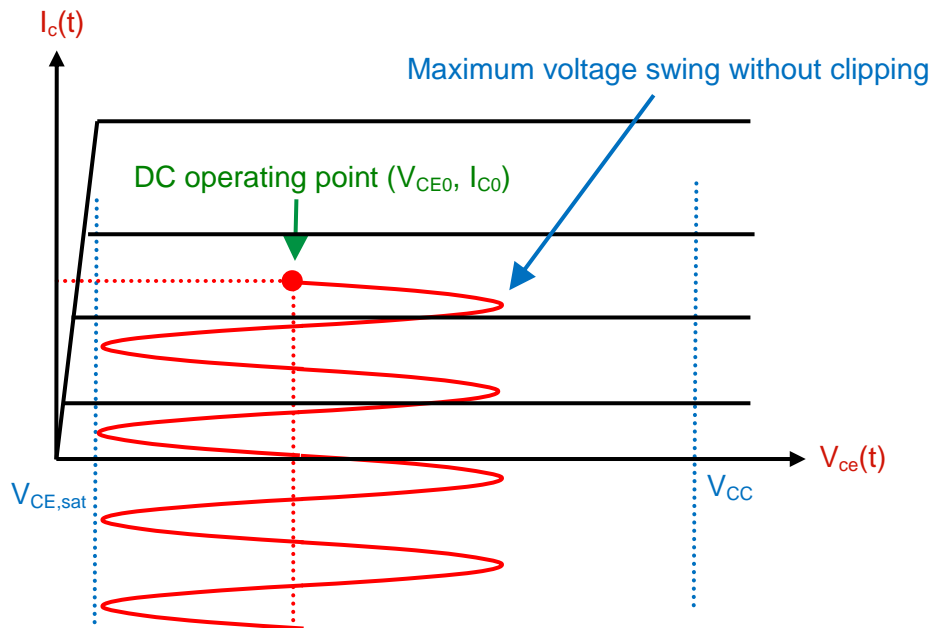
There is also a considerable spread of characteristics between devices of the same nominal type and even within the same batch. This leads to problems in the design of circuits using BJTs and greatly limits the usefulness of some amplifier circuits, such as the one we have just studied.

For instance, with our common-emitter amplifier, we have seen that the DC operating points and the small-signal voltage gain both depend on the value of β_F . The DC analysis of the amplifier was performed assuming $\beta_F = 100$, but this value may not be the exact value of the actual forward current gain of the BJT.

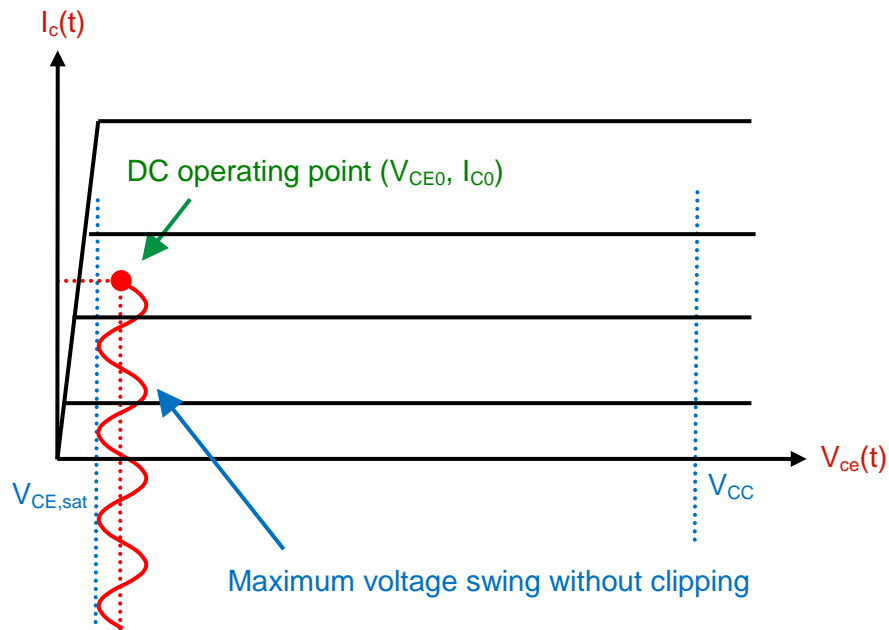
With $\beta_F = 100$, we found that $I_{C0} = \beta_F I_{B0} \approx 1.02 \text{ mA}$ and then $V_{CE0} = V_{CC} - R_C I_{C0} \approx 5.2 \text{ V}$, which leads to a DC operating point which is positioned midway between $V_{CE,sat}$ and V_{CC} . This is ideal since it allows the AC signal $v_{ce}(t)$ to swing by nearly $V_{CC}/2 = 5 \text{ volts}$ in either direction.



If β_F is actually equal to 150 instead of 100, we can easily see that $I_{C0} = \beta_F I_{B0} \approx 1.53 \text{ mA}$ and then $V_{CE0} = V_{CC} - R_C I_{C0} \approx 2.8 \text{ V}$, permitting a maximum negative excursion of approximately 2.5 V before the device enters saturation.



The extreme case is reached with $\beta_F = 200$ instead of 100. Here, the DC output voltage is only $V_{CE0} = V_{CC} - R_C I_{C0} \approx 0.41 \text{ V}$ which is at the edge of the saturation region, allowing no useful output swing in the negative direction.



For values of the forward current gain above 200, the BJT is driven hard into saturation and the circuit can no longer be used as an amplifier.

One apparent solution to this problem would be to measure the forward current gain of the transistor to be used and design the circuit appropriately. This is however not an attractive solution in a mass-production environment in which it would be impractical to design each circuit uniquely to match individual components. It would also cause severe problems when a device failed and needed to be replaced.

An alternative would be to select devices within a particular close range of parameters. This is possible but expensive, particularly if a very narrow spread is required.

The only practical solution to this problem is to design circuits that are not greatly affected by changes in the forward current gain of the transistors within them. There are techniques to do so, such as negative feedback for instance.

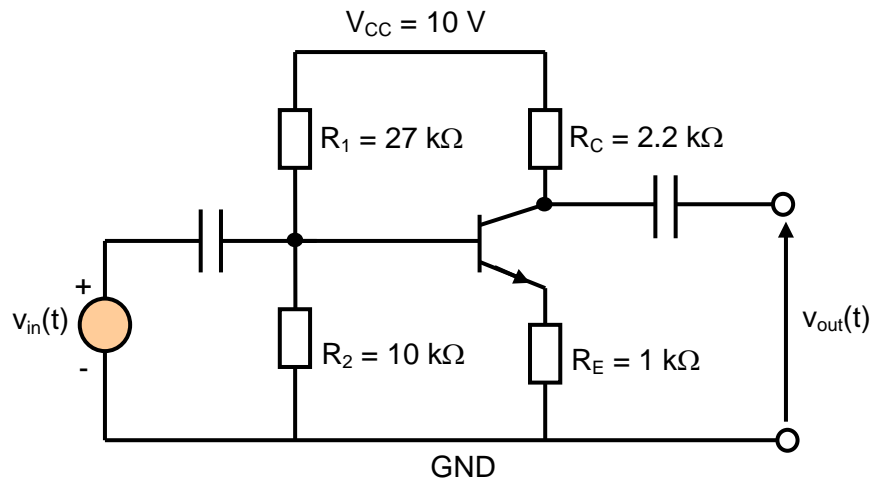
In Tutorials 3 and 4, we will study some amplifier circuits for which the biasing and/or the small-signal voltage gain do not depend on the value of β_F .

Tutorial 3 – Amplifier Circuits using BJTs (1)

Throughout this Tutorial, we use silicon BJTs with $\beta_F = 100$.

Question 1: Common-Emitter Amplifier with Negative Feedback

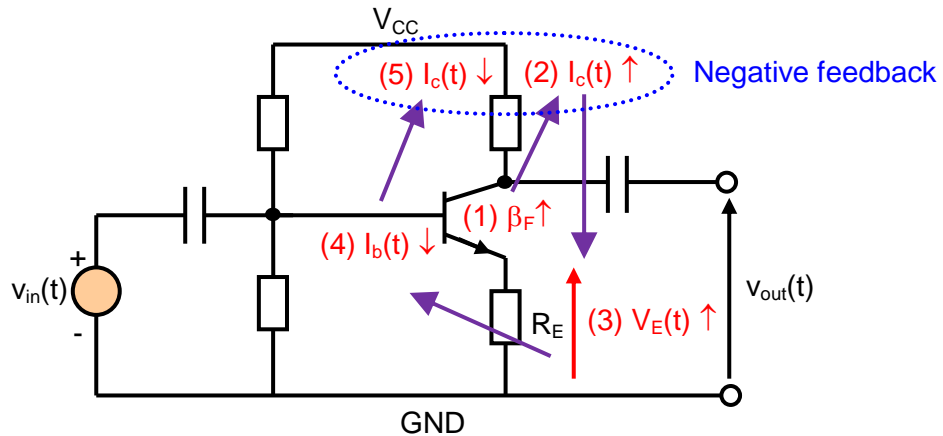
Consider the amplifier circuit depicted below.



The main difference with the amplifier circuit described in class is that we have added an emitter resistor R_E . The presence of this resistor makes various parameters of the circuit, such as the DC operating points and the small-signal voltage gain, more stable, i.e. less dependent on the transistor characteristics.

To understand the role played by R_E , let us consider a simple example. Assume that the collector current $I_c(t)$ increases for any reason (e.g., because of an increase in β_F). In this case, the voltage across R_E , which is proportional to $I_c(t)$ since $I_c(t) \approx I_e(t)$, also increases. This results in a decrease in the base current $I_b(t)$ that tends to cancel the initial increase in $I_c(t)$ since $I_c(t) = \beta_F I_b(t)$.

This process, called *negative feedback*, does provide stability to the operation of the circuit.

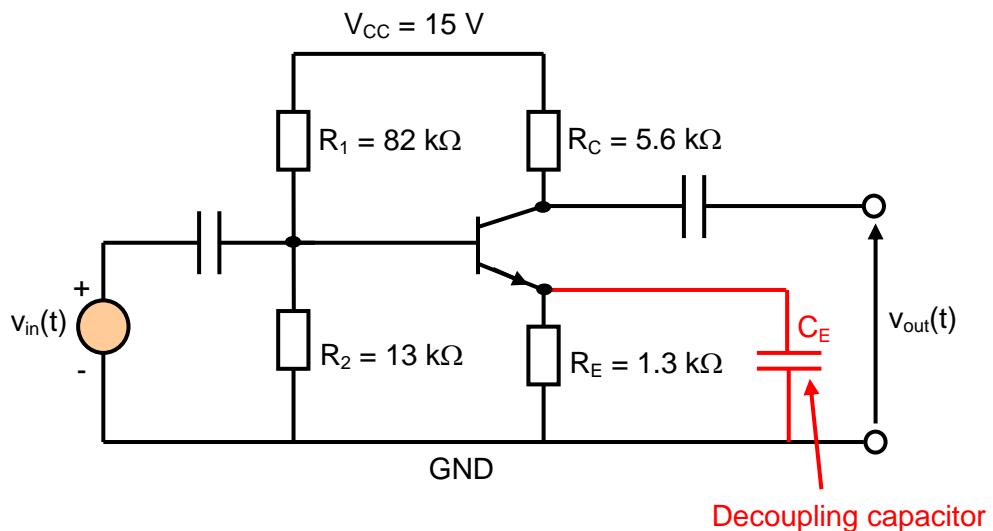


Perform the DC and AC analysis of this amplifier.

Compare your results with those you would obtain without negative feedback, i.e. by using the same circuit with no emitter resistance ($R_E = 0$). Note that, if $R_E = 0$, the amplifier considered here is a classical common-emitter amplifier similar to the one studied in the lecture notes.

Question 2: Common-Emitter Amplifier with Negative Feedback and Decoupling Capacitor

Consider the circuit depicted below.



In Question 1, we have seen that the use of negative feedback has several advantages. These include the stabilisation of circuit parameters (e.g., DC operating points) which become less affected by changes in the transistor's characteristics. However, this is achieved at the expense of a considerable reduction in voltage gain.

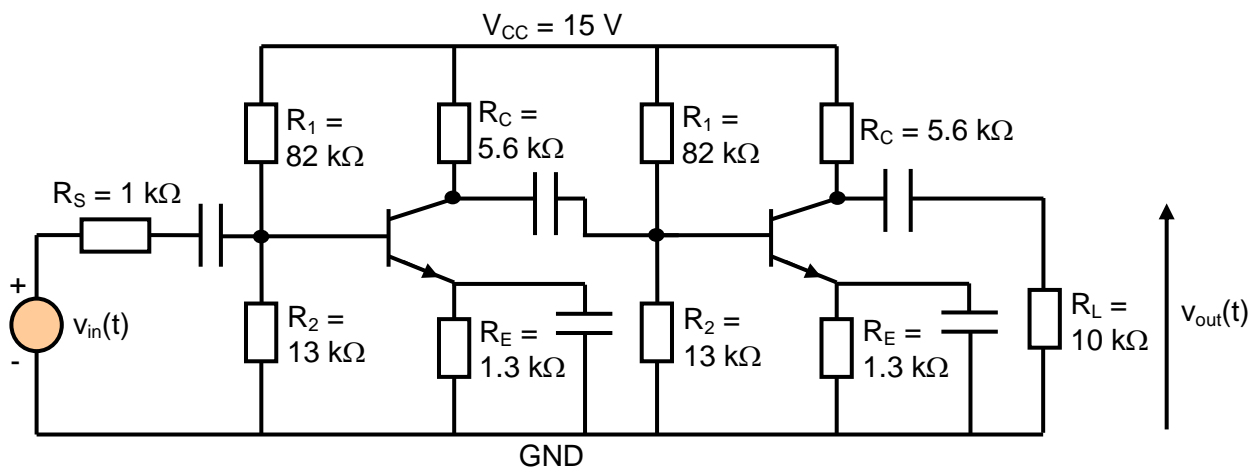
This fall in gain is a direct result of the negative feedback incorporated in the circuit. In some applications, this reduction in gain is unacceptable and a *decoupling capacitance* is used to reduce the amount of AC negative feedback while maintaining DC feedback. This increases the small-signal gain of the circuit but does not affect the DC feedback, which provides stability to the bias conditions of the circuit.

The decoupling capacitance C_E is placed in parallel with the emitter resistance R_E , providing a low-impedance path for AC signals from the emitter to ground, but having no effect on the steady bias voltages.

Perform the DC and AC analysis of the amplifier shown above. For the AC analysis, assume that the decoupling capacitance C_E is equivalent to a short circuit.

Question 3: Cascaded Amplifier

Consider the amplifier circuit depicted below.



This amplifier is obtained by *cascading* two common-emitter amplifiers strictly identical to that studied in Question 2. The load resistance is $R_L = 10 \text{ k}\Omega$, whereas the source resistance is $R_S = 1 \text{ k}\Omega$.

Compute the small-signal voltage gain $A_{V,C}$ of this *cascaded amplifier*.

What would be the value of $A_{V,C}$ if the load resistance R_L was equal to $10 \text{ }\Omega$ instead of $10 \text{ k}\Omega$?

Question 4: Our First Amplifier Revisited

Perform the DC and AC analysis of the common-emitter amplifier circuit designed in Question 2 of Tutorial 2.

Are the results you obtain coherent with the analysis performed in Question 2 of Tutorial 2?

Question 5: Design of a Common-Emitter Amplifier with Negative Feedback

Design a common-emitter amplifier with negative feedback.

The specifications of this circuit are as follows:

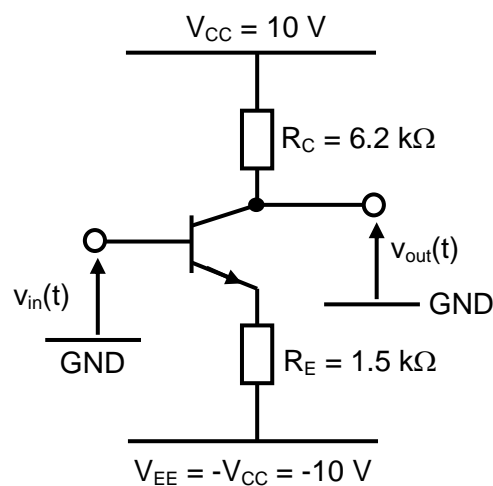
- Power supply $V_{CC} = 10 \text{ volts}$;
- Voltage gain $A_V \approx -9$;
- The maximum output voltage swing must be as large as possible;
- The input resistance r_{in} must not be too small;
- The output resistance r_{out} must not be too large.

Tutorial 4 – Amplifier Circuits using BJTs (2)

Throughout this Tutorial, we use silicon BJTs with $\beta_F = 100$.

Question 1: Common-Emitter Amplifier with Symmetrical Power Supplies

Consider the common-emitter amplifier circuit depicted below. Unlike the circuits we have studied so far, this amplifier uses two symmetrical power supplies $V_{CC} = 10$ volts and $V_{EE} = -10$ volts. This arrangement provides several advantages such as a simplification of the biasing circuitry and a significant increase in the input resistance.

