10

Fundamentals of the Metal–Oxide–Semiconductor Field-Effect Transistor

The single-junction semiconductor devices that we have considered, including the pn homojunction diode, can be used to produce rectifying current–voltage characteristics and to form electronic switching circuits. The transistor is a multijunction semiconductor device that, in conjunction with other circuit elements, is capable of current gain, voltage gain, and signal power gain. The basic transistor action is the control of current at one terminal by the voltage applied across the other two terminals of the device.

The **M**etal–**O**xide–**S**emiconductor **F**ield-**E**ffect **T**ransistor (MOSFET) is one of two major types of transistors. The fundamental physics of the MOSFET is developed in this chapter. The MOSFET is used extensively in digital circuit applications where, because of its small size, millions of devices can be fabricated in a single integrated circuit.

Two complementary configurations of MOS transistors, the n-channel MOSFET and the p-channel MOSFET, can be fabricated. Electronic circuit design becomes very versatile when the two types of devices are used in the same circuit. These circuits are referred to as complementary MOS (CMOS) circuits. ■

10.0 | PREVIEW

In this chapter, we will:

■ Study the characteristics of energy bands as a function of applied voltage in the metal–oxide–semiconductor structure known as the MOS capacitor. The MOS capacitor is the heart of the MOSFET.

■ Discuss the concept of surface inversion in the semiconductor of the MOS capacitor.

- Define and derive the expression for the threshold voltage, which is a basic parameter of the MOSFET.
- Discuss various physical structures of MOSFETs, including enhancement and depletion mode devices.
- Derive the ideal current–voltage relationship of the MOSFET.
- Develop the small-signal equivalent circuit of the MOSFET. This circuit is used to relate small-signal currents and voltages in analog circuits.
- Derive the frequency limiting factors of the MOSFET.

10.1 | THE TWO-TERMINAL MOS STRUCTURE

The heart of the MOSFET is the MOS capacitor shown in Figure 10.1. The metal may be aluminum or some other type of metal, although in many cases, it is actually a high-conductivity polycrystalline silicon that has been deposited on the oxide; however, the term metal is usually still used. The parameter t_{ox} in the figure is the thickness of the oxide and ϵ_{ox} is the permittivity of the oxide.

10.1.1 Energy-Band Diagrams

The physics of the MOS structure can be more easily explained with the aid of the simple parallel-plate capacitor. Figure 10.2a shows a parallel-plate capacitor with the top plate at a negative voltage with respect to the bottom plate. An insulator material separates the two plates. With this bias, a negative charge exists on the top plate, a positive charge exists on the bottom plate, and an electric field is induced between the two plates as shown. The capacitance per unit area for this geometry is

$$C' = \frac{\epsilon}{d} \tag{10.1}$$

where ϵ is the permittivity of the insulator and d is the distance between the two plates. The magnitude of the charge per unit area on either plate is
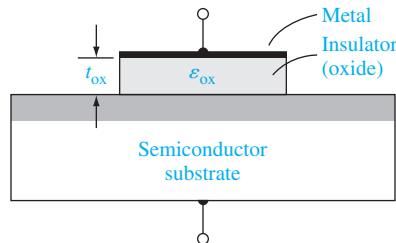
$$Q' = C'V \tag{10.2}$$

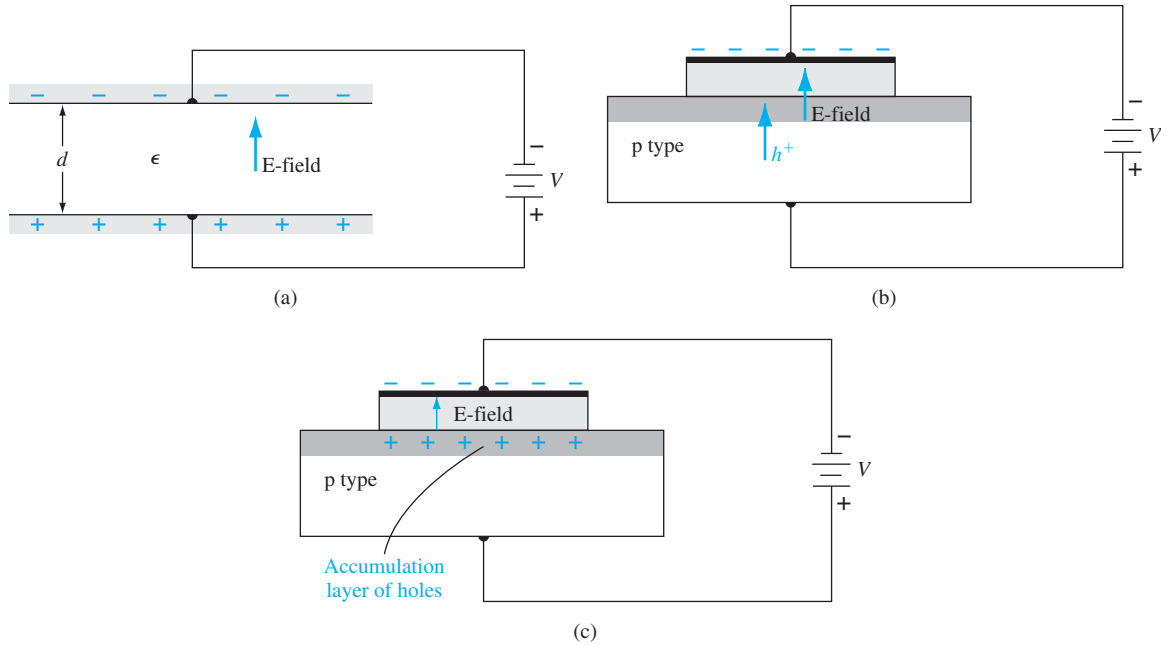**Figure 10.1** | The basic MOS capacitor structure.

Figure 10.2 | (a) A parallel-plate capacitor showing the electric field and conductor charges. (b) A corresponding MOS capacitor with a negative gate bias showing the electric field and charge flow. (c) The MOS capacitor with an accumulation layer of holes.

where the prime indicates charge or capacitance per unit area. The magnitude of the electric field is

$$E = \frac{V}{d} \tag{10.3}$$

Figure 10.2b shows a MOS capacitor with a p-type semiconductor substrate. The top metal gate is at a negative voltage with respect to the semiconductor substrate. From the example of the parallel-plate capacitor, we can see that a negative charge will exist on the top metal plate and an electric field will be induced with the direction shown in the figure. If the electric field were to penetrate into the semiconductor, the majority carrier holes would experience a force toward the oxide–semiconductor interface. Figure 10.2c shows the equilibrium distribution of charge in the MOS capacitor with this particular applied voltage. An *accumulation layer* of holes at the oxide–semiconductor junction corresponds to the positive charge on the bottom "plate" of the MOS capacitor.

Figure 10.3a shows the same MOS capacitor in which the polarity of the applied voltage is reversed. A positive charge now exists on the top metal plate and the induced electric field is in the opposite direction as shown. If the electric field penetrates the semiconductor in this case, majority carrier holes will experience a force away from the oxide–semiconductor interface. As the holes are pushed away from

the interface, a negative space charge region is created because of the fixed ionized acceptor atoms. The negative charge in the induced depletion region corresponds to the negative charge on the bottom "plate" of the MOS capacitor. Figure 10.3b shows the equilibrium distribution of charge in the MOS capacitor with this applied voltage.

The energy-band diagrams of the MOS capacitor with a p-type substrate for various gate biases are shown in Figure 10.4. Figure 10.4a shows the ideal case when zero bias is applied across the MOS device. The energy bands in the semiconductor are flat indicating no net charge exists in the semiconductor. This condition is known as *flat band* and is discussed in more detail later in the chapter.
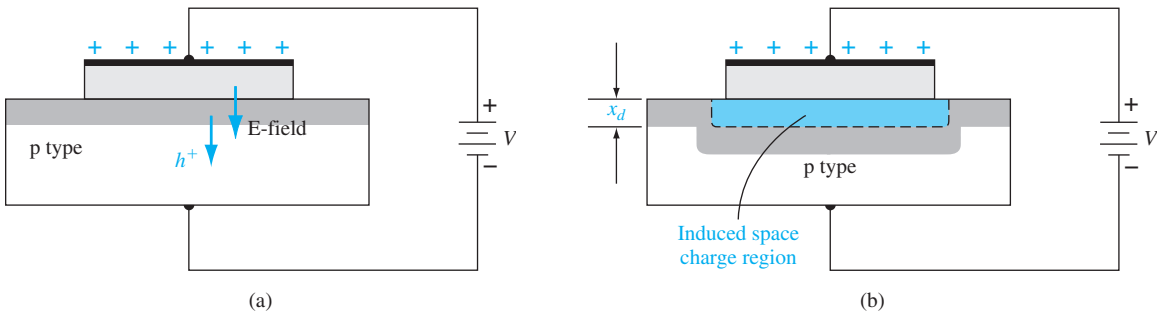

Figure 10.3 | The MOS capacitor with a moderate positive gate bias, showing (a) the electric field and charge flow and (b) the induced space charge region.
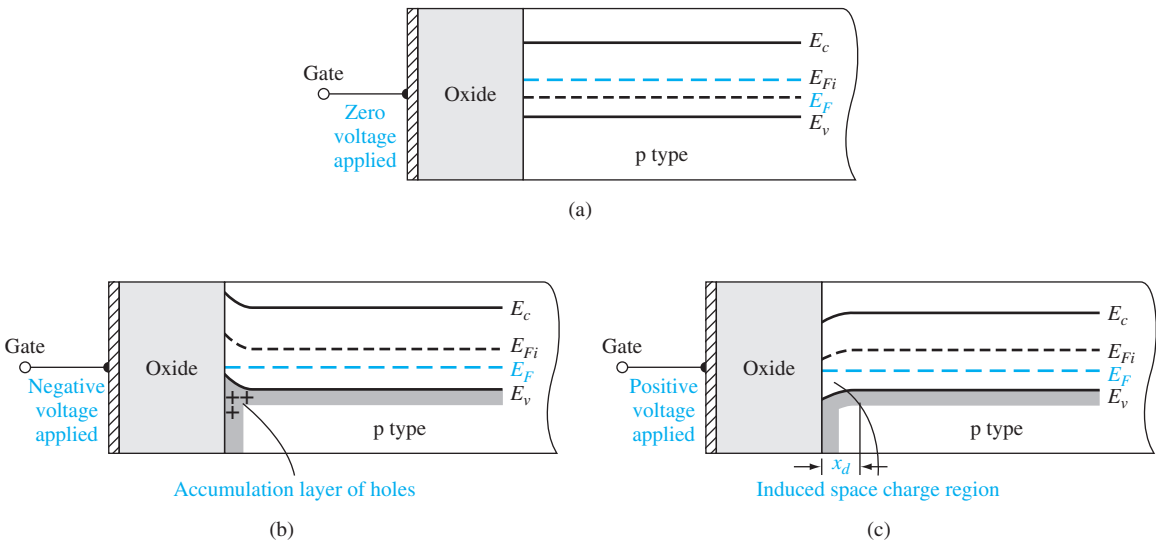

Figure 10.4 | The energy-band diagram of a MOS capacitor with a p-type substrate for (a) a zero applied gate bias showing the *ideal* case, (b) a negative gate bias, and (c) a moderate positive gate bias.

Figure 10.4b shows the energy-band diagram for the case when a negative bias is applied to the gate. (Remember that positive electron energy is plotted "upward" and positive voltage is plotted "downward.") The valence-band edge is closer to the Fermi level at the oxide–semiconductor interface than in the bulk material, which implies that there is an accumulation of holes. The semiconductor surface appears to be more p-type than the bulk material. The Fermi level is a constant in the semiconductor since the MOS system is in thermal equilibrium and there is no current through the oxide.

Figure 10.4c shows the energy-band diagram of the MOS system when a positive voltage is applied to the gate. The conduction- and valence-band edges bend as shown in the figure, indicating a space charge region similar to that in a pn junction. The conduction band and intrinsic Fermi levels move closer to the Fermi level. The induced space charge width is x_d.

Now consider the case when a still larger positive voltage is applied to the top metal gate of the MOS capacitor. We expect the induced electric field to increase in magnitude and the corresponding positive and negative charges on the MOS capacitor to increase. A larger negative charge in the MOS capacitor implies a larger induced space charge region and more band bending. Figure 10.5 shows such a condition. The intrinsic Fermi level at the surface is now below the Fermi level. The conduction band at the surface is now close to the Fermi level, whereas the valence band is close to the Fermi level in the bulk semiconductor. This result implies that the surface in the semiconductor adjacent to the oxide–semiconductor interface is n type. By applying a sufficiently large positive gate voltage, we have inverted the surface of the semiconductor from a p-type to an n-type semiconductor. We have created an *inversion layer* of electrons at the oxide–semiconductor interface.

In the MOS capacitor structure that we have just considered, we assumed a p-type semiconductor substrate. The same type of energy-band diagrams can be constructed for a MOS capacitor with an n-type semiconductor substrate. Figure 10.6a shows the MOS capacitor structure with a positive voltage applied to the top gate terminal. A positive charge exists on the top gate and an electric field is induced with the direction shown in the figure. An accumulation layer of electrons will be induced in the n-type substrate. The case when a negative voltage is applied to the top gate
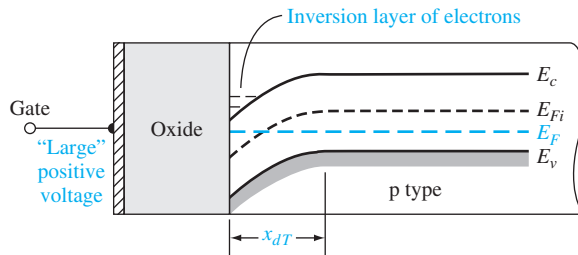

Figure 10.5 | The energy-band diagram of the MOS capacitor with a p-type substrate for a "large" positive gate bias.
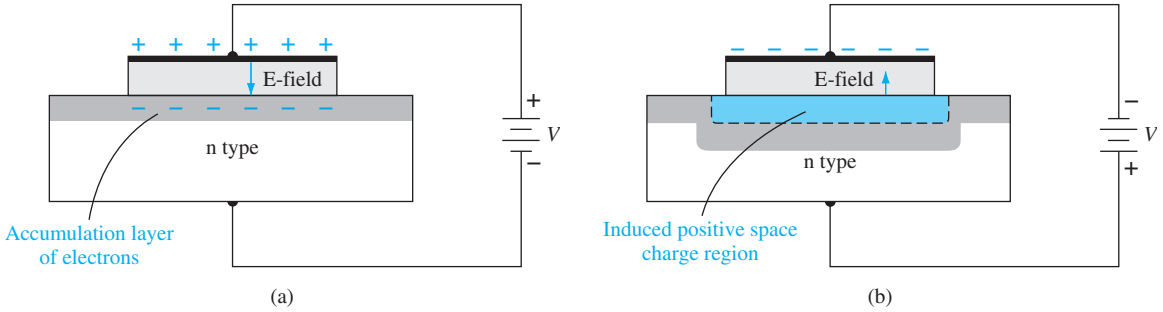
Figure 10.6 | The MOS capacitor with an n-type substrate for (a) a positive gate bias and (b) a moderate negative gate bias.

is shown in Figure 10.6b. A positive space charge region is induced in the n-type semiconductor in this situation.

The energy-band diagrams for this MOS capacitor with the n-type substrate are shown in Figure 10.7. Figure 10.7a shows the case when a positive voltage is applied to the gate and an accumulation layer of electrons is formed. Figure 10.7b shows the energy bands when a negative voltage is applied to the gate. The conduction and valence bands now bend upward indicating that a space charge region has been induced in the n-type substrate. Figure 10.7c shows the energy bands when a larger negative voltage is applied to the gate. The conduction and valence bands are bent even more and the intrinsic Fermi level has moved above the Fermi level. The valence band at the surface is now close to the Fermi level, whereas the conduction band is close to the Fermi level in the bulk semiconductor. This result implies that the semiconductor surface adjacent to the oxide–semiconductor interface is p type. By applying a sufficiently large negative voltage to the gate of the MOS capacitor, the semiconductor surface has been inverted from n type to p type. An inversion layer of holes has been induced at the oxide–semiconductor interface.

10.1.2   Depletion Layer Thickness

We may calculate the width of the induced space charge region adjacent to the oxide–semiconductor interface. Figure 10.8 shows the space charge region in a p-type semiconductor substrate. The potential ϕ_{fp} is the difference (in V) between E_{Fi} and E_F and is given by

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right)$$ (10.4)

where N_a is the acceptor doping concentration and n_i is the intrinsic carrier concentration.

The potential ϕ_s is called the surface potential; it is the difference (in V) between E_{Fi} measured in the bulk semiconductor and E_{Fi} measured at the surface. The surface
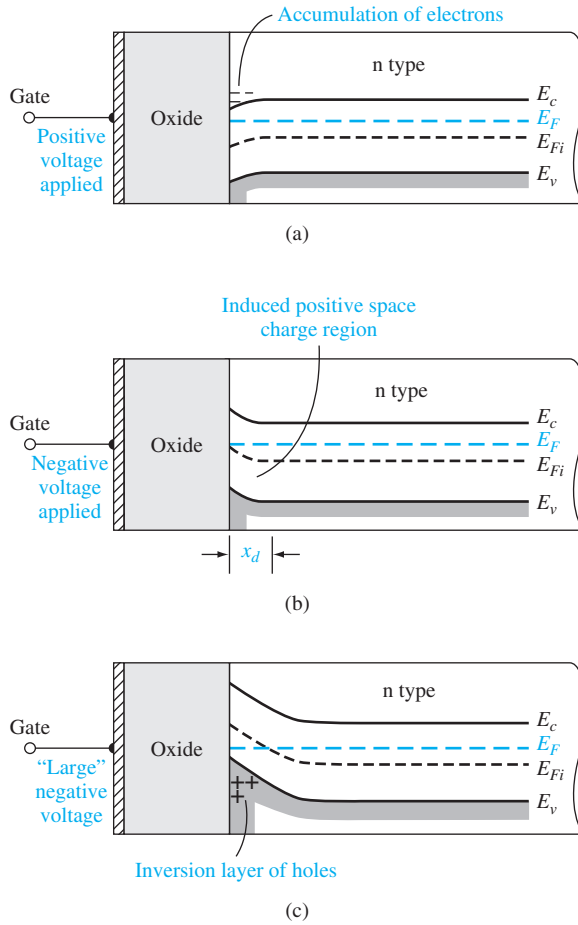
(a)

(b)

(c)

Figure 10.7 | The energy-band diagram of the MOS capacitor with an n-type substrate for (a) a positive gate bias, (b) a moderate negative bias, and (c) a "large" negative gate bias.

potential is the potential difference across the space charge layer. The space charge width can now be written in a form similar to that of a one-sided pn junction. We can write that

$$x_d = \left(\frac{2\epsilon_s \phi_s}{eN_a}\right)^{1/2}$$

(10.5)

where ϵ_s is the permittivity of the semiconductor. Equation (10.5) assumes that the abrupt depletion approximation is valid.
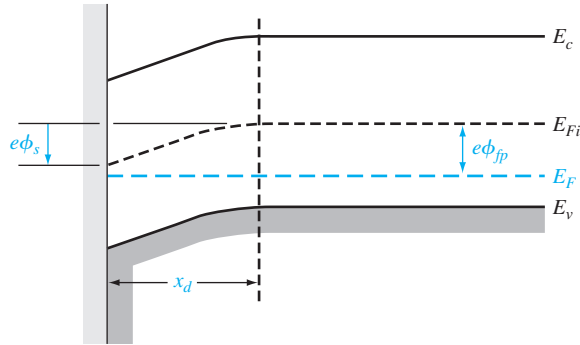
Figure 10.8 | The energy-band diagram in the p-type semiconductor, indicating surface potential.

Figure 10.9 shows the energy bands for the case in which $\phi_s = 2\phi_{fp}$. The Fermi level at the surface is as far above the intrinsic level as the Fermi level is below the intrinsic level in the bulk semiconductor. The electron concentration at the surface is the same as the hole concentration in the bulk material. This condition is known as the *threshold inversion point.* The applied gate voltage creating this condition is known as the *threshold voltage.* If the gate voltage increases above this threshold value, the conduction band will bend slightly closer to the Fermi level, but the change in the conduction band at the surface is now only a slight function of gate voltage. The electron concentration at the surface, however, is an exponential function of the surface potential. The surface potential may increase by a few (kT/e) volts, which will change the electron concentration by orders of magnitude, but the space charge width changes only slightly. In this case, then, the space charge region has essentially reached a maximum width.
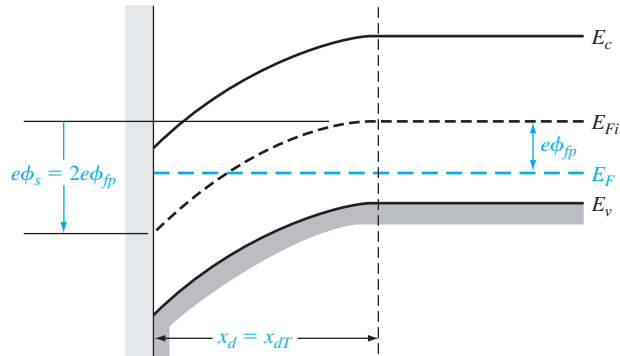

Figure 10.9 | The energy-band diagram in the p-type semiconductor at the threshold inversion point.

