Lecture 13: Transistors: overview

Contents

1 Introduction

pn junctions are examples of *2 terminal 1 junction* devices. There is one interface (junction) between the two components (can be same or different materials). A transistor is an example of a *3 terminal 2 junction* device. The name transistor is derived from the term *transfer resistance*. *The current (voltage) through two terminals is controlled by the current (voltage) through another pair of terminals.* Thus, a transistor essentially acts as a *switch*. Another feature of the transistor is the ability to *amplify* a signal between one pair of terminals by using an input signal at another pair of terminals. Thus, a transistor can also act as an amplifier. A *pn* junction, on the other hand acts as a rectifier i.e. conducts in only one direction.

2 Bipolar junction transistor

A bipolar junction transistor (BJT) consists of three differently doped regions. These can have the configuration of *npn* or *pnp* and the various layers can either be parallel or perpendicular to the surface. Consider a *pnp* BJT, with three differently doped regions.

1. **Emitter region** - this is usually a heavily doped region (p^+). The emitter 'emits' the carriers into the base.

2. **Base region** - this is a lightly doped n region. The base region is also physically thin so that carriers can pass through with minimal recombination.

3. **Collector region** - this is a p type region. The collector region has a larger width that the other two regions since charge is accumulated here from the base. .

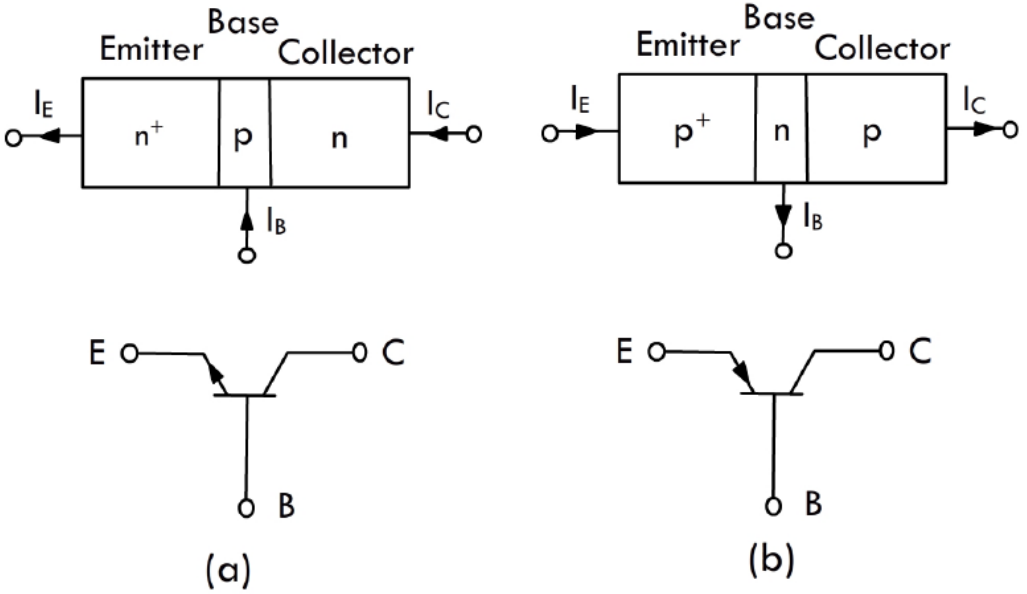
The symbols and nomenclature of the BJT transistor is shown in figure 2. Thus, a transistor consists of two pn junctions, each with its own depletion region.

1. Emitter-base junction - since the emitter is usually heavily doped, the depletion region lies almost entirely in the base.

2. Base-collector junction - the depletion region at this junction is usually divided between base and collector, since they are comparably doped.

There are three different configurations in which the BJT can function - common base, common emitter, and common collector. The circuit connections are summarized in figure 3.

Figure 2: Symbols and nomenclature of a (a) npn and (b) pnp transistor. The BJT consists of three regions, emitter, base, and collector. The emitter and collector are usually of one type of doping, while the base is another doping type



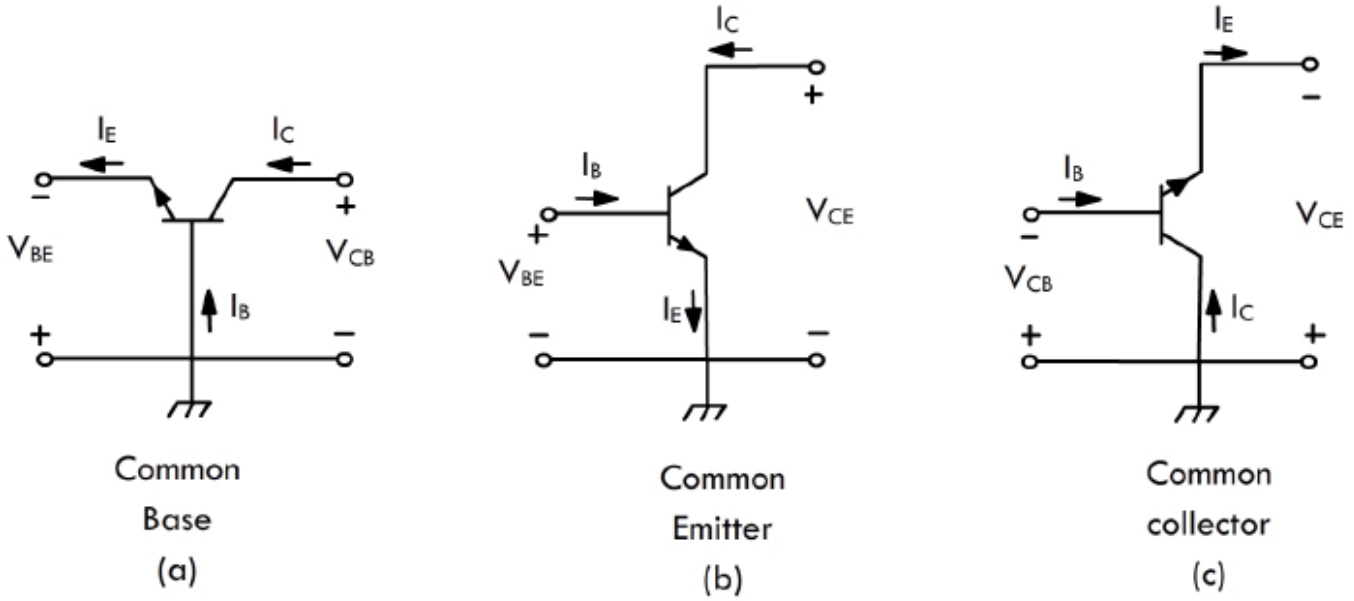


Figure 3: BJT transistor configurations - common (a) base, (b) emitter, and (c) collector configurations. Different configurations have different functionality in the circuit

2.1 Common base (CB) configuration

In this configuration the base is held common between the emitter and collector. This arrangement is summarized in figure 4. Looking at the circuit configuration, the emitter base junction is forward biased while the collected base junction is reverse biased. In a *pnp* transistor holes, which are the majority carriers in the emitter, are injected into the base. This constitutes the emitter current, I_E. These holes become minority carriers in the base, since it is doped n type. Some of these holes recombine with the electrons in the base. Hence, electrons are injected into the base to compensate for that and this forms the base current, I_B. The base collector junction is reverse biased so that the holes that do not recombine in the base are swept into the collector and form the collector current, I_c. This current is due to hole drift since the flow is due to the applied electric field. The ratio of the collector current to the emitter current is called the *current gain* or the *current transfer ratio* (α). Typical values of α are 0.99 - 0.999, i.e. only a small fraction is lost to recombination in the base. This is a consequence of the fact that the base region is physically thin and also lightly doped, so that the hole diffusion length is high. The total current in the circuit should be balanced, so that

$$I_E = I_C + I_B.$$

→ $I_E \approx I_C$ since I_B is very small

The transistor action arises in the CB configuration when the collector-base voltage (V_{CB}) is higher than the emitter-base voltage (V_{EB}). This leads to a net power gain in the device. A BJT is an example of a current controlled device since the current in the output circuit is controlled by the current and voltage in the input circuit.