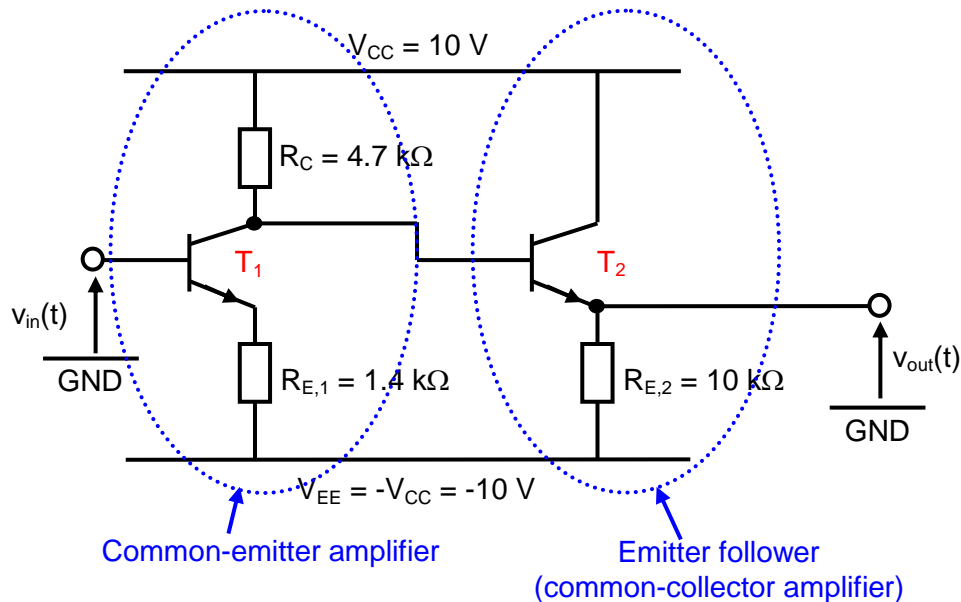


Perform the AC analysis of this amplifier. In particular, show that its input resistance is much larger than those of the amplifiers previously studied in Tutorial 3.

Question 2: Common-Emitter Amplifier with an Output Stage

Consider the amplifier circuit depicted below.

This circuit is designed by cascading a common-emitter amplifier similar to that studied in Question 1 with an *emitter follower* (also called *common-collector amplifier*) acting as an output stage.



The collector of T_2 is connected directly to the positive voltage supply. The input signal is applied between the base and ground, while the output signal is measured between the emitter and ground, which explains why such arrangement is called a common-collector amplifier.

Perform the AC analysis of this amplifier. In particular, show that its output resistance is much lower than that of a simple common-emitter amplifier (thanks to the use of a common-collector amplifier at the output stage).

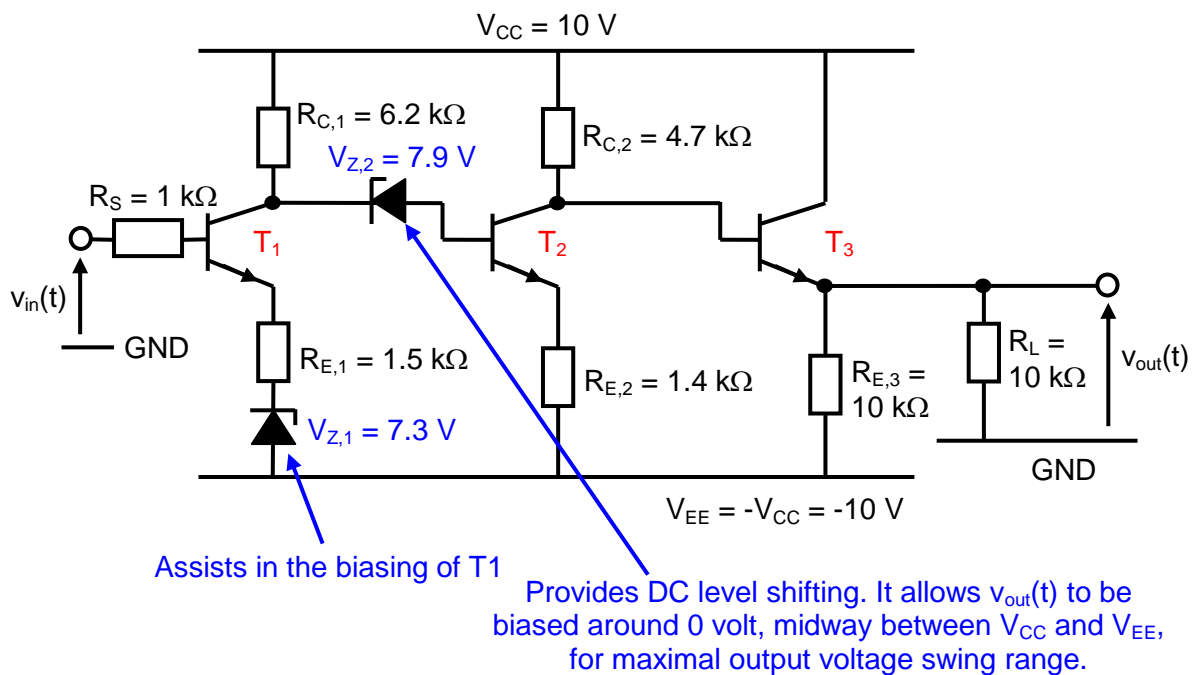
Question 3: Directly-Coupled (DC) Amplifier

The use of coupling and decoupling capacitors for designing amplifiers, although simple in principle, does have some disadvantages. First, each capacitor limits the low-frequency response of the amplifier (i.e., input signals with low frequencies cannot be amplified). Second, the presence of capacitors make the circuit more expensive and less suitable for the production of integrated circuits as capacitors require a large amount of area on a chip.

If care is taken with the design of the circuit, it is possible to avoid using capacitors between the stages of cascaded amplifiers by ensuring that the DC output voltage of one stage represents the correct biasing voltage for the next stage. This not only removes the need for coupling capacitors but also reduces the complexity of the biasing circuitry required.

A major advantage of this technique is that it removes the frequency limitations introduced by capacitive coupling, allowing the design of amplifiers that can be used at frequencies down to DC. These amplifiers are often referred to as *directly coupled (DC) amplifiers*.

Perform the DC and AC analysis of the directly coupled amplifier depicted below.



This amplifier is obtained by cascading the common-emitter amplifier considered in Question 1 and the common-emitter amplifier with an output stage studied in Question 2. The load resistance is $R_L = 10 \text{ k}\Omega$, whereas the source resistance is $R_S = 1 \text{ k}\Omega$.

What would be the value of the small-signal voltage gain if R_L was equal to $10 \text{ }\Omega$ instead of $10 \text{ k}\Omega$?

6. The Field-Effect Transistor (FET)

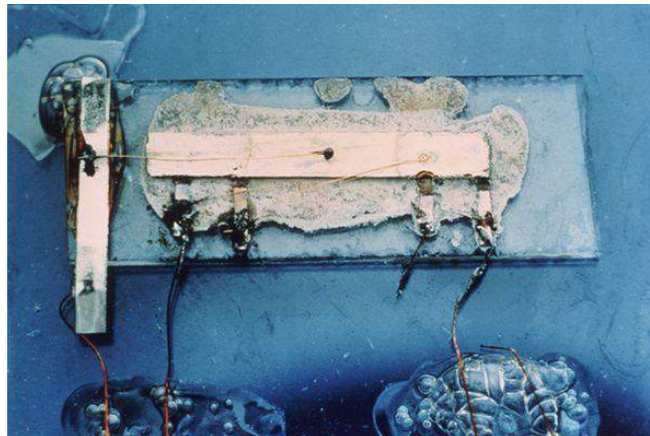
• A Historical Perspective

The basic principle behind the field-effect transistor (FET) was proposed in a patent by Austrian-Hungarian physicist Julius E Lilienfeld as early as 1925 and, independently, by German electrical engineer Oskar Heil in England in 1935. Insufficient knowledge of the materials and gate stability problems, however, delayed the practical usability of the device for a long time.

In 1947, John Bardeen and Walter Brattain, working at the Bell Labs (USA), succeeded in creating a point-contact transistor that achieved amplification. They indeed observed that when two gold point contacts were applied to a crystal of germanium, a signal was produced with the output power greater than the input. William Shockley, who was then leading the Solid State Physics Group at Bell Labs, understood the potential of such discovery and, over the next few months, worked to expand the knowledge of semiconductors. When Bell Labs attempted to patent their invention called "junction transistor", they found Lilienfeld already holding a patent which was worded in a way that would include all types of transistors. In any case, Bell Labs was able to work out an agreement with Lilienfeld. It was at that time that the Bell Labs version was given the name bipolar junction transistor, or simply junction transistor, and Lilienfeld's design took the name field-effect transistor.

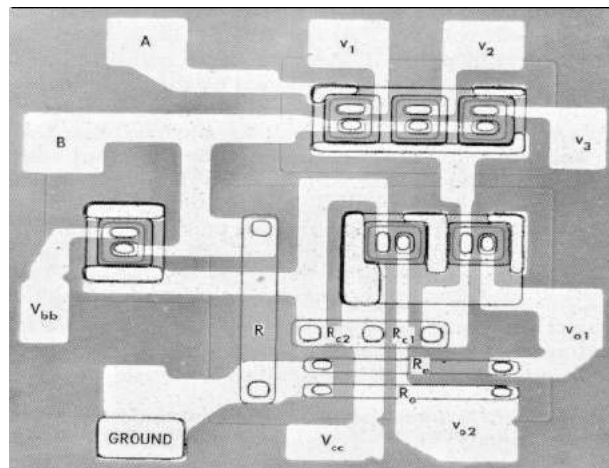
It is worth mentioning that, in 1948, the point-contact transistor was independently invented by German physicists Herbert Mataré and Heinrich Welker while working at the Compagnie des Freins et Signaux (France).

The next breakthrough was the invention of the integrated circuit in 1958 by the American physicist Jack Kilby (working at Texas Instruments). He was awarded the Nobel Prize in Physics for his discovery in 2000. His integrated circuit contained only a transistor and other components on a slice of germanium. An integrated circuit is basically an electronic circuit manufactured by the patterned diffusion of trace elements into the surface of a thin substrate of semiconductor material. Additional materials are deposited and patterned to form interconnections between semiconductor devices.



Jack Kilby's original integrated circuit

The first integrated circuits were made using BJTs. For instance, the first commercial digital integrated circuits were manufactured using the TTL (transistor-transistor logic) family, introduced in 1962, which offered a relatively high integration density at that time. The TTL logic family was so successful that it composed the largest fraction of the digital semiconductor market until the 1980s.



3-input ECL (emitter-coupled logic) logic gate, Motorola, 1966

During the 1950s and 1960s, several types of FETs were introduced. Among the most prominent ones were the junction FET (junction FET, proposed by William Schokley in 1952 and built by G Dacey and Ian Ross in 1953), the metal-semiconductor FET (MESFET, Carver Mead, 1966), and the metal-oxide-semiconductor FET (MOSFET, Dawon Kahng and Martin Atalla, 1960). Today,

the MOSFET is by far the most widely used type of transistor in integrated circuits. One can even reasonably say that it has been the undisputed king of semiconductors since the early 1990s. This is why we are going to focus throughout this chapter on the study of this particular transistor.

The first MOSFET was built by Dawon Kahng and Martin Atalla at Bell Labs in 1960. Operationally and structurally different from the BJT, the MOSFET was a particular type of metal-insulator-semiconductor FET using crystalline silicon (Si) for the semiconductor and a layer of silicon dioxide (SiO₂) for the insulator. What made the silicon MOSFET structure so special is that it did not generate localized electron traps at the Si-SiO₂ interface, and was thus inherently free from the trapping and scattering of carriers that had impeded the performance of earlier field-effect transistors.

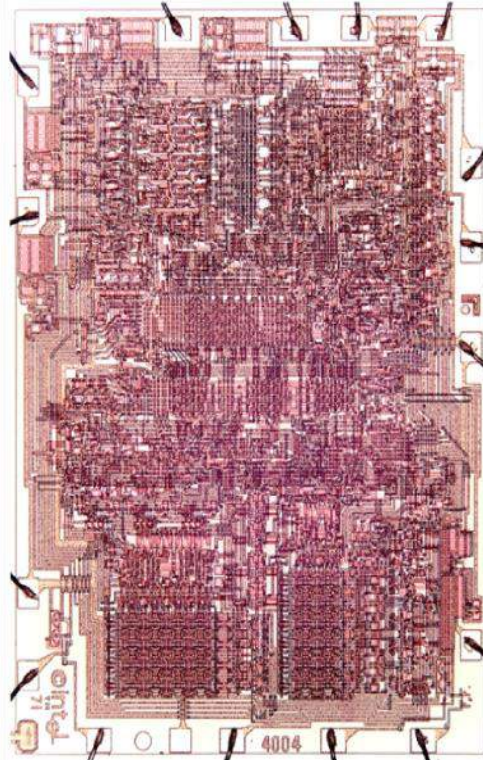
MOSFET digital circuits started to take off a bit later than their BJT counterparts. It is in fact during the early 1970s that digital circuits started being implemented using MOSFET transistors. For instance, the first commercially available microprocessor (the Intel 4004 introduced in 1971) was designed using about 2,300 P-channel MOSFETs (10- μ m manufacturing process). We can also mention the Motorola 68000 (1979) which was implemented using approximately 68,000 N-channel MOSFETs (4- μ m process).

Towards the beginning of the 1980s, complementary MOSFET (CMOS) logic, which employs both N-channel and P-channel MOSFETs, finally became the technology of choice for designing complex digital integrated circuits such as microprocessors, due to its very high integration density. Today, the CMOS logic family represents more than 99% of the digital market. All other logic families, including TTL, have now become history.

As an illustration of what can be achieved today using MOSFETs, the Oracle SPARC T4 microprocessor introduced in 2011 employs 855 million MOSFETs (40-nm process) on a chip of 403 mm² and has a clock frequency of 3 GHz. This is the CMOS technology, i.e. the MOSFET transistor, that has made the design of such extremely complex circuits possible.

The MOSFET's key advantages in digital circuits do not necessarily translate into supremacy in analogue circuits. The reason is that analogue and digital circuits draw upon different features of transistor behaviour. Digital circuits switch, spending most of their time outside the amplification

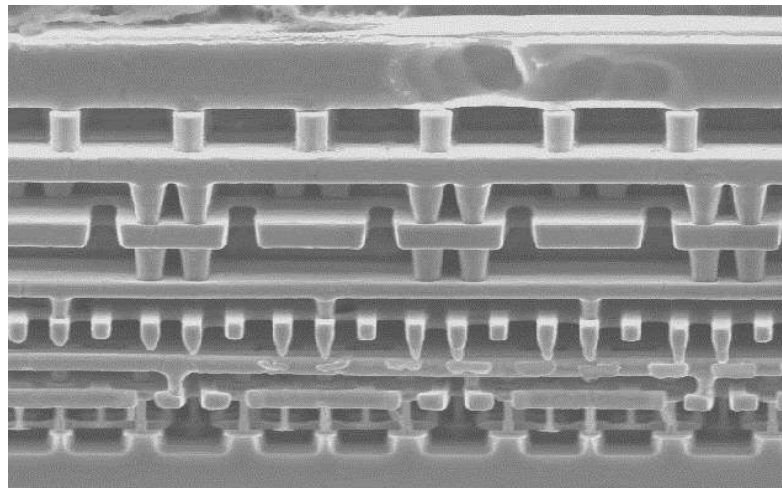
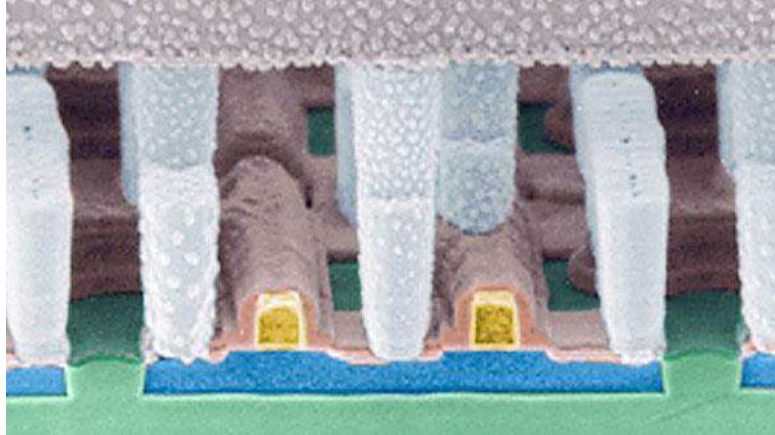
region. On the other hand, the performance of analogue circuits depend on transistor behaviour in the amplification region (remember the BJT logic inverter studied in Tutorial 2).



The Intel 4004 microprocessor (1971)

The BJT has traditionally been the analogue designer's transistor of choice for several reasons such as its high transconductance for a given current, high intrinsic gain, and high speed. However, at the present time, analogue integrated circuits are designed and fabricated in bipolar technology, in MOSFET technology, and in technologies that combine both types of devices in one manufacturing process (BiCMOS).