The maximum space charge width, x_{dT}, at this inversion transition point can be calculated from Equation (10.5) by setting $\phi_s = 2\phi_{fp}$. Then

$$x_{dT} = \left(\frac{4\epsilon_s \phi_{fp}}{eN_a}\right)^{1/2} \tag{10.6}$$

Objective: Calculate the maximum space charge width for a given semiconductor doping concentration.

EXAMPLE 10.1

Consider silicon at $T = 300$ K doped to $N_a = 10^{16}$ cm^{-3}. The intrinsic carrier concentration is $n_i = 1.5 \times 10^{10}$ cm^{-3}.

■ **Solution**

From Equation (10.4), we have

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259)\ln\left(\frac{10^{16}}{1.5 \times 10^{10}}\right) = 0.3473 \text{ V}$$

Then the maximum space charge width is

$$x_{dT} = \left[\frac{4\epsilon_s \phi_{fp}}{eN_a}\right]^{1/2} = \left[\frac{4(11.7)(8.85 \times 10^{-14})(0.3473)}{(1.6 \times 10^{-19})(10^{16})}\right]^{1/2}$$

or

$$x_{dT} \cong 0.30 \times 10^{-4} \text{ cm} = 0.30 \text{ }\mu\text{m}$$

■ **Comment**

The maximum induced space charge width is on the same order of magnitude as pn junction space charge widths.

■ **EXERCISE PROBLEM**

Ex 10.1 Consider an oxide-to-p-type silicon junction at $T = 300$ K. The impurity doping concentration in the silicon is $N_a = 2 \times 10^{15}$ cm^{-3}. Calculate the maximum space charge width. Does the space charge width increase or decrease as the p-type doping concentration decreases?

(Ans. $x_{dT} = 0.629$ μm, increase)

We have been considering a p-type semiconductor substrate. The same maximum induced space charge region width occurs in an n-type substrate. Figure 10.10 is the energy-band diagram at the threshold voltage with an n-type substrate. We can write

$$\phi_{fn} = V_t \ln\left(\frac{N_d}{n_i}\right) \tag{10.7}$$

and

$$x_{dT} = \left(\frac{4\epsilon_s \phi_{fn}}{eN_d}\right)^{1/2} \tag{10.8}$$

Note that we are always assuming the parameters ϕ_{fp} and ϕ_{fn} to be positive quantities.

Figure 10.11 is a plot of x_{dT} at $T = 300$ K as a function of doping concentration in silicon. The semiconductor doping can be either n type or p type.
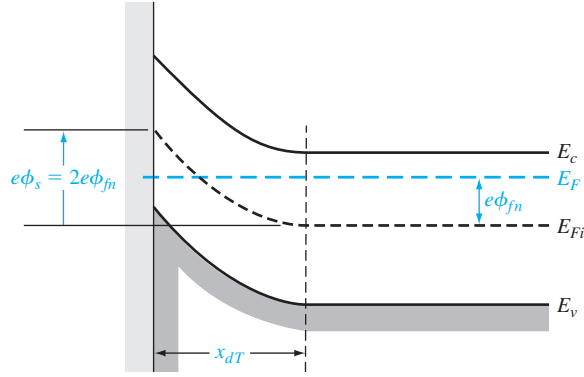
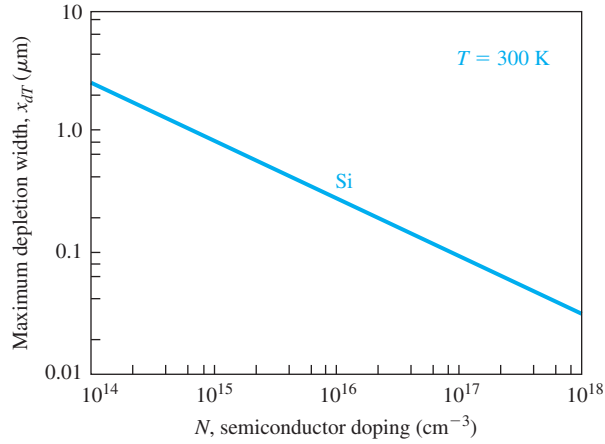**Figure 10.10 |** The energy-band diagram in the n-type semiconductor at the threshold inversion point.

Figure 10.11 | Maximum induced space charge region width versus semiconductor doping.

10.1.3 Surface Charge Density

From the results of Chapter 4, the electron concentration in the conduction band can be written in the form

$$n = n_i \exp\left[\frac{E_F - E_{Fi}}{kT}\right] \tag{10.9}$$

For a p-type semiconductor substrate, the electron inversion charge density can then be written as (see Figure 10.9)

$$n_s = n_i \exp\left[\frac{e(\phi_{fp} + \Delta\phi_s)}{kT}\right] = n_i \exp\left[\frac{\phi_{fp} + \Delta\phi_s}{V_t}\right] \tag{10.10a}$$

or

$$n_s = n_i \exp\left(\frac{\phi_{fp}}{V_t}\right) \cdot \exp\left(\frac{\Delta\phi_s}{V_t}\right) \tag{10.10b}$$

where $\Delta\phi_s$, is the surface potential greater than $2\phi_{fp}$.

We may note that

$$n_{st} = n_i \exp\left(\frac{\phi_{fp}}{V_t}\right) \tag{10.11}$$

where n_{st} is the surface charge density at the threshold inversion point. The electron inversion charge density can then be written as

$$n_s = n_{st} \exp\left(\frac{\Delta\phi_s}{V_t}\right) \tag{10.12}$$

Figure 10.12 shows the electron inversion charge density as a function of surface potential for the case when the threshold inversion charge density is $n_{st} = 10^{16}$ cm^{-3}. We may note that the inversion charge density increases by a factor of 10 with a 60-mV increase in surface potential. As discussed previously, the electron inversion charge density increases rapidly with small increases in surface potential, which means that the space charge width essentially reaches a maximum value.
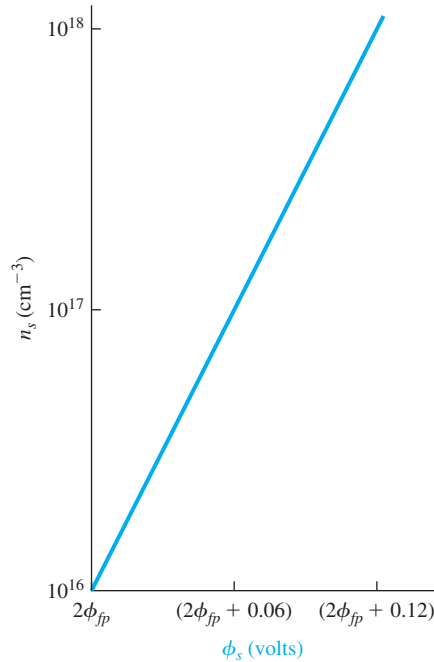

Figure 10.12 | Electron inversion charge density as a function of surface potential.

10.1.4 Work Function Differences

We have been concerned, so far, with the energy-band diagrams of the semiconductor material. Figure 10.13a shows the energy levels in the metal, silicon dioxide (SiO_2), and silicon relative to the vacuum level. The metal work function is ϕ_m and the electron affinity is χ. The parameter χ_i is the oxide electron affinity and, for SiO_2, $\chi_i = 0.9$ V.

Figure 10.13b shows the energy-band diagram of the entire MOS structure with zero gate voltage applied. The Fermi level is a constant through the entire system at thermal equilibrium. We may define ϕ'_m as a modified metal work function—the potential required to inject an electron from the metal into the conduction band of the oxide. Similarly, χ' is defined as a modified electron affinity. The voltage V_{ox0} is the potential drop across the oxide for zero applied gate voltage and is not necessarily zero because of the difference between ϕ_m and χ. The potential ϕ_{s0} is the surface potential for this case.

If we sum the energies from the Fermi level on the metal side to the Fermi level on the semiconductor side, we have

$$e\phi'_m + eV_{ox0} = e\chi' + \frac{E_g}{2} - e\phi_{s0} + e\phi_{fp} \qquad (10.13)$$

Equation (10.13) can be rewritten as

$$V_{ox0} + \phi_{s0} = -\left[\phi'_m - \left(\chi' + \frac{E_g}{2e} + \phi_{fp}\right)\right] \qquad (10.14)$$
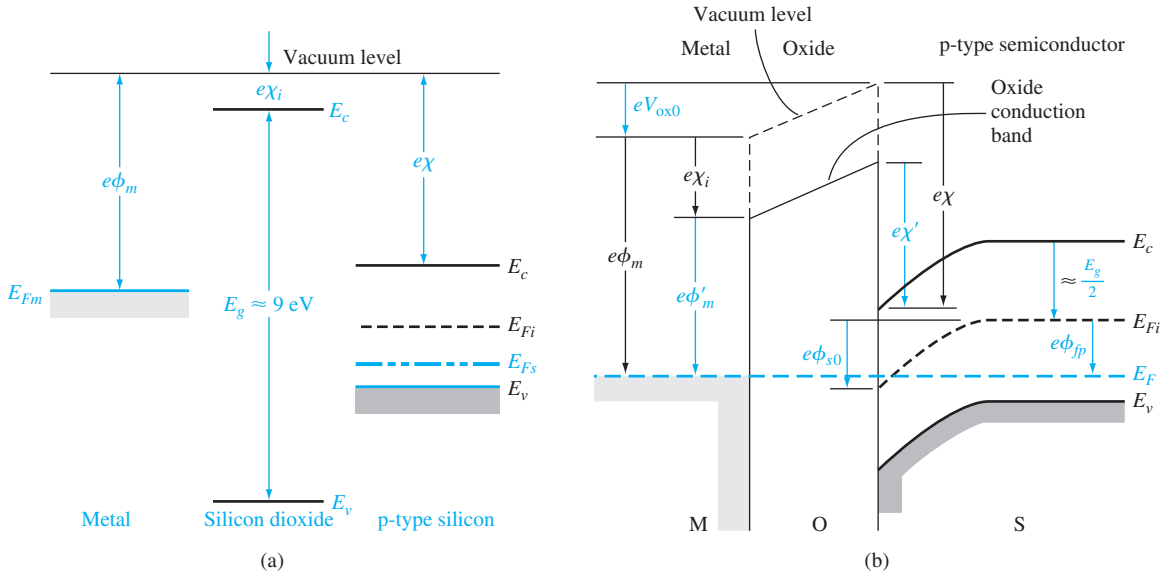

Figure 10.13 | (a) Energy levels in a MOS system prior to contact and (b) energy-band diagram through the MOS structure in thermal equilibrium after contact.

We can define a potential ϕ_{ms} as

$$\phi_{ms} \equiv \left[\phi'_m - \left(\chi' + \frac{E_g}{2e} + \phi_{fp} \right) \right] \qquad (10.15)$$

which is known as the metal–semiconductor work function difference.

Objective: Determine the metal–semiconductor work function difference, ϕ_{ms}, for a given MOS system and semiconductor doping.

EXAMPLE 10.2

For an aluminum–silicon dioxide junction, $\phi'_m = 3.20$ V and, for a silicon–silicon dioxide junction, $\chi' = 3.25$ V. We may assume that $E_g = 1.12$ V. Let the p-type doping be $N_a = 10^{15}$ cm^{-3}.

■ **Solution**

For silicon at $T = 300$ K, we may calculate ϕ_{fp} as

$$\phi_{fp} = V_t \ln \left(\frac{N_a}{n_i} \right) = (0.0259)\ln \left(\frac{10^{15}}{1.5 \times 10^{10}} \right) = 0.288 \text{ V}$$

Then the metal–semiconductor work function difference is

$$\phi_{ms} = \phi'_m - \left(\chi' + \frac{E_g}{2e} + \phi_{fp} \right) = 3.20 - (3.25 + 0.560 + 0.288)$$

or

$$\phi_{ms} = -0.898 \text{ V}$$

■ **Comment**

The value of ϕ_{ms} will become more negative as the doping of the p-type substrate increases.

■ **EXERCISE PROBLEM**

Ex 10.2 Repeat Example 10.2 for a semiconductor doping of $N_a = 10^{16}$ cm^{-3}.

(Ans. $\phi_{ms} = -0.957$ V)

Degenerately doped polysilicon deposited on the oxide is also often used as the metal gate. Figure 10.14a shows the energy-band diagram of a MOS capacitor with an n$^+$ polysilicon gate and a p-type substrate. Figure 10.14b shows the energy-band
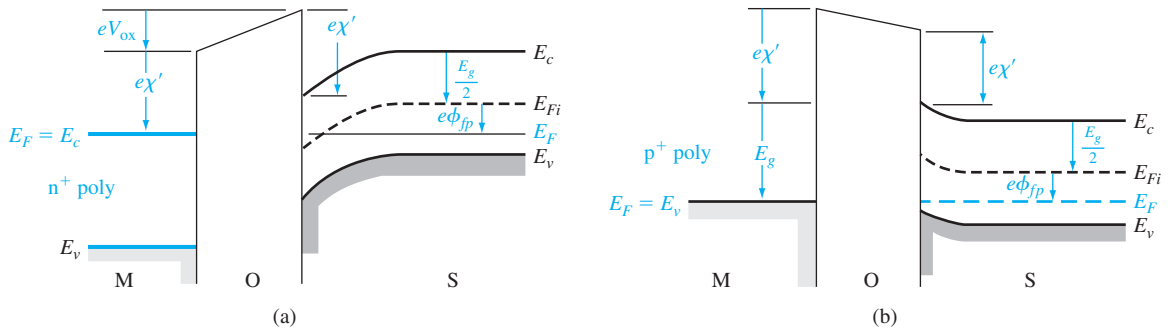

Figure 10.14 | Energy-band diagram through the MOS structure with a p-type substrate at zero gate bias for (a) an n$^+$ polysilicon gate and (b) a p$^+$ polysilicon gate.

diagram for the case of a p^+ polysilicon gate and the p-type silicon substrate. In the degenerately doped polysilicon, we will initially assume that $E_F = E_c$ for the n^+ case and $E_F = E_v$ for the p^+ case.

For the n^+ polysilicon gate, the metal–semiconductor work function difference can be written as

$$\phi_{ms} = \left[\chi' - \left(\chi' + \frac{E_g}{2e} + \phi_{fp}\right)\right] = -\left(\frac{E_g}{2e} + \phi_{fp}\right) \qquad (10.16)$$

and for the p^+ polysilicon gate, we have

$$\phi_{ms} = \left[\left(\chi' + \frac{E_g}{e}\right) - \left(\chi' + \frac{E_g}{2e} + \phi_{fp}\right)\right] = \left(\frac{E_g}{2e} - \phi_{fp}\right) \qquad (10.17)$$

However, for degenerately doped n^+ polysilicon and p^+ polysilicon, the Fermi level can be above E_c and below E_v, respectively, by 0.1 to 0.2 V. The experimental ϕ_{ms} values will then be slightly different from the values calculated by using Equations (10.16) and (10.17).

We have been considering a p-type semiconductor substrate. We may also have an n-type semiconductor substrate in a MOS capacitor. Figure 10.15 shows the energy-band diagram of the MOS capacitor with a metal gate and the n-type semiconductor substrate, for the case when a negative voltage is applied to the gate. The metal–semiconductor work function difference for this case is defined as

$$\phi_{ms} = \phi'_m - \left(\chi' + \frac{E_g}{2e} - \phi_{fn}\right) \qquad (10.18)$$
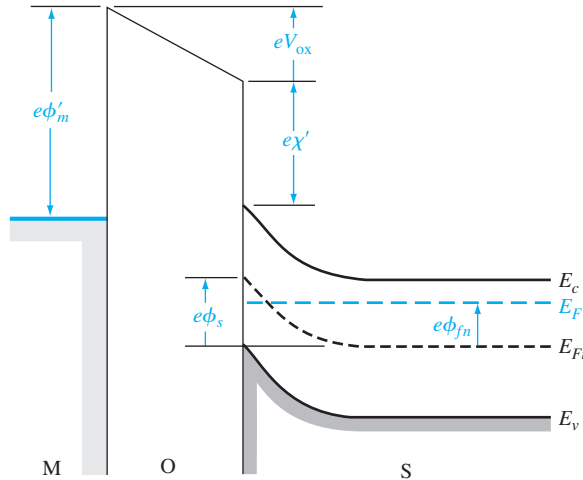

Figure 10.15 | Energy-band diagram through the MOS structure with an n-type substrate for a negative applied gate bias.
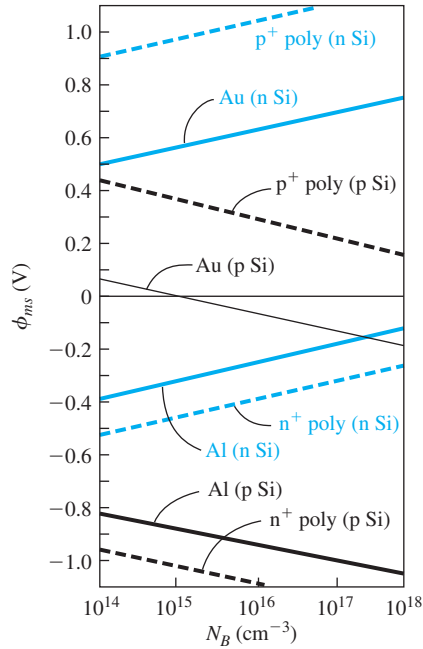
Figure 10.16 | Metal–semiconductor work function difference versus doping for aluminum, gold, and n⁻ and p⁻ polysilicon gates.
(*From Sze [17] and Werner [20].*)

where ϕ_{fn} is assumed to be a positive value. We will have similar expressions for n⁺ and p⁺ polysilicon gates.

Figure 10.16 shows the work function differences as a function of semiconductor doping for the various types of gates. We may note that the magnitudes of ϕ_{ms} for the polysilicon gates are somewhat larger than Equations (10.16) and (10.17) predict. This difference again is because the Fermi level is not equal to the conduction-band energy for the n⁺ gate and is not equal to the valence-band energy for the p⁺ gate. The metal–semiconductor work function difference becomes important in the flat-band and threshold voltage parameters discussed next.

10.1.5 Flat-Band Voltage

The *flat-band voltage* is defined as the applied gate voltage such that there is no band bending in the semiconductor and, as a result, zero net space charge in this region. Figure 10.17 shows this flat-band condition. Because of the work function difference and possible trapped charge in the oxide, the voltage across the oxide for this case is not necessarily zero.

We have implicitly been assuming that there is zero net charge density in the oxide material. This assumption may not be valid—a net fixed charge density,
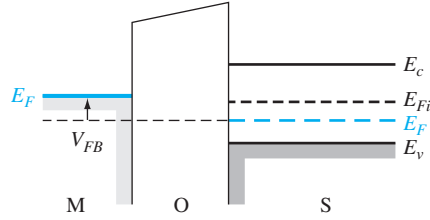
Figure 10.17 | Energy-band diagram of a
MOS capacitor at flat band.

usually positive, may exist in the insulator. The positive charge has been identi-
fied with broken or dangling covalent bonds near the oxide–semiconductor interface.
During the thermal formation of SiO_2, oxygen diffuses through the oxide and reacts
near the Si–SiO_2 interface to form the SiO_2. Silicon atoms may also break away from
the silicon material just prior to reacting to form SiO_2. When the oxidation process is
terminated, excess silicon may exist in the oxide near the interface, resulting in the
dangling bonds. The magnitude of this oxide charge seems, in general, to be a strong
function of the oxidizing conditions such as oxidizing ambient and temperature. The
charge density can be altered to some degree by annealing the oxide in an argon or
nitrogen atmosphere. However, the charge is rarely zero.

The net fixed charge in the oxide appears to be located fairly close to the oxide–
semiconductor interface. We will assume in the analysis of the MOS structure that an
equivalent trapped charge per unit area, Q'_{ss}, is located in the oxide directly adjacent
to the oxide–semiconductor interface. For the moment, we will ignore any other
oxide-type charges that may exist in the device. The parameter Q'_{ss} is usually given
in terms of number of electronic charges per unit area.

Equation (10.14), for zero applied gate voltage, can be written as

$$V_{ox0} + \phi_{s0} = -\phi_{ms} \tag{10.19}$$

If a gate voltage is applied, the potential drop across the oxide and the surface poten-
tial will change. We can then write

$$V_G = \Delta V_{ox} + \Delta \phi_s = (V_{ox} - V_{ox0}) + (\phi_s - \phi_{s0}) \tag{10.20}$$

Using Equation (10.19), we have

$$V_G = V_{ox} + \phi_s + \phi_{ms} \tag{10.21}$$

Figure 10.18 shows the charge distribution in the MOS structure for the flat-
band condition. There is zero net charge in the semiconductor, and we can assume
that an equivalent fixed surface charge density exists in the oxide. The charge density
on the metal is Q'_m, and from charge neutrality we have

$$Q'_m + Q'_{ss} = 0 \tag{10.22}$$

We can relate Q'_m to the voltage across the oxide by

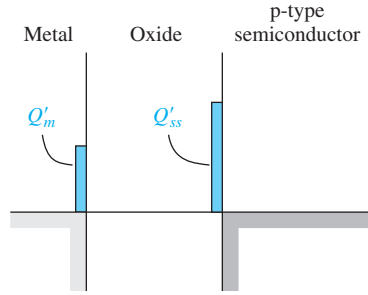$$V_{ox} = \frac{Q'_m}{C_{ox}} \tag{10.23}$$

Figure 10.18 | Charge distribution in a MOS capacitor at flat band.

where C_{ox} is the oxide capacitance per unit area.[1] Substituting Equation (10.22) into Equation (10.23), we have

$$V_{ox} = \frac{-Q'_{ss}}{C_{ox}} \qquad (10.24)$$

In the flat-band condition, the surface potential is zero, or $\phi_s = 0$. Then from Equation (10.21), we have

$$V_G = \boxed{V_{FB} = \phi_{ms} - \frac{Q'_{ss}}{C_{ox}}} \qquad (10.25)$$

Equation (10.25) is the flat-band voltage for this MOS device.