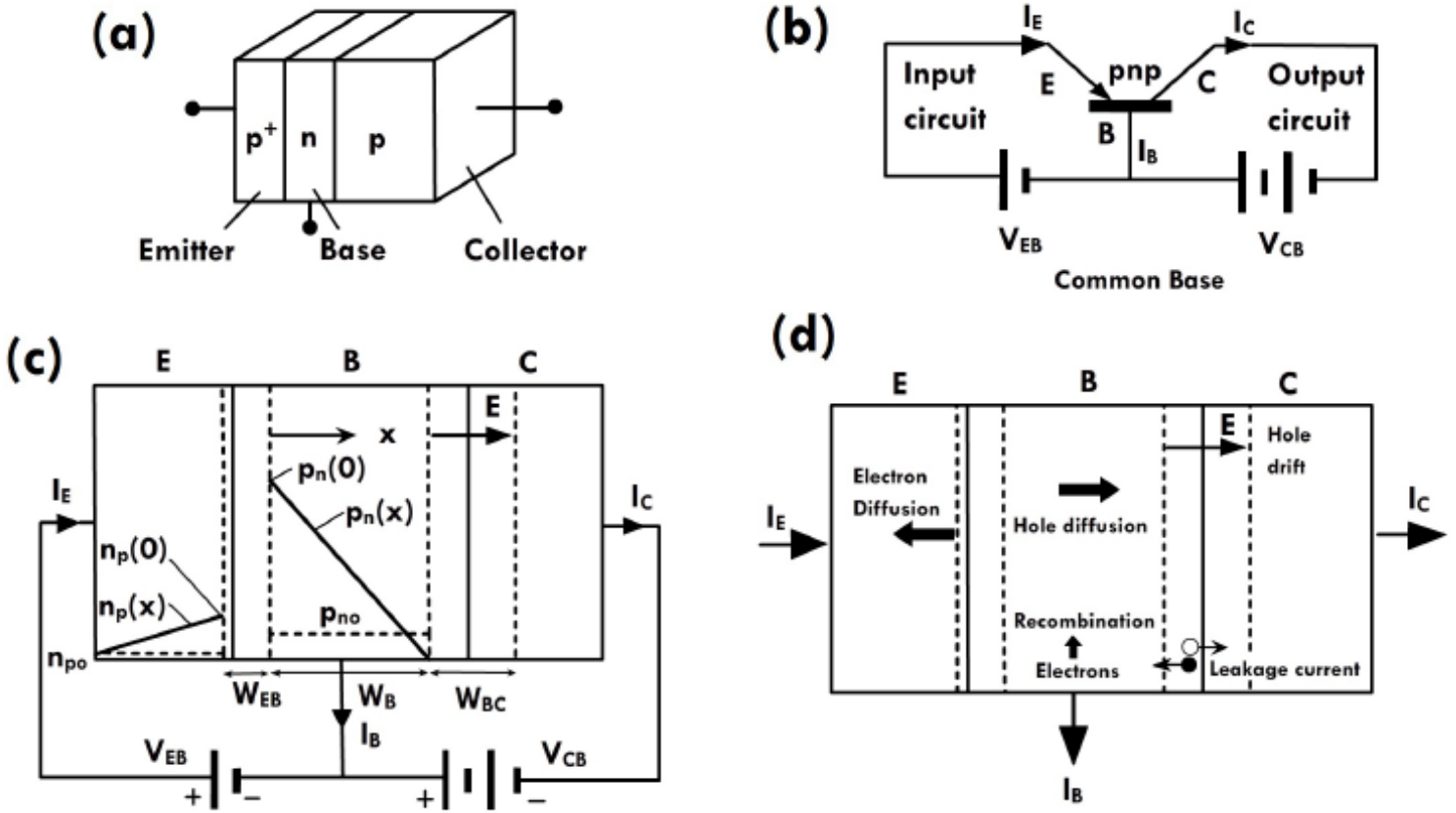


Figure 4: Summary of the common base BJT. (a) Schematic of the CB p^+np BJT (b) Circuit diagram showing the connections. (c) Schematic of the currents and concentration gradients. (d) The various diffusion and drift currents in the transistor.

2.2 Common emitter (CE) configuration

The circuit diagram for the CE configuration is summarized in figure 5. In this configuration, the emitter-base junction is forward biased. The base current is the input current and the collector current is the output. Hence, the transistor acts as a amplifier since a small base current is amplified into a larger collector current. This amplification is given by the current transfer ratio (α). If a time varying signal is applied to the base the amplified signal at the collector has the same time variation.

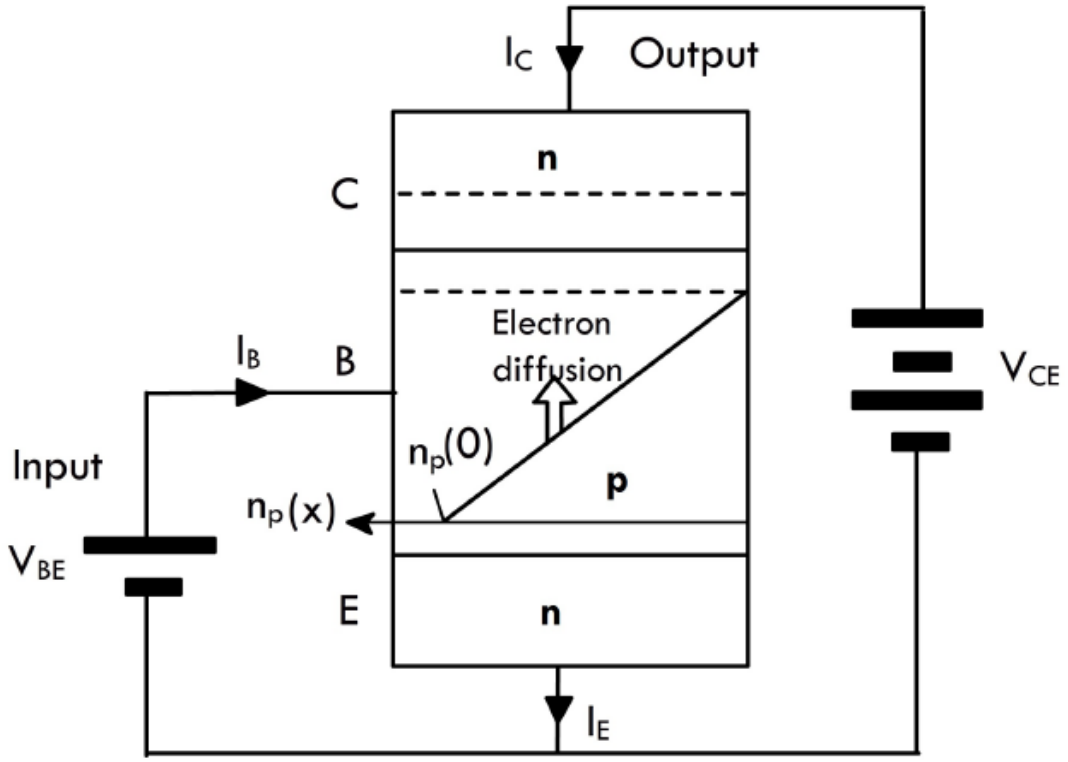
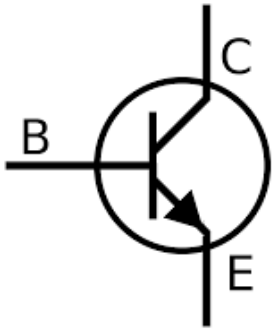
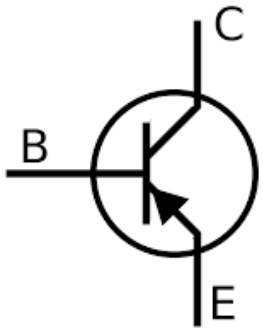