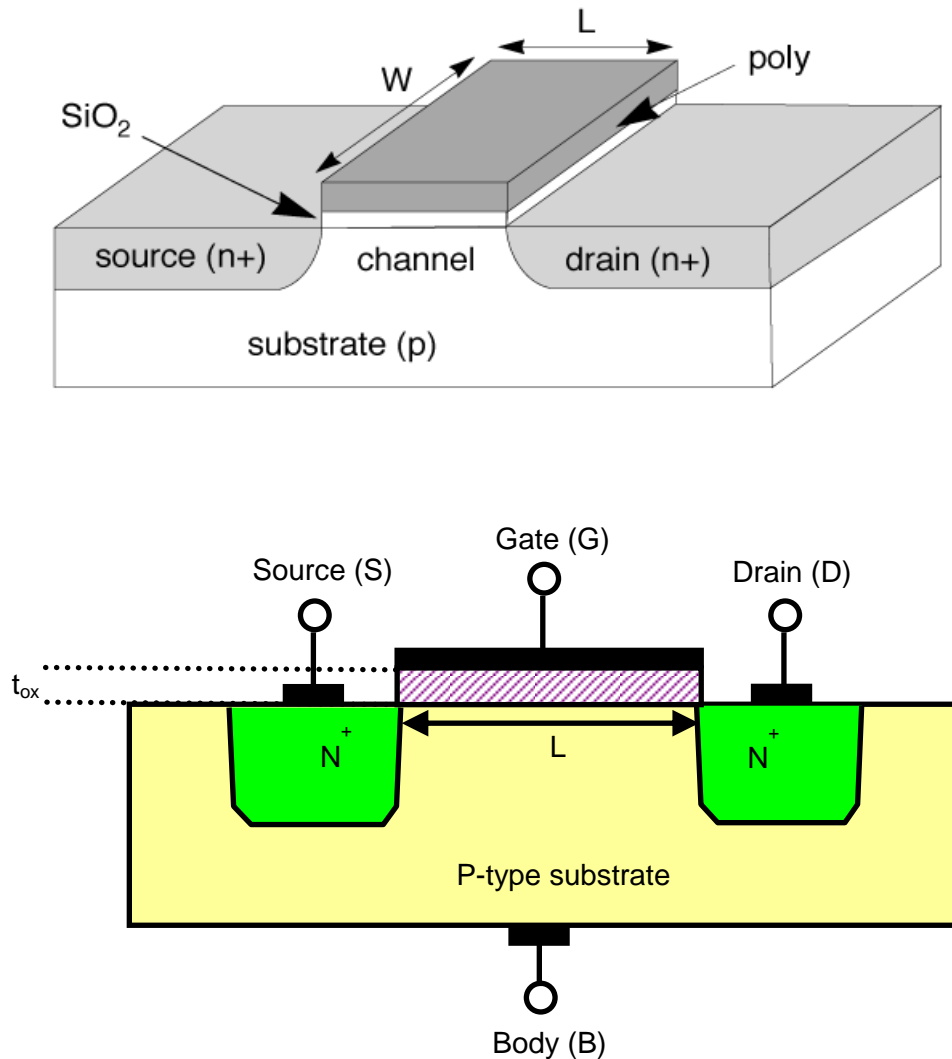
The necessity of combining complex digital functions with analogue functions on the same integrated circuit has resulted in an increased use of digital MOS technologies for analogue functions, particularly those functions such as digital-analogue conversion required for interfaces between analogue signals and digital systems. However, despite the strong economic incentive to use MOSFETs for both digital and analogue circuitry, bipolar technology will continue to be used in a wide range of applications requiring high-current drive capability and the highest levels of analogue performance.



Cross-sections of MOS integrated circuits

- **The Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET)**

A traditional metal–oxide–semiconductor (MOS) structure is obtained by growing a layer of silicon dioxide (SiO_2) above a silicon substrate and depositing a layer of metal or polycrystalline silicon (poly) on top of it. As silicon dioxide is a dielectric material (insulator), the MOS structure is thus equivalent to a capacitor, with one of the electrodes replaced by a semiconductor.



In the 3D view and the cross-section of an N-channel MOSFET (hereafter referred to as NMOS transistor) shown above, the parameters L and W denote the length and the width of the transistor, respectively. The thickness of the layer of SiO_2 is denoted as t_{ox} .

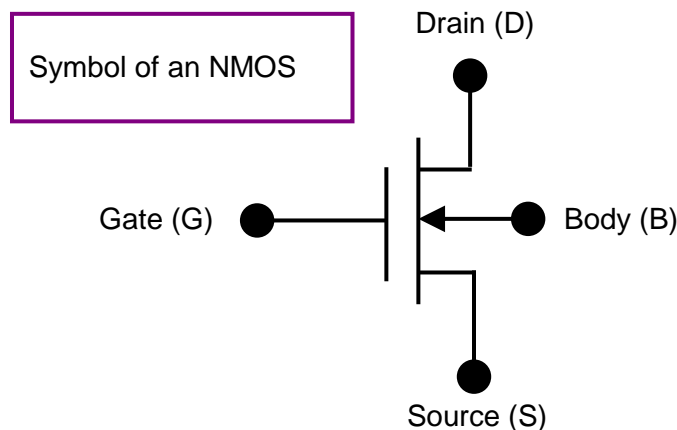
In this chapter, we are going to focus on the structure of an NMOS transistor, knowing that all explanations are also valid for a P-channel MOSFET (PMOS transistor).

To construct an NMOS transistor, we start from a lightly-doped P-type silicon substrate, called the body. In this P-type substrate, we implant two heavily-doped N-type regions, called the drain and the source. A thin layer of silicon dioxide (SiO_2), called the gate oxide, is used to cover the region that separates the drain and the source.

The gate oxide is itself covered by some metal or, traditionally, some conductive polycrystalline silicon (polysilicon, for short). The conductive material above the thin layer of SiO_2 forms the gate of the transistor.

It is worth mentioning that, unlike the BJT, a MOSFET has a symmetrical physical structure in the sense that the drain and source terminals are identical. In an NMOS transistor, the drain is defined by convention as the terminal connected to the highest voltage, whereas the source is the terminal connected to the lowest voltage. In other words, the current always flows from drain to source in an NMOS transistor.

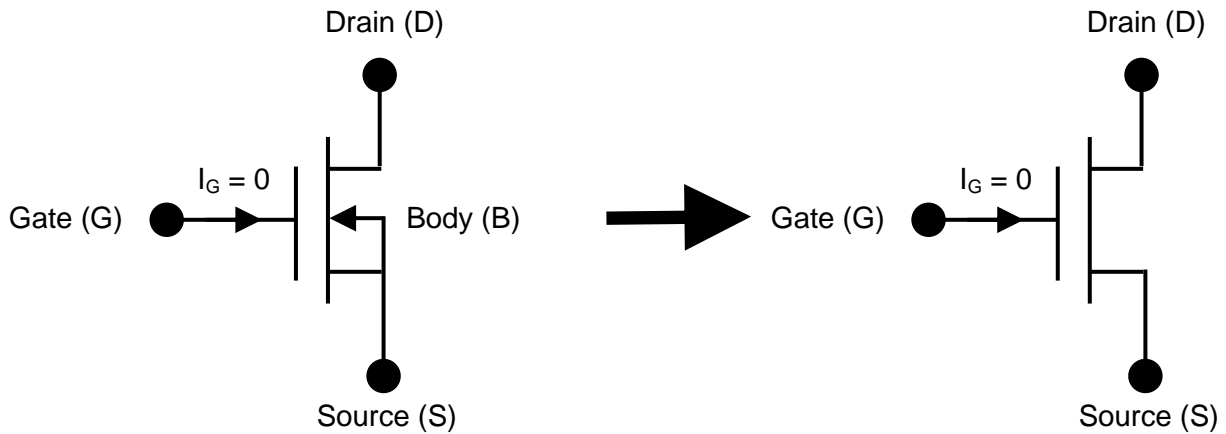
The MOSFET is a four-terminal device (gate, drain, source, and body). The symbol for the NMOS transistor is shown below.



The (DC) current I_G flowing through the gate is equal to zero due to the presence of the insulating layer of SiO_2 . This is a very important (and attractive) feature of MOSFETs since it essentially means that the input resistance of a MOSFET is infinite. Thus, using MOSFETs will allow us to design electronic circuits with a very high input resistance. This is actually one of the key advantages of MOSFETs over BJTs. Note that the infinite input resistance is a feature shared by all types of FET (MOSFET, JFET, MESFET...).

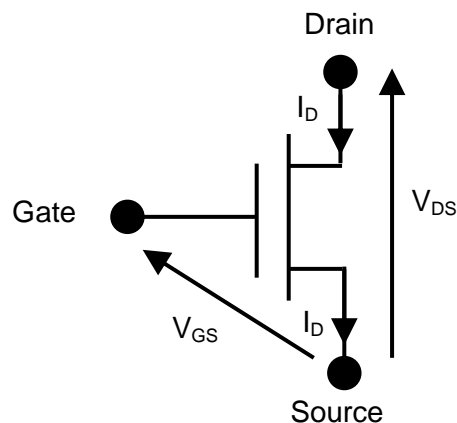
The voltage applied to the gate terminal determines if and how much current flows between the drain and the source. The body represents the fourth terminal of the transistor. Its function is actually secondary because it only serves to modulate the device characteristics and parameters.

Throughout these lecture notes, we will simply assume that the body terminal is connected to the source, and therefore does not need to be shown. Note that, in practice, this is often (but not always) the case.



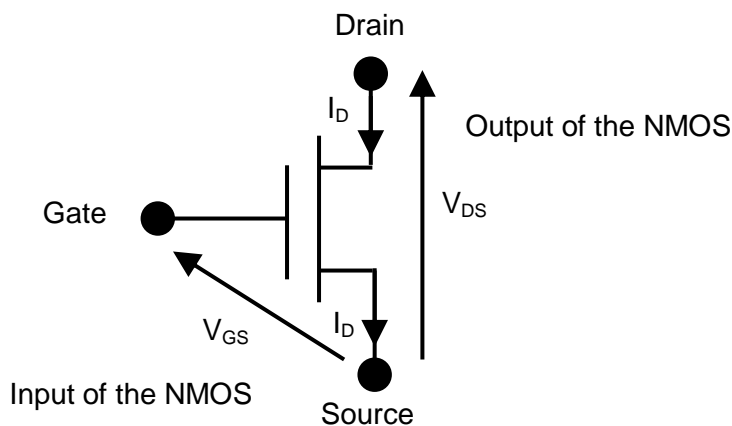
At the most superficial level, the transistor can be thought of as a switch. When the voltage applied to the gate is larger than a given value (the threshold voltage V_{TN} , not to be confused with the thermal voltage V_T), a conducting channel is formed between drain and source. In the presence of a voltage difference between the latter two ($V_{DS} > 0$), an electrical current I_D can then flow between them.

The conductivity of the channel depends on the gate voltage: the larger the voltage between gate and source (V_{GS}), the smaller the resistance of the conducting channel and the larger the current I_D . When the gate voltage is lower than the threshold voltage ($V_{GS} < V_{TN}$), no such conducting channel exists, and the switch is considered open.



• Common-Source Configuration of an NMOS Transistor

In most applications, the NMOS is viewed as a semiconductor device with an input and an output. Usually, the input parameter is the voltage V_{GS} , whereas the output parameters are the drain current I_D and the voltage V_{DS} . This particular arrangement is referred to as common-source configuration because the source terminal is common to both input and output.

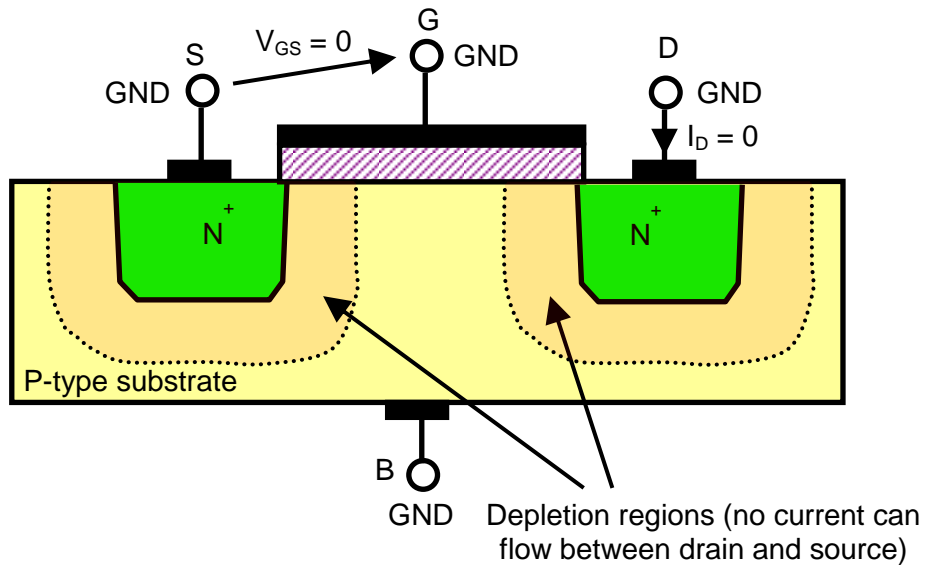


This is obviously very similar to the way we view a BJT in most electronic circuits. We therefore have the following analogies between BJT and MOSFET: source \equiv emitter, drain \equiv collector, gate \equiv base, and collector current \equiv drain current.

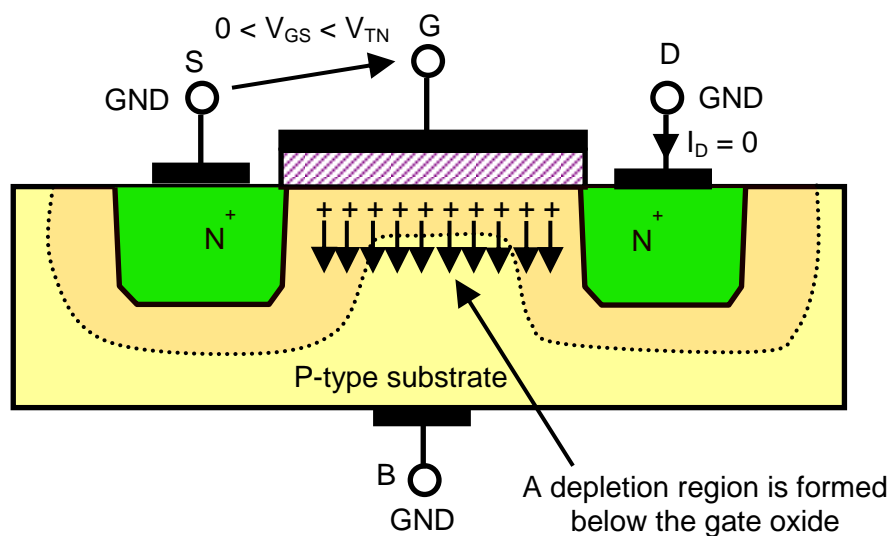
The main difference between both semiconductor devices is that the operation of the NMOS transistor is actually simpler than that of the bipolar transistor because, with an NMOS, we must only consider one input quantity (V_{GS}) since the input (gate) current I_G is always equal to zero. With a BJT, there are two input quantities (I_B and V_{BE}) that need to be taken into account.

• Current-Voltage Characteristics of an N-Channel MOSFET (NMOS transistor)

Consider first the case in which gate, drain, source, and body are connected to ground. The drain and source are connected by back-to-back PN junctions (substrate-source and substrate-drain). Under the mentioned conditions, both junctions have a bias of 0 volt and are therefore reverse-biased, which results in an extremely high resistance between drain and source.

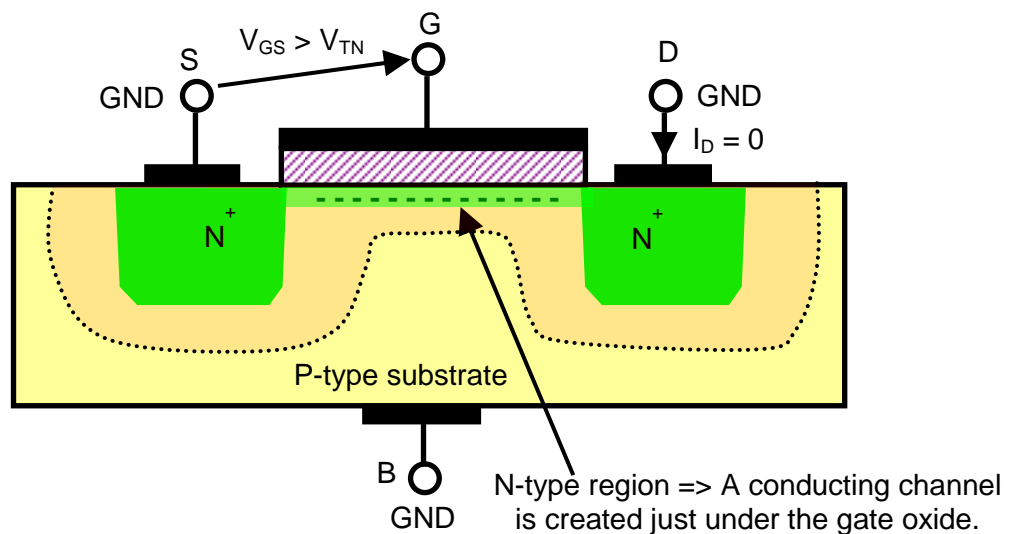


Assume now that a small positive voltage is applied to the gate (with respect to the source). We recall that the gate and substrate form the plates of a capacitor with the gate oxide as the dielectric (insulator). The positive gate voltage creates an electric field across the MOS capacitance which manifests itself initially by repelling holes (the majority carriers in the P-type silicon substrate). Hence, a depletion region is formed below the gate. This depletion region is similar to the one occurring in a PN junction.