Objective: Calculate the flat-band voltage for a MOS capacitor with a p-type semiconductor substrate.

EXAMPLE 10.3

Consider a MOS capacitor with a p-type silicon substrate doped to $N_a = 10^{16}$ cm^{-3}, a silicon dioxide insulator with a thickness of $t_{ox} = 20$ nm $= 200$ Å, and an n$^+$ polysilicon gate. Assume that $Q'_{ss} = 5 \times 10^{10}$ electronic charges per cm^2.

■ **Solution**
The work function difference, from Figure 10.16, is $\phi_{ms} \cong -1.1$V. The oxide capacitance is found to be

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{(3.9)(8.85 \times 10^{-14})}{200 \times 10^{-8}} = 1.726 \times 10^{-7} \text{ F/cm}^2$$

The equivalent oxide charge density is

$$Q'_{ss} = (5 \times 10^{10})(1.6 \times 10^{-19}) = 8 \times 10^{-9} \text{ C/cm}^2$$

The flat-band voltage is then determined to be

$$V_{FB} = \phi_{ms} - \frac{Q'_{ss}}{C_{ox}} = -1.1 \frac{-8 \times 10^{-9}}{1.726 \times 10^{-7}} = -1.15 \text{ V}$$

---

[1]Although we will, in general, use the primed notation for capacitance per unit area or charge per unit area, we will omit, for convenience, the prime on the oxide capacitance per unit area parameter.

■ **Comment**

The applied gate voltage required to achieve the flat-band condition for this p-type substrate is negative. If the amount of fixed oxide charge increases, the flat-band voltage becomes even more negative.

■ **EXERCISE PROBLEM**

Ex 10.3   Repeat Example 10.3 for a doping concentration of $N_a = 2 \times 10^{15}$ cm^{-3}, an oxide thickness of $t_{ox} = 4$ nm $= 40$ Å, and $Q'_{ss} = 2 \times 10^{10}$ electronic charges per cm^2. What is the value of the metal–semiconductor work function difference?

(Ans. $\phi_{ms} \cong -1.03$ V, $V_{FB} = -1.034$ V)

10.1.6   Threshold Voltage

The threshold voltage is defined as the applied gate voltage required to achieve the threshold inversion point. The threshold inversion point, in turn, is defined as the condition when the surface potential is $\phi_s = 2\phi_{fp}$ for the p-type semiconductor and $\phi_s = 2\phi_{fn}$ for the n-type semiconductor. These conditions are shown in Figures 10.9 and 10.10. The threshold voltage will be derived in terms of the electrical and geometrical properties of the MOS capacitor.

Figure 10.19 shows the charge distribution through the MOS device at the threshold inversion point for a p-type semiconductor substrate. The space charge width has reached its maximum value. We will assume that there is an equivalent oxide charge Q'_{ss} and the positive charge on the metal gate at threshold is Q'_{mT}. The prime on the charge terms indicates charge per unit area. Even though we are assuming that the surface has been inverted, we will neglect the inversion layer charge at this threshold inversion point. From conservation of charge, we can write

$$Q'_{mT} + Q'_{ss} = |Q'_{SD}(\text{max})| \tag{10.26}$$

where

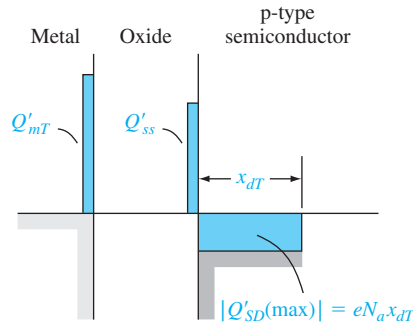$$\boxed{|Q'_{SD}(\text{max})| = eN_a x_{dT}} \tag{10.27}$$

Figure 10.19 | Charge distribution in a MOS capacitor with a p-type substrate at the threshold inversion point.
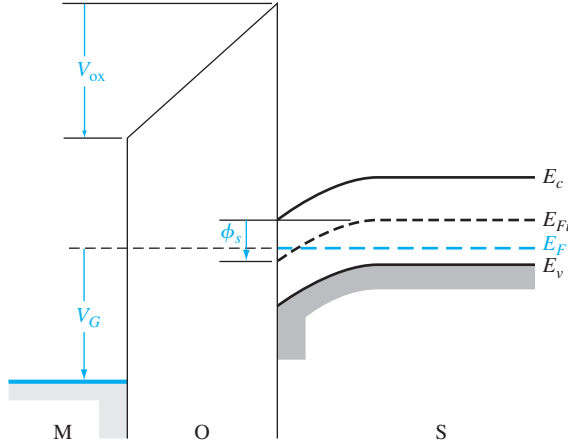
Figure 10.20 | Energy-band diagram through the MOS structure with a positive applied gate bias.

and is the magnitude of the maximum space charge density per unit area of the depletion region.

The energy-band diagram of the MOS system with an applied positive gate voltage is shown in Figure 10.20. As we mentioned, an applied gate voltage will change the voltage across the oxide and will change the surface potential. We had from Equation (10.20) that

$$V_G = \Delta V_{ox} + \Delta \phi_s = V_{ox} + \phi_s + \phi_{ms}$$

At threshold, we can define $V_G = V_{TN}$, where V_{TN} is the threshold voltage that creates the electron inversion layer charge. The surface potential is $\phi_s = 2\phi_{fp}$ at threshold, so Equation (10.20) can be written as

$$V_{TN} = V_{oxT} + 2\phi_{fp} + \phi_{ms} \tag{10.28}$$

where V_{oxT} is the voltage across the oxide at this threshold inversion point.

The voltage V_{oxT} can be related to the charge on the metal and to the oxide capacitance by

$$V_{oxT} = \frac{Q'_{mT}}{C_{ox}} \tag{10.29}$$

where again C_{ox} is the oxide capacitance per unit area. Using Equation (10.26), we can write

$$V_{oxT} = \frac{Q'_{mT}}{C_{ox}} = \frac{1}{C_{ox}} \left(|Q'_{SD}(\max)| - Q'_{ss} \right) \tag{10.30}$$

Finally, the threshold voltage can be written as

$$\boxed{V_{TN} = \frac{|Q'_{SD}(\max)|}{C_{ox}} - \frac{Q'_{ss}}{C_{ox}} + \phi_{ms} + 2\phi_{fp}} \tag{10.31a}$$

or

$$V_{TN} = \left(|Q'_{SD}(\max)| - Q'_{ss} \right) \left(\frac{t_{ox}}{\epsilon_{ox}} \right) + \phi_{ms} + 2\phi_{fp} \tag{10.31b}$$

Using the definition of flat-band voltage from Equation (10.25), we can also express the threshold voltage as

$$V_{TN} = \frac{|Q'_{SD}(\max)|}{C_{ox}} + V_{FB} + 2\phi_{fp} \qquad (10.31c)$$

For a given semiconductor material, oxide material, and gate metal, the threshold voltage is a function of semiconductor doping, oxide charge Q'_{ss}, and oxide thickness.

EXAMPLE 10.4

Objective: Calculate the threshold voltage of a MOS system using an aluminum gate.

Consider a p-type silicon substrate at $T = 300$ K doped to $N_a = 10^{15}$ cm^{-3}. Let $Q'_{ss} = 10^{10}$ cm^{-2}, $t_{ox} = 12$ nm $= 120$ Å, and assume the oxide is silicon dioxide.

■ **Solution**

From Figure 10.16, we find $\phi_{ms} \cong -0.88$ V. The other parameters are

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259) \ln\left(\frac{10^{15}}{1.5 \times 10^{10}}\right) = 0.2877 \text{ V}$$

and

$$x_{dT} = \left\{\frac{4\epsilon_s \phi_{fp}}{eN_a}\right\}^{1/2} = \left\{\frac{4(11.7)(8.85 \times 10^{-14})(0.2877)}{(1.6 \times 10^{-19})(10^{15})}\right\}^{1/2} = 8.63 \times 10^{-5} \text{ cm}$$

Then

$$|Q'_{SD}(\max)| = eN_a x_{dT} = (1.6 \times 10^{-19})(10^{15})(8.63 \times 10^{-5}) = 1.381 \times 10^{-8} \text{ C/cm}^2$$

The threshold voltage is now found to be

$$V_{TN} = (|Q'_{SD}(\max)| - Q'_{ss})\left(\frac{t_{ox}}{\epsilon_{ox}}\right) + \phi_{ms} + 2\phi_{fp}$$

$$= [(1.381 \times 10^{-8}) - (10^{10})(1.6 \times 10^{-19})] \cdot \left[\frac{120 \times 10^{-8}}{(3.9)(8.85 \times 10^{-14})}\right]$$

$$+ (-0.88) + 2(0.2877)$$

or

$$V_{TN} = -0.262 \text{ V}$$

■ **Comment**

In this example, the semiconductor is fairly lightly doped, which, in conjunction with the positive charge in the oxide and the work function difference, is sufficient to induce an electron inversion layer charge even with zero applied gate voltage. This condition makes the threshold voltage negative.

■ **EXERCISE PROBLEM**

Ex 10.4　Determine the metal–semiconductor work function difference and the threshold voltage for a silicon MOS device at $T = 300$ K for the following parameters: p$^+$ polysilicon gate, $N_a = 2 \times 10^{16}$ cm^{-3}, $t_{ox} = 8$ nm $= 80$ Å, and $Q'_{ss} = 2 \times 10^{10}$ cm^{-2}.

(Ans. $\phi_{ms} \cong +0.28$ V, $V_{TN} = +1.16$ V)

A negative threshold voltage for a p-type substrate implies a depletion mode device. A negative voltage must be applied to the gate in order to make the inversion
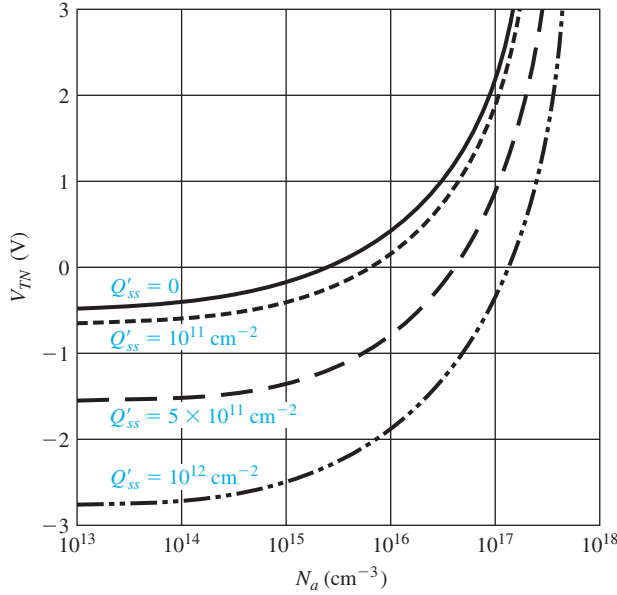
Figure 10.21 | Threshold voltage of an n-channel MOSFET versus the p-type substrate doping concentration for various values of oxide trapped charge ($t_{ox} = 500$ Å, aluminum gate).

layer charge equal to zero, whereas a positive gate voltage will induce a larger inversion layer charge.

Figure 10.21 is a plot of the threshold voltage V_{TN} as a function of the acceptor doping concentration for various positive oxide charge values. We may note that the p-type semiconductor must be somewhat heavily doped in order to obtain an enhancement mode device.

The previous derivation of the threshold voltage assumed a p-type semiconductor substrate. The same type of derivation can be done with an n-type semiconductor substrate, where a negative gate voltage can induce an inversion layer of holes at the oxide–semiconductor interface.

Figure 10.15 shows the energy-band diagram of the MOS structure with an n-type substrate and with an applied negative gate voltage. The threshold voltage for this case can be derived and is given by

$$V_{TP} = \left(-|Q'_{SD}\,(\text{max})| - Q'_{ss}\right)\left(\frac{t_{ox}}{\epsilon_{ox}}\right) + \phi_{ms} - 2\phi_{fn} \qquad (10.32)$$

where

$$\phi_{ms} = \phi'_m - \left(\chi' + \frac{E_g}{2e} - \phi_{fn}\right) \qquad (10.33a)$$

$$|Q'_{SD}\,(\text{max})| = eN_d x_{dT} \qquad (10.33b)$$

$$x_{dT} = \left\{\frac{4\epsilon_s\phi_{fn}}{eN_d}\right\}^{1/2} \qquad (10.33c)$$
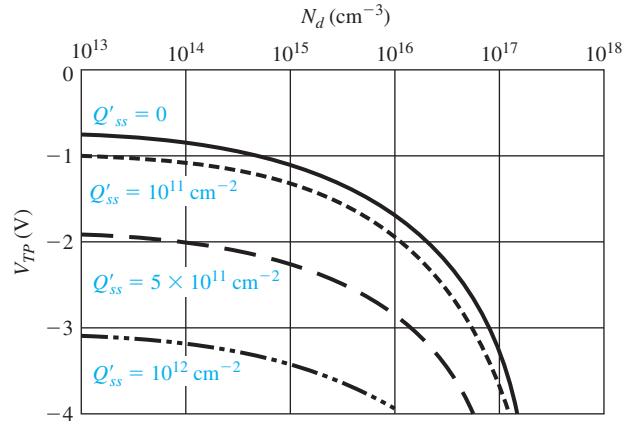
Figure 10.22 | Threshold voltage of a p-channel MOSFET versus the n-type substrate doping concentration for various values of oxide trapped charge (t_{ox} = 500 Å, aluminum gate).

and

$$\phi_{fn} = V_t \ln \left(\frac{N_d}{n_i} \right) \tag{10.33d}$$

We may note that x_{dT} and ϕ_{fn} are defined as positive quantities. We may also note that the notation of V_{TP} is the threshold voltage that will induce an inversion layer of holes. We will later drop the N and P subscript notation on the threshold voltage, but, for the moment, the notation may be useful for clarity.

Figure 10.22 is a plot of V_{TP} versus doping concentration for several values of Q'_{ss}. We may note that, for all values of positive oxide charge, this MOS capacitor is always an enhancement mode device. As the Q'_{ss} charge increases, the threshold voltage becomes more negative, which means that it takes a larger applied gate voltage to create the inversion layer of holes at the oxide–semiconductor interface.

DESIGN EXAMPLE 10.5

Objective: Determine the gate material and design the semiconductor doping concentration to yield a specified threshold voltage.

Consider a MOS device with silicon dioxide and an n-type silicon substrate. The oxide thickness is t_{ox} = 12 nm = 120 Å and the oxide charge is $Q'_{ss} = 2 \times 10^{10}$ cm^{-2}. The threshold voltage is to be approximately V_{TP} = −0.3 V.

■ **Solution**

The solution to this design problem is not straight forward, since the doping concentration parameter appears in the terms ϕ_{fn}, x_{dT}, Q'_{SD} (max), and ϕ_{ms}. The threshold voltage, then, is a nonlinear function of N_d. We resort to trial and error to obtain a solution.

Figure 10.22 shows the threshold voltage for an aluminum gate system. Since the required threshold voltage in this problem is less negative than the values shown in Figure 10.22,

we require a metal–semiconductor work function difference value that is more positive than for the aluminum gate. We therefore choose a p^+ polysilicon gate.

Consider a doping concentration of $N_d = 10^{17}$ cm^{-3}. From Figure 10.16, the metal–semiconductor work function difference is $\phi_{ms} \cong +1.1$ V. The remaining parameters are found to be

$$\phi_{fn} = V_t \ln\left(\frac{N_d}{n_i}\right) = (0.0259)\ln\left(\frac{10^{17}}{1.5 \times 10^{10}}\right) = 0.407 \text{ V}$$

$$x_{dT} = \left(\frac{4\epsilon_s \phi_{fn}}{eN_d}\right)^{1/2} = \left\{\frac{4(11.7)(8.85 \times 10^{-14})(0.407)}{(1.6 \times 10^{-19})(10^{17})}\right\}^{1/2}$$

$$= 1.026 \times 10^{-5} \text{ cm}$$

$$|Q'_{SD}(\text{max})| = eN_d x_{dT} = (1.6 \times 10^{-19})(10^{17})(1.026 \times 10^{-5})$$

$$= 1.642 \times 10^{-7} \text{ C/cm}^2$$

The threshold voltage is determined to be

$$V_{TP} = [-|Q'_{SD}(\text{max})| - Q'_{ss}] \cdot \left(\frac{t_{ox}}{\epsilon_{ox}}\right) + \phi_{ms} - 2\phi_{fn}$$

or

$$V_{TP} = \frac{[-(1.642 \times 10^{-7}) - (2 \times 10^{10})(1.6 \times 10^{-19})] \cdot (120 \times 10^{-8})}{(3.9)(8.85 \times 10^{-14})}$$
$$+1.1 - 2(0.407)$$

which yields

$$V_{TP} = -0.296 \text{ V} \cong -0.3 \text{ V}$$

■ Comment
The negative threshold voltage, with the n-type substrate, implies that this MOS capacitor is an enhancement mode device. The inversion layer charge is zero with zero applied gate voltage, and a negative gate voltage must be applied to induce the hole inversion charge.

■ EXERCISE PROBLEM
Ex 10.5 Consider a MOS capacitor with silicon dioxide and an n-type silicon substrate at $T = 300$ K with the following parameters: p^+ polysilicon gate, $N_d = 2 \times 10^{16}$ cm^{-3}, $t_{ox} = 20$ nm $= 200$ Å, and $Q'_{ss} = 5 \times 10^{10}$ cm^{-2}. Determine the threshold voltage. Is the capacitor an enhancement mode or depletion mode device?

(Ans. $V_{TP} = -0.12$ V, enhancement mode)

TEST YOUR UNDERSTANDING

TYU 10.1 (a) Consider an oxide-to-n-type silicon junction at $T = 300$ K. The impurity doping concentration in the silicon is $N_d = 8 \times 10^{15}$ cm^{-3}. Calculate the maximum space charge width in the silicon. (b) Repeat part (a) for a doping concentration of $N_d = 4 \times 10^{16}$ cm^{-3}.

[Ans. (a) $0.332 \ \mu$m; (b) $0.158 \ \mu$m]

TYU 10.2 Consider an n^+ polysilicon gate in a MOS structure with a p-type silicon sub-strate. The doping concentration of the silicon is $N_a = 3 \times 10^{16}$ cm^{-3}. Using Equation (10.16), find the value of ϕ_{ms}.

(Ans. $\phi_{ms} = -0.936$ V)

TYU 10.3 Repeat TYU 10.2 for a p^+ polysilicon gate using Equation (10.17).

(Ans. $\phi_{ms} = +0.184$ V)

TYU 10.4 Consider the MOS capacitor described in Exercise TYU 10.3. The oxide thickness is $t_{ox} = 16$ nm $= 160$ Å and the oxide charge density is $Q'_{ss} = 8 \times 10^{10}$ cm^{-2}. Determine the flat-band voltage.

(Ans. $+0.125$ V)

TYU 10.5 Consider an n^+ polysilicon gate on silicon dioxide with a p-type silicon sub-strate doped to $N_a = 3 \times 10^{16}$ cm^{-3}. Assume $Q'_{ss} = 5 \times 10^{10}$ cm^{-2}. Determine the required oxide thickness such that the threshold voltage is $V_{TN} = +0.65$ V.

(Ans. $t_{ox} = 45.2$ nm $= 452$ Å)

10.2 | CAPACITANCE–VOLTAGE CHARACTERISTICS

As mentioned previously, the MOS capacitor structure is the heart of the MOSFET. A great deal of information about the MOS device and the oxide–semiconductor interface can be obtained from the capacitance versus voltage or C–V characteristics of the device. The capacitance of a device is defined as

$$C = \frac{dQ}{dV} \tag{10.34}$$

where dQ is the magnitude of the differential change in charge on one plate as a function of the differential change in voltage dV across the capacitor. The capacitance is a small-signal or ac parameter and is measured by superimposing a small ac voltage on an applied dc gate voltage. The capacitance, then, is measured as a function of the applied dc gate voltage.

10.2.1 Ideal C–V Characteristics

First we will consider the ideal C–V characteristics of the MOS capacitor and then discuss some of the deviations that occur from these idealized results. We will initially assume that there is zero charge trapped in the oxide and also that there is no charge trapped at the oxide–semiconductor interface.

