11

Metal–Oxide–Semiconductor Field-Effect Transistor: Additional Concepts

In this chapter we present additional concepts that are commonly encountered in metal–oxide–semiconductor field-effect transistors (MOSFETs). These concepts include nonideal effects, small device geometry, breakdown, threshold voltage adjustment by ion implantation, and radiation effects. Although there are a multitude of details that become important when fabricating MOSFETs in ICs, we are able to consider only a few here. Many additional details can be found in more advanced texts. ■

11.0 | PREVIEW

In this chapter, we will:

■ Describe and analyze subthreshold conduction, which is the phenomenon whereby current is induced in the channel before the defined threshold voltage is reached.

■ Analyze channel length modulation, which is a characteristic of short-channel lengths and leads to a finite output resistance.

■ Consider the effects of a decrease in carrier mobility due to increasing gate voltage.

■ Analyze the effects of carrier saturation velocity. Carriers can easily reach their saturation velocity in short-channel devices.

■ Discuss MOSFET scaling, which describes how various parameters must be changed as device size is decreased.

■ Consider the deviations in threshold voltage due to small geometry devices, including short channel length devices and small channel width devices.

- Describe and analyze various voltage breakdown mechanisms in MOSFETs.
- Describe and analyze the technique of threshold voltage adjustment by ion implantation.
- Consider the introduction of trapped oxide charges by ionizing radiation and hot electron effects.

11.1 | NONIDEAL EFFECTS

As with any semiconductor device, the experimental characteristics of MOSFETs deviate to some degree from the ideal relations that have been theoretically derived using the various assumptions and approximations. In this section, we consider five effects that cause deviations from the assumptions used in the ideal derivations. These effects are subthreshold conduction, channel length modulation, mobility variations, velocity saturation, and ballistic transport.

11.1.1   Subthreshold Conduction

The ideal current–voltage relationship predicts zero drain current when the gate-to-source voltage is less than or equal to the threshold voltage. Experimentally, I_D is not zero when $V_{GS} \leq V_T$. Figure 11.1 shows a comparison between the ideal characteristic that was derived, and the experimental results. The drain current, which exists for $V_{GS} \leq V_T$, is known as the *subthreshold current*.
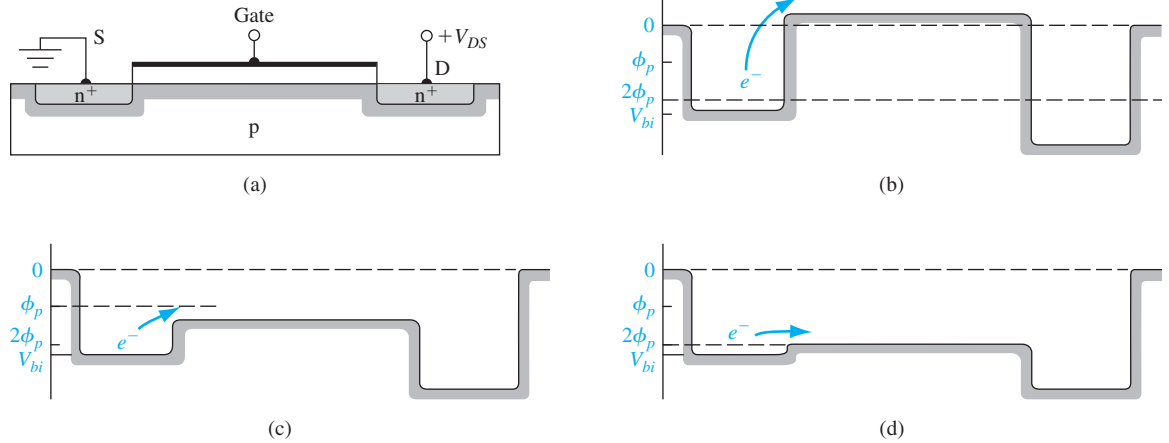
Figure 11.2 shows the energy-band diagram of an MOS structure with a p-type substrate biased so that $\phi_s < 2\phi_{fp}$. At the same time, the Fermi level is closer to the conduction band than the valence band, so the semiconductor surface develops the
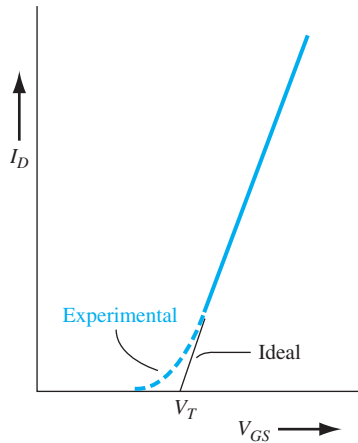
Figure 11.1 | Comparison of ideal and experimental plots of $\sqrt{I_D}$ versus V_{GS}.
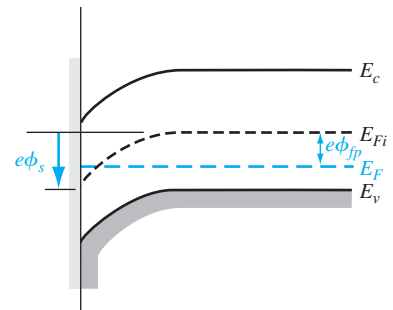
Figure 11.2 | Energy-band diagram when $\phi_{fp} < \phi_s < 2\phi_{fp}$.