As the gate voltage increases, the potential at the silicon surface reaches a critical value at some point, and the semiconductor surface inverts to N-type material. This point marks the onset of a phenomenon known as strong inversion.

Further increases in the gate voltage results in additional electrons (the minority carriers in the P-type silicon substrate) in the thin inversion layer directly under the gate oxide. Hence, an N-type channel is formed between the source and drain regions, the conductivity of which is dependent on the gate-source voltage V_{GS} .



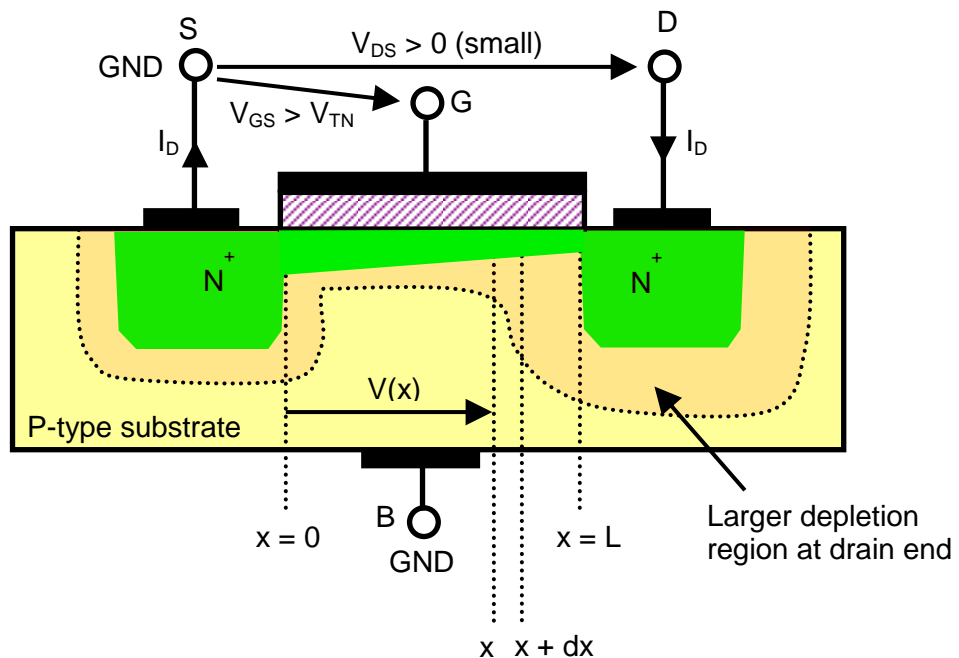
The value of V_{GS} where strong inversion occurs (and a conducting channel is created) is called the threshold voltage V_{TN} . Typically, V_{TN} is equal to few tenths of a volt. For instance, for a digital circuit with a supply voltage $V_{DD} = 5\text{ V}$, we would have $V_{TN} \approx 0.8\text{ V}$, whereas for $V_{DD} = 2.5\text{ V}$, we would rather have $V_{TN} \approx 0.4\text{ V}$. Note that the value of the threshold voltage can be adjusted during the manufacturing process depending on the requirements of the circuit designer.

This parameter V_{TN} is a function of several components, most of which are material constants, such as the difference in work function between gate and substrate materials, the oxide thickness, the Fermi voltage, the charge of impurities trapped at the surface between channel and gate oxide, and the dosage of ions implanted for threshold adjustment. It can also be shown that V_{TN} is dependent on the voltage V_{SB} between the source and body terminals.

We can now define the first mode of operation of an NMOS: The transistor is in the cut-off mode when $V_{GS} < V_{TN}$. In such case, there is no conducting channel and no current can flow from drain to source, which leads to $I_D = 0$.

Assume now that $V_{GS} > V_{TN}$ and that a small voltage V_{DS} is applied between drain and source. The voltage difference causes a current I_D to flow from drain to source.

At a point x along the channel, the voltage with respect to the source is denoted as $V(x)$. The gate-to-channel voltage V_{Gx} , i.e. the voltage across the MOS capacitance, at that point x is therefore given by $V_{Gx} = V_{GS} + V_{Sx} = V_{GS} - V_{xS} = V_{GS} - V(x)$.



The charge $dQ(x)$ in the channel between points x and $(x + dx)$ is given by

$$dQ(x) = C_{ox} W dx (V_{GS} - V(x) - V_{TN}),$$

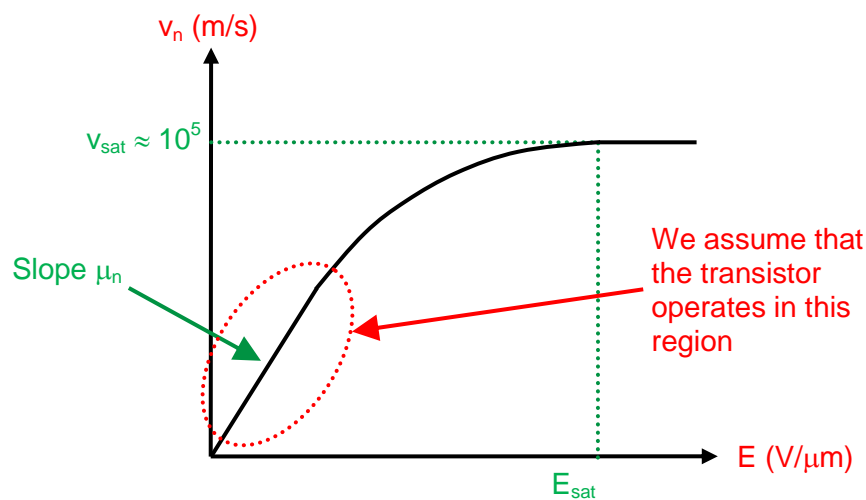
where C_{ox} designates the capacitance per unit area presented by the gate oxide, and is expressed as $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$, where ϵ_{ox} denotes the permittivity of the gate oxide ($\epsilon_{ox} = 3.5 \times 10^{-11}$ F/m).

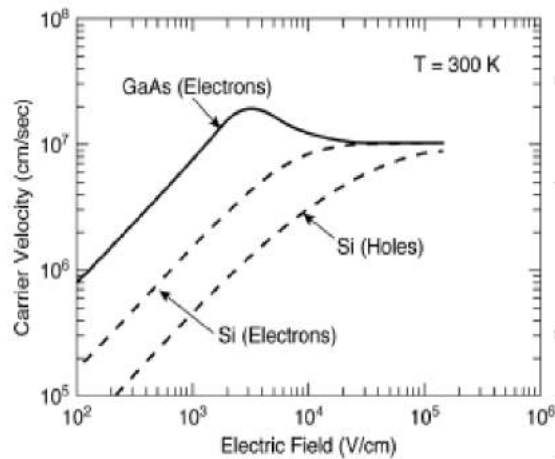
The current flowing from drain to source is given by $I_D = \frac{dQ(x)}{dt} = \frac{dQ(x)}{dx} \frac{dx}{dt} = \frac{dQ(x)}{dx} v_n(x)$, where $v_n(x)$ designates the drift velocity of the electrons (which are the carriers in the channel) at point x . Hence, we can write:

$$I_D = C_{ox} W (V_{GS} - V(x) - V_{TN}) v_n(x).$$

The drift velocity of the electrons $v_n(x)$ at point x is proportional to the horizontal electric field $E(x)$ (created by the applied voltage V_{DS}) in the channel at the same point x provided that the electric field is small ($E(x) \ll E_{sat}$).

We can thus write $I_D = \mu_n E(x) C_{ox} W (V_{GS} - V(x) - V_{TN})$, where the constant μ_n designates the mobility of the electrons in the channel. The mobility depends on both the temperature and the doping level but is almost constant for a wide range of normally used doping levels. Also, μ_n is sometimes called the surface mobility for electrons because the channel forms at the surface of the silicon. Typical values range from about $500 \text{ cm}^2/\text{V}\cdot\text{s}$ to about $700 \text{ cm}^2/\text{V}\cdot\text{s}$, which are much less than the mobility of electrons in the bulk of the silicon (about $1400 \text{ cm}^2/\text{V}\cdot\text{s}$) because surface defects not present in the bulk impede the flow of electrons in MOS transistors.





The electric field $E(x)$ at point x is related to the voltage $V(x)$ at the same point through the well-known expression

$$E(x) = \frac{dV(x)}{dx}.$$

As a result, we obtain the following expression: $I_D dx = \mu_n dV(x) C_{ox} W (V_{GS} - V(x) - V_{TN})$.

By integrating this equation over the length L of the channel, an expression of the current I_D can be obtained:

$$\int_0^L I_D dx = \mu_n C_{ox} W \int_0^{V_{DS}} (V_{GS} - V(x) - V_{TN}) dV(x)$$

$$\Rightarrow I_D = \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_{TN})V(x) - \frac{V(x)^2}{2} \right]_0^{V_{DS}}$$

$$\Rightarrow I_D = \mu_n C_{ox} \frac{W}{L} \left(V_{GS} - V_{TN} - \frac{V_{DS}}{2} \right) V_{DS}.$$

The mode of operation where this equation holds is called the linear mode of operation because the NMOS transistor is somewhat equivalent to a voltage-controlled resistance in this mode.

In fact, for $V_{DS} \ll 2(V_{GS} - V_{TN})$, we even observe a perfectly linear dependence between V_{DS} and I_D , and we can write

$$I_D \approx \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TN}) V_{DS}.$$

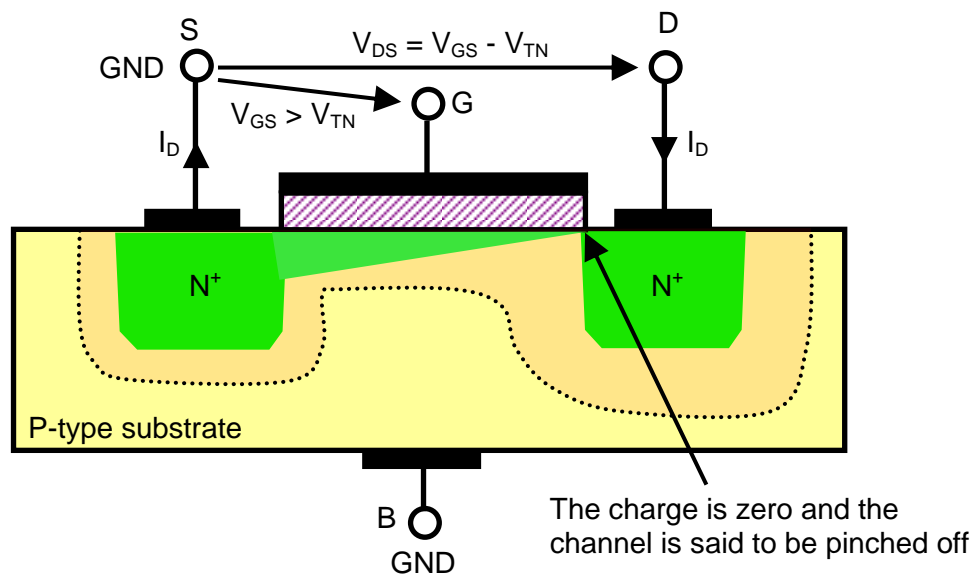
This indicates that, for small values of V_{DS} , the transistor is equivalent to a resistance R_{NMOS} given by

$$R_{NMOS} = \frac{V_{DS}}{I_D} \approx \left[\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TN}) \right]^{-1}.$$

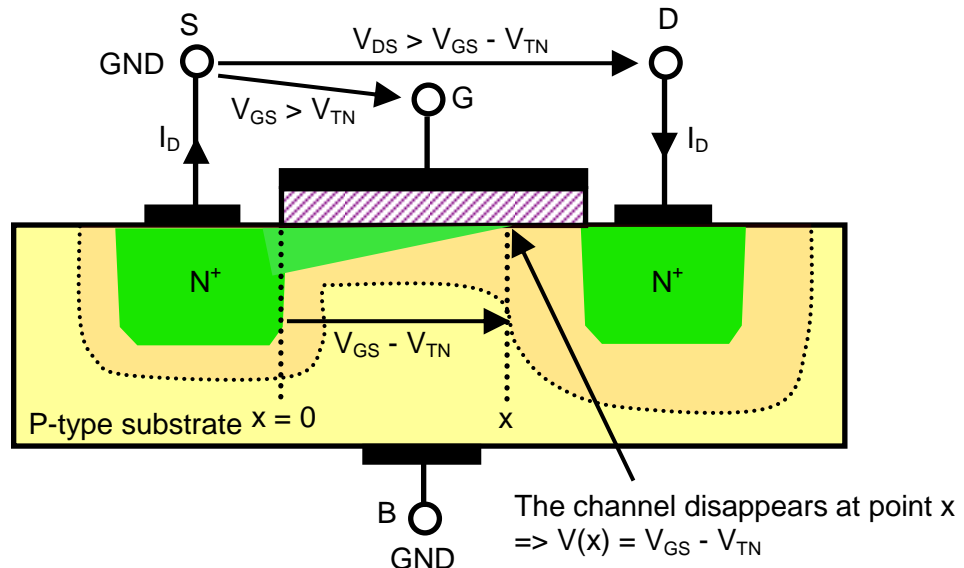
To simplify the equations, we can introduce the parameter $k_n = \mu_n C_{ox} \frac{W}{L}$ called gain factor of the transistor. A typical value of k_n for $W/L = 10$, $t_{ox} = 10$ nm, $\mu_n = 570$ cm²/V.s, and $\epsilon_{ox} = 3.5 \times 10^{-11}$ F/m is $k_n = 570 \times 10^{-4} \times \frac{3.5 \times 10^{-11}}{10 \times 10^{-9}} \times 10 \approx 2$ mA/V².

Note that the designer can change the value of the gain factor k_n by simply changing the ratio W/L , which is very convenient in practice.

As the value of V_{DS} is further increased, the conducting channel narrows at the drain end ($dQ(L) \rightarrow 0$). When V_{DS} reaches the value $V_{GS} - V_{TN}$, the gate-to-channel voltage $V_{GD} = V_{GS} - V_{DS}$ at the drain end becomes equal to the threshold voltage V_{TN} , thus meaning that the channel is no longer connected to the drain. In other words, the channel disappears. This phenomenon is called pinch-off.



As V_{DS} increases beyond $V_{GS} - V_{TN}$, the pinch-off point moves closer to the source. The voltage difference over the conducting channel from the pinch-off point to the source remains fixed at $V_{GS} - V_{TN}$. As a result, the drain current does no longer depend on V_{DS} , but only on the voltage across the channel which is equal to $V_{GS} - V_{TN}$. In other words, the drain current I_D no longer increases with V_{DS} .



When $V_{DS} > V_{GS} - V_{TN}$, the NMOS transistor is said to be in the saturation mode of operation. The expression of the constant drain current is obtained by replacing V_{DS} by $V_{GS} - V_{TN}$ in the current expression derived earlier:

$$I_D = k_n \left(V_{GS} - V_{TN} - \frac{V_{DS}}{2} \right) V_{DS}, \text{ for } V_{DS} = V_{GS} - V_{TN}$$

$$\Rightarrow I_D = k_n \left(V_{GS} - V_{TN} - \frac{V_{GS} - V_{TN}}{2} \right) (V_{GS} - V_{TN})$$

$$\Rightarrow I_D = \frac{k_n}{2} (V_{GS} - V_{TN})^2.$$

We can notice the square dependency of the constant drain current with respect to V_{GS} .

As a summary, an NMOS transistor has three modes of operations:

- **First mode of operation:** The transistor is in the cut-off mode when $V_{GS} < V_{TN}$. In this case, the drain current I_D is equal to zero and the transistor is equivalent to an open switch since the drain is not connected to the source.

- **Second mode of operation:** The transistor is in the linear mode when $V_{GS} \geq V_{TN}$ and $V_{DS} < V_{GS} - V_{TN}$. In this case, the drain current I_D is given by

$$I_D = k_n \left(V_{GS} - V_{TN} - \frac{V_{DS}}{2} \right) V_{DS}.$$

In this mode, the transistor resembles a voltage-controlled resistance given by

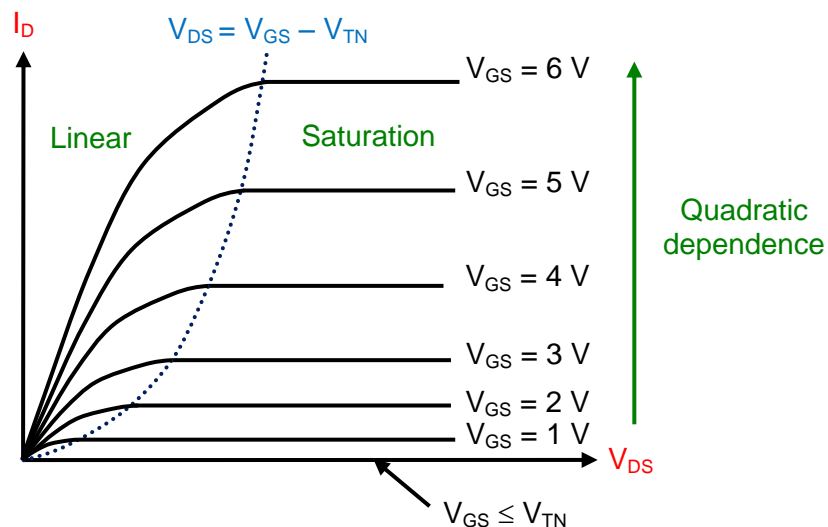
$$R_{NMOS} = \frac{dV_{DS}}{dI_D} = \frac{1}{k_n (V_{GS} - V_{TN} - V_{DS})}.$$

- **Third mode of operation:** The transistor is in the saturation mode when $V_{GS} \geq V_{TN}$ and $V_{DS} \geq V_{GS} - V_{TN}$. In this case, the drain current I_D is given by

$$I_D = \frac{k_n}{2} (V_{GS} - V_{TN})^2.$$

In this mode, the transistor is equivalent to a voltage-controlled current source. The saturation mode is traditionally used for designing MOS linear amplifiers in analogue electronics.