Figure 5: A common emitter configuration for the npn BJT. This configuration is used when a transistors is to be used as an amplifier. A small variation in base current is amplified at the collector, which acts as the out-put

Transistors symbols and basic circuits



NPN
Bipolar Junction
Transistor



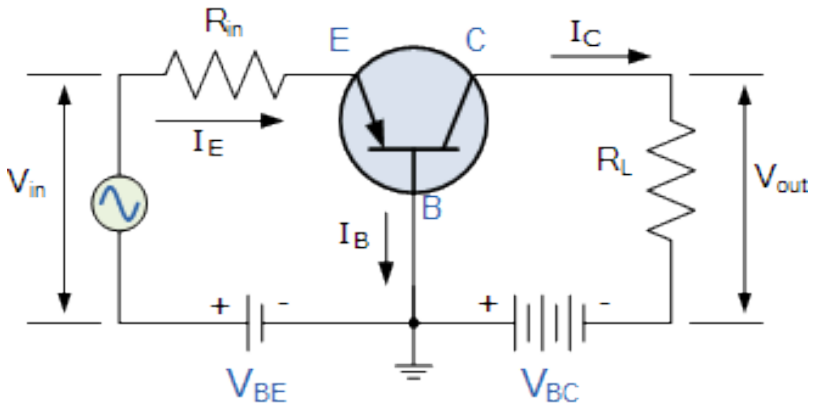
PNP
Bipolar Junction
Transistor

BJT is a current controlled device, For BJT configurations

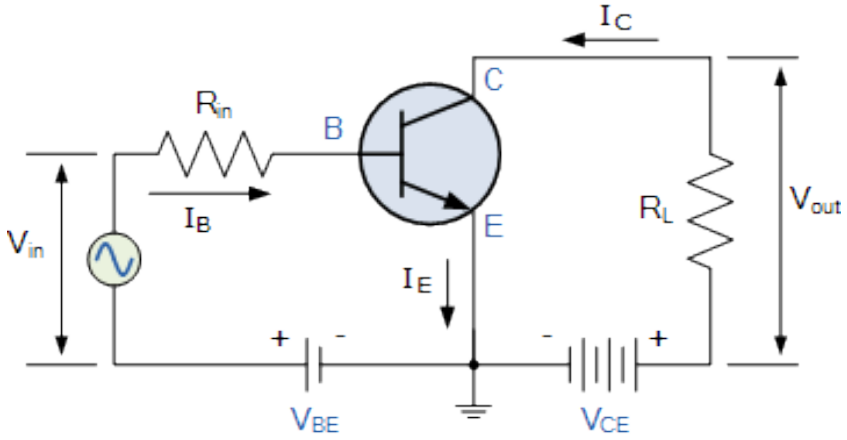
$$\text{Alpha},(\alpha) = \frac{I_C}{I_E} \quad \text{and} \quad \text{Beta},(\beta) = \frac{I_C}{I_B}$$

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

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



Common Base (CB) configuration for pnp transistor



Common Emitter (CB) configuration for npn transistor

3 Junction field effect transistor

In the BJT there are distinct emitter, base, and collector regions, forming 2 *pn* junctions. These regions exist even in the absence of an external bias, since they are created during the fabrication process. The device is current controlled, since carrier injection into the base and loss due to recombination decides the gain. In a field effect transistor (FET) current flow takes place through a channel in the device. The channel is either already fabricated in the device (junction FET) or is created by application of an external potential (metal oxide FET) and disappears when the bias is removed. The reason for calling these *'field effect'* transistors is that the current flow depends on the width of the channel which is controlled by the external potential (electric field). Thus FETs are *voltage controlled* devices.

The basic structure of the JFET is shown in figure 6. The schematic picture of the device provides a greater understanding of the role of the channel. A JFET consists of 3 regions where electrical connections are made i.e. the *source*, *drain*, and *gate*. The channel provides the pathway for carrier transport from the source to the drain, while the gate bias (sign and magnitude)

affects the channel width and hence its resistance. By considering the behavior of the channel for different gate biases, it is possible to obtain the I-V characteristics of the JFET.

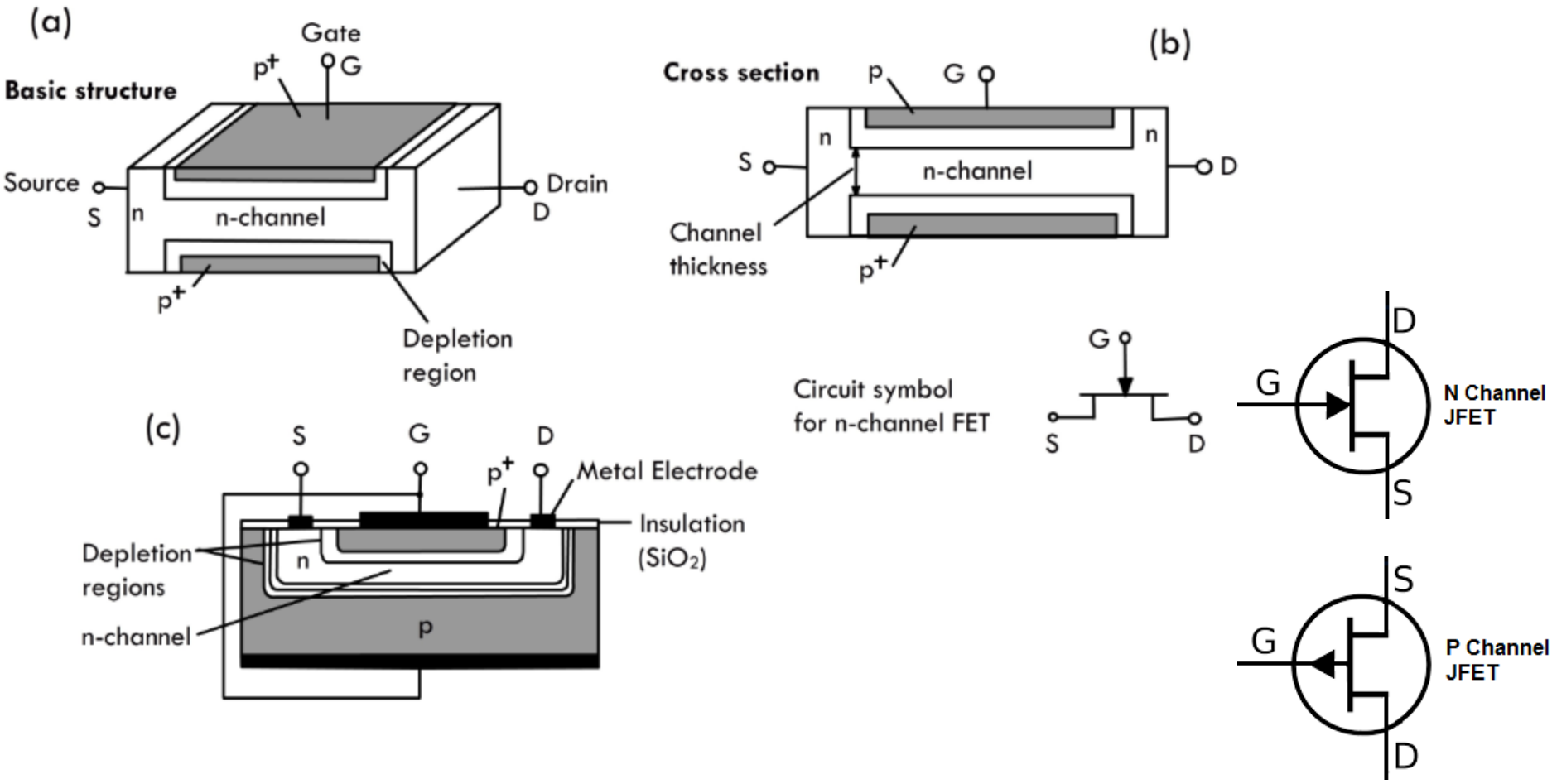


Figure 6: A junction field effect transistor (JFET). (a) Three dimensional representation of the JFET. (b) Cross section of the ideal JFET, showing the n-channel and the transistor symbol. (c) A practical implementation of the JFET. The electrical leads are on top and they are separated by using SiO as the insulator.