There are three operating conditions of interest in the MOS capacitor: accumulation, depletion, and inversion. Figure 10.23a shows the energy-band diagram of a MOS capacitor with a p-type substrate for the case when a negative voltage is applied to the gate, inducing an accumulation layer of holes in the semiconductor at the oxide–semiconductor interface. A small differential change in voltage across the MOS structure will cause a differential change in charge on the metal gate and also in the hole accumulation charge, as shown in Figure 10.23b. The differential changes in charge density occur at the edges of the oxide, as in a parallel-plate capacitor. The
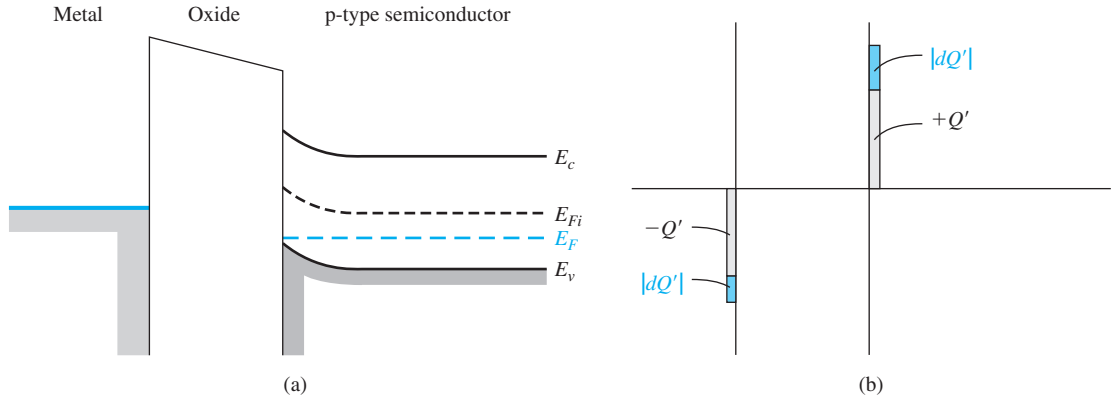
Figure 10.23 | (a) Energy-band diagram through a MOS capacitor for the accumulation mode. (b) Differential charge distribution at accumulation for a differential change in gate voltage.
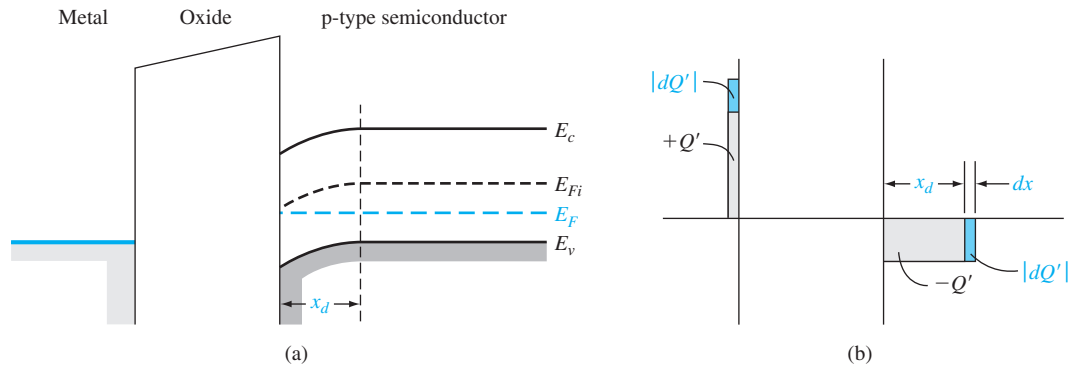

Figure 10.24 | (a) Energy-band diagram through a MOS capacitor for the depletion mode. (b) Differential charge distribution at depletion for a differential change in gate voltage.

capacitance C' per unit area of the MOS capacitor for this accumulation mode is just the oxide capacitance, or

$$C'(\text{acc}) = C_{\text{ox}} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}}} \tag{10.35}$$

Figure 10.24a shows the energy-band diagram of the MOS device when a small positive voltage is applied to the gate, inducing a space charge region in the semiconductor; Figure 10.24b shows the charge distribution through the device for this condition. The oxide capacitance and the capacitance of the depletion region are in series. A small differential change in voltage across the capacitor will cause a differential change in the space charge width. The corresponding differential changes

in charge densities are shown in the figure. The total capacitance of the series combination is

$$\frac{1}{C'(\text{depl})} = \frac{1}{C_{ox}} + \frac{1}{C'_{SD}}$$ (10.36a)

or

$$C'(\text{depl}) = \frac{C_{ox}C'_{SD}}{C_{ox} + C'_{SD}}$$ (10.36b)

Since $C_{ox} = \epsilon_{ox}/t_{ox}$ and $C'_{SD} = \epsilon_s/x_d$, Equation (10.36b) can be written as

$$C'(\text{depl}) = \frac{C_{ox}}{1 + \dfrac{C_{ox}}{C'_{SD}}} = \frac{\epsilon_{ox}}{t_{ox} + \left(\dfrac{\epsilon_{ox}}{\epsilon_s}\right)x_d}$$ (10.37)

As the space charge width increases, the total capacitance $C'(\text{depl})$ decreases.

We had defined the threshold inversion point to be the condition when the maximum depletion width is reached, but there is essentially zero inversion charge density. This condition will yield a minimum capacitance C'_{min}, which is given by

$$C'_{min} = \frac{\epsilon_{ox}}{t_{ox} + \left(\dfrac{\epsilon_{ox}}{\epsilon_s}\right)x_{dT}}$$ (10.38)

Figure 10.25a shows the energy-band diagram of this MOS device for the inversion condition. In the ideal case, a small incremental change in the voltage across the MOS capacitor will cause a differential change in the inversion layer charge density. The space charge width does not change. If the inversion charge can respond to the change in capacitor voltage as indicated in Figure 10.25b, then the capacitance is again just the oxide capacitance, or

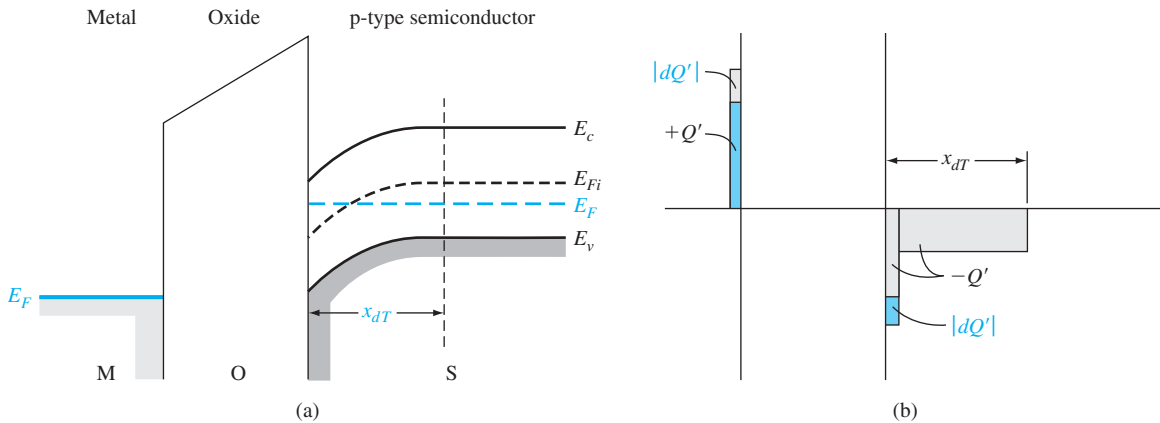$$C'(\text{inv}) = C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$$ (10.39)

Figure 10.25 | (a) Energy-band diagram through a MOS capacitor for the inversion mode. (b) Differential charge distribution at inversion for a low-frequency differential change in gate voltage.
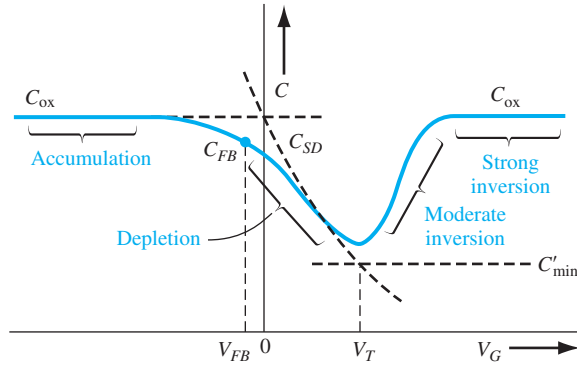
Figure 10.26 | Ideal low-frequency capacitance versus gate voltage of a MOS capacitor with a p-type substrate. Individual capacitance components are also shown.

Figure 10.26 shows the ideal capacitance versus gate voltage, or C–V, character-istics of the MOS capacitor with a p-type substrate. The three dashed segments cor-respond to the three components C_{ox}, C'_{SD}, and C'_{min}. The solid curve is the ideal net capacitance of the MOS capacitor. Moderate inversion, which is indicated in the figure, is the transition region between the point when only the space charge density changes with gate voltage and when only the inversion charge density changes with gate voltage.

The point on the curve that corresponds to the flat-band condition is of interest. The flat-band condition occurs between the accumulation and depletion conditions. The capacitance at flat band is given by

$$C'_{FB} = \frac{\epsilon_{ox}}{t_{ox} + \left(\frac{\epsilon_{ox}}{\epsilon_s}\right)\sqrt{\left(\frac{kT}{e}\right)\left(\frac{\epsilon_s}{eN_a}\right)}} \qquad (10.40)$$

We may note that the flat-band capacitance is a function of oxide thickness as well as semiconductor doping. The general location of this point on the C–V plot is shown in Figure 10.26.

Objective: Calculate C_{ox}, C'_{min}, and C'_{FB} for a MOS capacitor.

EXAMPLE 10.6

Consider a p-type silicon substrate at $T = 300$ K doped to $N_a = 10^{16}$ cm^{-3}.

The oxide is silicon dioxide with a thickness of $t_{ox} = 18$ nm $= 180$ Å, and the gate is aluminum.

■ **Solution**

The oxide capacitance is

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{(3.9)(8.85 \times 10^{-14})}{180 \times 10^{-8}} = 1.9175 \times 10^{-7} \text{ F/cm}^2$$

To find the minimum capacitance, we need to calculate

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259) \ln\left(\frac{10^{16}}{1.5 \times 10^{10}}\right) = 0.3473 \text{ V}$$

and

$$x_{dT} = \left\{ \frac{4\epsilon_s \phi_{fp}}{eN_a} \right\}^{1/2} = \left\{ \frac{4(11.7)(8.85 \times 10^{-14})(0.3473)}{(1.6 \times 10^{-19})(10^{16})} \right\}^{1/2}$$

$$\cong 0.30 \times 10^{-4} \text{ cm}$$

Then

$$C'_{min} = \frac{\epsilon_{ox}}{t_{ox} + \left(\frac{\epsilon_{ox}}{\epsilon_s} \right) x_{dT}} = \frac{(3.9)(8.85 \times 10^{-14})}{180 \times 10^{-8} + \left(\frac{3.9}{11.7} \right)(0.30 \times 10^{-4})}$$

$$= 2.925 \times 10^{-8} \text{ F/cm}^2$$

We may note that

$$\frac{C'_{min}}{C_{ox}} = \frac{2.925 \times 10^{-8}}{1.9175 \times 10^{-7}} = 0.1525$$

The flat-band capacitance is

$$C'_{FB} = \frac{\epsilon_{ox}}{t_{ox} + \left(\frac{\epsilon_{ox}}{\epsilon_s} \right) \sqrt{\frac{V_t \epsilon_s}{eN_a}}}$$

$$= \frac{(3.9)(8.85 \times 10^{-14})}{180 \times 10^{-8} + \left(\frac{3.9}{11.7} \right) \sqrt{\frac{(0.0259)(11.7)(8.85 \times 10^{-14})}{(1.6 \times 10^{-19})(10^{16})}}}$$

$$= 1.091 \times 10^{-7} \text{ F/cm}^2$$

We also note that

$$\frac{C'_{FB}}{C_{ox}} = \frac{1.091 \times 10^{-7}}{1.9175 \times 10^{-7}} = 0.569$$

■ **Comment**

The ratios of C'_{min}/C_{ox} and C'_{FB}/C_{ox} are typical values obtained in C–V plots.

■ **EXERCISE PROBLEM**

Ex 10.6 Consider a MOS capacitor with the following parameters: n+ polysilicon gate, $N_a = 3 \times 10^{16}$ cm^{-3}, $t_{ox} = 8$ nm = 80 Å, and $Q'_{ss} = 2 \times 10^{10}$ cm^{-2}. Determine the ratios C'_{min}/C_{ox} and C'_{FB}/C_{ox}.

(Ans. $C'_{min}/C_{ox} = 0.504$, $C'_{FB}/C_{ox} = 0.118$)

If we assume the oxide capacitance per unit area is $C_{ox} = 1.9175 \times 10^{-7}$ F/cm^2 and the channel length and width are $L = 2$ μm and $W = 20$ μm, respectively, then the total gate oxide capacitance is

$$C_{oxT} = C_{ox} LW = (1.9175 \times 10^{-7})(2 \times 10^{-4})(20 \times 10^{-4})$$

$$= 7.67 \times 10^{-14} \text{ F} = 0.0767 \text{ pF} = 76.7 \text{ fF}$$

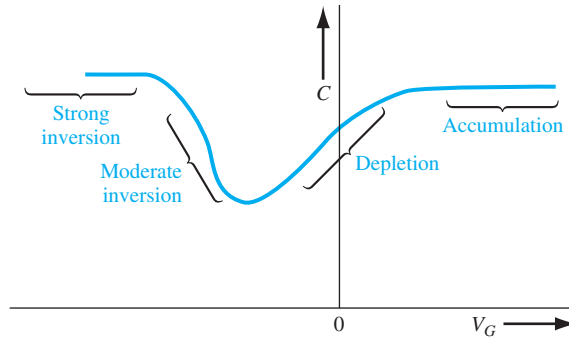The total oxide capacitance in a typical MOS device is quite small.

Figure 10.27 | Ideal low-frequency capacitance versus gate voltage of a MOS capacitor with an n-type substrate.

The same type of ideal *C–V* characteristics is obtained for a MOS capacitor with an n-type substrate by changing the sign of the voltage axis. The accumulation condition is obtained for a positive gate bias and the inversion condition is obtained for a negative gate bias. This ideal curve is shown in Figure 10.27.

10.2.2 Frequency Effects

Figure 10.25a shows the MOS capacitor with a p-type substrate and biased in the inversion condition. We have argued that a differential change in the capacitor voltage in the ideal case causes a differential change in the inversion layer charge density. However, we must consider the source of electrons that produces a change in the inversion charge density.

There are two sources of electrons that can change the charge density of the inversion layer. The first source is by diffusion of minority carrier electrons from the p-type substrate across the space charge region. This diffusion process is the same as that in a reverse-biased pn junction that generates the ideal reverse saturation current. The second source of electrons is by thermal generation of electron–hole pairs within the space charge region. This process is again the same as that in a reverse-biased pn junction generating the reverse-biased generation current. Both of these processes generate electrons at a particular rate. The electron concentration in the inversion layer, then, cannot change instantaneously. If the ac voltage across the MOS capacitor changes rapidly, the change in the inversion layer charge will not be able to respond. The *C–V* characteristics will then be a function of the frequency of the ac signal used to measure the capacitance.

In the limit of a very high frequency, the inversion layer charge will not respond to a differential change in capacitor voltage. Figure 10.28 shows the charge distribution in the MOS capacitor with a p-type substrate. At a high-signal frequency, the differential change in charge occurs at the metal and in the space charge width in the semiconductor. The capacitance of the MOS capacitor is then C'_{min}, which we discussed earlier.

The high-frequency and low-frequency limits of the *C–V* characteristics are shown in Figure 10.29. In general, high frequency corresponds to a value on the order of 1 MHz and low frequency corresponds to values in the range of 5 to 100 Hz. Typically, the high-frequency characteristics of the MOS capacitor are measured.
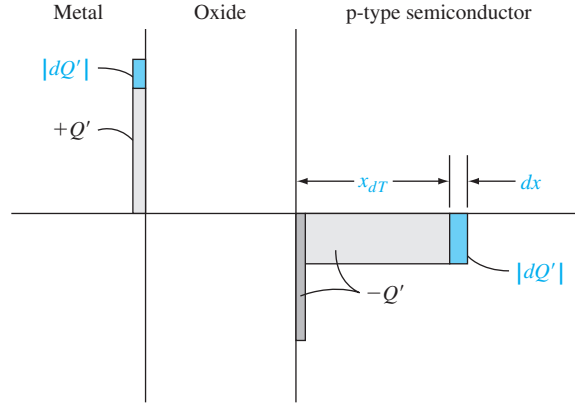
Figure 10.28 | Differential charge distribution at inversion for a high-frequency differential change in gate voltage.
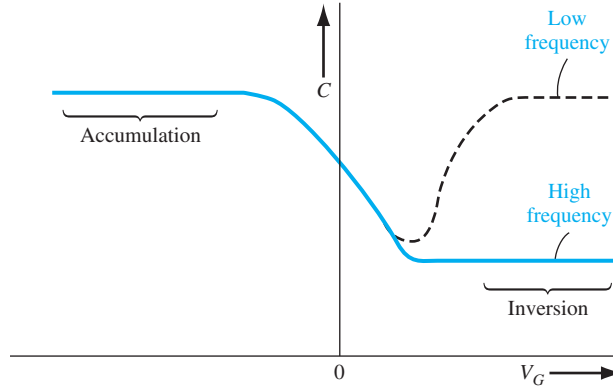

Figure 10.29 | Low-frequency and high-frequency capacitance versus gate voltage of a MOS capacitor with a p-type substrate.

10.2.3 Fixed Oxide and Interface Charge Effects

In all of the discussion concerning C–V characteristics so far, we have assumed an ideal oxide in which there are no fixed oxide or oxide–semiconductor interface charges. These two types of charges will change the C–V characteristics.

We previously discussed how the fixed oxide charge affects the threshold voltage. This charge will also affect the flat-band voltage. The flat-band voltage from Equation (10.25) is given by

$$V_{FB} = \phi_{ms} - \frac{Q'_{ss}}{C_{ox}}$$

where Q'_{ss} is the equivalent fixed oxide charge and ϕ_{ms} is the metal–semiconductor work function difference. The flat-band voltage shifts to more negative voltages for a positive fixed oxide charge. Since the oxide charge is not a function of gate voltage, the curves
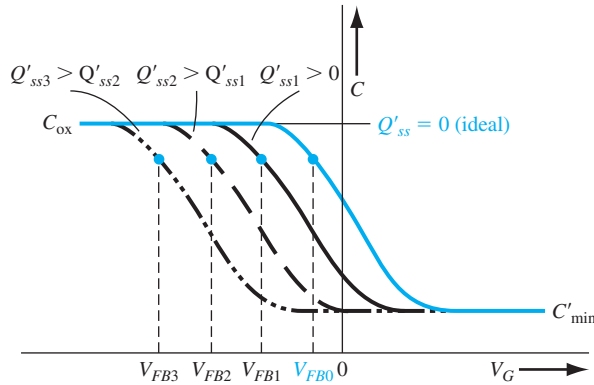
Figure 10.30 | High-frequency capacitance versus gate voltage of a MOS capacitor with a p-type substrate for several values of effective trapped oxide charge.
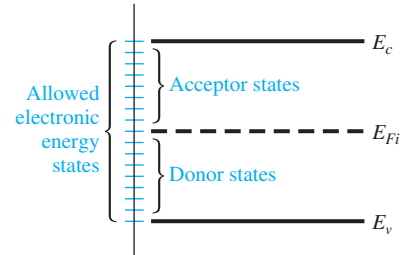

Figure 10.31 | Schematic diagram showing interface states at the oxide–semiconductor interface.

show a parallel shift with oxide charge, and the shape of the C–V curves remains the same as the ideal characteristics. Figure 10.30 shows the high-frequency characteristics of a MOS capacitor with a p-type substrate for several values of fixed positive oxide charge.

The C–V characteristics can be used to determine the equivalent fixed oxide charge. For a given MOS structure, ϕ_{ms} and C_{ox} are known, so the ideal flat-band voltage and flat-band capacitance can be calculated. The experimental value of flat-band voltage can be measured from the C–V curve, and the value of fixed oxide charge can then be determined. The C–V measurements are a valuable diagnostic tool to characterize a MOS device. This characterization is especially useful in the study of radiation effects on MOS devices, for example.

We first encountered oxide–semiconductor interface states in Chapter 9 in the discussion of Schottky barrier diodes. Figure 10.31 shows the energy-band diagram of a semiconductor at the oxide–semiconductor interface. The periodic nature of the semiconductor is abruptly terminated at the interface so that allowed electronic energy levels will exist within the forbidden bandgap. These allowed energy states are referred to as interface states. Charge can flow between the semiconductor and interface states, in contrast to the fixed oxide charge. The net charge in these interface states is a function of the position of the Fermi level in the bandgap.

In general, acceptor states exist in the upper half of the bandgap and donor states exist in the lower half of the bandgap. An acceptor state is neutral if the Fermi level is below the state and becomes negatively charged if the Fermi level is above the state. A donor state is neutral if the Fermi level is above the state and becomes positively charged if the Fermi level is below the state. The charge of the interface states is then a function of the gate voltage applied across the MOS capacitor.