Output characteristic of an NMOS transistor



The output characteristic of an NMOS transistor somewhat looks like that of an NPN bipolar junction transistor. The saturation mode for a BJT corresponds to the linear mode of an NMOS, whereas the forward active mode for an NPN BJT ($V_{CE} > V_{CE,sat}$) is analogous to the saturation mode for an NMOS ($V_{DS} \geq V_{GS} - V_{TN}$).

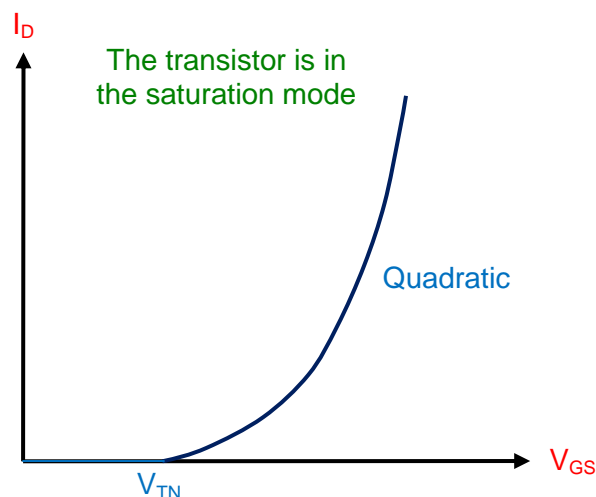
Remember that we found the following analogies between BJT and MOSFET: source \equiv emitter, drain \equiv collector, gate \equiv base, and collector current \equiv drain current. Therefore, there are also analogies between the input voltages V_{BE} and V_{GS} , the output voltages V_{CE} and V_{DS} , and the output currents I_C and I_D .

As for the transfer characteristic linking the output current (I_C , I_D) and the input voltage (V_{BE} , V_{GS}), we notice that the drain current I_D varies like the square of the voltage ($V_{GS} - V_{TN}$) in an NMOS transistor, while the collector current I_C varies exponentially with the voltage V_{BE} in a BJT.

This indicates that a given variation of the input voltage (V_{BE} , V_{GS}) will result in a larger variation of the output current (I_C , I_D) with a BJT than with a MOSFET. It thus seems that the BJT is inherently more suitable than the MOSFET for the design of high-gain voltage amplifiers.

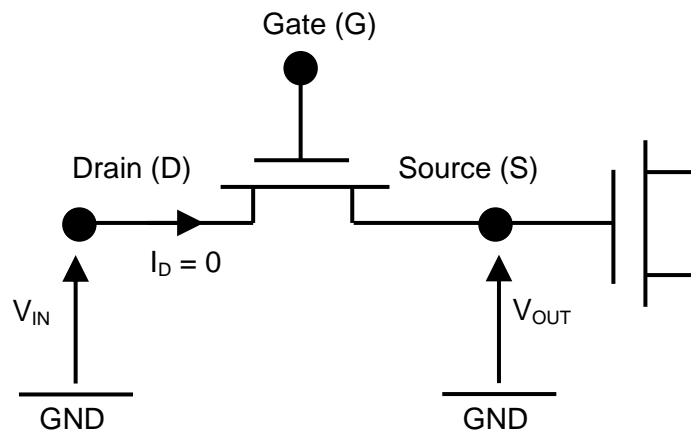
However, the MOSFET, like any other types of FET, has an infinite input resistance (recall that the gate current is always equal to zero), which is not the case with the BJT. Therefore, if we want to design a circuit with a very large input resistance, it will be preferable to employ a FET rather than a BJT.

Transfer characteristic of an NMOS transistor



• **The NMOS Transistor as a Switch**

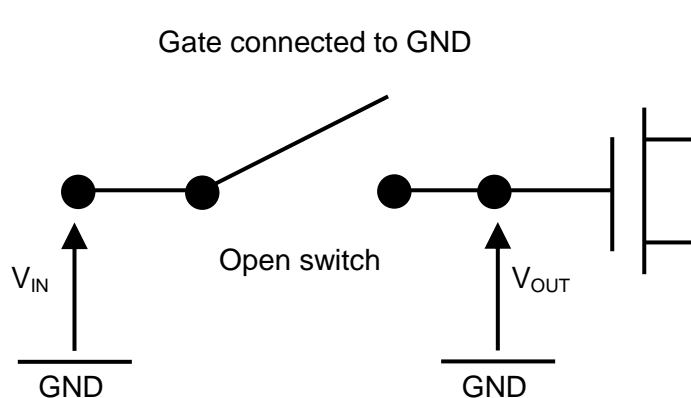
Consider the circuit below. We propose to show that the NMOS transistor is equivalent to an imperfect switch. We will assume that the voltage applied on the gate can be equal to either 0 V (GND) or V_{DD} (which is the supply voltage, i.e. the highest possible voltage in the circuit).



(1) The gate is connected to ground.

In this case, we can write $V_{GS} = 0 - V_{OUT} = -V_{OUT}$ and $V_{DS} = V_{IN} - V_{OUT}$.

Since all voltage values in the circuit range from 0 to V_{DD} , $-V_{OUT}$ is negative and thus smaller than the threshold voltage V_{TN} (which is always positive for an NMOS transistor, V_{TN} is typically equal to few tenths of a volt). This means that $V_{GS} < V_{TN}$. The NMOS transistor is in the cut-off mode of operation and thus equivalent to an open switch since the drain is not connected to the source.



(2) The gate is connected to the supply voltage V_{DD} .

In this case, we can write $V_{GS} = V_{DD} - V_{OUT}$ and $V_{DS} = V_{IN} - V_{OUT}$.

The transistor is in the cut-off mode when $V_{GS} < V_{TN}$, i.e. $V_{OUT} > V_{DD} - V_{TN}$. In this case, the transistor is equivalent to an open switch since the drain is not connected to the source.

The transistor is in the linear mode when $V_{GS} \geq V_{TN}$, i.e. $V_{OUT} \leq V_{DD} - V_{TN}$, and $V_{IN} - V_{OUT} < V_{DD} - V_{OUT} - V_{TN}$, i.e. $V_{IN} < V_{DD} - V_{TN}$. In this case, the drain current I_D is given by

$$I_D = k_n \left(V_{DD} - V_{OUT} - V_{TN} - \frac{V_{IN} - V_{OUT}}{2} \right) (V_{IN} - V_{OUT}) = 0$$

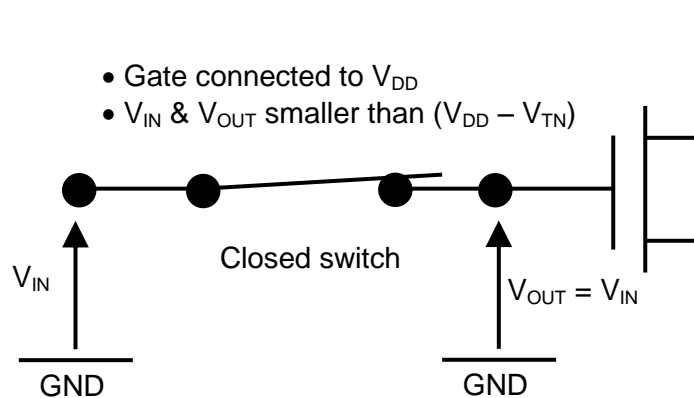
$$\Rightarrow \left(V_{DD} - V_{TN} - \frac{V_{IN}}{2} - \frac{V_{OUT}}{2} \right) (V_{IN} - V_{OUT}) = 0.$$

There are two possible solutions to this equation:

- (a) $V_{DD} - V_{TN} - \frac{V_{IN}}{2} - \frac{V_{OUT}}{2} = 0 \Rightarrow V_{OUT} = 2(V_{DD} - V_{TN}) - V_{IN}$.
- (b) $V_{IN} - V_{OUT} = 0 \Rightarrow V_{OUT} = V_{IN}$.

The solution (a) must be discarded as it is incompatible with the conditions necessary for the transistor to be in the linear mode. On the other hand, the solution (b) is compatible with those conditions.

Our conclusion is that, when $V_{OUT} \leq V_{DD} - V_{TN}$ and $V_{IN} < V_{DD} - V_{TN}$, we have $V_{OUT} = V_{IN}$ and the transistor is thus equivalent to a closed switch.



Note that this closed switch exhibits a resistance R given by

$$R = \frac{dV_{DS}}{dI_D} \Big|_{V_{DS} = V_{IN} - V_{OUT}, V_{GS} = V_{DD} - V_{OUT}} = \frac{1}{k_n (V_{DD} - V_{TN} - V_{IN})}$$

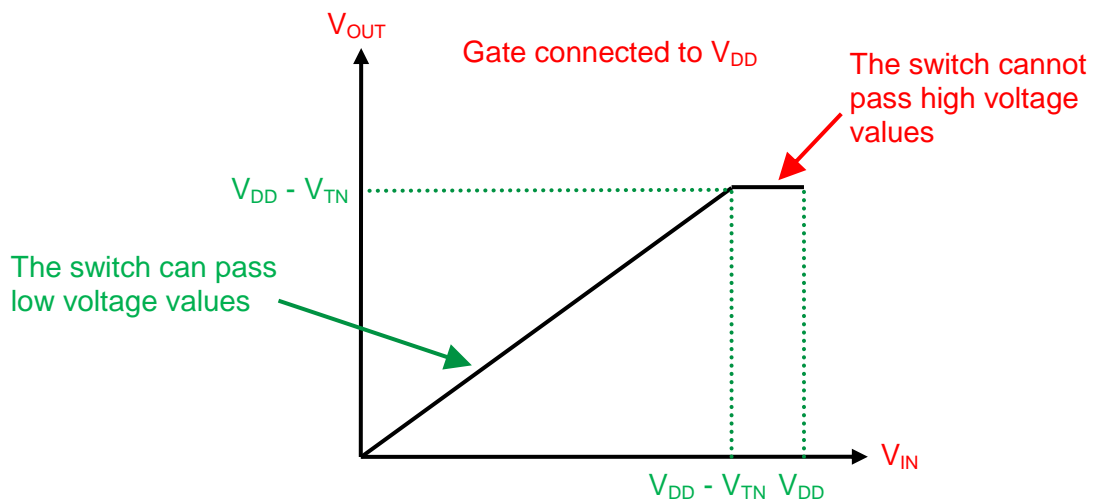
There is no voltage drop across this resistance because the current flowing through the switch is equal to zero. Therefore, the presence of a non-zero resistance R is not an issue (as long as R does not become infinite).

The transistor is in the saturation mode when $V_{GS} \geq V_{TN}$, i.e. $V_{OUT} \leq V_{DD} - V_{TN}$, and $V_{IN} - V_{OUT} \geq V_{DD} - V_{OUT} - V_{TN}$, i.e. $V_{IN} \geq V_{DD} - V_{TN}$. In this case, the drain current I_D is given by

$$I_D = \frac{k_n}{2} (V_{DD} - V_{OUT} - V_{TN})^2 = 0 \Rightarrow V_{OUT} = V_{DD} - V_{TN}$$

We thus conclude that, when $V_{OUT} \leq V_{DD} - V_{TN}$ and $V_{IN} \geq V_{DD} - V_{TN}$, we have $V_{OUT} = V_{DD} - V_{TN}$ and the switch output is then constant.

As a summary, an NMOS transistor can be thought of as a switch controlled by its gate voltage. The switch is closed when the voltage applied on the gate is high (e.g., equal to V_{DD}), and is open when this gate voltage is low (e.g., equal to 0 V). However, this switch is imperfect because, when closed, it cannot pass voltage values that are above $V_{DD} - V_{TN}$.

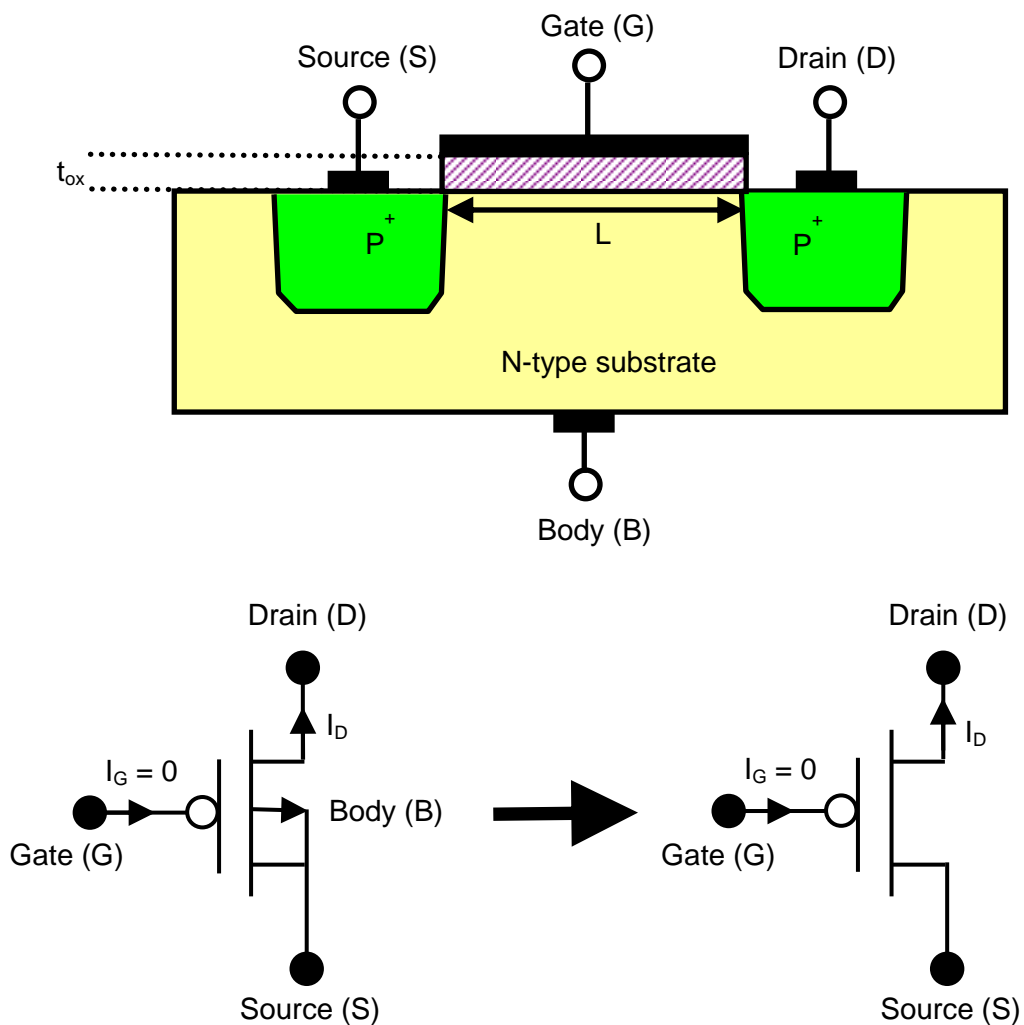


This two-state operation (open switch/closed switch) makes MOS transistors particularly amenable to digital electronics since it matches the concepts of binary digital logic.

• **Current-Voltage Equations for a P-Channel MOSFET (PMOS transistor)**

Throughout this chapter, we have extensively studied the N-channel MOSFET. In many circuits, it is also necessary to employ P-channel MOSFETs, hereafter referred to as PMOS transistors. For instance, since the 1980s, the dominant technology in the world of digital electronics has been the CMOS (complementary) logic family that makes use of both NMOS and PMOS transistors on the same silicon chip.

The cross-section and the symbol of a PMOS transistor are depicted below.



The behaviour of the PMOS is very similar to that of an NMOS transistor, except that the polarities of the various voltages and currents are reversed. In other words, for a PMOS, the voltages V_{TP} (threshold voltage), V_{GS} , and V_{DS} are negative, whereas the drain current I_D flows from source to drain, thus implying that the source is always connected to the highest voltage while the drain is connected to the lowest voltage ($V_{DS} < 0$).

A PMOS transistor has three modes of operation:

- **First mode of operation:** The transistor is in the cut-off mode when $V_{GS} > V_{TP}$ ($V_{GS} < 0$, $V_{TP} < 0$). In this case, there is no conducting channel between drain and source, and the drain current I_D is thus equal to zero. The transistor is equivalent to an open switch since the drain is not connected to the source.