Consider a simple system where the gate is shorted with respect to the source. So V_{GS} is zero and the drain is biased positively with respect to the source, so that $V_{DS} > 0$. For a n channel, this leads to electrons flowing from the source to the drain (biasing would be reversed for a p channel). The current flow in this scenario, with increasing V_{DS}, is summarized in figure 7. The p regions (source and drain) in the transistors are heavily doped so that the depletion region falls in the n side and the channel width is the region between the 2 depletion regions. In the bias condition shown in figure 7, the region between the source and the p is forward biased while the region between the drain and p is reverse biased. With increase in V_{DS} the current flow in the channel increases but at the same time the channel starts to narrow near the drain side. Ultimately, beyond a certain value of V_{DS}, the channel is *pinched off* near the drain end. Hence, there is no increase in current, for a small increase in voltage, since there is injection of electrons in the pinched off region and these get swept into the drain. This is shown in the I-V characteristics of the JFET (with V_{GS} zero) in figure 8. The current initially increases with V_{GS} and then gets saturated after a certain voltage when pinch off occurs.

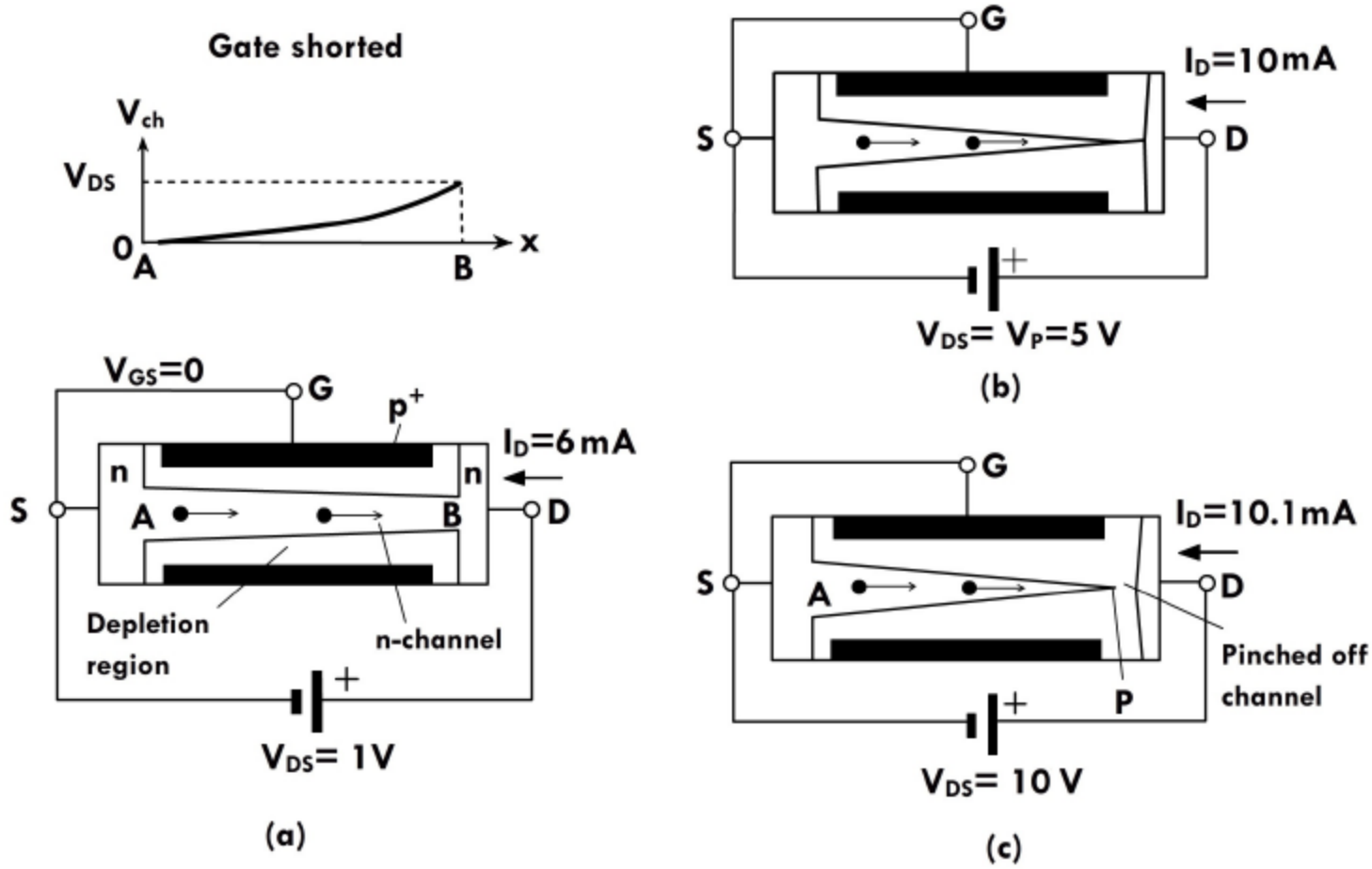


Figure 7: Carrier flow in a n channel with the gate shorted (V_{GS} =0). (a) With low V_{DS} a current flows through the channel and increase with increasing V_{DS} (b) With further increase in V_{DS} the channel pinches off near the drain, since the drain gate junction is reverse biased. c) After pinch-off, there is no further increase in current, reaching a saturation.