Figure 11.3 | (a) Cross section along channel length of n-channel MOSFET. Energy-band diagrams along channel length at (b) accumulation, (c) weak inversion, and (d) inversion.

characteristics of a lightly doped n-type material. We would expect, then, to observe some conduction between the n$^+$ source and drain contacts through this weakly inverted channel. The condition for $\phi_{fp} < \phi_s < 2\phi_{fp}$ is known as *weak inversion*.

Figure 11.3 shows the surface potential along the length of the channel at accumulation, weak inversion, and threshold for the case when a small drain voltage is applied. The bulk p-substrate is assumed to be at zero potential. Figure 11.3b, c shows the accumulation and weak inversion cases. There is a potential barrier between the n$^+$ source and channel region which the electrons must overcome in order to generate a channel current. A comparison of these barriers with those in pn junctions would suggest that the channel current is an exponential function of V_{GS}. In the inversion mode, shown in Figure 11.3d, the barrier is so small that we lose the exponential dependence, since the junction is more like an ohmic contact.

The actual derivation of the subthreshold current is beyond the scope of this chapter. We can write that

$$I_D(\text{sub}) \propto \left[\exp\left(\frac{eV_{GS}}{kT} \right) \right] \cdot \left[1 - \exp\left(\frac{-eV_{DS}}{kT} \right) \right] \quad\quad (11.1)$$

If V_{DS} is larger than a few (kT/e) volts, then the subthreshold current is independent of V_{DS}.

Figure 11.4 shows the exponential behavior of the subthreshold current for several body-to-source voltages. Also shown on the curves are the threshold voltage values. Ideally, a change in gate voltage on approximately 60 mV produces an order of magnitude change in the subthreshold current. A detailed analysis of the subthreshold condition shows that the slope of the I_D versus V_{DS} curve is a function of the semiconductor doping and is also a function of the interface state density. The
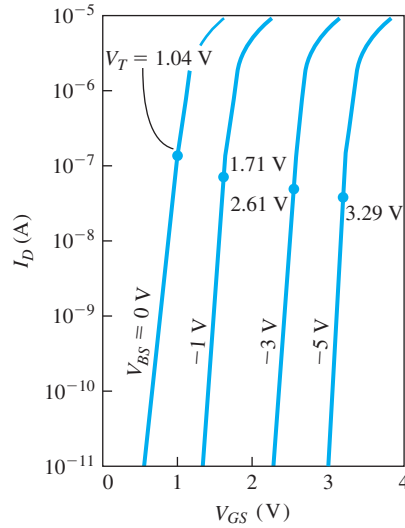
Figure 11.4 | Subthreshold current–voltage characteristics for several values of substrate voltage (the threshold voltage is indicated on each curve). *(From Schroder [17].)*

measurement of the slope of these curves has been used to experimentally determine the oxide–semiconductor interface state density.

If a MOSFET is biased at or even slightly below the threshold voltage, the drain current is not zero. The subthreshold current may add significantly to power dissipation in a large-scale integrated circuit in which hundreds or thousands of MOSFETs are used. The circuit design must include the subthreshold current or ensure that the MOSFET is biased sufficiently below the threshold voltage in the "off" state.

11.1.2 Channel Length Modulation

We assumed in the derivation of the ideal current–voltage relationship that the channel length L was a constant. However, when the MOSFET is biased in the saturation region, the depletion region at the drain terminal extends laterally into the channel, reducing the effective channel length. Since the depletion region width is bias dependent, the effective channel length is also bias dependent and is modulated by the drain-to-source voltage. This channel length modulation effect is shown in Figure 11.5 for an n-channel MOSFET.