Figure 10.32a shows the energy-band diagram in a p-type semiconductor of a MOS capacitor biased in the accumulation condition. In this case, there is a net positive charge trapped in the donor states. Now let the gate voltage change to produce the energy-band diagram shown in Figure 10.32b. The Fermi level corresponds to
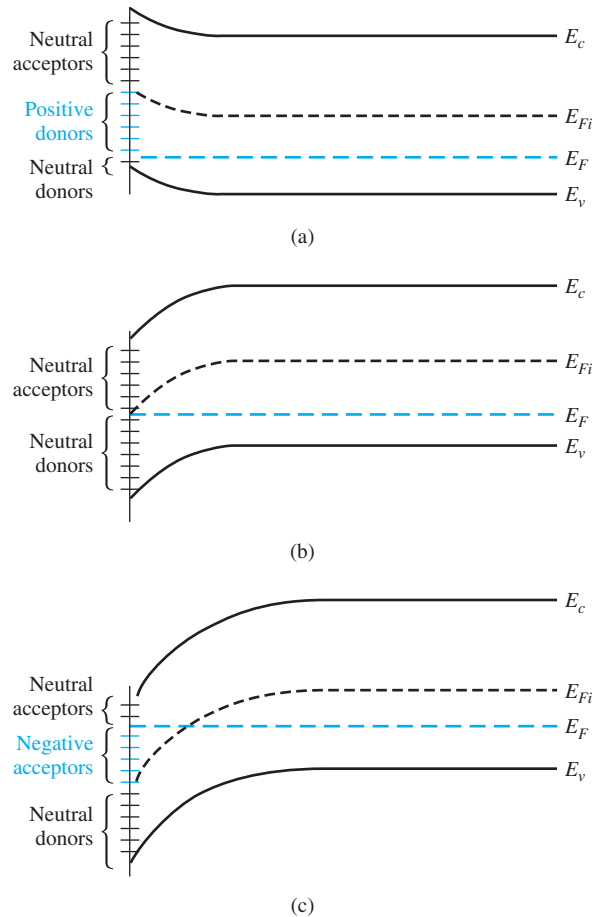
Figure 10.32 | Energy-band diagram in a p-type semiconductor showing the charge trapped in the interface states when the MOS capacitor is biased (a) in accumulation, (b) at midgap, and (c) at inversion.

the intrinsic Fermi level at the surface; thus, all interface states are neutral. This particular bias condition is known as *midgap*. Figure 10.32c shows the condition at inversion in which there is now a net negative charge in the acceptor states.

The net charge in the interface states changes from positive to negative as the gate voltage sweeps from the accumulation, depletion, to the inversion condition. We noted that the *C–V* curves shifted in the negative gate voltage direction due to positive fixed oxide charge. When interface states are present, the amount and direction of the shift change as we sweep through the gate voltage, since the amount and sign of the interface trapped charge change. The *C–V* curves now become "smeared out" as shown in Figure 10.33.

Again, the *C–V* measurements can be used as a diagnostic tool in semiconductor device process control. For a given MOS device, the ideal *C–V* curve can be determined.
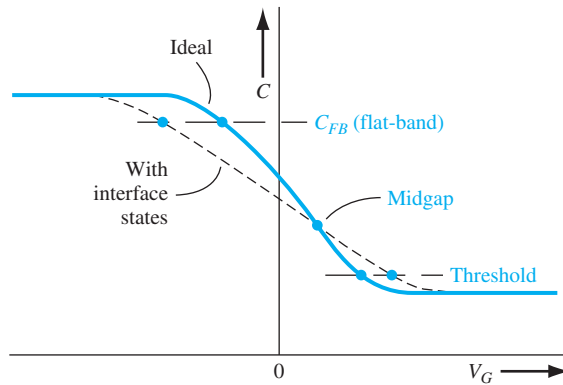
Figure 10.33 | High-frequency C–V characteristics of a
MOS capacitor showing effects of interface states.

Any "smearing out" in the experimental curve indicates the presence of interface states
and any parallel shift indicates the presence of fixed oxide charge. The amount of
smearing out can be used to determine the density of interface states. These types of
measurement are extremely useful in the study of radiation effects on MOS devices.

10.3 | THE BASIC MOSFET OPERATION

The current in a MOSFET is due to the flow of charge in the inversion layer or channel
region adjacent to the oxide–semiconductor interface. We have discussed the creation
of the inversion layer charge in enhancement-type MOS capacitors. We may also have
depletion-type devices in which a channel already exists at zero gate voltage.

10.3.1 MOSFET Structures

There are four basic MOSFET device types. Figure 10.34 shows an n-channel
enhancement mode MOSFET. Implicit in the enhancement mode notation is the idea
that the semiconductor substrate is not inverted directly under the oxide with zero
gate voltage. A positive gate voltage induces the electron inversion layer, which then
"connects" the n-type source and the n-type drain regions. The source terminal is
the source of carriers that flow through the channel to the drain terminal. For this
n-channel device, electrons flow from the source to the drain so the conventional
current will enter the drain and leave the source. The conventional circuit symbol for
this n-channel enhancement mode device is also shown in the figure.

Figure 10.35 shows an n-channel depletion mode MOSFET. An n-channel region
exists under the oxide with 0 V applied to the gate. However, we have shown that the
threshold voltage of a MOS device with a p-type substrate may be negative; this means
that an electron inversion layer already exists with zero gate voltage applied. Such a de-
vice is also considered to be a depletion mode device. The n-channel shown in the figure
can be an electron inversion layer or an intentionally doped n region. The conventional
circuit symbol for the n-channel depletion mode MOSFET is also shown in the figure.

Figures 10.36a, b show a p-channel enhancement mode MOSFET and a p-channel
depletion mode MOSFET. In the p-channel enhancement mode device, a negative
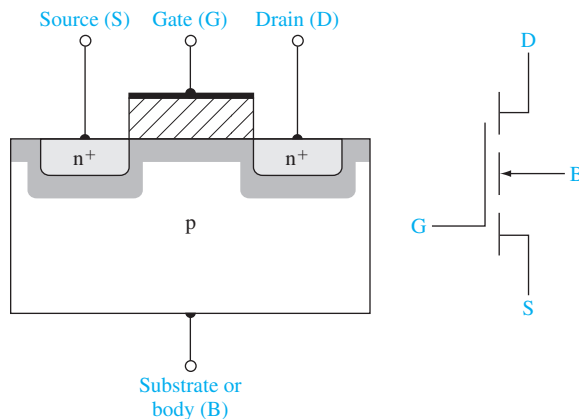
Figure 10.34 | Cross section and circuit symbol for an n-channel enhancement mode MOSFET.
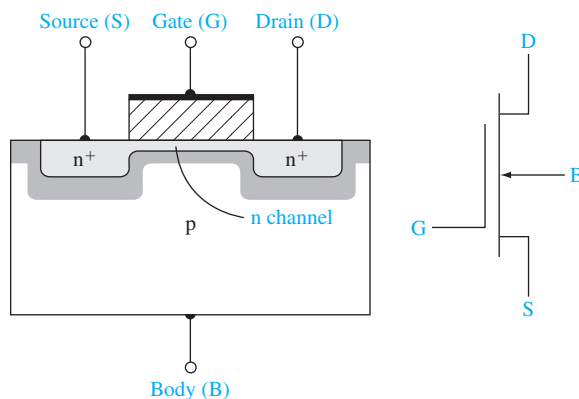

Figure 10.35 | Cross section and circuit symbol for an n-channel depletion mode MOSFET.

gate voltage must be applied to create an inversion layer of holes that will "connect" the p-type source and drain regions. Holes flow from the source to the drain, so the conventional current will enter the source and leave the drain. A p-channel region exists in the depletion mode device even with zero gate voltage. The conventional circuit symbols are shown in the figure.

10.3.2 Current–Voltage Relationship—Concepts

Figure 10.37a shows an n-channel enhancement mode MOSFET with a gate-to-source voltage that is less than the threshold voltage and with only a very small drain-to-source voltage. The source and substrate, or body, terminals are held at ground potential. With this bias configuration, there is no electron inversion layer, the drain-to-substrate pn junction is reverse biased, and the drain current is zero (disregarding pn junction leakage currents).
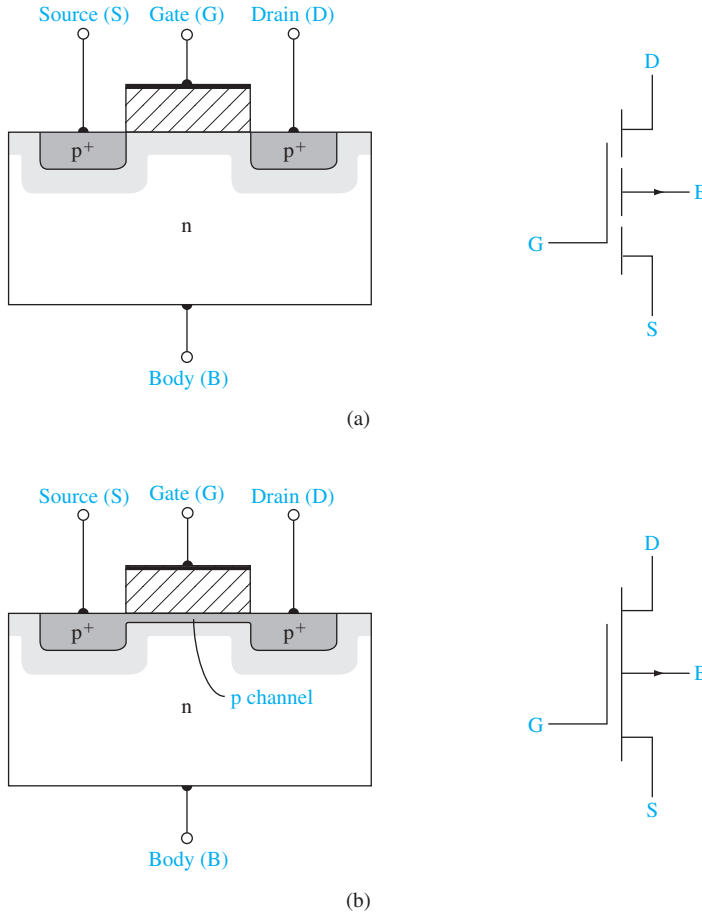
Figure 10.36 | Cross section and circuit symbol for (a) a p-channel enhancement mode MOSFET and (b) a p-channel depletion mode MOSFET.

Figure 10.37b shows the same MOSFET with an applied gate voltage such that $V_{GS} > V_T$. An electron inversion layer has been created so that when a small drain voltage is applied, the electrons in the inversion layer will flow from the source to the positive drain terminal. The conventional current enters the drain terminal and leaves the source terminal. In this ideal case, there is no current through the oxide to the gate terminal.

For small V_{DS} values, the channel region has the characteristics of a resistor, so we can write

$$I_D = g_d V_{DS} \tag{10.41}$$

where g_d is defined as the channel conductance in the limit as $V_{DS} \to 0$. The channel conductance is given by

$$g_d = \frac{W}{L} \cdot \mu_n |Q_n'| \tag{10.42}$$

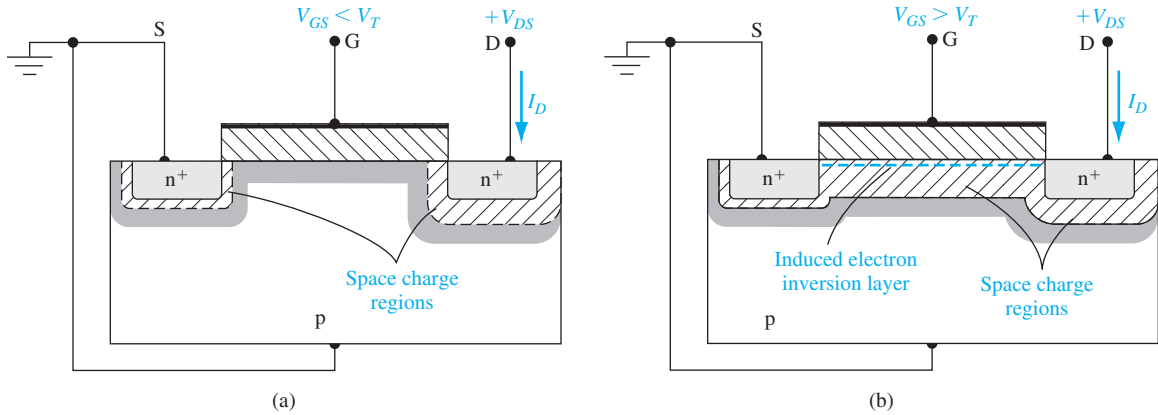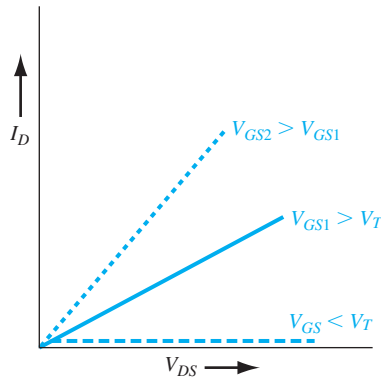**Figure 10.37** | The n-channel enhancement mode MOSFET (a) with an applied gate voltage $V_{GS} < V_T$ and (b) with an applied gate voltage $V_{GS} > V_T$.

Figure 10.38 | I_D versus V_{DS} characteristics for small values of V_{DS} at three V_{GS} voltages.

where μ_n is the mobility of the electrons in the inversion layer and $|Q_n'|$ is the magnitude of the inversion layer charge per unit area. The inversion layer charge is a function of the gate voltage; thus, the basic MOS transistor action is the modulation of the channel conductance by the gate voltage. The channel conductance, in turn, determines the drain current. We will initially assume that the mobility is a constant; we will discuss mobility effects and variations in the next chapter.

The I_D versus V_{DS} characteristics, for small values of V_{DS}, are shown in Figure 10.38. When $V_{GS} < V_T$, the drain current is zero. As V_{GS} becomes larger than V_T, channel inversion charge density increases, which increases the channel conductance. A larger value of g_d produces a larger initial slope of the I_D versus V_{DS} characteristic as shown in the figure.

Figure 10.39a shows the basic MOS structure for the case when $V_{GS} > V_T$ and the applied V_{DS} voltage is small. The thickness of the inversion channel layer in the
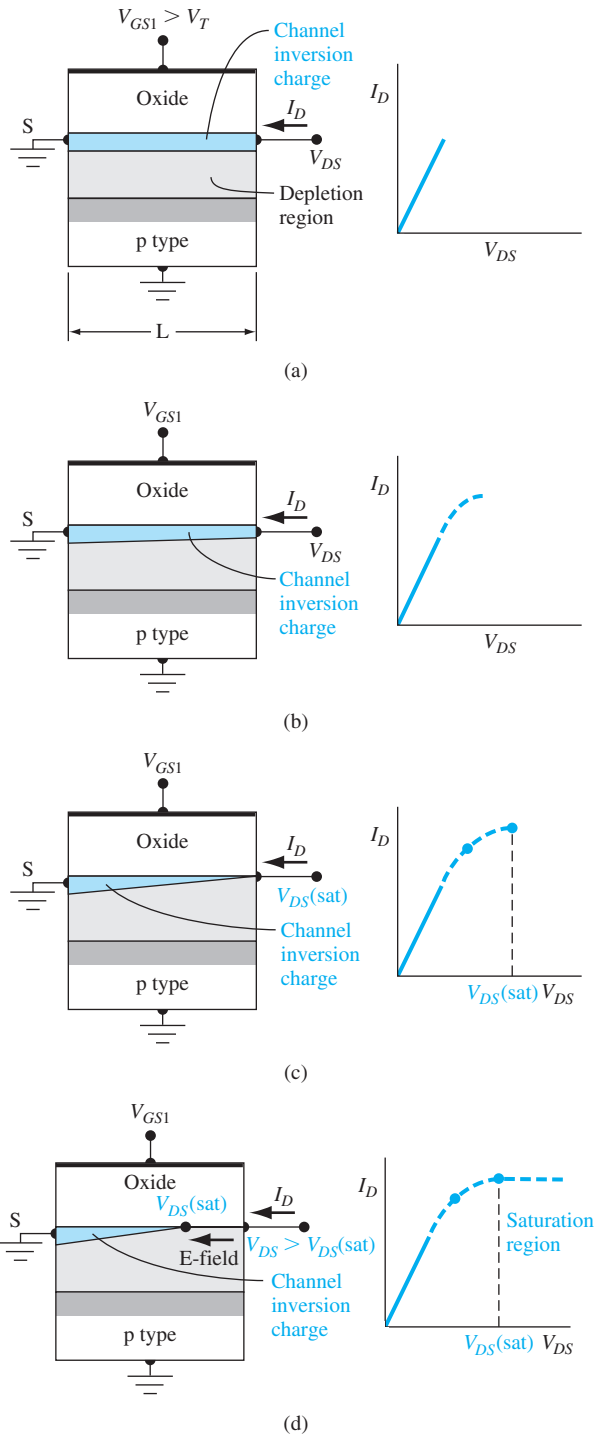
Figure 10.39 | Cross section and I_D versus V_{DS} curve when $V_{GS} < V_T$ for (a) a small V_{DS} value, (b) a larger V_{DS} value, (c) a value of $V_{DS} = V_{DS}(\text{sat})$, and (d) a value of $V_{DS} > V_{DS}(\text{sat})$.

figure qualitatively indicates the relative charge density, which is essentially constant along the entire channel length for this case. The corresponding I_D versus V_{DS} curve is shown in the figure.

Figure 10.39b shows the situation when the V_{DS} value increases. As the drain voltage increases, the voltage drop across the oxide near the drain terminal decreases, which means that the induced inversion charge density near the drain also decreases. The incremental conductance of the channel at the drain decreases, which then means that the slope of the I_D versus V_{DS} curve will decrease. This effect is shown in the I_D versus V_{DS} curve in the figure.

When V_{DS} increases to the point where the potential drop across the oxide at the drain terminal is equal to V_T, the induced inversion charge density is zero at the drain terminal. This effect is schematically shown in Figure 10.39c. At this point, the incremental conductance at the drain is zero, which means that the slope of the I_D versus V_{DS} curve is zero. We can write

$$V_{GS} - V_{DS}(\text{sat}) = V_T \qquad (10.43a)$$

or

$$\boxed{V_{DS}(\text{sat}) = V_{GS} - V_T} \qquad (10.43b)$$

where $V_{DS}(\text{sat})$ is the drain-to-source voltage producing zero inversion charge density at the drain terminal.

When V_{DS} becomes larger than the $V_{DS}(\text{sat})$ value, the point in the channel at which the inversion charge is just zero moves toward the source terminal. In this case, electrons enter the channel at the source, travel through the channel toward the drain, and then, at the point where the charge goes to zero, the electrons are injected into the space charge region where they are swept by the E-field to the drain contact. If we assume that the change in channel length ΔL is small compared to the original length L, then the drain current will be a constant for $V_{DS} > V_{DS}(\text{sat})$. The region of the I_D versus V_{DS} characteristic is referred to as the *saturation region.* Figure 10.39d shows this region of operation.

When V_{GS} changes, the I_D versus V_{DS} curve will change. We saw that, if V_{GS} increases, the initial slope of I_D versus V_{DS} increases. We can also note from Equation (10.43b) that the value of $V_{DS}(\text{sat})$ is a function of V_{GS}. We can generate the family of curves for this n-channel enhancement mode MOSFET as shown in Figure 10.40.

Figure 10.41 shows an n-channel depletion mode MOSFET. If the n-channel region is actually an induced electron inversion layer created by the metal–semiconductor work function difference and fixed charge in the oxide, the current–voltage characteristics are exactly the same as we have discussed, except that V_T is a negative quantity. We may also consider the case when the n-channel region is actually an n-type semiconductor region. In this type of device, a negative gate voltage will induce a space charge region under the oxide, reducing the thickness of the n-channel region. The reduced thickness decreases the channel conductance, which reduces the drain current. A positive gate voltage will create an electron accumulation layer, which increases the drain current. One basic requirement for this device is that the channel
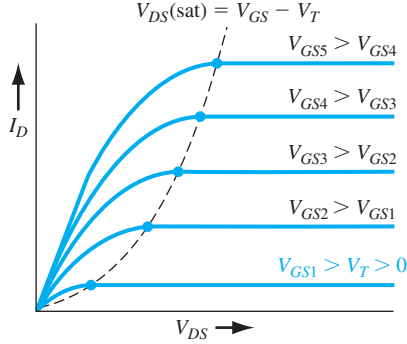
Figure 10.40 | Family of I_D versus V_{DS} curves for an n-channel enhancement mode MOSFET.
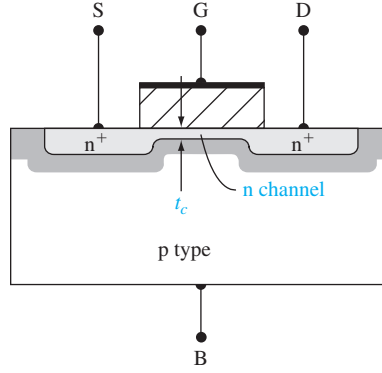
Figure 10.41 | Cross section of an n-channel depletion mode MOSFET.

thickness t_c must be less than the maximum induced space charge width in order to be able to turn the device off. The general I_D versus V_{DS} family of curves for an n-channel depletion mode MOSFET is shown in Figure 10.42.

In the next section, we derive the ideal current–voltage relation for the n-channel MOSFET. In the nonsaturation region, we obtain

$$I_D = \frac{W\mu_n C_{ox}}{2L}\left[2(V_{GS} - V_T)V_{DS} - V_{DS}^2\right] \tag{10.44a}$$

which can be written as

$$I_D = \frac{k_n'}{2}\cdot\frac{W}{L}\cdot\left[2(V_{GS} - V_T)V_{DS} - V_{DS}^2\right] \tag{10.44b}$$

or

$$I_D = K_n\left[2(V_{GS} - V_T)V_{DS} - V_{DS}^2\right] \tag{10.44c}$$

Figure 10.42 | Family of I_D versus V_{DS} curves for an n-channel depletion mode MOSFET.

