Parameter	BJT	MOSFET
Full form	BJT stands for Bipolar Junction Transistor.	MOSFET stands for Metal Oxide Semiconductor Field Effect Transistor.
Definition	BJT is a three-terminal semiconductor device used for switching and amplification of signals.	MOSFET is a four-terminal semiconductor device which is used for switching applications.
Types	Based on the construction, BJTs are classified into two types: NPN and PNP.	Based on the construction and operation, the MOSFETs are classified into four types: P-channel enhancement MOSFET, N-channel enhancement MOSFET, P-channel depletion MOSFET and N-channel depletion MOSFET.
Terminals	BJT has three terminals viz. emitter, base and collector.	MOSFET has four terminals, i.e., source, drain, gate and body (or substrate).
Charge carriers	In BJT, both electrons and holes act as charge carriers.	In MOSFET, either electrons or holes act as charge carriers depending on the type of channel between source and drain.
Polarity	BJT is a bipolar device.	MOSFET is a unipolar device.
Controlling quantity	BJT is a current controlled device.	MOSFET is a voltage controlled device.
Input impedance	BJT has low input impedance.	MOSFET has relatively high input impedance.
Temperature coefficient	BJT has negative temperature coefficient.	MOSFET has positive temperature coefficient.
Switching frequency	The switching frequency BJT is low.	For MOSFET, the switching frequency is relatively high.
Power consumption	BJT consumes more power than MOSFET.	The power consumed by a MOSFET is less than BJT
Applications	BJT is preferred for the low current applications. It is widely used as **amplifiers**, **oscillators**, and **electronic switches**.	MOSFET is suitable for high power applications. It is used in **power supplies**, etc.

3.3.1 A First Glance at the Device

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

At the most superficial level, the transistor can be considered to be a switch. When a voltage is applied to the gate that is larger than a given value called the *threshold voltage* V_T, a conducting channel is formed between drain and source. In the presence of a voltage difference between the latter two, current flows between them. The conductivity of the channel is modulated by the gate voltage— the larger the voltage difference between gate and source, the smaller the resistance of the conducting channel and the larger the current.

When the gate voltage is lower than the threshold, no such channel exists, and the switch is considered open.

Two types of MOSFET devices can be identified. The NMOS transistor consists of n^+ drain and source regions, embedded in a p-type substrate. The current is carried by electrons moving through an n-type channel between source and drain. This is in contrast with the pn-junction diode, where current is carried by both holes and electrons. MOS devices can also be made by using an n-type substrate and p^+ drain and source regions. In such a transistor, current is carried by holes moving through a p-type channel. The device is called a p-channel MOS, or PMOS transistor. In a complementary MOS technology (CMOS), both devices are present. The cross-section of a contemporary dual-well CMOS process was presented in Chapter 2, and is repeated here for convenience (Figure 3.11).

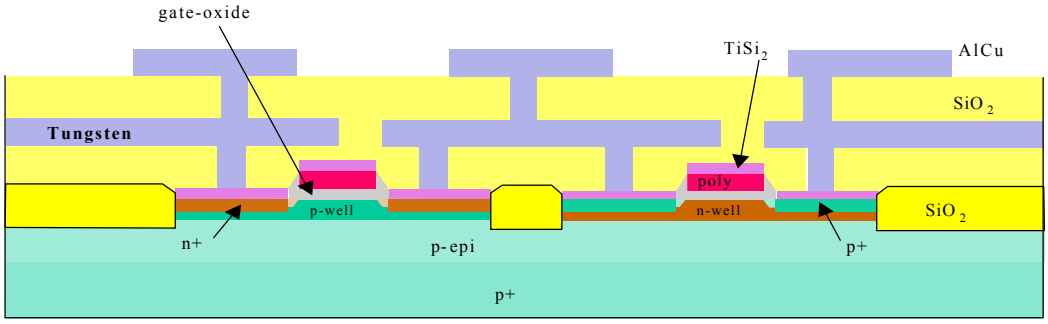
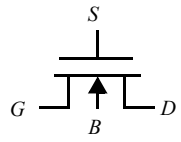
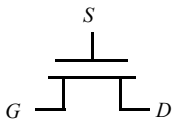


Figure 3.11 Cross-section of contemporary dual-well CMOS process.