The depletion width extending into the p-region of a pn junction under zero bias can be written as

$$x_p = \sqrt{\frac{2\epsilon_s \phi_{fp}}{eN_a}} \tag{11.2}$$
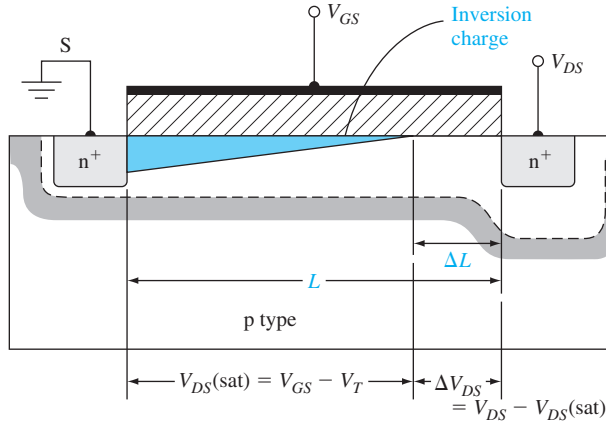
Figure 11.5 | Cross-section of an n-channel MOSFET showing the channel length modulation effect.

For a one-sided n^+p junction, essentially all of the applied reverse-biased voltage is across the low-doped p region. The space charge width of the drain–substrate junction is approximately

$$x_p = \sqrt{\frac{2\epsilon_s}{eN_a}(\phi_{fp} + V_{DS})} \tag{11.3}$$

However, the space charge region defined by ΔL, as shown in Figure 11.5, does not begin to form until $V_{DS} > V_{DS}(\text{sat})$. As a first approximation, we can write that ΔL is the total space charge width minus the space charge width that exists when $V_{DS} = V_{DS}(\text{sat})$, or

$$\Delta L = \sqrt{\frac{2\epsilon_s}{eN_a}}\left[\sqrt{\phi_{fp} + V_{DS}(\text{sat}) + \Delta V_{DS}} - \sqrt{\phi_{fp} + V_{DS}(\text{sat})}\right] \tag{11.4}$$

where

$$\Delta V_{DS} = V_{DS} - V_{DS}(\text{sat}) \tag{11.5}$$

The applied drain-to-source voltage is V_{DS} and we are assuming that $V_{DS} > V_{DS}(\text{sat})$.

As a second approximation at determining ΔL, we can consider Figure 11.6 and revisit the one-dimensional Poisson's equation. The electric field E_{sat} is the lateral electric field at the point where the inversion layer charge is pinched off. Neglecting any charges that exist due to current, we can write

$$\frac{dE}{dx} = \frac{\rho(x)}{\epsilon_s} \tag{11.6}$$

where $\rho(x) = -eN_a$ and is a constant for a uniformly doped substrate. Integrating Equation (11.6) and applying the boundary conditions give the electric field in the space charge region defined by ΔL:

$$E = -\frac{eN_a x}{\epsilon_s} - E_{\text{sat}} \tag{11.7}$$
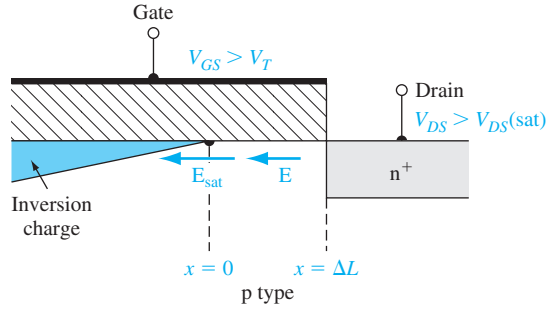
Figure 11.6 | Expanded view of cross section near the drain terminal of an n-channel MOSFET showing the channel length modulation effect.

The potential in this region is

$$\phi(x) = -\int \mathrm{E}dx = \frac{eN_a x^2}{2\epsilon_s} + \mathrm{E}_{sat}x + C_1 \tag{11.8}$$

where C_1 is a constant of integration. The boundary conditions are $\phi(x = 0) = V_{DS}(\text{sat})$ and $\phi(x = \Delta L) = V_{DS}$. Substituting these boundary conditions into Equation (11.8), we obtain

$$V_{DS} = \frac{eN_a(\Delta L)^2}{2\epsilon_s} + \mathrm{E}_{sat}(\Delta L) + V_{DS}(\text{sat}) \tag{11.9}$$

Solving for ΔL, we can write

$$\Delta L = \sqrt{\frac{2\epsilon_s}{eN_a}} \left[\sqrt{\phi_{sat} + [V_{DS} - V_{DS}(\text{sat})]} - \sqrt{\phi_{sat}} \right] \tag{11.10}$$

where

$$\phi_{sat} = \frac{2\epsilon_s}{eN_a} \cdot \left(\frac{\mathrm{E}_{sat}}{2} \right)^2$$

In general, the value of E_{sat} is in the range $10^4 < \mathrm{E}_{sat} < 2 \times 10^5$ V/cm.

Other models used to determine ΔL include the negative charges due to the drain current and also include two-dimensional effects. These models are not considered here.