The parameter $k'_n = \mu_n C_{ox}$ is called the *process conduction parameter* for the n-channel MOSFET and has units of A/V^2. The parameter $K_n = (W\mu_n C_{ox})/2L = (k'_n/2) \cdot (W/L)$ is called the *conduction parameter* for the n-channel MOSFET and also has units of A/V^2.
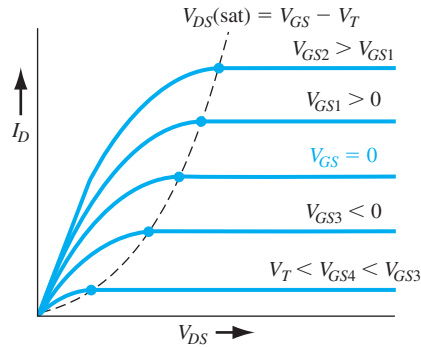
When the transistor is biased in the saturation region, the ideal currernt–voltage relation is given by

$$I_D = \frac{W\mu_n C_{ox}}{2L}(V_{GS} - V_T)^2 \qquad (10.45\text{a})$$

which can be written as

$$I_D = \frac{k'_n}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_T)^2 \qquad (10.45\text{b})$$

or

$$I_D = K_n(V_{GS} - V_T)^2 \qquad (10.45\text{c})$$

In general, for a given technology, the process conduction parameter, k'_n, is a constant. From Equations (10.44b) and (10.45b), then, we see that the design of a MOSFET, in terms of current capability, is determined by the width-to-length parameter.

The operation of a p-channel device is the same as that of the n-channel device, except the charge carrier is the hole and the conventional current direction and voltage polarities are reversed.

*10.3.3 Current–Voltage Relationship—Mathematical Derivation

In the previous section, we qualitatively discussed the current–voltage characteristics. In this section, we derive the mathematical relation between the drain current, the gate-to-source voltage, and the drain-to-source voltage. Figure 10.43 shows the geometry of the device that we use in this derivation.

In this analysis, we make the following assumptions:

1. The current in the channel is due to drift rather than diffusion.
2. There is no current through the gate oxide.
3. A gradual channel approximation is used in which $\partial E_y/\partial y \gg \partial E_x/\partial x$. This approximation means that E_x is essentially a constant.
4. Any fixed oxide charge is an equivalent charge density at the oxide–semiconductor interface.
5. The carrier mobility in the channel is constant.

We start the analysis with Ohm's law, which can be written as

$$J_x = \sigma E_x \qquad (10.46)$$

where σ is the channel conductivity and E_x is the electric field along the channel created by the drain-to-source voltage. The channel conductivity is given by $\sigma = e\mu_n n(y)$, where μ_n is the electron mobility and $n(y)$ is the electron concentration in the inversion layer.
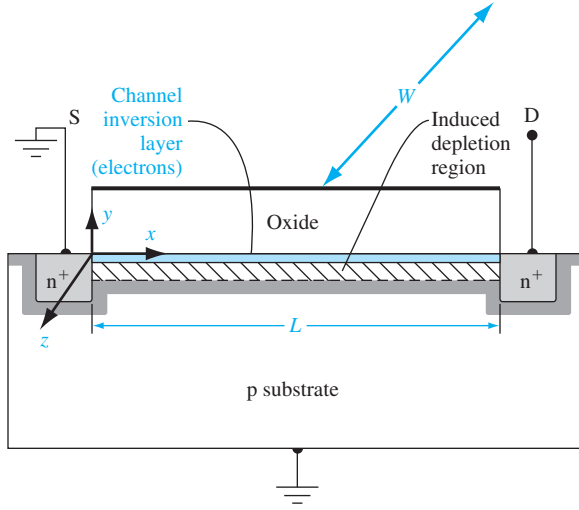
Figure 10.43 | Geometry of a MOSFET for I_D versus V_{DS} derivation.

The total channel current is found by integrating J_x over the cross-sectional area in the y and z directions. Then

$$I_x = \int_y \int_z J_x \, dy \, dz \tag{10.47}$$

We may write that

$$Q'_n = -\int en(y) \, dy \tag{10.48}$$

where Q'_n is the inversion layer charge per unit area and is a negative quantity for this case.

Equation (10.47) then becomes

$$I_x = -W\mu_n Q'_n E_x \tag{10.49}$$

where W is the channel width, the result of integrating over z.

Two concepts we use in the currernt–voltage derivation are charge neutrality and Gauss's law. Figure 10.44 shows the charge densities through the device for $V_{GS} > V_T$. The charges are all given in terms of charge per unit area. Using the concept of charge neutrality, we can write

$$Q'_m + Q'_{ss} + Q'_n + Q'_{SD}(\text{max}) = 0 \tag{10.50}$$

The inversion layer charge and induced space charge are negative for this n-channel device.

Gauss's law can be written as

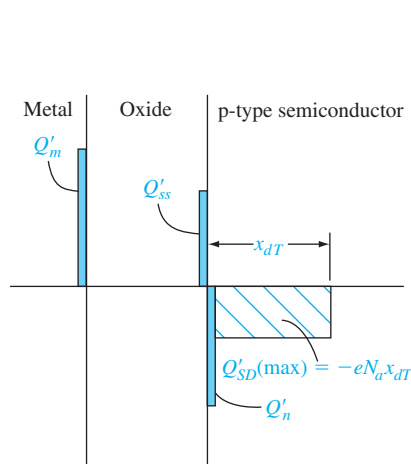$$\oint_s \epsilon E_n \, dS = Q_T \tag{10.51}$$

Figure 10.44 | Charge distribution in the n-channel enhancement mode MOSFET for $V_{GS} > V_T$.
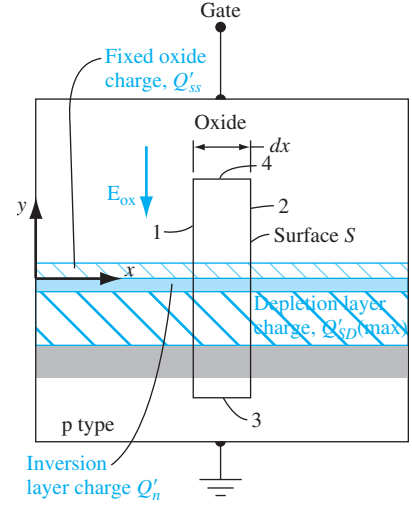

Figure 10.45 | Geometry for applying Gauss's law.

where the integral is over a closed surface. Q_T is the total charge enclosed by the surface, and E_n is the outward directed normal component of the electric field crossing the surface S. Gauss's law is applied to the surface defined in Figure 10.45. Since the surface must be enclosed, we must take into account the two end surfaces in the *x-y* plane. However, there is no *z*-component of the electric field so these two end surfaces do not contribute to the integral of Equation (10.51).

Now consider the surfaces labeled 1 and 2 in Figure 10.45. From the gradual channel approximation, we assume that E_x is essentially a constant along the channel length. This assumption means that E_x into surface 2 is the same as E_x out of surface 1. Since the integral in Equation (10.51) involves the outward component of the E-field, the contributions of surfaces 1 and 2 cancel each other. Surface 3 is in the neutral p region, so the electric field is zero at this surface.

Surface 4 is the only surface that contributes to Equation (10.51). Taking into account the direction of the electric field in the oxide, Equation (10.51) becomes

$$\oint_s \epsilon E_n \, dS = -\epsilon_{ox} E_{ox} W \, dx = Q_T \tag{10.52}$$

where ϵ_{ox} is the permittivity of the oxide. The total charge enclosed is

$$Q_T = [Q'_{ss} + Q'_n + Q'_{SD}(\max)] W \, dx \tag{10.53}$$

Combining Equations (10.52) and (10.53), we have

$$-\epsilon_{ox} E_{ox} = Q'_{ss} + Q'_n + Q'_{SD}(\max) \tag{10.54}$$

We now need an expression for E_{ox}. Figure 10.46a shows the oxide and channel. We assume that the source is at ground potential. The voltage V_x is the potential in the channel at a point x along the channel length. The potential difference across the oxide at x is a function of V_{GS}, V_x, and the metal–semiconductor work function difference.
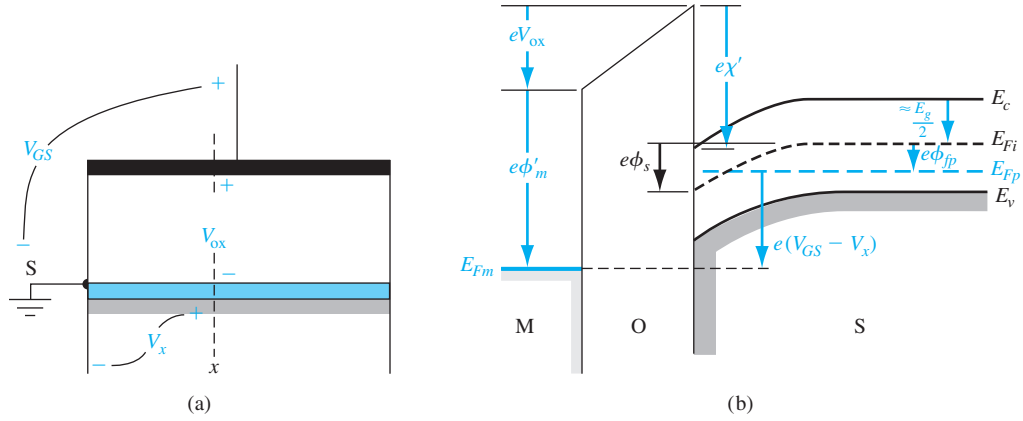
Figure 10.46 | (a) Potentials at a point x along the channel. (b) Energy-band diagram through the MOS structure at the point x.

The energy-band diagram through the MOS structure at point x is shown in Figure 10.46b. The Fermi level in the p-type semiconductor is E_{Fp} and the Fermi level in the metal is E_{Fm}. We have

$$E_{Fp} - E_{Fm} = e(V_{GS} - V_x) \tag{10.55}$$

Considering the potential barriers, we can write

$$V_{GS} - V_x = (\phi'_m + V_{ox}) - \left(\chi' + \frac{E_g}{2e} - \phi_s + \phi_{fp}\right) \tag{10.56}$$

which can also be written as

$$V_{GS} - V_x = V_{ox} + 2\phi_{fp} + \phi_{ms} \tag{10.57}$$

where ϕ_{ms} is the metal–semiconductor work function difference, and $\phi_s = 2\phi_{fp}$ for the inversion condition.

The electric field in the oxide is

$$E_{ox} = \frac{V_{ox}}{t_{ox}} \tag{10.58}$$

Combining Equations (10.54), (10.57), and (10.58), we find that

$$-\epsilon_{ox} E_{ox} = \frac{-\epsilon_{ox}}{t_{ox}}[(V_{GS} - V_x) - (\phi_{ms} + 2\phi_{fp})]$$

$$= Q'_{ss} + Q'_n + Q'_{SD}(\max) \tag{10.59}$$

The inversion charge density, Q'_n, from Equation (10.59) can be substituted into Equation (10.49) and we obtain

$$I_x = -W\mu_n C_{ox} \frac{dV_x}{dx}[(V_{GS} - V_x) - V_T] \tag{10.60}$$

where $E_x = -dV_x/dx$ and V_T is the threshold voltage defined by Equation (10.31b).

We can now integrate Equation (10.60) over the length of the channel. We have

$$\int_0^L I_x \, dx = -W\mu_n C_{ox} \int_{V_x(0)}^{V_x(L)} [(V_{GS} - V_T) - V_x] \, dV_x \tag{10.61}$$

We are assuming a constant mobility μ_n. For the n-channel device, the drain current enters the drain terminal and is a constant along the entire channel length. Letting $I_D = -I_x$, Equation (10.61) becomes

$$I_D = \frac{W\mu_n C_{ox}}{2L} [2(V_{GS} - V_T)V_{DS} - V_{DS}^2] \tag{10.62}$$

Equation (10.62) is valid for $V_{GS} \geq V_T$ and for $0 \leq V_{DS} \leq V_{DS}(\text{sat})$.

Equation (10.62) can also be written as

$$I_D = \frac{k_n'}{2} \cdot \frac{W}{L} \cdot [2(V_{GS} - V_T)V_{DS} - V_{DS}^2] = K_n[2(V_{GS} - V_T)V_{DS} - V_{DS}^2] \tag{10.63}$$

where k_n' is the process conduction parameter and K_n is the conduction parameter. These parameters are described and defined in Equations (10.44b) and (10.44c).

Figure 10.47 shows plots of Equation (10.62) as a function of V_{DS} for several values of V_{GS}. We can find the value of V_{DS} at the peak current value from $\partial I_D / \partial V_{DS} = 0$. Then, using Equation (10.62), the peak current occurs when

$$V_{DS} = V_{GS} - V_T \tag{10.64}$$

This value of V_{DS} is just $V_{DS}(\text{sat})$, the point at which saturation occurs. For $V_{DS} > V_{DS}(\text{sat})$, the ideal drain current is a constant and is equal to

$$I_D(\text{sat}) = \frac{W\mu_n C_{ox}}{2L} [2(V_{GS} - V_T)V_{DS}(\text{sat}) - V_{DS}^2(\text{sat})] \tag{10.65}$$

Using Equation (10.64) for $V_{DS}(\text{sat})$, Equation (10.65) becomes

$$I_D(\text{sat}) = \frac{W\mu_n C_{ox}}{2L} (V_{GS} - V_T)^2 \tag{10.66}$$

Equation (10.66) can also be written as

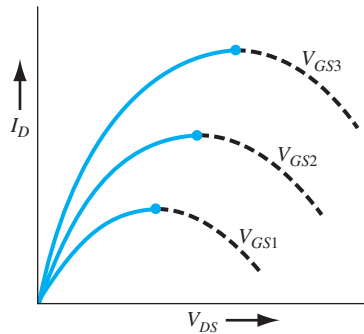$$I_D = \frac{k_n'}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_T)^2 = K_n (V_{GS} - V_T)^2 \tag{10.67}$$

Figure 10.47 | Plots of I_D versus V_{DS} from Equation (10.62).

Equation (10.62) is the ideal current–voltage relationship of the n-channel MOSFET in the nonsaturation region for $0 \leq V_{DS} \leq V_{DS}(\text{sat})$, and Equation (10.66) is the ideal current–voltage relationship of the n-channel MOSFET in the saturation region for $V_{DS} \geq V_{DS}(\text{sat})$. These *I–V* expressions were explicitly derived for an n-channel enhancement mode device. However, these same equations apply to an n-channel depletion mode MOSFET in which the threshold voltage V_T is a negative quantity.

Objective: Design the width of a MOSFET such that a specified current is induced for a given applied bias.

EXAMPLE 10.7

Consider an ideal n-channel MOSFET with parameters $L = 1.25$ μm, $\mu_n = 650$ cm²/V-s, $C_{ox} = 6.9 \times 10^{-8}$ F/cm², and $V_T = 0.65$ V. Design the channel width W such that $I_D(\text{sat}) = 4$ mA for $V_{GS} = 5$ V.

■ **Solution**

For the transition biased in the saturation region, we have, from Equation (10.66),

$$I_D(\text{sat}) = \frac{W \mu_n C_{ox}}{2L} (V_{GS} - V_T)^2$$

or

$$4 \times 10^{-3} = \frac{W(650)(6.9 \times 10^{-8})}{2(1.25 \times 10^{-4})} \cdot (5 - 0.65)^2 = 3.39 \ W$$

Then

$$W = 11.8 \ \mu\text{m}$$

■ **Comment**

The current capability of a MOSFET is directly proportional to the channel width W. The current handling capability can be increased by increasing W.

■ **EXERCISE PROBLEM**

Ex 10.7 The parameters of an n-channel silicon MOSFET are $\mu_n = 650$ cm²/V-s, $t_{ox} = 8$ nm = 80 Å, $W/L = 12$, and $V_T = 0.40$ V. If the transistor is biased in the saturation region, find the drain current for (*a*) $V_{GS} = 0.8$ V, (*b*) $V_{GS} = 1.2$ V, and (*c*) $V_{GS} = 1.6$ V.

[Ans. (*a*) 0.269 mA; (*b*) 1.077 mA; (*c*) 2.423 mA]

We can use the *I–V* relations to experimentally determine the mobility and threshold voltage parameters. From Equation (10.62), we can write, for very small values of V_{DS},

$$I_D = \frac{W \mu_n C_{ox}}{L} (V_{GS} - V_T) V_{DS} \quad\quad (10.68)$$

Figure 10.48a shows a plot of Equation (10.68) as a function of V_{GS} for constant V_{DS}. A straight line is fitted through the points. The deviation from the straight line at low values of V_{GS} is due to subthreshold conduction and the deviation at higher values of V_{GS}
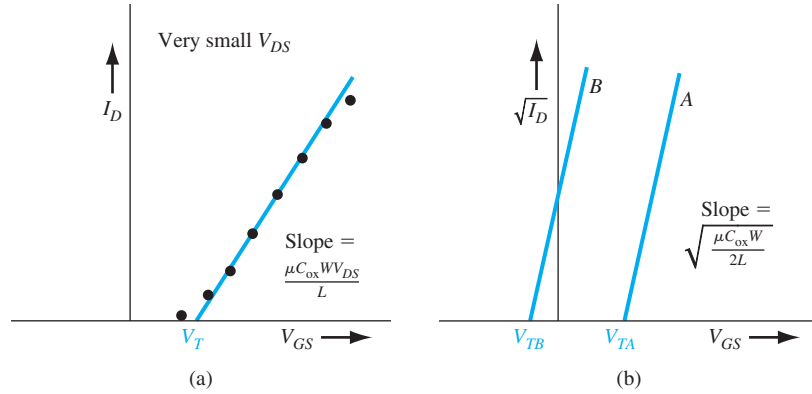
Figure 10.48 | (a) I_D versus V_{GS} (for small V_{DS}) for enhancement mode MOSFET. (b) Ideal $\sqrt{I_D}$ versus V_{GS} in saturation region for enhancement mode (curve A) and depletion mode (curve B) n-channel MOSFETs.

is due to mobility being a function of gate voltage. Both of these effects will be considered in the next chapter. The extrapolation of the straight line to zero current gives the threshold voltage, and the slope is proportional to the inversion carrier mobility.

   Now consider the case when the transistor is biased in the saturation region. If we take the square root of Equation (10.66), we obtain

$$\sqrt{I_D(\text{sat})} = \sqrt{\frac{W\mu_n C_{ox}}{2L}}(V_{GS} - V_T) \tag{10.69}$$

Figure 10.48b is a plot of Equation (10.69). In the ideal case, we can obtain the same information from both curves. However, as we will see in the next chapter, the threshold voltage may be a function of V_{DS} in short-channel devices. Since Equation (10.69) applies to devices biased in the saturation region, the V_T parameter in this equation may differ from the extrapolated value determined in Figure 10.48a. In general, the nonsaturation current–voltage characteristics will produce the more reliable data.

EXAMPLE 10.8

Objective: Determine the inversion carrier mobility from experimental results.

   Consider an n-channel MOSFET with $W = 15 \ \mu\text{m}$, $L = 2 \ \mu\text{m}$, and $C_{ox} = 6.9 \times 10^{-8} \ \text{F/cm}^2$. Assume that the drain current in the nonsaturation region for $V_{DS} = 0.10$ V is $I_D = 35 \ \mu\text{A}$ at $V_{GS} = 1.5$ V and $I_D = 75 \ \mu\text{A}$ at $V_{GS} = 2.5$ V.

■ **Solution**

From Equation (10.68), we can write

$$I_{D2} - I_{D1} = \frac{W\mu_n C_{ox}}{L}(V_{GS2} - V_{GS1})V_{DS}$$

so that

$$75 \times 10^{-6} - 35 \times 10^{-6} = \left(\frac{15}{2}\right)\mu_n(6.9 \times 10^{-8})(2.5 - 1.5)(0.10)$$

which yields

$$\mu_n = 773 \text{ cm}^2/\text{V-s}$$

We can then determine

$$V_T = 0.625 \text{ V}$$

■ Comment

The mobility of carriers in the inversion layer is less than that in the bulk semiconductor due to the surface scattering effect. We will discuss this effect in the next chapter.

■ EXERCISE PROBLEM

Ex 10.8 An n-channel silicon MOSFET has the following parameters: $W = 6 \ \mu\text{m}$, $L = 1.5 \ \mu\text{m}$, and $t_{ox} = 8 \text{ nm} = 80 \text{ Å}$. When the transistor is biased in the saturation region, the drain current is $I_D(sat) = 0.132$ mA at $V_{GS} = 1.0$ V and $I_D(sat) = 0.295$ mA at $V_{GS} = 1.25$ V. Determine the electron mobility and the threshold voltage.