- **Second mode of operation:** The transistor is in the linear mode when $V_{GS} \leq V_{TP}$ and $V_{DS} > V_{GS} - V_{TP}$ ($V_{GS} < 0$, $V_{TP} < 0$, $V_{DS} < 0$). In this case, the drain current I_D is given by

$$I_D = k_p \left(V_{GS} - V_{TP} - \frac{V_{DS}}{2} \right) V_{DS},$$

where $k_p = \mu_p C_{ox} \frac{W}{L}$ is called the gain factor of the transistor. A typical value of k_p for $W/L = 10$,

$t_{ox} = 10 \text{ nm}$, $\mu_p = 260 \text{ cm}^2/\text{V}\cdot\text{s}$, and $\epsilon_{ox} = 3.5 \times 10^{-11} \text{ F/m}$ is $k_p = 260 \times 10^{-4} \times \frac{3.5 \times 10^{-11}}{10 \times 10^{-9}} \times 10 \approx 0.9$

mA/V^2 . The parameter μ_p designates the mobility of holes in the channel (since the current carriers are holes in a PMOS transistor). Note that the mobility of holes is lower than that of electrons. Due to the lower mobility of holes, the gain factor of a PMOS transistor is therefore smaller than that of its equivalent NMOS transistor (with identical physical dimensions). In fact, this explains why PMOS transistors do actually operate slower than NMOS transistors.

In this mode, the transistor resembles a voltage-controlled resistance given by

$$R_{PMOS} = \frac{dV_{SD}}{dI_D} = -\frac{dV_{DS}}{dI_D} = -\frac{1}{k_p (V_{GS} - V_{TP} - V_{DS})}.$$

- **Third mode of operation:** The transistor is in the saturation mode when $V_{GS} \leq V_{TP}$ and $V_{DS} \leq V_{GS} - V_{TP}$ ($V_{GS} < 0$, $V_{TP} < 0$, $V_{DS} < 0$). In this case, the drain current I_D is given by

$$I_D = \frac{k_p}{2} (V_{GS} - V_{TP})^2.$$

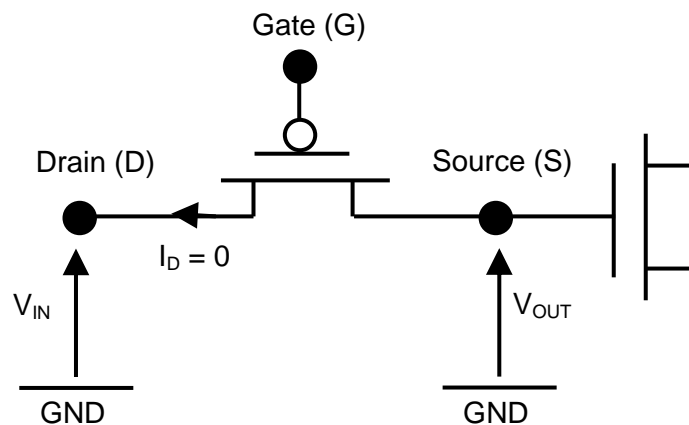
In this mode, the transistor is equivalent to a voltage-controlled current source.

We will have a closer look at PMOS transistors in Tutorial 5.

Tutorial 5 – MOSFET Transistors

Question 1

Consider the circuit depicted below, which shows a PMOS transistor used as a switch. All voltage values in the circuit vary from 0 (GND) to V_{DD} (supply voltage).



We recall the three modes of operation for a PMOS transistor:

- **First mode of operation:** The transistor is in the cut-off mode when $V_{GS} > V_{TP}$ ($V_{GS} < 0$, $V_{TP} < 0$). In this case, there is no conducting channel between drain and source, and the drain current I_D is thus equal to zero. The transistor is equivalent to an open switch since the drain is not connected to the source.

- **Second mode of operation:** The transistor is in the linear mode when $V_{GS} \leq V_{TP}$ and $V_{DS} > V_{GS} - V_{TP}$ ($V_{GS} < 0$, $V_{TP} < 0$, $V_{DS} < 0$). In this case, the drain current I_D is given by

$$I_D = k_p \left(V_{GS} - V_{TP} - \frac{V_{DS}}{2} \right) V_{DS},$$

where $k_p = \mu_p C_{ox} \frac{W}{L}$ is called the gain factor of the transistor. In this mode, the transistor resembles a voltage-controlled resistance given by

$$R_{PMOS} = -\frac{dV_{DS}}{dI_D} = -\frac{1}{k_p (V_{GS} - V_{TP} - V_{DS})}.$$

- *Third mode of operation*: The transistor is in the saturation mode when $V_{GS} \leq V_{TP}$ and $V_{DS} \leq V_{GS} - V_{TP}$ ($V_{GS} < 0$, $V_{TP} < 0$, $V_{DS} < 0$). In this case, the drain current I_D is given by

$$I_D = \frac{k_p}{2} (V_{GS} - V_{TP})^2.$$

In this mode, the transistor is equivalent to a voltage-controlled current source.

Show that, when the gate is connected to V_{DD} , the PMOS transistor is equivalent to an open switch.

Show that, when the gate is connected to GND, the PMOS transistor is equivalent to an imperfect closed switch.

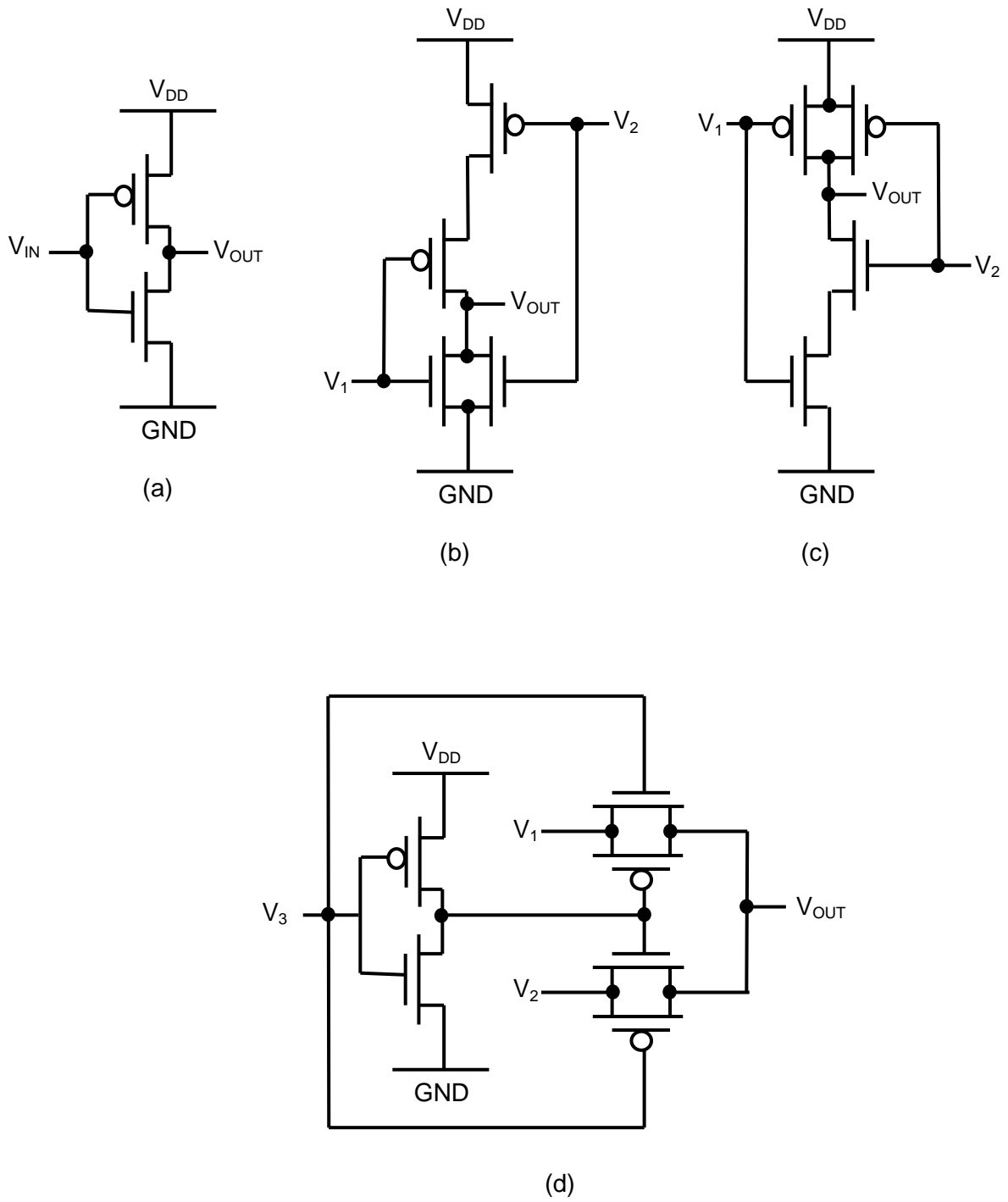
Question 2

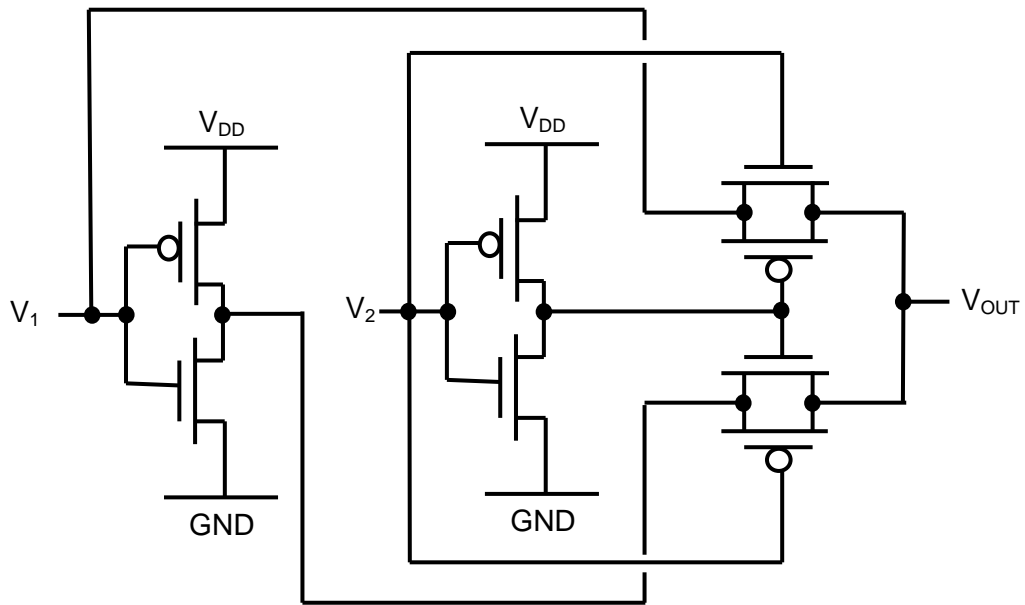
Design, using two MOS transistors, a switch that can be considered as perfect. When closed, this switch should be able to pass all voltage values ranging from 0 to V_{DD} , where V_{DD} designates the voltage supply.

Assume that the current flowing through the switch is equal to zero, which implies that the switch output is connected to the gate(s) of other MOSFETs.

Question 3

Consider the five CMOS (complementary MOS) circuits depicted below.





(e)

For each circuit, determine the output voltage V_{out} , assuming that the input voltages (denoted as V_{IN} , V_1 , V_2 and V_3) can be either low (equal to 0 volt) or high (equal to the supply voltage V_{DD}). The output of each circuit is connected to the gates of MOSFETs.

Determine the function performed by each circuit.

In each case, what is the value of the current drawn from the power supply? Comment on your result.

7. Design of Linear Amplifiers using MOSFETs

The design methodology for linear amplifiers using FETs is identical to that used for bipolar linear amplifiers. In both cases, we need to bias the transistor to make sure that it operates in the proper mode (forward active for BJTs, saturation for FETs) and then we need to consider the small-signal model of this transistor to analyze the performance of the amplifier thus obtained.

Hereafter, we are going to focus on the design of linear amplifiers using N-channel MOSFETs, often referred to as NMOS transistors.

• Biasing of MOS Transistors

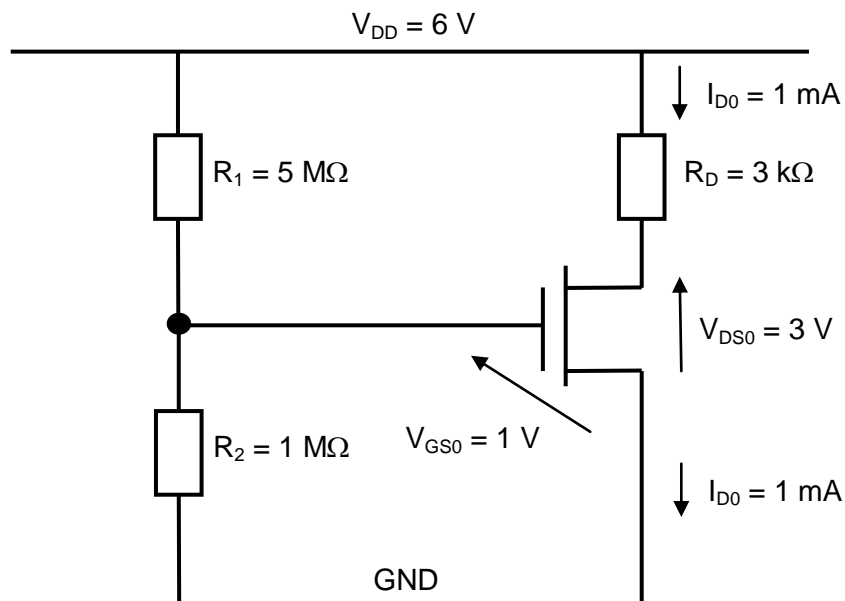
MOS transistors have three different modes of operation. The mode that is used to design MOS amplifiers corresponds to the constant-current region of the output characteristic (as was also the case with BJT amplifiers). This mode is the saturation mode of operation. A MOS transistor must remain in the saturation mode when used to design an amplifier, whatever the value of the input signal to be amplified. Biasing an amplifier circuit allows us to make sure that the MOS transistor is in the saturation mode.

Biasing a circuit consists of applying a constant (DC) voltage across the circuit (supply voltage) to ensure that the MOS transistor(s) inside this circuit always operate in the saturation mode.

Below is an example of biasing circuit for an amplifier using an NMOS transistor with a threshold voltage $V_{TN} = 0.5 \text{ V}$ and a gain factor $k_n = 8 \text{ mA/V}^2$. We assume that this transistor is in the saturation mode of operation. We will have to check later whether or not this assumption is right.

Due to the voltage divider formed by R_1 and R_2 , we can easily see that $V_{GS0} = \frac{R_2}{R_1 + R_2} V_{DD} = 1 \text{ V}$,

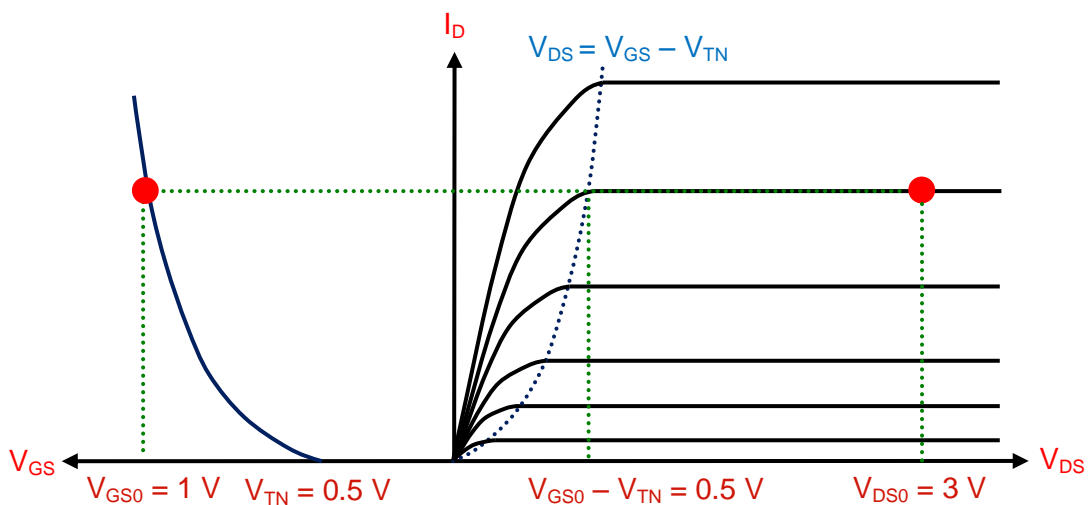
and thus $V_{GS0} - V_{TN} = 0.5 \text{ V}$. Since $V_{GS0} > V_{TN}$, the NMOS transistor is either in the linear mode or the saturation mode.