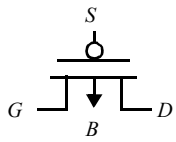
Circuit symbols for the various MOS transistors are shown in Figure 3.12. As mentioned earlier, the transistor is a four-port device with gate, source, drain, and body terminals (Figures a and c). Since the body is generally connected to a dc supply that is identical for all devices of the same type (GND for NMOS, V_{dd} for PMOS), it is most often not shown on the schematics (Figures b and d). **If the fourth terminal is not shown, it is assumed that the body is connected to the appropriate supply.**



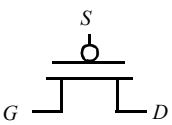
(a) NMOS transistor as 4-terminal device



(b) NMOS transistor as 3-terminal device



(a) PMOS transistor as 4-terminal device



(d) PMOS transistor as 3-terminal device

Figure 3.12 Circuit symbols for MOS transistors.

3.3.2 The MOS Transistor under Static Conditions

In the derivation of the static model of the MOS transistor, we concentrate on the NMOS device. All the arguments made are valid for PMOS devices as well as will be discussed at the end of the section.

The Threshold Voltage

Consider first the case where $V_{GS} = 0$ and drain, source, and bulk 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 0 V bias and can be considered off, which results in an extremely high resistance between drain and source.

Assume now that a positive voltage is applied to the gate (with respect to the source), as shown in Figure 3.13. The gate and substrate form the plates of a capacitor with the gate oxide as the dielectric. The positive gate voltage causes positive charge to accumulate on the gate electrode and negative charge on the substrate side. The latter manifests itself initially by repelling mobile holes. Hence, a depletion region is formed below the gate. This depletion region is similar to the one occurring in a *pn*-junction diode. Consequently, similar expressions hold for the width and the space charge per unit area. Compare these expressions to Eq. (3.4) and Eq. (3.5).

$$W_d = \sqrt{\frac{2\varepsilon_{si}\phi}{qN_A}} \tag{3.14}$$

and

$$Q_d = \sqrt{2qN_A\varepsilon_{si}\phi} \tag{3.15}$$

with N_A the substrate doping and ϕ the voltage across the depletion layer (i.e., the potential at the oxide-silicon boundary).

As the gate voltage increases, the potential at the silicon surface at some point reaches a critical value, where the semiconductor surface inverts to *n*-type material. This

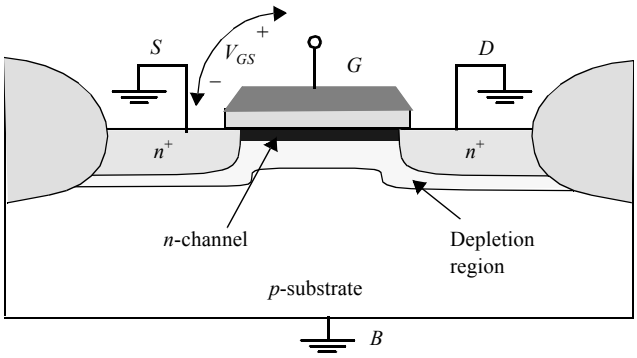


Figure 3.13 NMOS transistor for positive V_{GS}, showing depletion region and induced channel.

point marks the onset of a phenomenon known as *strong inversion* and occurs at a voltage equal to twice the *Fermi Potential* (Eq. (3.16)) ($\phi_F \approx -0.3$ V for typical *p*-type silicon substrates):

$$\phi_F = -\phi_T \ln(\frac{N_A}{n_i}) \quad (3.16)$$

$n_i \rightarrow$ intrinsic carrier concentration i.e. the no of electrones and holes that are thermally generated in the material

Further increases in the gate voltage produce no further changes in the depletion-layer width, but result in additional electrons in the thin inversion layer directly under the oxide. These are drawn into the inversion layer from the heavily doped *n*+ source region. Hence, a continuous *n*-type channel is formed between the source and drain regions, the conductivity of which is modulated by the gate-source voltage.

In the presence of an inversion layer, the charge stored in the depletion region is fixed and equals

$$Q_{B0} = \sqrt{2qN_A\varepsilon_{si}|-2\phi_F|} \quad (3.17)$$

This picture changes somewhat in case a substrate bias voltage V_{SB} is applied (V_{SB} is normally positive for *n*-channel devices). This causes the surface potential required for strong inversion to increase and to become $|-2\phi_F + V_{SB}|$. The charge stored in the depletion region now is expressed by Eq. (3.18)

$$Q_B = \sqrt{2qN_A\varepsilon_{si}(|-2\phi_F + V_{SB}|)} \quad (3.18)$$

The value of V_{GS} where strong inversion occurs is called the *threshold voltage* V_T. V_T is a function of several components, most of which are material constants such as the difference in work-function between gate and substrate material, 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. From the above arguments, it has become clear that the source-bulk voltage V_{SB} has an impact on the threshold. as well. Rather than relying on a complex (and hardly accurate) analytical expression for the threshold, we rely on an empirical parameter called V_{T0}, which is the threshold voltage for $V_{SB} = 0$, and is mostly a function of the manufacturing process. The threshold voltage under different body-biasing conditions can then be determined in the following manner,

$$V_T = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|}) \quad (3.19)$$

The parameter γ (gamma) is called the *body-effect coefficient*, and expresses the impact of changes in V_{SB}. Observe that the threshold voltage has a **positive** value for a typical **NMOS** device, while it is **negative** for a normal **PMOS** transistor.