(Ans. $\mu_n \approx 600 \text{ cm}^2/\text{V-s}$, $V_T = 0.495 \text{ V}$)

The current–voltage relationship of a p-channel device can be obtained by the same type of analysis. Figure 10.49 shows a p-channel enhancement mode MOSFET. The voltage polarities and current direction are the reverse of those in the n-channel device. We may note the change in the subscript notation for this device. For the current direction shown in the figure, the I–V relation for the p-channel MOSFET biased in the nonsaturation region is

$$I_D = \frac{W \mu_p C_{ox}}{2L} \left[2(V_{SG} + V_T)V_{SD} - V_{SD}^2 \right] \tag{10.70}$$

Equation (10.70) is valid for $0 \le V_{SD} \le V_{SD}(sat)$.
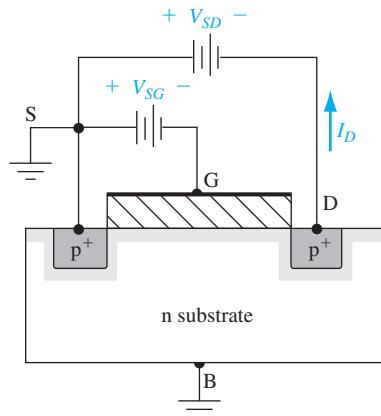

Figure 10.49 | Cross section and bias configuration for a p-channel enhancement mode MOSFET.

Equation (10.70) can also be written as

$$I_D = \frac{k_p'}{2} \cdot \frac{W}{L} \cdot \left[2(V_{SG} + V_T)V_{SD} - V_{SD}^2\right] = K_p\left[2(V_{SG} + V_T)V_{SD} - V_{SD}^2\right] \quad (10.71)$$

where $k_p' = \mu_p C_{ox}$ is the *process conduction parameter* for the p-channel MOSFET and $K_p = (W\mu_p C_{ox})/(2L) = (k_p'/2) \cdot (W/L)$ is the *conduction parameter.*

When the transistor is biased in the saturation region, the *I–V* relation is given by

$$I_D(\text{sat}) = \frac{W\mu_p C_{ox}}{2L}(V_{SG} + V_T)^2 \quad (10.72)$$

Equation (10.72) is valid for $V_{SD} \geq V_{SD}\,(sat)$.

Equation (10.72) can also be written as

$$I_D = \frac{k_p'}{2} \cdot \frac{W}{L} \cdot (V_{SG} + V_T)^2 = K_p(V_{SG} + V_T)^2 \quad (10.73)$$

The source-to-drain saturation voltage is given by

$$V_{SD}(sat) = V_{SG} + V_T \quad (10.74)$$

Note the change in the sign in front of V_T and note that the mobility is now the mobility of the holes in the hole inversion layer charge. Keep in mind that V_T is negative for a p-channel enhancement mode MOSFET and positive for a depletion mode p-channel device.

One assumption we made in the derivation of the currernt–voltage relationship was that the charge neutrality condition given by Equation (10.50) was valid over the entire length of the channel. We implicitly assumed that $Q_{SD}'(\text{max})$ was constant along the length of the channel. The space charge width, however, varies between source and drain due to the drain-to-source voltage; it is widest at the drain when $V_{DS} > 0$. A change in the space charge density along the channel length must be balanced by a corresponding change in the inversion layer charge. An increase in the space charge width means that the inversion layer charge is reduced, implying that the drain current and drain-to-source saturation voltage are less than the ideal values. The actual saturation drain current may be as much as 20 percent less than the predicted value due to this bulk charge effect.

10.3.4 Transconductance

The MOSFET transconductance is defined as the change in drain current with respect to the corresponding change in gate voltage, or

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \quad (10.75)$$

The transconductance is sometimes referred to as the transistor gain.

If we consider an n-channel MOSFET operating in the nonsaturation region, then using Equation (10.62), we have

$$g_{mL} = \frac{\partial I_D}{\partial V_{GS}} = \frac{W \mu_n C_{ox}}{L} \cdot V_{DS} \tag{10.76}$$

The transconductance increases linearly with V_{DS} but is independent of V_{GS} in the nonsaturation region.

The I–V characteristics of an n-channel MOSFET in the saturation region are given by Equation (10.66). The transconductance in this region of operation is given by

$$g_{ms} = \frac{\partial I_D(\text{sat})}{\partial V_{GS}} = \frac{W \mu_n C_{ox}}{L}(V_{GS} - V_T) \tag{10.77}$$

In the saturation region, the transconductance is a linear function of V_{GS} and is independent of V_{DS}.

The transconductance is a function of the geometry of the device as well as of carrier mobility and threshold voltage. The transconductance increases as the width of the device increases, and it also increases as the channel length and oxide thickness decrease. In the design of MOSFET circuits, the size of the transistor, in particular the channel width W, is an important engineering design parameter.

10.3.5 Substrate Bias Effects

In all of our analyses so far, the substrate, or body, has been connected to the source and held at ground potential. In MOSFET circuits, the source and body may not be at the same potential. Figure 10.50a shows an n-channel MOSFET and the associated double-subscripted voltage variables. The source-to-substrate pn junction must always be zero or reverse biased, so V_{SB} must always be greater than or equal to zero.
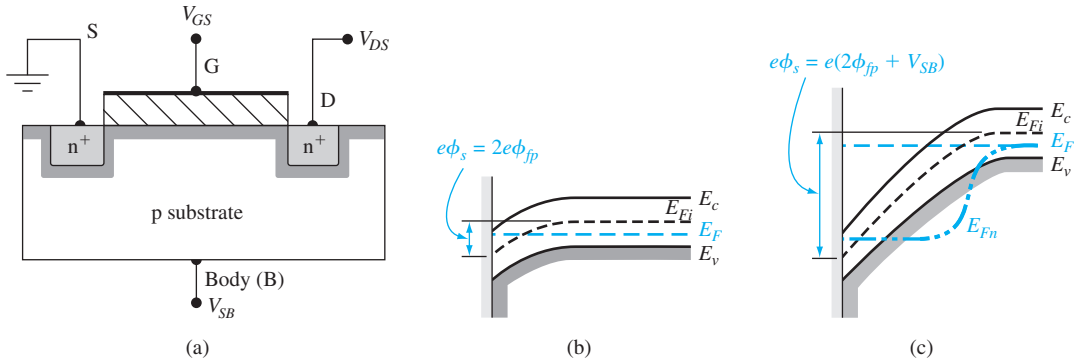

Figure 10.50 | (a) Applied voltages on an n-channel MOSFET. (b) Energy-band diagram at inversion point when $V_{SB} = 0$. (c) Energy-band diagram at inversion point when $V_{SB} > 0$ is applied.

If $V_{SB} = 0$, threshold is defined as the condition when $\phi_s = 2\phi_{fp}$ as we discussed previously and as shown in Figure 10.50b. When $V_{SB} > 0$ the surface will still try to invert when $\phi_s = 2\phi_{fp}$. However, these electrons are at a higher potential energy than are the electrons in the source. The newly created electrons will move laterally and flow out of the source terminal. When $\phi_s = 2\phi_{fp} + V_{SB}$, the surface reaches an equilibrium inversion condition. The energy-band diagram for this condition is shown in Figure 10.50c. The curve represented as E_{Fn} is the Fermi level from the p substrate through the reverse-biased source–substrate junction to the source contact.

The space charge region width under the oxide increases from the original x_{dT} value when a reverse-biased source–substrate junction voltage is applied. With an applied $V_{SB} > 0$, there is more charge associated with this region. Considering the charge neutrality condition through the MOS structure, the positive charge on the top metal gate must increase to compensate for the increased negative space charge in order to reach the threshold inversion point. So when $V_{SB} > 0$, the threshold voltage of the n-channel MOSFET increases.

When $V_{SB} = 0$, we had

$$Q'_{SD}\,(\text{max}) = -eN_a x_{dT} = -\sqrt{2e\epsilon_s\,N_a(2\phi_{fp})} \qquad (10.78)$$

When $V_{SB} > 0$, the space charge width increases and we now have

$$Q'_{SD} = -eN_a x_d = -\sqrt{2e\epsilon_s N_a(2\phi_{fp} + V_{SB})} \qquad (10.79)$$

The change in the space charge density is then

$$\Delta Q'_{SD} = -\sqrt{2e\epsilon_s\,N_a}\left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}\right] \qquad (10.80)$$

To reach the threshold condition, the applied gate voltage must be increased. The change in threshold voltage can be written as

$$\Delta V_T = -\frac{\Delta Q'_{SD}}{C_{ox}} = \frac{\sqrt{2e\epsilon_s N_a}}{C_{ox}}\left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}\right] \qquad (10.81)$$

where $\Delta V_T = V_T(V_{SB} > 0) - V_T(V_{SB} = 0)$. We may note that V_{SB} must always be positive so that, for the n-channel device, ΔV_T is always positive. The threshold voltage of the n-channel MOSFET will increase as a function of the source–substrate junction voltage.

From Equation (10.81), we may define

$$\gamma = \frac{\sqrt{2e\epsilon_s N_a}}{C_{ox}} \qquad (10.82)$$

where γ is defined as the *body-effect coefficient*. Equation (10.81) may then be written as

$$\Delta V_T = \gamma\left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}\right] \qquad (10.83)$$

EXAMPLE 10.9

Objective: Calculate the body-effect coefficient and the change in the threshold voltage due to an applied source-to-body voltage.

Consider an n-channel silicon MOSFET at $T = 300$ K. Assume the substrate is doped to $N_a = 3 \times 10^{16}$ cm^{-3} and assume the oxide is silicon dioxide with a thickness of $t_{ox} = 20$ nm $= 200$ Å. Let $V_{SB} = 1$ V.

■ **Solution**

We can calculate that

$$\phi_{fp} = V_t \ln \left(\frac{N_a}{n_i} \right) = (0.0259) \ln \left(\frac{3 \times 10^{16}}{1.5 \times 10^{10}} \right) = 0.3758 \text{ V}$$

and

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{(3.9)(8.85 \times 10^{-14})}{200 \times 10^{-8}} = 1.726 \times 10^{-7} \text{ F/cm}^2$$

From Equation (10.82), we find the body-effect coefficient to be

$$\gamma = \frac{\sqrt{2e\epsilon_s N_a}}{C_{ox}} = \frac{[2(1.6 \times 10^{-19})(11.7)(8.85 \times 10^{-14})(3 \times 10^{16})]^{1/2}}{1.726 \times 10^{-7}}$$

or

$$\gamma = 0.5776 \text{ V}^{1/2}$$

The change in threshold voltage for $V_{SB} = 1$ V is found to be

$$\Delta V_T = \gamma \left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}} \right]$$
$$= (0.5776) \left[\sqrt{2(0.3758) + 1} - \sqrt{2(0.3758)} \right]$$
$$= (0.5776)[1.3235 - 0.8669] = 0.264 \text{ V}$$

■ **Comment**

Figure 10.51 shows plots of $\sqrt{I_D(sat)}$ versus V_{GS} for various applied values of V_{SB}. The original threshold voltage is assumed to be $V_{TO} = 0.64$ V.

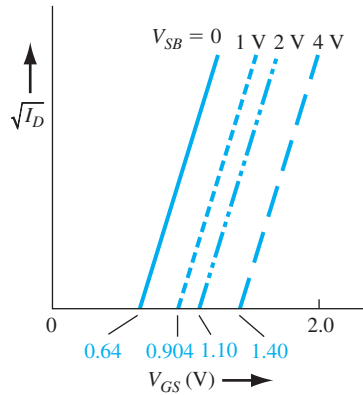**Figure 10.51** | Plots of $\sqrt{I_D}$ versus V_{GS} at several values of V_{SB} for an n-channel MOSFET.

■ **EXERCISE PROBLEM**

Ex 10.9 A silicon MOSFET has the following parameters: $N_a = 10^{16}$ cm^{-3} and $t_{ox} = 12$ nm $= 120$ Å. Calculate (*a*) the body-effect coefficient and (*b*) the change in threshold voltage for (*i*) $V_{SB} = 1$ and (*ii*) $V_{SB} = 2$ V.

[Ans. (*a*) $\gamma = 0.200$ V$^{1/2}$; (*b*) (*i*) $\Delta V_T = 0.0937$ V, (*ii*) $\Delta V_T = 0.162$ V]

If a body or substrate bias is applied to a p-channel device, the threshold voltage is shifted to more negative values. Because the threshold voltage of a p-channel enhancement mode MOSFET is negative, a body voltage will increase the applied negative gate voltage required to create inversion. The same general observation was made for the n-channel MOSFET.

TEST YOUR UNDERSTANDING

TYU 10.6 The silicon n-channel MOSFET described in Exercise Problem Ex 10.7 is to be redesigned by changing the W/L ratio such that $I_D = 100$ μA when the transistor is biased in the saturation regin with $V_{GS} = 1.0$ V.

(Ans. $W/L = 1.98$)

TYU 10.7 The parameters of a p-channel MOSFET are $\mu_p = 310$ cm^2/V-s, $t_{ox} = 220$ Å, $W/L = 60$, and $V_T = -0.40$ V. If the transistor is biased in the saturation region, find the drain current for $V_{SG} = 1$, 1.5, and 2 V.

(Ans. $I_D = 0.526$, 1.77 and 3.74 mA)

TYU 10.8 The p-channel MOSFET in TYU 10.7 is to be redesigned by changing the (W/L) ratio such that $I_D = 200$ μA when the transistor is biased in the saturation region with $V_{SG} = 1.25$ V.

(Ans. $W/L = 11.4$)

TYU 10.9 Repeat Exercise Problem Ex 10.9 for a substrate impurity doping concentration of $N_a = 10^{15}$ cm^{-3}.

[Ans. (*a*) $\gamma = 0.0633$ V$^{1/2}$; (*b*) (*i*) $\Delta V_T = 0.0314$ V, (*ii*) $\Delta V_T = 0.0536$ V]

10.4 | FREQUENCY LIMITATIONS

In many applications, the MOSFET is used in a linear amplifier circuit. A small-signal equivalent circuit for the MOSFET is needed in order to mathematically analyze the electronic circuit. The equivalent circuit contains capacitances and resistances that introduce frequency effects. We initially develop a small-signal equivalent circuit and then discuss the physical factors that limit the frequency response of the MOSFET. A transistor cutoff frequency, which is a figure of merit, is then defined and an expression derived for this factor.

10.4.1 Small-Signal Equivalent Circuit

The small-signal equivalent circuit of the MOSFET is constructed from the basic MOSFET geometry. A model based on the inherent capacitances and resistances
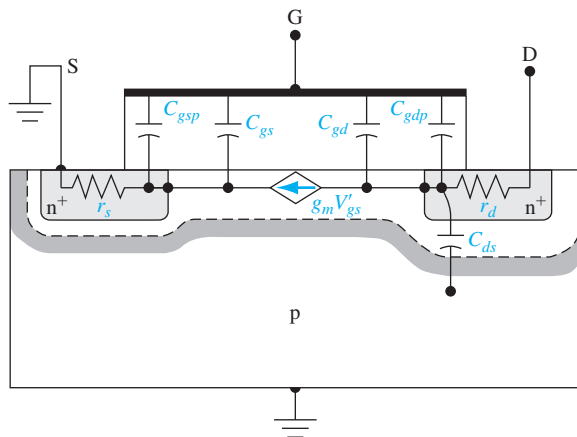
Figure 10.52 | Inherent resistances and capacitances in the n-channel MOSFET structure.

within the transistor structure, along with elements that represent the basic device equations, is shown in Figure 10.52. One simplifying assumption we will make in the equivalent circuit is that the source and substrate are both tied to ground potential.

Two of the capacitances connected to the gate are inherent in the device. These capacitances are C_{gs} and C_{gd}, which represent the interaction between the gate and the channel charge near the source and drain terminals, respectively. The remaining two gate capacitances, C_{gsp} and C_{gdp}, are parasitic or overlap capacitances. In real devices, the gate oxide will overlap the source and drain contacts because of tolerance or fabrication factors. As we will see, the drain overlap capacitance—C_{gdp}, in particular—will lower the frequency response of the device. The parameter C_{ds} is the drain-to-substrate pn junction capacitance, and r_s and r_d are the series resistances associated with the source and drain terminals. The small-signal channel current is controlled by the internal gate-to-source voltage through the transconductance.

The small-signal equivalent circuit for the n-channel common-source MOSFET is shown in Figure 10.53. The voltage V'_{gs} is the internal gate-to-source voltage that
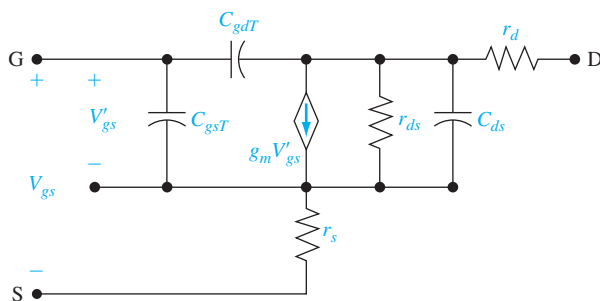

Figure 10.53 | Small-signal equivalent circuit of a common-source n-channel MOSFET.

controls the channel current. The parameters C_{gsT} and C_{gdT} are the total gate-to-source and total gate-to-drain capacitances. One parameter, r_{ds}, shown in Figure 10.53, is not shown in Figure 10.52. This resistance is associated with the slope I_D versus V_{DS}. In the ideal MOSFET biased in the saturation region, I_D is independent of V_{DS} so that r_{ds} would be infinite. In short-channel-length devices, in particular, r_{ds} is finite because of channel length modulation, which we will consider in the next chapter.

A simplified small-signal equivalent circuit valid at low frequency is shown in Figure 10.54. The series resistances, r_s and r_d, have been neglected, so the drain current is essentially only a function of the gate-to-source voltage through the transconductance. The input gate impedance is infinite in this simplified model.

The source resistance r_s can have a significant effect on the transistor characteristics. Figure 10.55 shows a simplified, low-frequency equivalent circuit including r_s but neglecting r_{ds}. The drain current is given by

$$I_d = g_m V'_{gs} \tag{10.84}$$

and the relation between V_{gs} and V'_{gs} can be found from

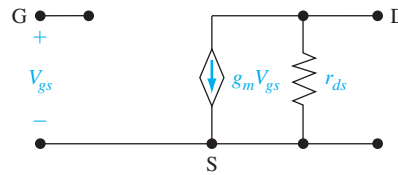$$V_{gs} = V'_{gs} + (g_m V'_{gs})r_s = (1 + g_m r_s)V'_{gs} \tag{10.85}$$

Figure 10.54 | Simplified, low-frequency small-signal equivalent circuit of a common-source n-channel MOSFET.
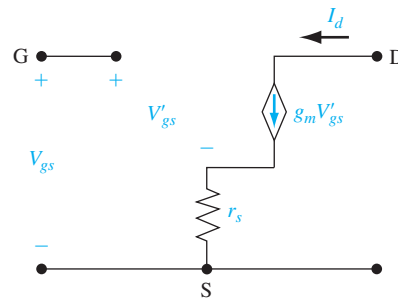
Figure 10.55 | Simplified, low-frequency small-signal equivalent circuit of common-source n-channel MOSFET including source resistance r_s.

The drain current from Equation (10.84) can now be written as

$$I_d = \left(\frac{g_m}{1 + g_m r_s}\right) V_{gs} = g'_m V_{gs} \tag{10.86}$$

The source resistance reduces the effective transconductance or transistor gain.

The equivalent circuit of the p-channel MOSFET is exactly the same as that of the n-channel except that all voltage polarities and current directions are reversed. The same capacitances and resistances that are in the n-channel model apply to the p-channel model.

10.4.2 Frequency Limitation Factors and Cutoff Frequency

There are two basic frequency limitation factors in the MOSFET. The first factor is the channel transit time. If we assume that carriers are traveling at their saturation drift velocity v_{sat}, then the transit time is $\tau_t = L/v_{sat}$ where L is the channel length. If $v_{sat} = 10^7$ cm/s and $L = 1$ μm, then $\tau_t = 10$ ps, which translates into a maximum frequency of 100 GHz. This frequency is much larger than the typical maximum frequency response of a MOSFET. The transit time of carriers through the channel is usually not the limiting factor in the frequency responses of MOSFETs.

The second limiting factor is the gate or capacitance charging time. If we neglect r_s, r_d, r_{ds}, and C_{ds}, the resulting equivalent small-signal circuit is shown in Figure 10.56 where R_L is a load resistance.

The input gate impedance in this equivalent circuit is no longer infinite. Summing currents at the input gate node, we have

$$I_i = j\omega C_{gsT} V_{gs} + j\omega C_{gdT}(V_{gs} - V_d) \tag{10.87}$$

where I_i is the input current. Likewise, summing currents at the output drain node, we have

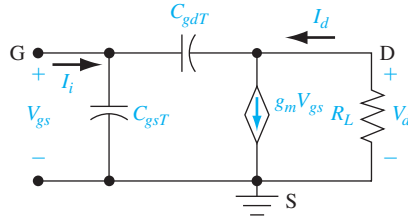$$\frac{V_d}{R_L} + g_m V_{gs} + j\omega C_{gdT}(V_d - V_{gs}) = 0 \tag{10.88}$$

Figure 10.56 | High-frequency small-signal equivalent circuit of common-source n-channel MOSFET.
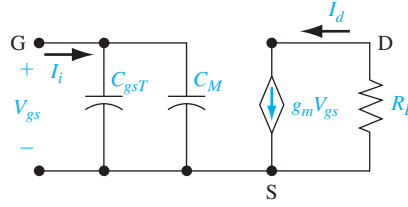
Figure 10.57 | Small-signal equivalent circuit including Miller capacitance.