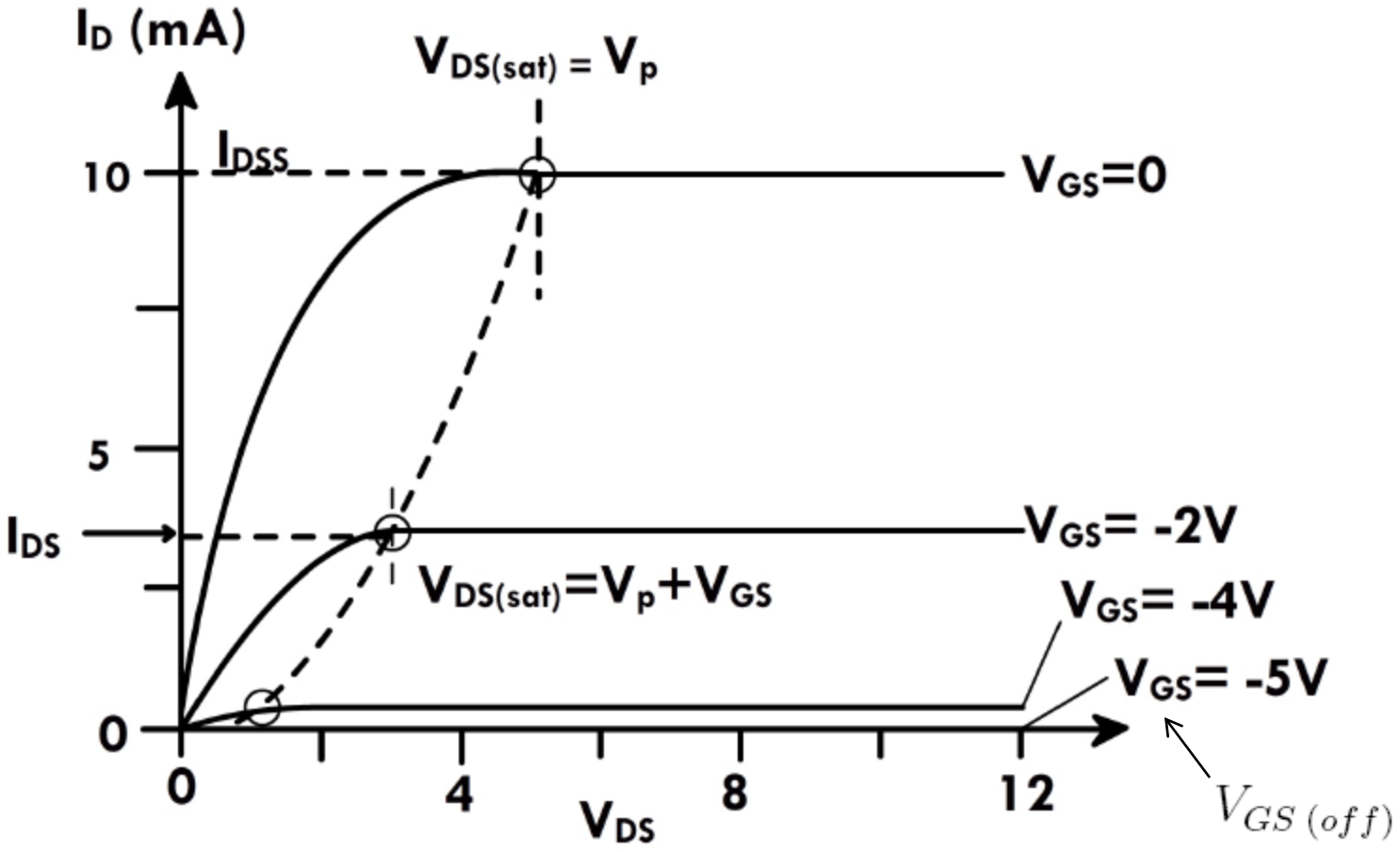


Figure 8: IV characteristics of the JFET for different V_{GS}. The current is highest when the gate is shorted. Applying a negative bias at the gate reduces the width of the channel and reduces the channel conductivity.

In a JFET device with the gate shorted, the channel width is determined by the dopant concentrations which also determines the voltage where pinch-off occurs. But by applying a potential to the gate it is possible to change the width of the channel. This is responsible for the transistor action, since the current across the two terminals (source and drain) is controlled by the voltage across two other terminals (source and gate). In a n channel JFET if $V_{GS} > 0$ then the channel width will slightly increase but device modulation is not achieved (current control does not happen). On the other hand, if the gate is biased negatively with respect to the source, $V_{GS} < 0$, the channel width is reduced so that pinch off occurs earlier and the voltage where pinch-off occurs is determined the magnitude of V_{GS}. The channel behavior for a negatively biased gate is summarized in figure 9.

Negatively biased gate

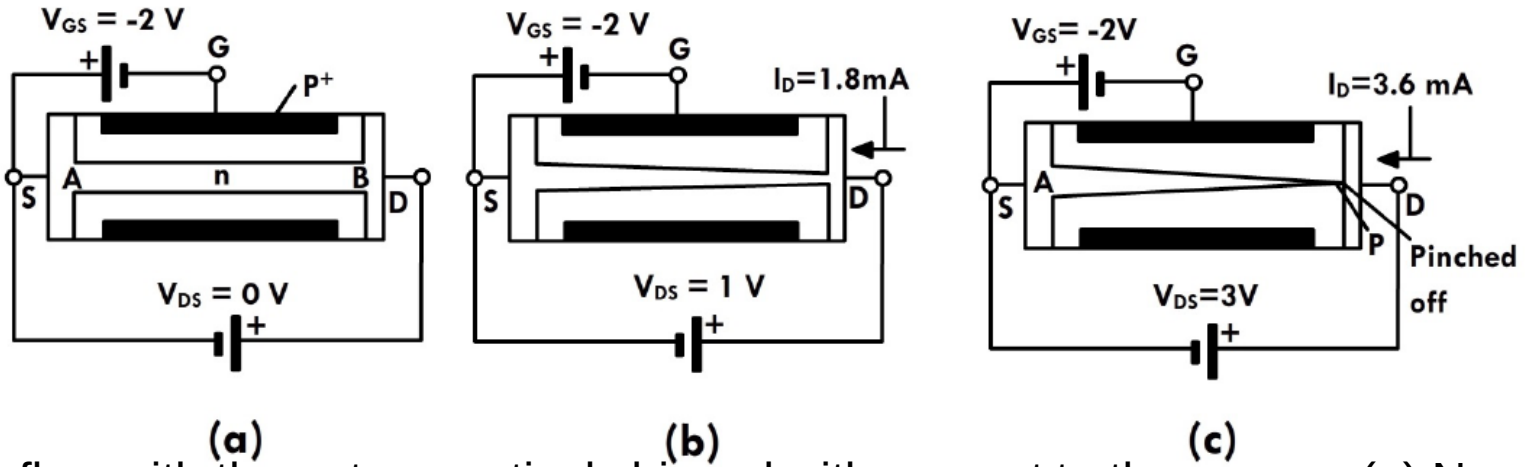


Figure 9: Carrier flow with the gate negatively biased with respect to the source. (a) No V_{DS}, n-channel is narrower than a shorted gate (b) With positive V_{DS}, current flows occurs with channel narrowing near the drain. (c) Pinch-off happens at high V_{DS} (lower than that for $V_{GS} = 0$)

Thus, with increasing value of V_{GS} (negatively biased), the V_{DS} at which pinch-off occurs comes earlier and also the current through the channel decreases. This is reflected in the I-V characteristics plot, shown in figure 8. With increasingly negative V_{GS}, the channel current decreases. There is also a critical value of V_{GS} where the pinch off occurs before the device is in operation. This happens when the depletion width becomes wide enough that the channel is completely destroyed. In this scenario, no current flows through the channel, except for small leakage current due to thermal generation of the channel, except for small leakage current due to thermal generation of carriers. This is shown in figure 10. The drain current (I_D) in the channel depends on the value of V_{GS} (field effect) by the expression

$$I_{DS} = I_{DSS}\,[1 - (\frac{V_{GS}}{V_{GS\,(off)}})]^2 \tag{1}$$

where I_{DSS} is the drain current with the gate shorted and $V_{GS\,(off)}$ is the gate voltage when the channel is destroyed. This behavior is plotted in figure 11.

Figure 10: At large negative bias (when $V_{GS}= -V_P = V_{GS(off)}$) the channel is completely destroyed (closed). There is only a small leakage current, similar to the reverse saturation current in the pn junction.