The Fermi potential, also known as the electrochemical potential or the Fermi level, is a crucial concept in solid-state physics and plays a significant role in understanding the behavior of materials, particularly semiconductors and conductors. Here's a simplified explanation:

Imagine a "sea" of electrons filling all the available energy levels in a material. The Fermi potential represents the energy level at which this sea is half-filled. In other words, it indicates the energy level where there's a 50% probability of finding an electron.

Factors influencing the Fermi potential:

- **Temperature:** As temperature increases, the sea of electrons gets agitated, and the Fermi potential generally increases.
- **Doping:** Adding dopant atoms (impurities) can shift the Fermi potential. In n-type semiconductors, it moves closer to the conduction band, while in p-type semiconductors, it moves closer to the valence band.
- **Material properties:** The Fermi potential varies depending on the inherent properties of the material, such as its bandgap and electron affinity.

Significance of the Fermi potential:

- **Understanding conductivity:** The Fermi potential helps determine whether a material is an insulator, semiconductor, or conductor. In insulators, the Fermi level lies deep within the bandgap, while in conductors, it resides within the conduction band, allowing for easy electron flow.
- **Device modeling:** The Fermi potential is essential for modeling the behavior of electronic devices like transistors, where it influences factors like threshold voltage and current flow.

The equation in the image is:
$VT = VTO + \gamma(-2\sqrt{\Phi F} + VSB - 1 - 2\Phi F)$ (3.19)
where:

- VT is the threshold voltage, which is the value of the gate-to-source voltage (VGS) where strong inversion occurs.
- VTO is the threshold voltage for VSB = 0, which is mostly a function of the manufacturing process.
- γ (gamma) is the body-effect coefficient, which expresses the impact of changes in VSB.
- VSB is the source-to-bulk voltage.
- ΦF is the bulk Fermi potential.

The equation is used to determine the threshold voltage of a MOSFET (metal-oxide-semiconductor field-effect transistor) under different body-biasing conditions. Body biasing is a technique that is used to control the threshold voltage of a MOSFET by applying a voltage to the bulk terminal.
The equation is empirical, meaning that it is based on experimental data rather than being derived from first principles. This is because the exact relationship between VGS, VSB, and VT is complex and depends on a number of factors, such as the material properties of the device and the manufacturing process.
The body-effect coefficient (γ) is typically negative for NMOS devices and positive for PMOS devices. This means that the threshold voltage of an NMOS device will decrease as VSB becomes more positive, while the threshold voltage of a PMOS device will increase as VSB becomes more positive.
I hope this explanation is helpful!

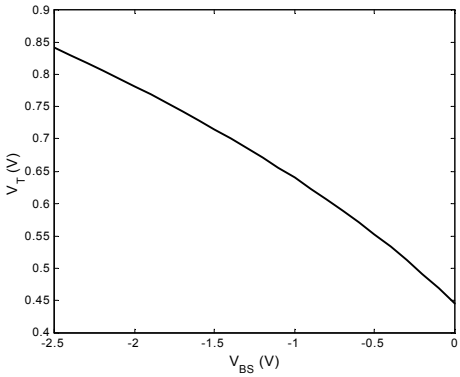


Figure 3.14 Effect of body-bias on theshold.

The effect of the well bias on the threshold voltage of an NMOS transistor is plotted in for typical values of $|-2\phi_F|$ = 0.6 V and γ = 0.4 $V^{0.5}$. A negative bias on the well or substrate causes the threshold to increase from 0.45 V to 0.85 V. Note also that V_{SB} always has to be larger than -0.6 V in an NMOS. If not, the source-body diode becomes forward biased, which deteriorates the transistor operation.

Example 3.5 Threshold Voltage of a PMOS Transistor

An PMOS transistor has a threshold voltage of -0.4 V, while the body-effect coefficient equals -0.4. Compute the threshold voltage for V_{SB} = -2.5 V. $2\phi_F$ = 0.6 V.