Combining Equations (10.87) and (10.88) to eliminate the voltage variable V_d, we can determine the input current as

$$I_i = j\omega \left[C_{gsT} + C_{gdT} \left(\frac{1 + g_m R_L}{1 + j\omega R_L C_{gdT}} \right) \right] V_{gs} \tag{10.89}$$

Normally, $\omega R_L C_{gdT}$ is much less than unity; therefore, we may neglect the $(j\omega R_L C_{gdT})$ term in the denominator. Equation (10.89) then simplifies to

$$I_i = j\omega [C_{gsT} + C_{gdT}(1 + g_m R_L)] V_{gs} \tag{10.90}$$

Figure 10.57 shows the equivalent circuit with the equivalent input impedance described by Equation (10.90). The parameter C_M is the Miller capacitance and is given by

$$C_M = C_{gdT}(1 + g_m R_L) \tag{10.91}$$

The serious effect of the drain overlap capacitance now becomes apparent. When the transistor is operating in the saturation region, C_{gd} essentially becomes zero, but C_{gdp} is a constant. This parasitic capacitance is multiplied by the gain of the transistor and can become a significant factor in the input impedance.

The cutoff frequency f_T is defined to be the frequency at which the magnitude of the current gain of the device is unity, or when the magnitude of the input current I_i is equal to the ideal load current I_d. From Figure 10.57, we can see that

$$I_i = j\omega (C_{gsT} + C_M) V_{gs} \tag{10.92}$$

and the ideal load current is

$$I_d = g_m V_{gs} \tag{10.93}$$

The magnitude of the current gain is then

$$\left| \frac{I_d}{I_i} \right| = \frac{g_m}{2\pi f(C_{gsT} + C_M)} \tag{10.94}$$

Setting the magnitude of the current gain equal to unity at the cutoff frequency, we find

$$f_T = \frac{g_m}{2\pi (C_{gsT} + C_M)} = \frac{g_m}{2\pi C_G} \tag{10.95}$$

where C_G is the equivalent input gate capacitance.

In the ideal MOSFET, the overlap or parasitic capacitances, C_{gsp} and C_{gdp}, are zero. Also, when the transistor is biased in the saturation region, C_{gd} approaches zero and C_{gs} is approximately $C_{ox}WL$. The transconductance of the ideal MOSFET biased in the saturation region and assuming a constant mobility is given by Equation (10.77) as

$$g_{ms} = \frac{W\mu_n C_{ox}}{L}(V_{GS} - V_T)$$

Then, for this ideal case, the cutoff frequency is

$$f_T = \frac{g_m}{2\pi C_G} = \frac{\frac{W\mu_n C_{ox}}{L}(V_{GS} - V_T)}{2\pi(C_{ox}WL)} = \frac{\mu_n(V_{GS} - V_T)}{2\pi L^2} \qquad (10.96)$$

Objective: Calculate the cutoff frequency of an ideal MOSFET with a constant mobility.   **EXAMPLE 10.10**

Assume that the electron mobility in an n-channel device is $\mu_n = 400$ cm²/V-s and that the channel length is $L = 4$ μm. Also assume that $V_T = 1$ V and $V_{GS} = 3$ V.

■ **Solution**

From Equation (10.96), the cutoff frequency is

$$f_T = \frac{\mu_n(V_{GS} - V_T)}{2\pi L^2} = \frac{400(3-1)}{2\pi (4 \times 10^{-4})^2} = 796 \text{ MHz}$$

■ **Comment**

In an actual MOSFET, the effect of the parasitic capacitance will substantially reduce the cutoff frequency from that calculated in this example.

■ **EXERCISE PROBLEM**

Ex 10.10   An n-channel silicon MOSFET has the following parameters: $\mu_n = 420$ cm²/V-s, $t_{ox} = 18$ nm $= 180$ Å, $L = 1.2$ μm, $W = 24$ μm, and $V_T = 0.4$ V. The transistor is biased in the saturation region at $V_{GS} = 1.5$ V. Determine the cutoff frequency.
(Ans. $f_T = 5.11$ GHz)

TEST YOUR UNDERSTANDING

TYU 10.10   Consider the n-channel MOSFET described in Exercise Problem Ex 10.10. The transistor is connected to an effective load resistance of $R_L = 100$ kΩ. Calculate the ratio of Miller capacitance C_M to gate-to-drain capacitance C_{gdT}.
(Ans. 178)

*10.5 | THE CMOS TECHNOLOGY

The primary objective of this book is to present the basic physics of semiconductor materials and devices without considering in detail the various fabrication processes; this important subject is left to other books. However, there is one MOS technology

that is used extensively, for which the basic fabrication techniques must be considered in order to understand essential characteristics of these devices and circuits. The MOS technology we consider briefly is the complementary MOS, or CMOS, process.

We have considered the physics of both n-channel and p-channel enhancement mode MOSFETs. Both devices are used in a CMOS inverter, which is the basis of CMOS digital logic circuits. The dc power dissipation in a digital circuit can be reduced to very low levels by using a complementary p-channel and n-channel pair.

It is necessary to form electrically isolated p- and n-substrate regions in an integrated circuit to accommodate the n- and p-channel transistors. The p-well process has been a commonly used technique for CMOS circuits. The process starts with a fairly low doped n-type silicon substrate in which the p-channel MOSFET will be fabricated. A diffused p region, called a p well, is formed in which the n-channel MOSFET will be fabricated. In most cases, the p-type substrate doping level must be larger than the n-type substrate doping level to obtain the desired threshold voltages. The larger p doping can easily compensate the initial n doping to form the p well. A simplified cross section of the p-well CMOS structure is shown in Figure 10.58a. The notation FOX stands for field oxide, which is a relatively thick oxide separating the devices. The FOX prevents either the n or p substrate from becoming inverted and helps maintain isolation between the two devices. In practice, additional processing steps must be included; for example, providing connections so that the p well and n substrate can be electrically connected to the appropriate voltages. The n substrate
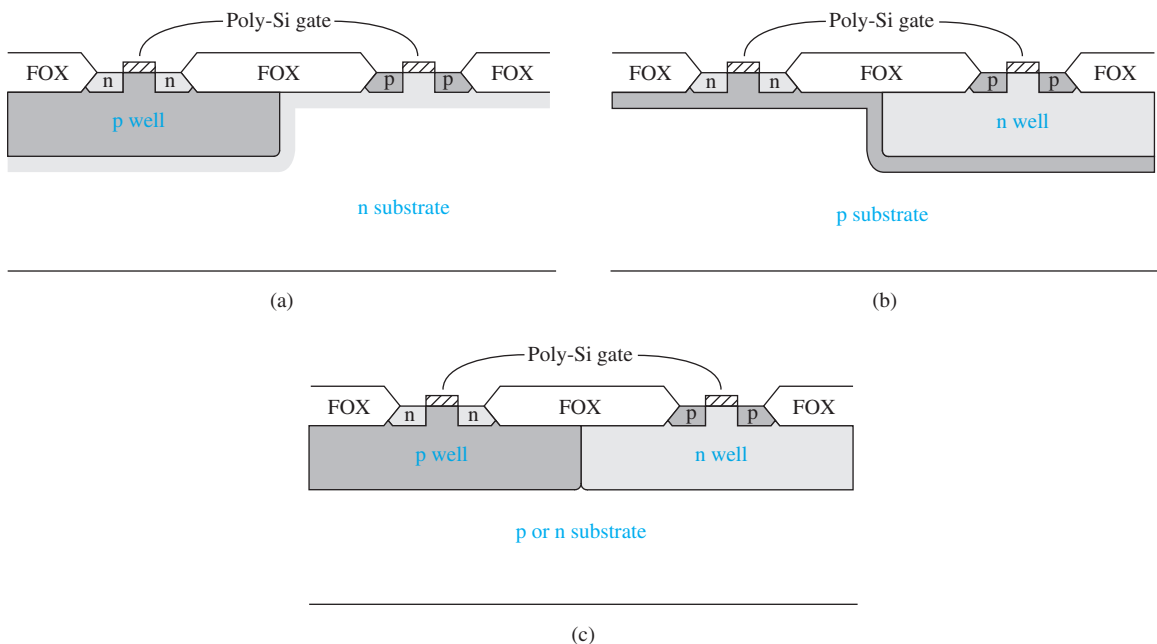

Figure 10.58 | CMOS structures: (a) p well, (b) n well, and (c) twin well.
(From Yang [22].)

must always be at a higher potential than the p well; therefore, this pn junction will always be reverse biased.

With ion implantation now being extensively used for threshold voltage control, both the n-well CMOS process and twin-well CMOS process can be used. The n-well CMOS process, shown in Figure 10.58b, starts with an optimized p-type substrate that is used to form the n-channel MOSFETs. (The n-channel MOSFETs, in general, have superior characteristics, so this starting point should yield excellent n-channel devices.) The n well is then added, in which the p-channel devices are fabricated. The n-well doping can be controlled by ion implantation.

The twin-well CMOS process, shown in Figure 10.58c, allows both the p-well and n-well regions to be optimally doped to control the threshold voltage and trans-conductance of each transistor. The twin-well process allows a higher packing density because of self-aligned channel stops.

One major problem in CMOS circuits has been latch-up. *Latch-up* refers to a high-current, low-voltage condition that may occur in a four-layer pnpn structure. Figure 10.59a shows the circuit of a CMOS inverter and Figure 10.59b shows a simplified integrated circuit layout of the inverter circuit. In the CMOS layout, p^+ source to n substrate to p well to n^+ source forms such a four-layer structure.

The equivalent circuit of this four-layer structure is shown in Figure 10.60. The silicon-controlled rectifier action involves the interaction of the parasitic pnp and npn bipolar transistors. Bipolar transistors are discussed in Chapter 12. The npn transistor corresponds to the vertical n^+-source to p-well to n-substrate structure and the pnp transistor corresponds to the lateral p-well to n-substrate to p^+-source structure. Under normal CMOS operation, both parasitic bipolar transistors are cut off. However, under certain conditions, avalanche breakdown may occur in the p-well to n-substrate junction, driving both bipolar transistors into saturation. This high-current, low-voltage condition—latch-up—can sustain itself by positive feedback. The condition can prevent the CMOS circuit from operating and can also cause permanent damage and burnout of the circuit.

Latch-up can be prevented if the product $\beta_n \beta_p$ is less than unity at all times, where β_n and β_p are the common-emitter current gains of the npn and pnp parasitic
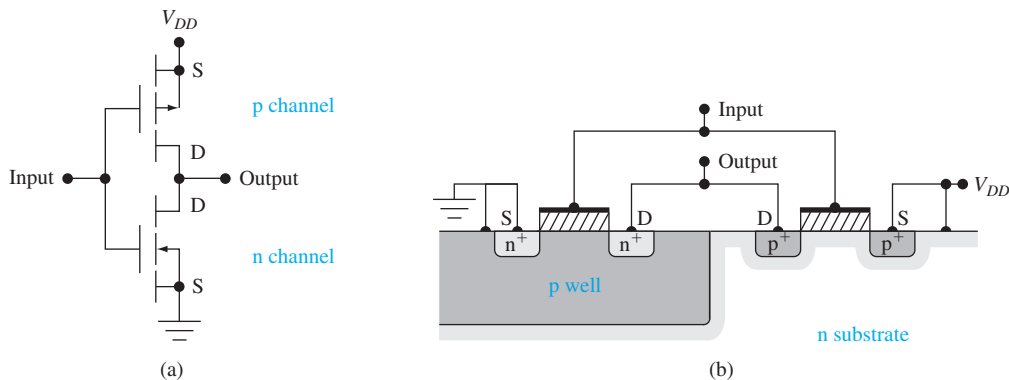

Figure 10.59 | (a) CMOS inverter circuit. (b) Simplified integrated circuit cross section of CMOS inverter.
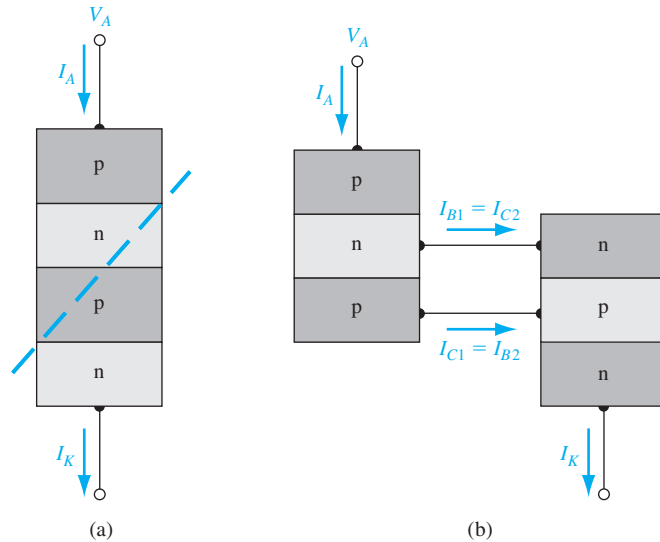
Figure 10.60 | (a) The splitting of the basic pnpn structure. (b) The two-transistor equivalent circuit of the four-layered pnpn device.

bipolar transistors, respectively. One method of preventing latch-up is to "kill" the minority carrier lifetime. Minority carrier lifetime degradation can be accomplished by gold doping or neutron irradiation, either of which introduces deep traps within the semiconductor. The deep traps increase the excess minority carrier recombination rate and reduce current gain. A second method of preventing latch-up is by using proper circuit layout techniques. If the two bipolar transistors can be effectively decoupled, then latch-up can be minimized or prevented. The two parasitic bipolar transistors can also be decoupled by using a different fabrication technology. The silicon-on-insulator technology, for example, allows the n-channel and the p-channel MOSFETs to be isolated from each other by an insulator. This isolation decouples the parasitic bipolar transistors.

10.6 | SUMMARY

- The fundamental physics and characteristics of the metal-oxide–semiconductor field-effect transistor (MOSFET) have been considered in this chapter.
- The heart of the MOSFET is the MOS capacitor. The energy bands in the semiconductor adjacent to the oxide–semiconductor interface bend, depending upon the voltage applied to the gate.
- An inversion layer of electrons can be created at the oxide–semiconductor surface in a p-type semiconductor by applying a sufficiently positive gate voltage, and an inversion layer of holes can be created at the oxide–semiconductor surface in an n-type semiconductor by applying a sufficiently negative gate voltage.
- The threshold voltage is the applied gate voltage required to reach the threshold inversion point. The flat-band voltage was defined and discussed.

■ The n-channel MOSFET, both enhancement mode and depletion mode, and the p-channel MOSFET, both enhancement mode and depletion mode, were described.

■ The basic transistor action is the modulation of the current at the drain terminal by the gate-to-source voltage.

■ The ideal MOSFET current–voltage relations were derived.

■ The body-effect coefficient was defined and discussed. The expression for the shift in threshold voltage due to the body effect was derived.

■ A small-signal equivalent circuit of the MOSFET was developed.

■ Various physical factors in the MOSFET that affect the frequency limitations were discussed. An expression for the cutoff frequency was developed.

■ The CMOS technology was briefly considered.

GLOSSARY OF IMPORTANT TERMS

accumulation layer charge   The induced charge directly under an oxide that is in excess of the thermal-equilibrium majority carrier concentration.

channel conductance   The ratio of drain current to drain-to-source voltage in the limit as $V_{DS} \rightarrow 0$.

channel conductance modulation   The process whereby the channel conductance varies with gate-to-source voltage.

CMOS   Complementary MOS; the technology that uses both p- and n-channel devices in an electronic circuit fabricated in a single semiconductor chip.

conduction parameter   The multiplying coefficient of the voltage terms to obtain the MOSFET drain current.

cutoff frequency   The signal frequency at which the input ac gate current is equal to the output ac drain current.

depletion mode MOSFET   The type of MOSFET in which a gate voltage must be applied to turn the device off.

enhancement mode MOSFET   The type of MOSFET in which a gate voltage must be applied to turn the device on.

equivalent fixed oxide charge   The effective fixed charge in the oxide, Q'_{ss}, directly adjacent to the oxide–semiconductor interface.

field-effect   The phenomenon by which an electric field perpendicular to the surface of a semiconductor can modulate the conductance.

flat-band voltage   The gate voltage that must be applied to create the flat-band condition in which there is no space charge region in the semiconductor under the oxide.

interface states   The allowed electronic energy states within the bandgap energy at the oxide–semiconductor interface.

inversion layer charge   The induced charge directly under the oxide, which is the opposite type compared with the semiconductor doping.

inversion layer mobility   The mobility of carriers in the inversion layer.

metal–semiconductor work function difference   The parameter ϕ_{ms}, a function of the difference between the metal work function and semiconductor electron affinity.

oxide capacitance   The ratio of oxide permittivity to oxide thickness, which is the capacitance per unit area, C_{ox}.

process conduction parameter   The product of carrier mobility and oxide capacitance.

saturation   The condition in which the inversion charge density is zero at the drain and the drain current is no longer a function of the drain-to-source voltage.

strong inversion   The condition in which the inversion charge density is larger than the magnitude of the semiconductor doping concentration.

threshold inversion point   The condition in which the inversion charge density is equal in magnitude to the semiconductor doping concentration.

threshold voltage   The gate voltage that must be applied to achieve the threshold inversion point.

transconductance   The ratio of an incremental change in drain current to the corresponding incremental change in gate voltage.

weak inversion   The condition in which the inversion charge density is less than the magnitude of the semiconductor doping concentration.

CHECKPOINT

After studying this chapter, the reader should have the ability to:

■   Sketch the energy-band diagrams in the semiconductor of the MOS capacitor under various bias conditions.

■   Describe the process by which an inversion layer of charge is created in a MOS capacitor.

■   Discuss the reason the space charge width reaches a maximum value once the inversion layer is formed.

■   Discuss what is meant by the metal–semiconductor work function difference and why this value is different between aluminum, n^+ polysilicon, and p^+ polysilicon gates.

■   Describe what is meant by flat-band voltage.

■   Define threshold voltage.

■   Sketch the C–V characteristics of a MOS capacitor with p-type and n-type semiconductor substrates under high-frequency and low-frequency conditions.

■   Discuss the effects of fixed trapped oxide charge and interface states on the C–V characteristics.

■   Sketch the cross sections of n-channel and p-channel MOSFET structures.

■   Explain the basic operation of the MOSFET.

■   Discuss the I–V characteristics of the MOSFET when biased in the nonsaturation and saturation regions.

■   Describe the substrate bias effects on the threshold voltage.

■   Sketch the small-signal equivalent circuit, including capacitances, of the MOSFET, and explain the physical origin of each capacitance.

■   Discuss the condition that defines the cutoff frequency of a MOSFET.

■   Sketch the cross section of a CMOS structure.

■   Discuss what is meant by latch-up in a CMOS structure.

REVIEW QUESTIONS

1.   Sketch the energy-band diagrams in a MOS capacitor with an n-type substrate in accumulation, depletion, and inversion modes.

2.   Describe what is meant by an inversion layer of charge. Describe how an inversion layer of charge can be formed in a MOS capacitor with a p-type substrate.

3. Why does the space charge region in the semiconductor of a MOS capacitor reach a maximum width once the inversion layer is formed?

4. Define surface potential. Does the surface potential change significantly with gate voltage once threshold is reached?

5. Sketch the energy-band diagram through a MOS structure with a p-type substrate and an n^+ polysilicon gate under zero bias.

6. Define the flat-band voltage. Sketch the energy-band diagram in a MOS capacitor at flat band.

7. Define the threshold voltage. What is the surface potential at the threshold voltage?

8. Sketch the C–V characteristics of a MOS capacitor with an n-type substrate under the low-frequency condition. How do the characteristics change for the high-frequency condition?

9. Indicate the approximate capacitance at flat band on the C–V characteristic of a MOS capacitor with a p-type substrate under the high-frequency condition.

10. What is the effect on the C–V characteristics of a MOS capacitor with a p-type substrate if the amount of positive trapped oxide charge increases?

11. Qualitatively sketch the inversion charge density in the channel region when the transistor is biased in the nonsaturation region. Repeat for the case when the transistor is biased in the saturation region.

12. Define $V_{DS}(\text{sat})$.

13. Define enhancement mode and depletion mode for both n-channel and p-channel devices.

14. Sketch the charge distribution through a MOS capacitor with a p-type substrate when biased in the inversion mode. Write the charge neutrality equation.

15. Discuss why the threshold voltage changes when a reverse-biased source-to-substrate voltage is applied to a MOSFET.

PROBLEMS

(*Note:* In the following problems, assume the semiconductor and oxide in the MOS system are silicon and silicon dioxide, respectively, and assume the temperature is $T = 300$ K unless otherwise stated. Use Figure 10.16 to determine the metal–semiconductor work function difference.)

Section 10.1   The Two-Terminal MOS Structure

10.1   The dc charge distributions of four ideal MOS capacitors are shown in Figure P10.1. For each case: (*a*) Is the semiconductor n or p type? (*b*) Is the device biased in the accumulation, depletion, or inversion mode? (*c*) Draw the energy-band diagram in the semiconductor region.

10.2   (*a*) Calculate the maximum space charge width x_{dT} and the maximum space charge density $|Q'_{SD}(\text{max})|$ in a MOS capacitor with a p-type silicon substrate at $T = 300$ K for doping concentrations of (*i*) $N_a = 7 \times 10^{15}$ cm^{-3} and (*ii*) $N_a = 3 \times 10^{16}$ cm^{-3}. (*b*) Repeat part (*a*) for $T = 350$ K.

10.3   (*a*) Consider a MOS capacitor at $T = 300$ K with an n-type silicon substrate. Determine the silicon doping concentration such that $|Q'_{SD}(\text{max})| = 1.25 \times 10^{-8}$ C/cm^{-2}. (*b*) What is the surface potential that results in the maximum space charge width?