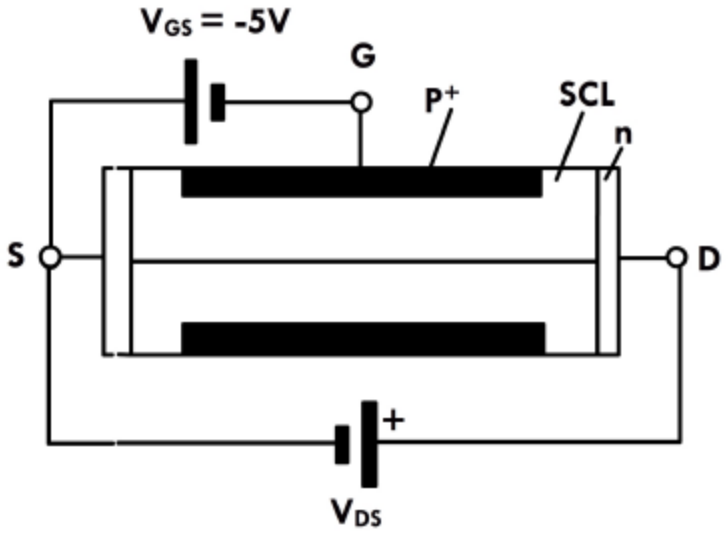
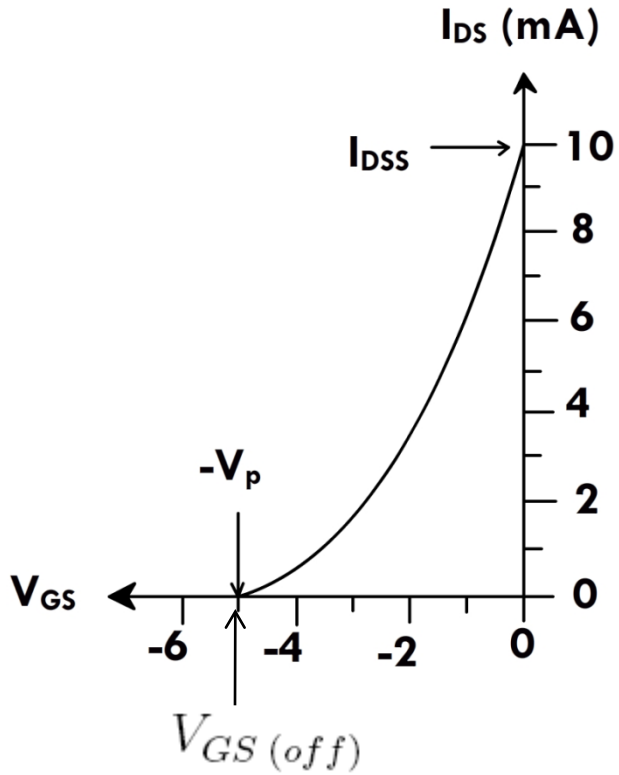


Figure 11: Drain current vs. gate source voltage in a JFET. With increasingly negative VGS, the drain current reduces until it becomes zero. Maximum current (I_{DSS}) is when the gate is shorted.



4 Metal oxide semiconductor FET

The metal oxide semiconductor FET (MOSFET) is a particular type of FET where the channel for carrier conduction is created by the applied electric field. In this way MOSFET is different from the JFET, where the channel is already present in the material, and the applied electric field is used to narrow the channel to achieve transistor action. To understand the formation of the channel, under the effect of the applied field, consider a parallel plate capacitor arrangement, involving two metals, as shown in figure 12. Metals have a high electron density (typically 10^{21} to 10^{22} cm^{-3}). Hence the charges reside on the surface with minimal field penetration into the bulk.

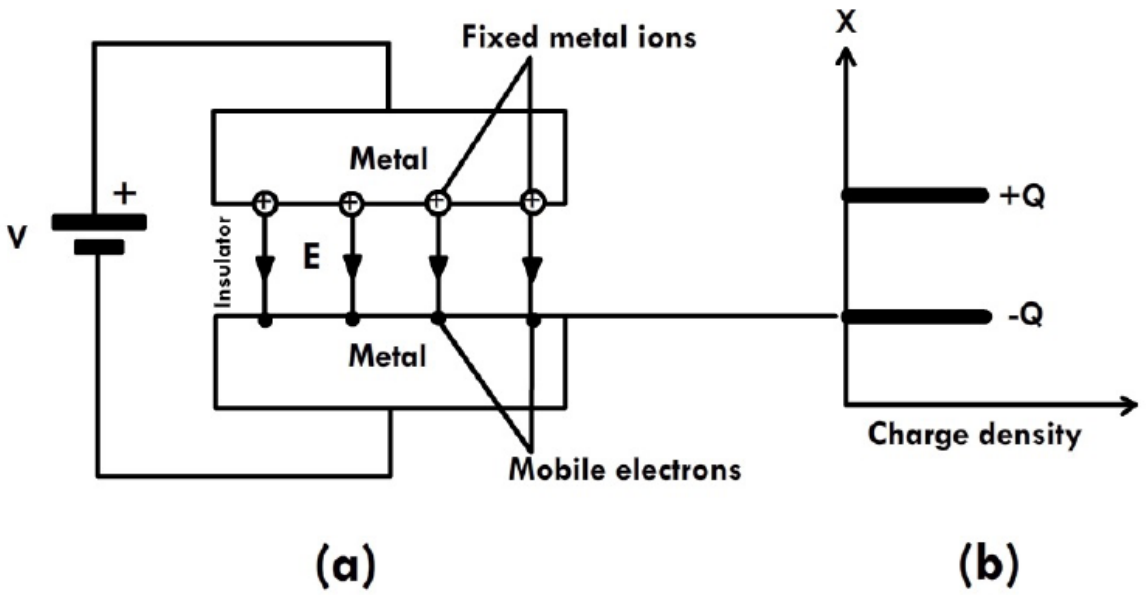


Figure 12: Parallel plate capacitors with two metals, separated by an insulator. (a) One metal plate has a net positive charge on the surface and the other has a net negative charge. (b) The excess charges reside on the surface and do not penetrate in the bulk.

Consider a situation when one of the metals is replaced by a p-type semiconductor. This scenario is shown in figure 13. The metal is connected to the positive terminal and the p type semiconductor is connected to the negative terminal. So a net positive charge resides on the metal surface. To maintain charge neutrality, a net negative charge must reside on the semiconductor. But the charge density in a semiconductor is much smaller than a metal (typically 10^{16} to 10^{18} cm^{-3}), so that in the semiconductor the charge not only resides on the semiconductor but also penetrates to a certain depth within the bulk, as shown in figure 13(a). Given that it is a p type semiconductor and the excess charges are electrons there is thus a *lowering* of the hole concentration at a region near the surface. This is called a **depletion region**. With increase in applied voltage, the positive charge on the metal increases and the width of the depletion region in the semiconductor increases. Correspondingly, the electron concentration on the surface of the semiconductor increases. There is a certain voltage, called **threshold voltage**, V_{th}, above which there forms a region near the semiconductor surface where the electron concentration is higher than the hole concentration. This is called the **inversion region**. A *n channel* is then created near the surface of a p type semiconductor by the application of external potential. The formation of the inversion region, is shown in figure 13(b). In the inversion region, $n > p$, while in the depletion region, $p > n$ but $p \ll N_A$.