Once the value of V_{GS0} is known, we can determine the DC drain current by using the expression

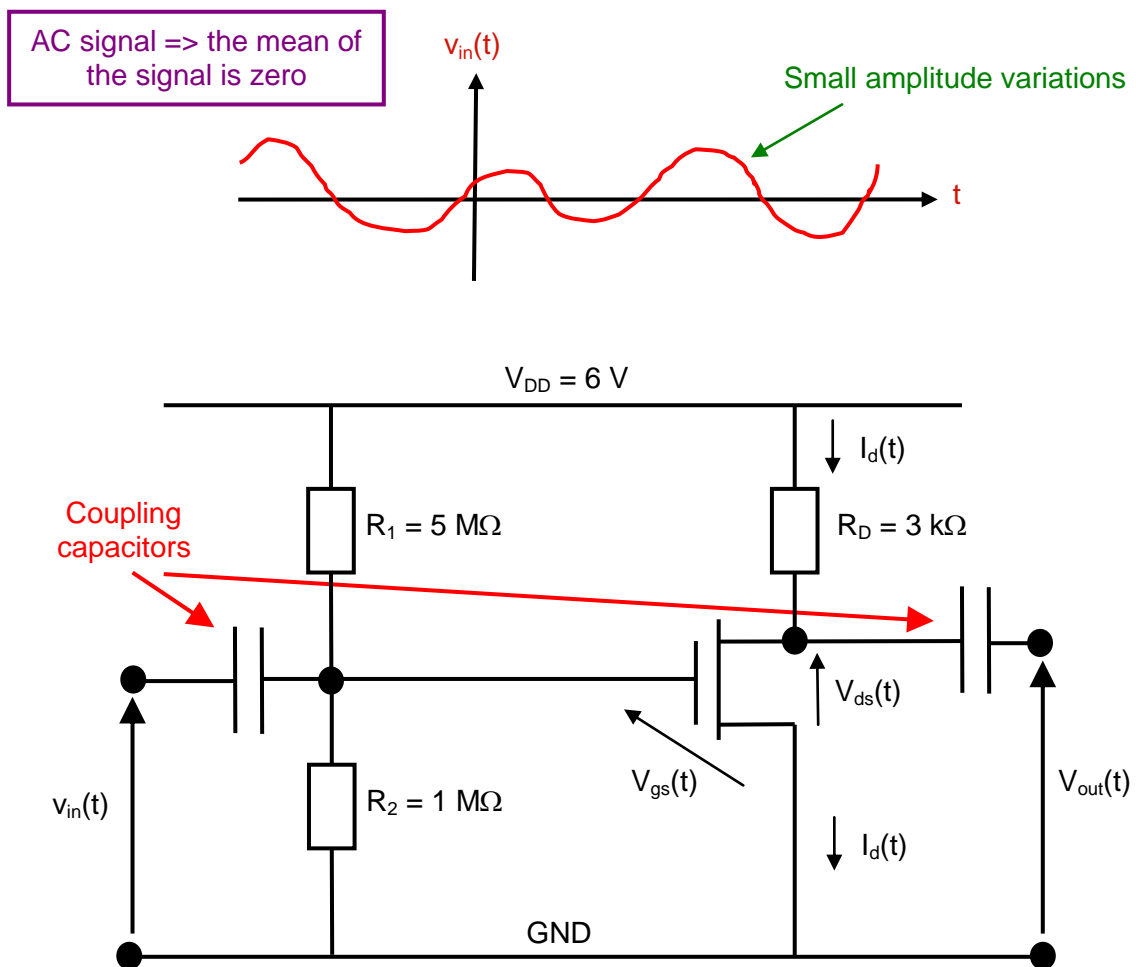
$$I_{D0} = \frac{k_n}{2} (V_{GS0} - V_{TN})^2. \text{ We find } I_{D0} = 1 \text{ mA.}$$

Finally, we can compute the value of the DC voltage V_{DS0} as $V_{DS0} = V_{DD} - R_D I_{D0} = 3 \text{ V}$. Since $V_{DS0} > V_{GS0} - V_{TN}$, we conclude that our assumption was right and the NMOS transistor is indeed in the saturation mode.



• Design of a Common-Source Amplifier

Our NMOS transistor is properly biased and can therefore be used to amplify an input signal $v_{in}(t)$. We assume that $v_{in}(t)$ represents a voltage variation that needs to be amplified. We can now study the complete voltage amplifier circuit which is depicted below. Note that this is a common-source configuration. The input voltage $v_{in}(t)$ is an AC signal with small amplitude variations.



The presence of both coupling capacitors protects the biasing circuit from any external influence. In other words, any change in the external environment of the amplifier circuit does not have any effect on the DC operating points.

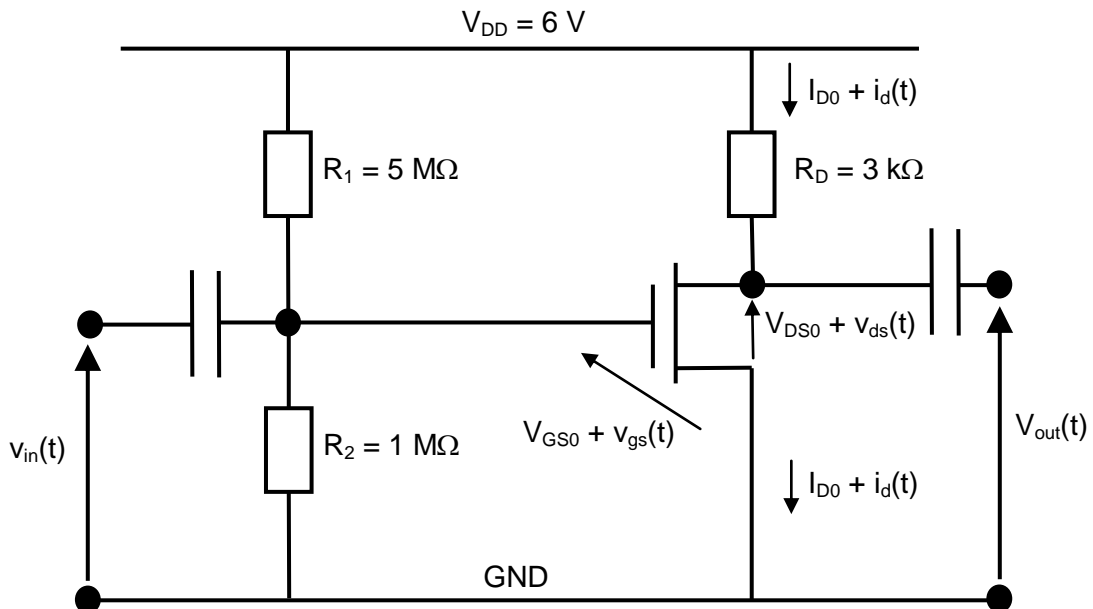
If there was a DC component in $v_{in}(t)$, it would be filtered out by the capacitance and would not affect the biasing of the circuit either. It is worthwhile mentioning that, as a result, the amplifier described here cannot amplify a DC signal.

We assume that both coupling capacitances are sufficiently high to be considered as short circuits (perfect wires) at the frequency of the input signal.

A variation of the input signal $v_{in}(t)$ will induce corresponding variations of the various voltages and currents inside the circuit around their DC values. For instance, a small variation of $v_{in}(t)$ will induce a corresponding variation of the drain current around its mean value $I_{D0} = 1 \text{ mA}$.

Any signal $X(t)$ in the circuit is therefore the sum of a DC component X_0 (mean value of $X(t)$) and an AC component $x(t)$, i.e. $X(t) = X_0 + x(t)$. The quantity X_0 was determined by the DC analysis of the circuit. In particular, we have:

- $I_d(t) = I_{D0} + i_d(t)$;
- $V_{gs}(t) = V_{GS0} + v_{gs}(t)$;
- $V_{ds}(t) = V_{DS0} + v_{ds}(t)$.



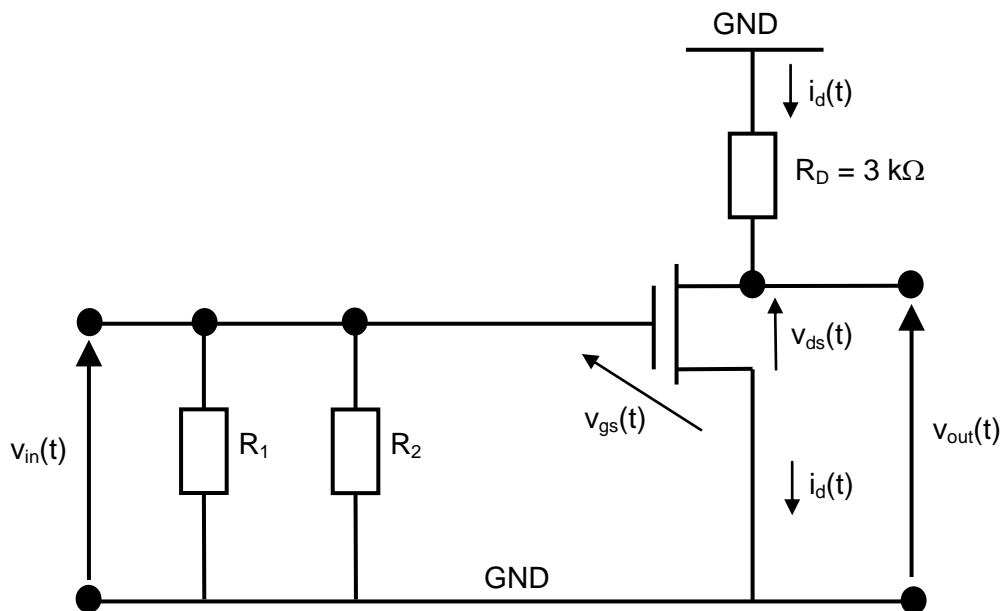
Note that the relationship between the drain current $i_d(t)$ and the voltage $V_{gs}(t)$ is quadratic, and therefore not linear. In other words, the amplification process will introduce some distortion of the signal, which may be a problem in some applications. However, this distortion can generally be considered negligible if the variation $v_{gs}(t)$ of the gate-source voltage around its DC value V_{GS0} is sufficiently small (since quadratic curve \approx straight line over small intervals).

The magnitude of $v_{in}(t)$ is considered to be sufficiently small so that the variation of $V_{ds}(t)$ around V_{DS0} allows the MOS transistor to remain in the saturation mode of operation.

The voltage gain of this amplifier is defined as $A_v = \frac{v_{out}(t)}{v_{in}(t)}$.

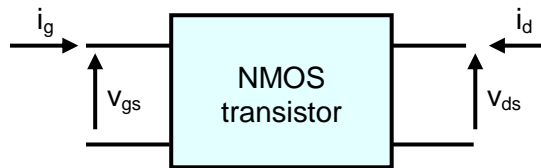
In order to compute the voltage gain of this amplifier, we now need to focus on the AC operation of the amplifier circuit. From an AC perspective, the amplifier circuit is equivalent to the circuit depicted below.

Remember that the supply voltage V_{DD} is a voltage equal to zero (ground) for the AC operation of the amplifier, thus implying that the resistance R_1 is now connected between the gate of the transistor and the ground.

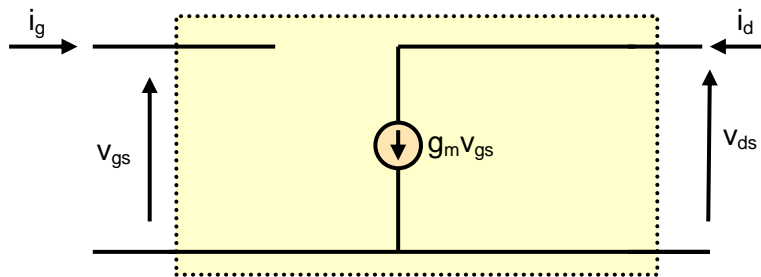


• **Small-Signal Model of a MOSFET**

In order to perform the AC analysis of the amplifier, we need to replace the NMOS transistor with an equivalent model valid for a small-amplitude input signal (“small signal”). The NMOS transistor is now seen as a “black box” with two inputs and two outputs.



We can easily show that the (generic) small-signal model of a MOSFET is as follows:



The transconductance g_m , expressed in siemens (S), of the NMOS transistor is given by

$$g_m = \frac{i_d}{v_{gs}} = \left. \frac{dI_D}{dV_{GS}} \right|_{V_{GS0}, I_{D0}}$$

Around the DC operating point, we can use the expression of the drain current valid in the saturation mode:

$$I_D = \frac{k_n}{2} (V_{GS} - V_{TN})^2$$

We thus obtain $g_m = k_n (V_{GS0} - V_{TN})$. For the amplifier consider here, we have $g_m = 4 \text{ mS}$.

At this stage, it is interesting to compare the transconductance of a MOSFET with that of a BJT.

The transconductance of a BJT is given by

$$g_m = \frac{i_c}{V_{be}} = \frac{i_c}{i_b} \frac{i_b}{V_{be}} = \frac{\beta_F}{h_{11}} \approx \beta_F \frac{I_{B0}}{V_T} = \frac{I_{C0}}{V_T}.$$

The BJT transconductance is thus proportional to the DC collector current I_{C0} , and we have

$$\frac{g_m}{I_{C0}} = \frac{1}{V_T}, \text{ where } V_T \text{ designates the thermal voltage } (V_T \approx 25 \text{ mV at room temperature}).$$

As for the MOS transconductance, we can also show that it depends on the drain current I_{D0} . In

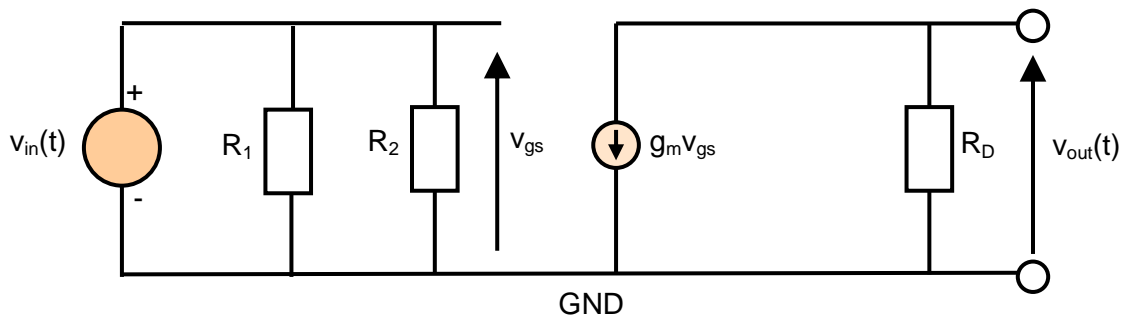
fact, the ratio $\frac{g_m}{I_{D0}}$ is given by

$$\frac{g_m}{I_{D0}} = \frac{k_n (V_{GS0} - V_{TN})}{\frac{k_n}{2} (V_{GS0} - V_{TN})^2} = \frac{2}{V_{GS0} - V_{TN}}.$$

Since the term $V_{GS0} - V_{TN}$ is typically much larger than the thermal voltage V_T , we conclude that, for a given DC output current I_{C0} or I_{D0} , the BJT transconductance tends to be much larger than the MOS transconductance. Therefore, a given variation of the input voltage (v_{be} , v_{gs}) tends to produce a much larger variation of the output current (i_c , i_d) with a BJT than with a MOSFET. In other words, bipolar transistors seem to be inherently more suitable than MOSFETS for the design of high-gain amplifiers.

• **Small-Signal Analysis of our Common-Source Amplifier**

We can now replace the NMOS transistor with its small-signal model in order to determine the expressions of the voltage gain, input resistance, and output resistance of our common-source amplifier.



We recall the values of the various parameters: $R_1 = 5 \text{ M}\Omega$, $R_2 = 1 \text{ M}\Omega$, $g_m = 4 \text{ mS}$, and $R_D = 3 \text{ k}\Omega$.

We can show that $v_{\text{out}}(t) = -R_D g_m v_{\text{gs}}$. Therefore, we obtain $v_{\text{out}}(t) = -g_m R_D v_{\text{in}}(t)$.

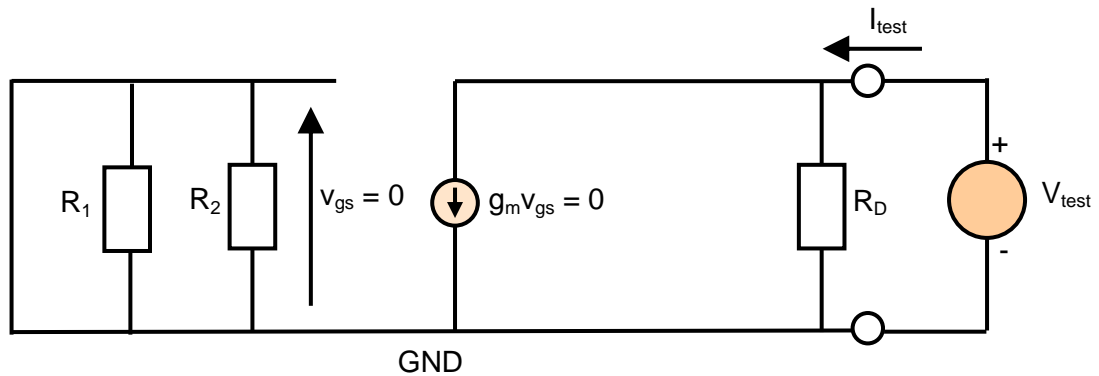
The voltage gain of our common-source amplifier is thus given by

$$A_v = \frac{v_{\text{out}}(t)}{v_{\text{in}}(t)} = -g_m R_D = -12.$$

This voltage gain is significantly lower than those typically achieved with a common-emitter bipolar amplifier (because of the lower transconductances).