Using Eq. (3.19), we obtain V_T(-2.5 V) = -0.4 - 0.4 × ((2.5+0.6)$^{0.5}$ - 0.6$^{0.5}$) V = -0.79 V, which is twice the threshold under zero-bias conditions!

Resistive Operation

Assume now that $V_{GS} > V_T$ 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 (Figure 3.15). Using a simple analysis, a first-order expression of the current as a function of V_{GS} and V_{DS} can be obtained.

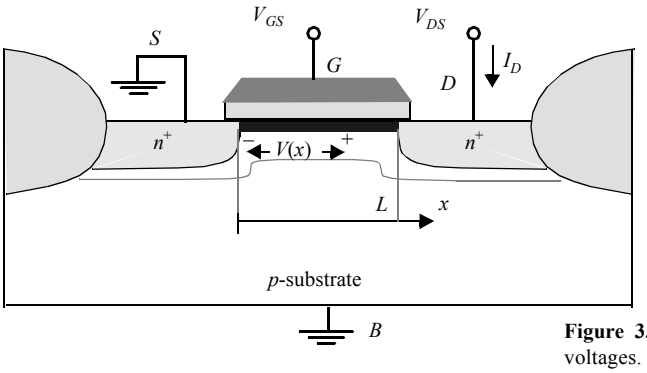


Figure 3.15 NMOS transistor with bias voltages.

At a point x along the channel, the voltage is $V(x)$, and the gate-to-channel voltage at that point equals $V_{GS} - V(x)$. Under the assumption that this voltage exceeds the threshold voltage all along the channel, the induced channel charge per unit area at point x can be computed.

$$Q_i(x) = -C_{ox}[V_{GS} - V(x) - V_T] \tag{3.20}$$

C_{ox} stands for the capacitance per unit area presented by the gate oxide, and equals

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \tag{3.21}$$

with $\varepsilon_{ox} = 3.97 \times \varepsilon_o = 3.5 \times 10^{-11}$ F/m the oxide permittivity, and t_{ox} is the thickness of the oxide. The latter which is 10 nm (= 100 Å) or smaller for contemporary processes. For an oxide thickness of 5 nm, this translates into an oxide capacitance of 7 fF/μm^2.

The current is given as the product of the drift velocity of the carriers υ_n and the available charge. Due to charge conservation, it is a constant over the length of the channel. W is the width of the channel in a direction perpendicular to the current flow.

$$I_D = -\upsilon_n(x) Q_i(x) W \tag{3.22}$$

The electron velocity is related to the electric field through a parameter called the *mobility* μ_n (expressed in m^2/V·s). The mobility is a complex function of crystal structure, and local electrical field. In general, an empirical value is used.

$$\upsilon_n = -\mu_n \xi(x) = \mu_n \frac{dV}{dx} \tag{3.23}$$

Combining Eq. (3.20) – Eq. (3.23) yields

$$I_D dx = \mu_n C_{ox} W (V_{GS} - V - V_T) dV \tag{3.24}$$

Integrating the equation over the length of the channel L yields the voltage-current relation of the transistor.

$$I_D = k'_n \frac{W}{L}\left[(V_{GS} - V_T)V_{DS} - \frac{{V_{DS}}^2}{2}\right] = k_n\left[(V_{GS} - V_T)V_{DS} - \frac{{V_{DS}}^2}{2}\right] \tag{3.25}$$

k'_n is called the *process transconductance parameter* and equals

$$k'_n = \mu_n C_{ox} = \frac{\mu_n \varepsilon_{ox}}{t_{ox}} \tag{3.26}$$

The product of the process transconductance k'_n and the (W/L) ratio of an (NMOS) transistor is called the *gain factor* k_n of the device. For smaller values of V_{DS}, the quadratic factor in Eq. (3.25) can be ignored, and we observe a linear dependence between V_{DS} and I_D. The operation region where Eq. (3.25) holds is hence called the *resistive* or *linear* region. One of its main properties is that it displays a continuous conductive channel between source and drain regions.