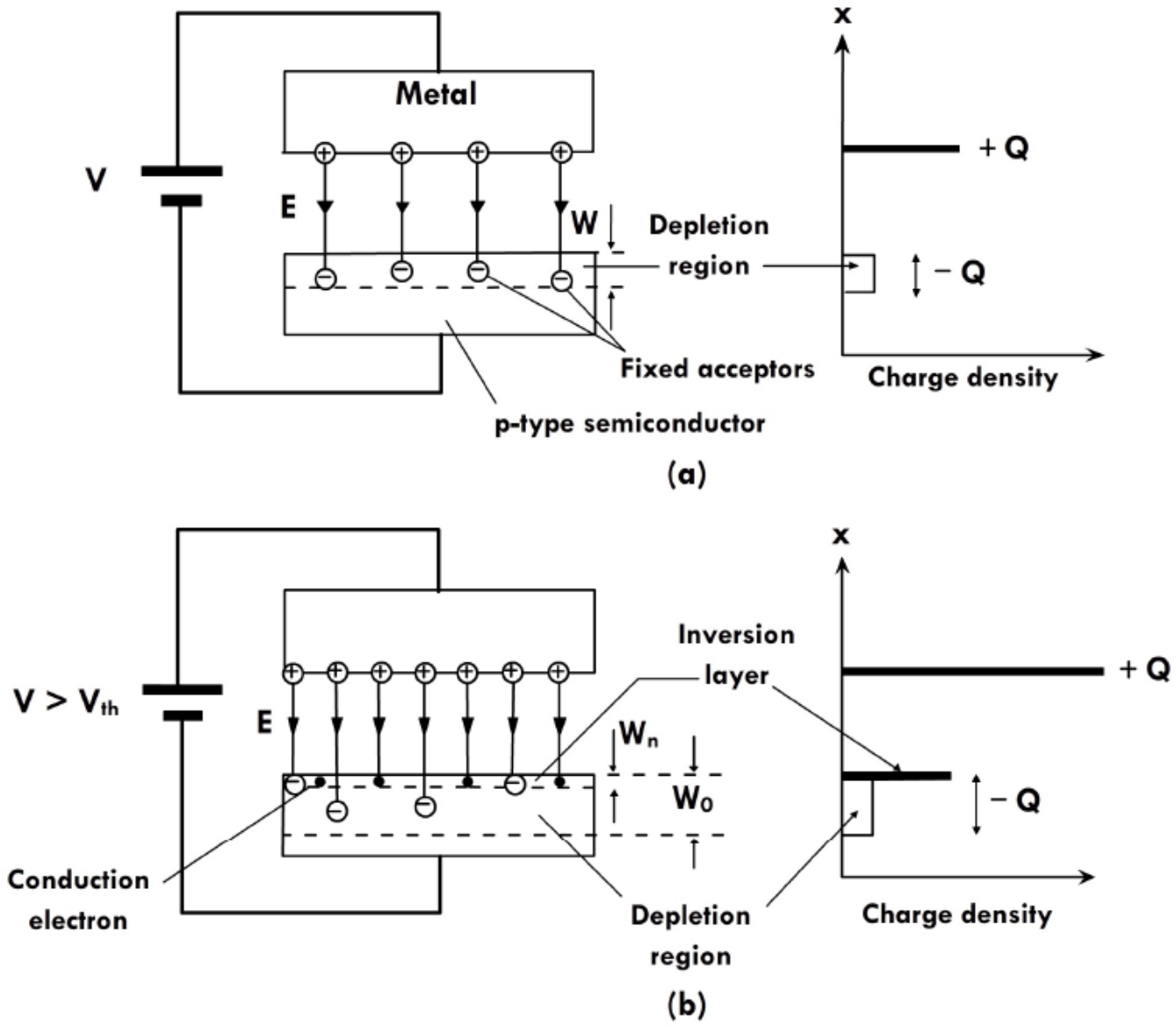
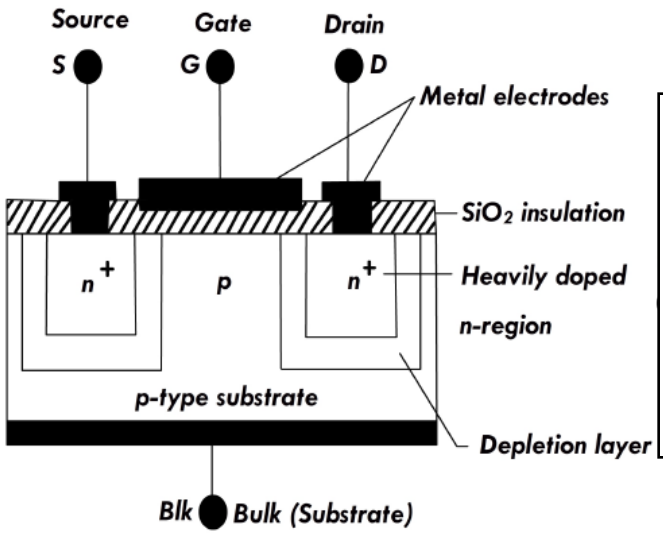


Figure 13: Metal insulator semiconductor setup. Because of the difference in charge density between the metal and semiconductor, charges penetrate into the bulk of the semiconductor creating (a) Depletion region (b) Inversion (n-channel) and depletion at higher voltages

The basic structure of the MOSFET is shown in figure 14. The figure shows a *npn* MOSFET with three electrical connections, source, drain and gate,



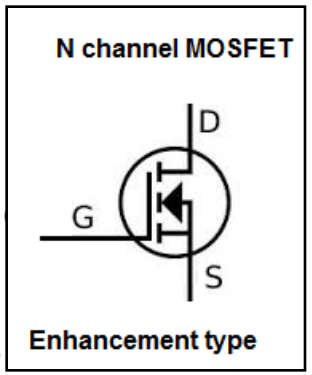


Figure 14: MOSFET basic structure with device symbol. There is a source, gate, and drain. The source and drain are connected to heavily doped n+ regions. The gate is separated from the p semiconductor by an insulator and is used to form the n-channel

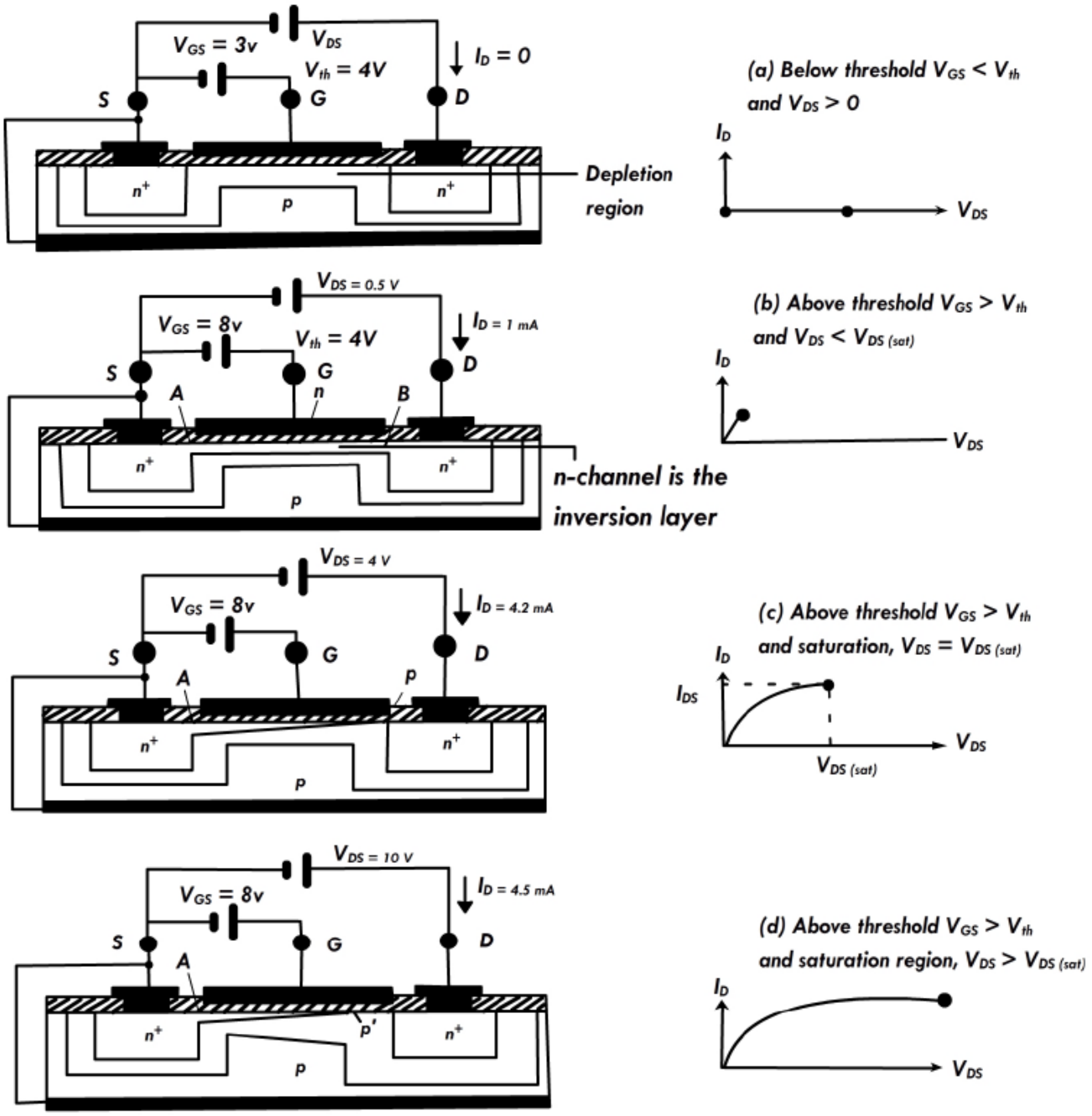
similar to a JFET. This structure is called an *enhancement MOSFET*. The bulk of the semiconductor is p type with two heavily doped n^+ regions near the source and drain. Thus, two pn^+ junctions are formed, with the depletion region lying mostly in the p side. The gate is usually made of metal or more recently, heavily doped poly-Si (with high electrical conductivity) and it is separated from the semiconductor by an insulator. The most common insulator is SiO_2, which is the reason for the name *metal-oxide-semiconductor*, though high k dielectrics based on Hafnium, have replaced the simple silicon oxide. Electrical connections are made to the source, drain, and gate at different biases similar to the JFET.