The input resistance is given by $r_{\text{in}} = \frac{R_1 R_2}{R_1 + R_2} \approx 833 \text{ k}\Omega$. The input resistance of a MOS amplifier is typically much higher than that of a BJT amplifier due to the infinite input resistance of MOSFETs.

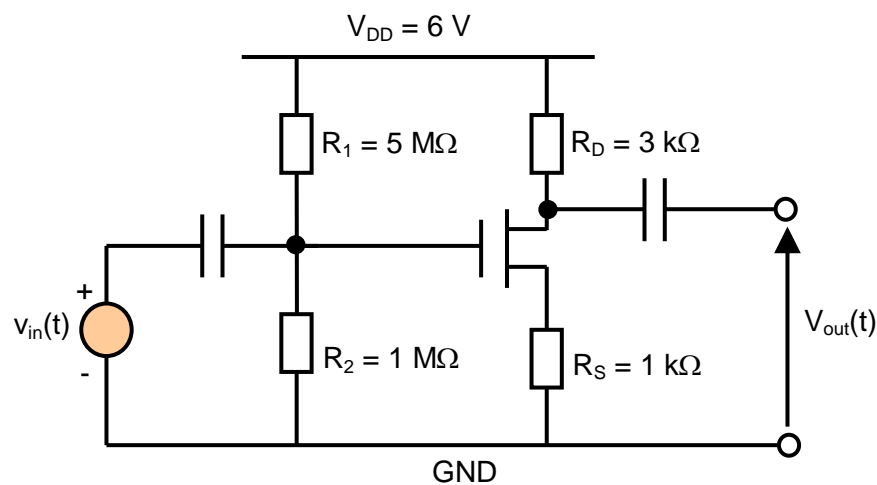
The output resistance is given by $r_{\text{out}} = R_D = 3 \text{ k}\Omega$. There is here no noticeable difference between common-emitter bipolar amplifiers and common-source MOS amplifiers.



Tutorial 6 – Amplifier Circuits using MOSFETs

Question 1

Consider the amplifier circuit depicted below.

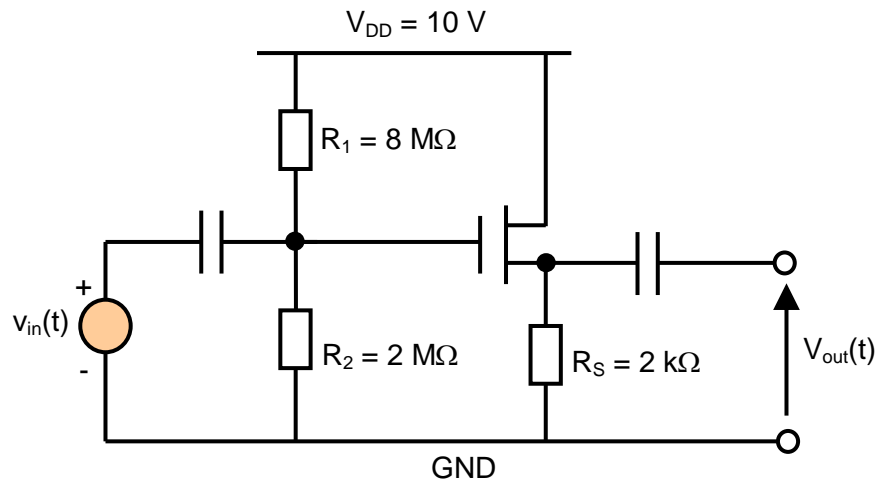


The only difference with the amplifier circuit described in the lecture notes is that we have added a source resistance R_S . The presence of this source resistance implements a negative feedback effect.

Perform the DC and AC analysis of this amplifier. Assume the use of an NMOS transistor with a threshold voltage $V_{TN} = 0.5\text{ V}$ and a gain factor $k_n = 8\text{ mA/V}^2$.

Question 2

Consider the amplifier circuit depicted below.



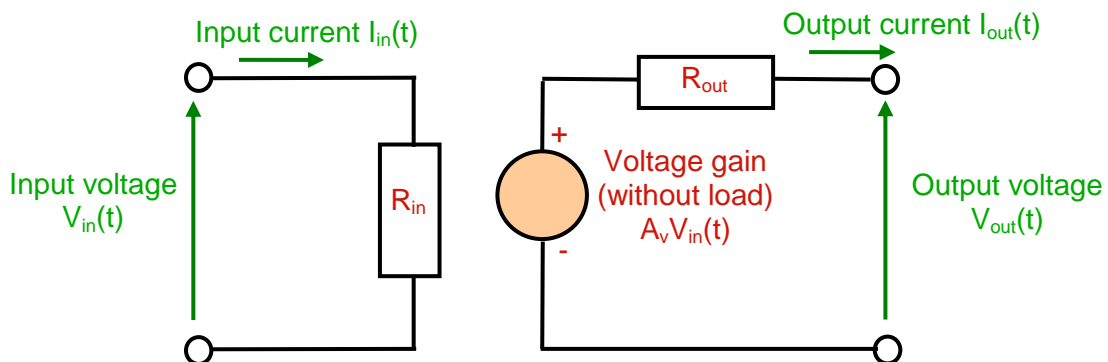
The drain is connected directly to the voltage supply. The input signal is applied between the gate and ground, while the output signal is measured between the source and ground. This arrangement is a common-drain amplifier.

Perform the DC and AC analysis of this amplifier. Assume the use of an NMOS transistor with a threshold voltage $V_{TN} = 1 \text{ V}$ and a gain factor $k_n = 8 \text{ mA/V}^2$.

8. Operational Amplifiers

• Basic Concept

The operational amplifier (op-amp) is an “ultimate” voltage amplifier in the sense that its characteristics are ideal (at least in theory). We recall that the equivalent circuit of a voltage amplifier is as follows:



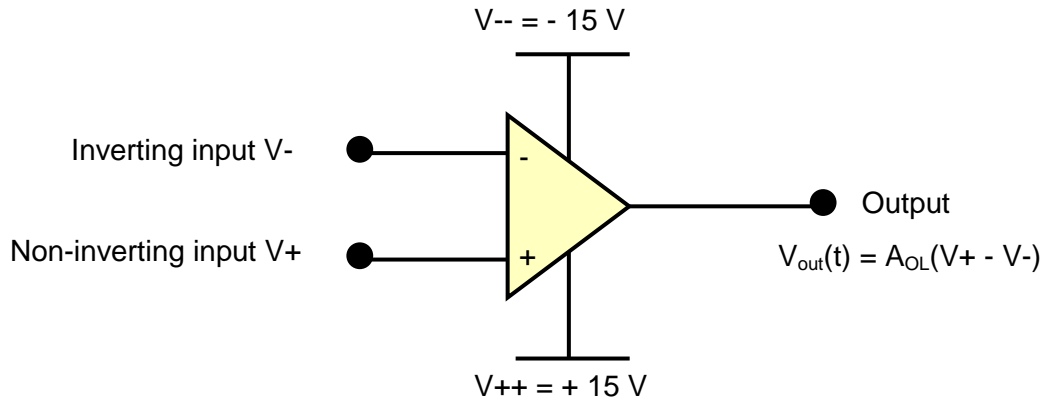
The characteristics of the op-amp are:

- Infinite input resistance (very high in practice);
- Zero output resistance (very small in practice);
- Infinite voltage gain (very high in practice).

The op-amp was initially designed to perform mathematical operations. Although now superseded in that area by the digital computer, op-amps remain a common feature of analogue electronic circuits.

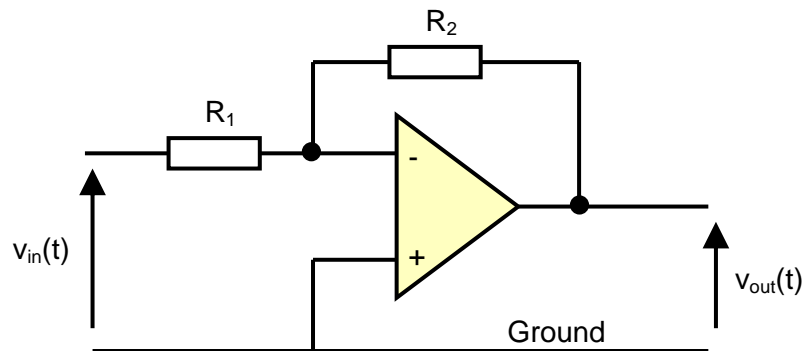
The open-loop voltage gain A_{OL} , for a typical op-amp, ranges from 10^3 to 10^6 . Such open-loop gain is too large to be useful since noise would cause the circuit to clip. Feedback must thus be used to control and stabilize the amplifier gain. Stabilization is obtained by feeding the output back into the input (feedback loop). In this way, the closed-loop gain does not depend on the amplifier characteristics.

The figure below shows a complete diagram of an op-amp. Typical values for the supply voltages are $V_{++} = 15\text{ V}$ and $V_{--} = -15\text{ V}$. The positive and negative voltage supplies are necessary to allow the amplification of both positive and negative signals without special biasing.



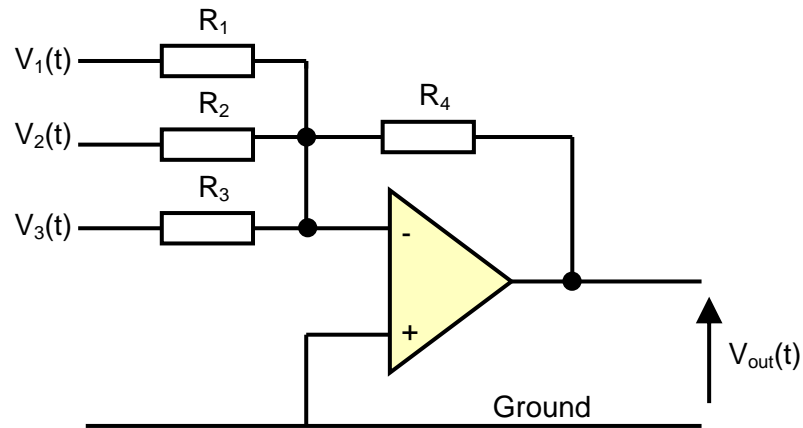
We are now going to study several “classical” circuits that employ operational amplifiers.

- **Inverting Amplifier**



The gain, G , of such device is given by $G = \frac{V_{out}(t)}{V_{in}(t)} = -\frac{R_2}{R_1}$.

• Summing Amplifier



The output signal is given by $V_{\text{out}}(t) = -\left(\frac{R_4}{R_1}\right)V_1(t) - \left(\frac{R_4}{R_2}\right)V_2(t) - \left(\frac{R_4}{R_3}\right)V_3(t)$.

If $R_1 = R_2 = R_3 = R$, we have $V_{\text{out}}(t) = -\frac{R_4}{R} [V_1(t) + V_2(t) + V_3(t)]$, and the output voltage is then proportional to the sum of the input voltages.

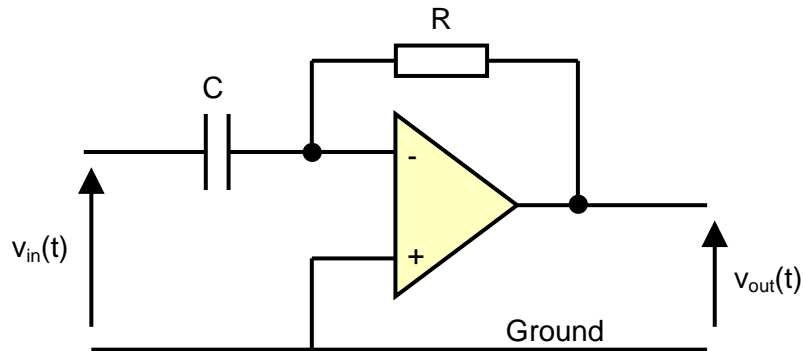
For only one input and a constant reference voltage V_{ref} , we obtain

$$V_{\text{out}}(t) = -\left(\frac{R_4}{R_1}\right)V_1(t) - \left(\frac{R_4}{R_2}\right)V_{\text{ref}},$$

where the second term represents an offset voltage. This provides a convenient method for obtaining an output signal with any required voltage offset.

• Differentiation Circuit

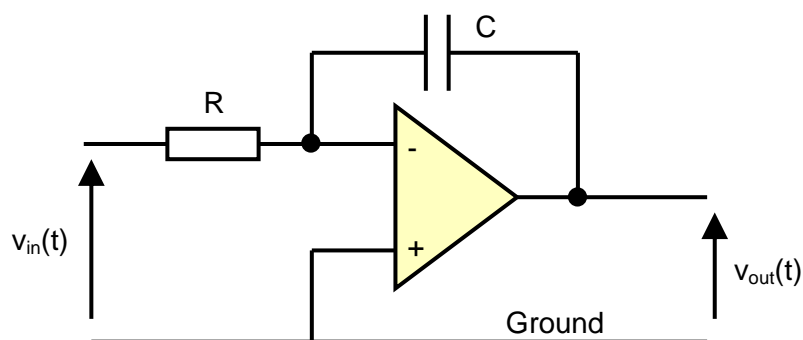
To obtain a differentiation circuit, we replace the input resistor of the inverting amplifier with a capacitor.



The output signal is expressed as $V_{out}(t) = -RC \frac{dV_{in}(t)}{dt}$, and is therefore proportional to the derivative of the input signal.

• Integration Circuit

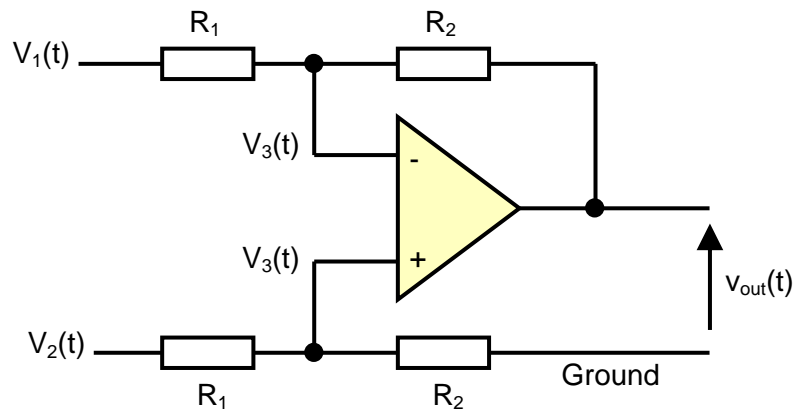
Integration is obtained by reversing the resistor and the capacitor. The capacitor is now in the feedback loop.



We have $\frac{dV_{out}(t)}{dt} = -\frac{1}{RC} V_{in}(t) \rightarrow V_{out}(t) = -\frac{1}{RC} \int V_{in}(t) dt$. The output signal is proportional to the integral of the input signal.

• **Differential Amplifiers**

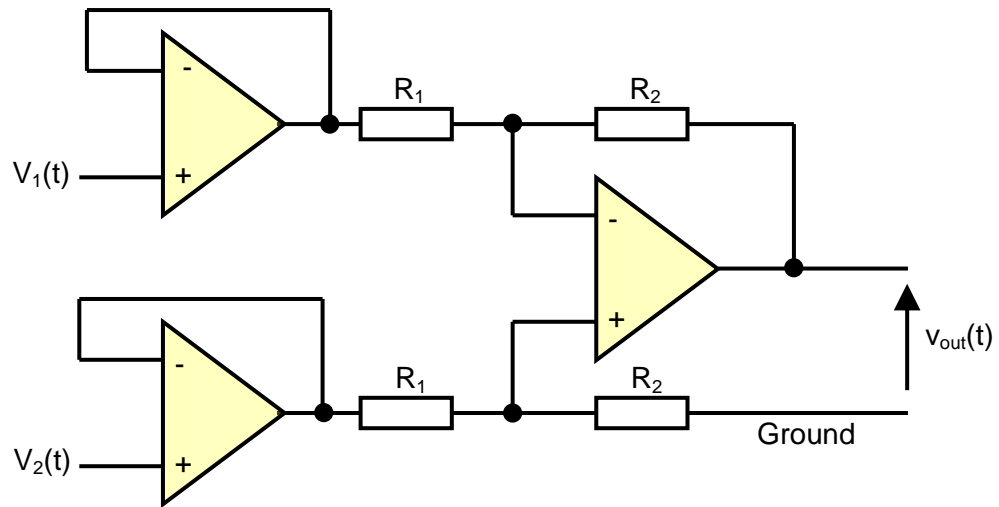
The differential amplifier amplifies the difference between two input signals (-) and (+). It forms the central basis of more sophisticated instrumentation amplifier circuits.



The voltage divider rule tells us that $V_3(t) = \frac{R_2 V_2(t)}{R_1 + R_2}$.

In addition, we have $\frac{V_1(t) - V_3(t)}{R_1} + \frac{V_{out}(t) - V_3(t)}{R_2} = 0$, which yields $V_{out}(t) = \frac{R_2}{R_1} [V_2(t) - V_1(t)]$.

Such differential amplifier is usually limited in its performance by its low input impedance. Two buffer amplifiers are commonly added to remove this limitation as follows:



- END -