NOTICE: The W and L parameters in Eq. (3.25) represent the *effective channel width and length* of the transistor. These values differ from the dimensions *drawn* on the layout due to effects such as lateral diffusion of the source and drain regions (L), and the encroachment of the isolating field oxide (W). In the remainder of the text, W and L will

always stand for the effective dimensions, while a *d* subscript will be used to indicate the drawn size. The following expressions related the two parameters, with ΔW and ΔL parameters of the manufacturing process:

$$W = W_d - \Delta W$$
$$L = L_d - \Delta L \tag{3.27}$$

The Saturation Region

As the value of the drain-source voltage is further increased, the assumption that the channel voltage is larger than the threshold all along the channel ceases to hold. This happens when $V_{GS} - V(x) < V_T$. At that point, the induced charge is zero, and the conducting channel disappears or is *pinched off*. This is illustrated in Figure 3.16, which shows (in an

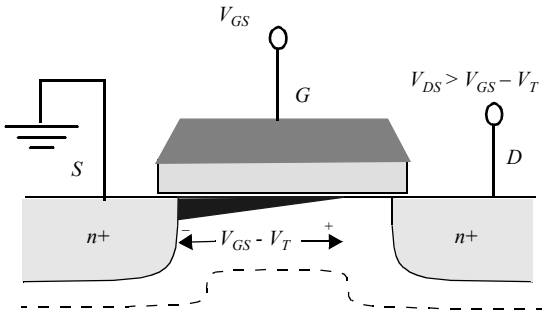


Figure 3.16 NMOS transistor under pinch-off conditions.

exaggerated fashion) how the channel thickness gradually is reduced from source to drain until pinch-off occurs. No channel exists in the vicinity of the drain region. Obviously, for this phenomenon to occur, it is essential that the pinch-off condition be met at the drain region, or

$$V_{GS} - V_{DS} \leq V_T. \tag{3.28}$$

Under those circumstances, the transistor is in the *saturation* region, and Eq. (3.25) no longer holds. The voltage difference over the induced channel (from the pinch-off point to the source) remains fixed at $V_{GS} - V_T$, and consequently, the current remains constant (or saturates). Replacing V_{DS} by $V_{GS} - V_T$ in Eq. (3.25) yields the drain current for the saturation mode. It is worth observing that, to a first agree, the current is no longer a function of V_{DS}. Notice also the *squared dependency* of the drain current with respect to the control voltage V_{GS}.

$$I_D = \frac{k'_n}{2}\frac{W}{L}(V_{GS} - V_T)^2 \tag{3.29}$$

Channel-Length Modulation

The latter equation seems to suggest that the transistor in the saturation mode acts as a perfect current source — or that the current between drain and source terminal is a constant, independent of the applied voltage over the terminals. This not entirely correct. The effective length of the conductive channel is actually modulated by the applied V_{DS}: increasing V_{DS} causes the depletion region at the drain junction to grow, reducing the length of the effective channel. As can be observed from Eq. (3.29), the current increases when the length factor L is decreased. A more accurate description of the current of the MOS transistor is therefore given in Eq. (3.30).

$$I_D = I_D'(1 + \lambda V_{DS}) \tag{3.30}$$

with I_D' the current expressions derived earlier, and λ an empirical parameter, called the *channel-length modulation*. Analytical expressions for λ have proven to be complex and inaccurate. λ varies roughly with the inverse of the channel length. In shorter transistors, the drain-junction depletion region presents a larger fraction of the channel, and the channel-modulation effect is more pronounced. It is therefore advisable to resort to long-channel transistors if a high-impedance current source is needed.

Velocity Saturation

The behavior of transistors with very short channel lengths (called *short-channel devices*) deviates considerably from the resistive and saturated models, presented in the previous paragraphs. The main culprit for this deficiency is the *velocity saturation* effect. Eq. (3.23) states that the velocity of the carriers is proportional to the electrical field, independent of the value of that field. In other words, the carrier mobility is a constant. However, at high field strengths, the carriers fail to follow this linear model. In fact, when the electrical field along the channel reaches a critical value ξ_c, the velocity of the carriers tends to saturate due to scattering effects (collisions suffered by the carriers). This is illustrated in Figure 3.17.

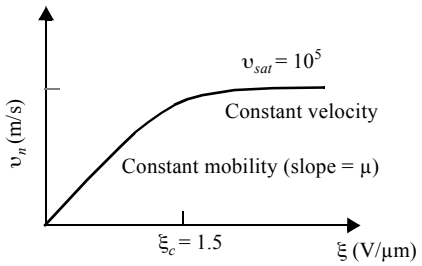


Figure 3.17 Velocity-saturation effect.

For *p*-type silicon, the critical field at which electron saturation occurs is around 1.5×10^6 V/m (or 1.5 V/μm), and the saturation velocity υ_{sat} approximately equals 10^5 m/s. This means that in an NMOS device with a channel length of 1μm, only a couple of volts between drain and source are needed to reach the saturation point. This condition is easily met in current short-channel devices. Holes in an *n*-type silicon saturate at the same velocity, although a higher electrical field is needed to achieve saturation. Velocity-saturation effects are hence less pronounced in PMOS transistors.