4.1 MOSFET I-V characteristics

The I-V characteristics of the MOSFET is summarized in figure 15. In a *npn* MOSFET, the gate is biased positive with respect to the source ($V_{GS} > 0$). This reverse bias causes electrons to accumulate at the the p-type semiconductor oxide interface. Below the threshold voltage for inversion (V_{th}),there is no n-channel created and any current is due to thermally generated carriers and is negligible. When the gate voltage exceeds V_{th}, an n-channel is created. With the drain is biased positive with respect to the source ($V_{DS} > 0$), electrons flow from source to drain, through the n-channel. With increasing V_{DS}, given that the drain is reverse biased with respect to the p-type semiconductor, similar to the JFET, pinch-off occurs near the drain and the current is saturated. The typical I-V characteristics is shown in figure 16.

The graph is similar to that of the JFET, shown in figure 8. The difference lines in the behavior by changing the gate source voltage. In the JFET, the channel already exists, and the role of V_{GS} is to narrow the channel and limit conduction. In the MOSFET the channel is created by V_{GS} and its role is to increase conduction (below pinch off). Both JFET and MOSFET are voltage controlled devices. The gate voltage acts as a switch to *turn on and off* the device, providing the transistor action.

Figure 15: MOSFET I-V characteristics. (a) Below the threshold voltage, there is only an depletion region and no current at any V_{DS}. (b) When inversion is achieved (by increasing V_{GS}) an n-channel is created and current increases with applied V_{DS}.(c) After a certain point ($V_{DS(sat)}$), pinch-off occurs because voltage difference between gate and drain decrease → reducing inversion layer at B. (d) After pinch-off the current reaches a saturation



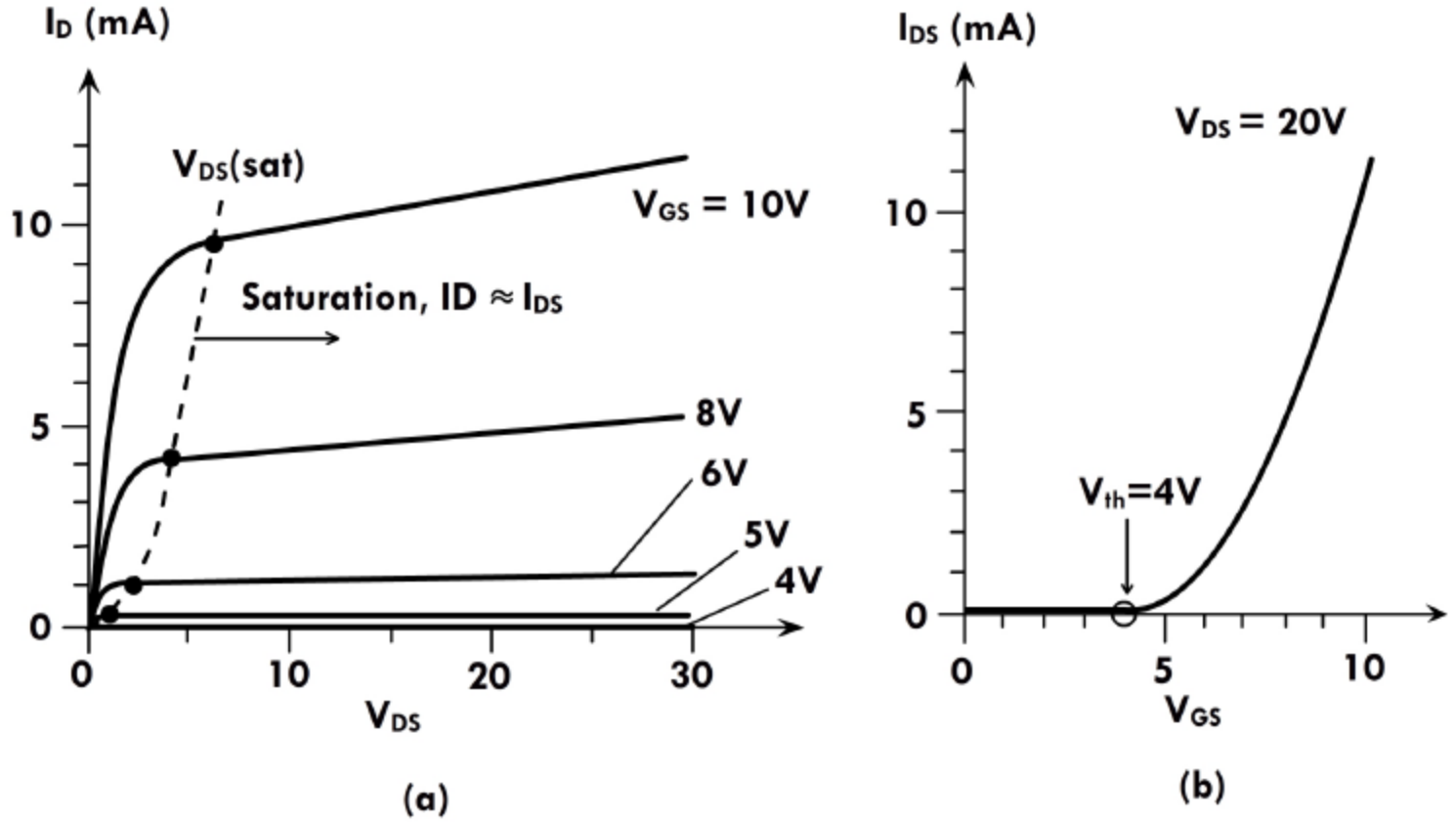


Figure 16: (a) MOSFET I-V characteristics, I_D vs. V_{DS} for varying gate voltages, V_{GS} and (b) I_{DS} vs. V_{GS} for a given V_{DS}. In JFET, the applied gate voltage narrows the channel for conduction while in MOSFET, the applied gate voltage makes the channel wider