Since the drain current is inversely proportional to the channel length, we may write

$$I_D' = \left(\frac{L}{L - \Delta L} \right) I_D \tag{11.11}$$

where I_D' is the actual drain current and I_D is the ideal drain current. Since ΔL is a function of V_{DS}, I_D' is now also a function of V_{DS} even though the transistor is biased in the saturation region.
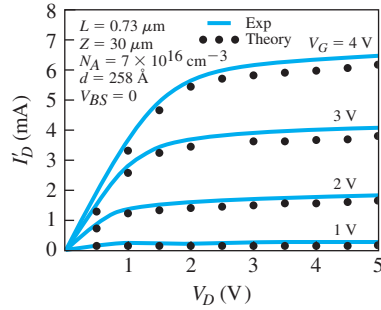
Figure 11.7 | Current–voltage characteristics of a MOSFET showing short-channel effects.
(From Sze [22].)

Since I_D' is now a function of V_{DS}, the output resistance is no longer infinite. The drain current in the saturation region can be written as

$$I_D' = \frac{k_n'}{2} \cdot \frac{W}{L} \cdot [(V_{GS} - V_T)^2 (1 + \lambda V_{DS})] \qquad (11.12)$$

where λ is the *channel length modulation parameter.*

The output resistance is given by

$$r_o = \left(\frac{\partial I_D'}{\partial V_{DS}}\right)^{-1} = \left\{\frac{k_n'}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_T)^2 \cdot \lambda\right\}^{-1} \qquad (11.13a)$$

Since λ is normally small, Equation (11.13a) can be written as

$$r_o \cong \frac{1}{\lambda I_D} \qquad (11.13b)$$

Figure 11.7 shows some typical I_D' versus V_{DS} curves with positive slopes in the saturation region due to channel length modulation. As the MOSFET dimensions become smaller, the change in the channel length ΔL becomes a larger fraction of the original length L, and the channel length modulation becomes more severe.

Objective: Determine the increase in drain current due to short channel modulation.

EXAMPLE 11.1

Consider an n-channel MOSFET with a substrate doping concentration of $N_a = 2 \times 10^{16}$ cm^{-3}, a threshold voltage of $V_T = 0.4$ V, and a channel length of $L = 1$ μm. The device is biased at $V_{GS} = 1$ V and $V_{DS} = 2.5$ V. Determine the ratio of actual drain current compared to the ideal value.

■ **Solution**

We find

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259) \ln\left(\frac{2 \times 10^{16}}{1.5 \times 10^{10}}\right) = 0.3653 \text{ V}$$

$$V_{DS}(\text{sat}) = V_{GS} - V_T = 1.0 - 0.4 = 0.6 \text{ V}$$

and

$$\Delta V_{DS} = V_{DS} - V_{DS}(\text{sat}) = 2.5 - 0.6 = 1.9 \text{ V}$$

Using Equation (11.4), we determine

$$\Delta L = \sqrt{\frac{2\epsilon_s}{eN_a}} \left[\sqrt{\phi_{fp} + V_{DS}(\text{sat}) + \Delta V_{DS}} - \sqrt{\phi_{fp} + V_{DS}(\text{sat})} \right]$$

$$= \sqrt{\frac{2(11.7)(8.85 \times 10^{-14})}{(1.6 \times 10^{-19})(2 \times 10^{16})}} \left[\sqrt{0.3653 + 0.6 + 1.9} - \sqrt{0.3653 + 0.6} \right]$$

$$= 1.807 \times 10^{-5} \text{ cm}$$

or

$$\Delta L = 0.1807 \ \mu\text{m}$$

Then

$$\frac{I'_D}{I_D} = \frac{L}{L - \Delta L} = \frac{1}{1 - 0.1807} = 1.22$$

■ **Comment**

The actual drain current increases as the effective channel length decreases when the transistor is biased in the saturation region.

■ **EXERCISE PROBLEM**

Ex 11.1   An n-channel MOSFET has the same properties as described in Example 11.1 except for the channel length. The transistor is biased at $V_{GS} = 0.8$ V and $V_{DS} = 2.5$ V. Find the minimum channel length such that the ratio of actual drain current to the ideal drain current due to channel length modulation is no larger than 1.35.

(Ans. $L = 0.698 \ \mu$m)

11.1.3   Mobility Variation

In the derivation of the ideal *I–V* relationship, we explicitly assumed that the mobility was a constant. However, this assumption must be modified for two reasons. The first effect to be considered is the variation in mobility with gate voltage. The second reason for a mobility variation is that the effective carrier mobility decreases as the carrier approaches the velocity saturation limit. This effect is discussed in the next section.

The inversion layer charge is induced by a vertical electric field, which is shown in Figure 11.8 for an n-channel device. A positive gate voltage produces a force on the electrons in the inversion layer toward the surface. As the electrons travel through the channel toward the drain, they are attracted to the surface, but then are repelled by localized coulombic forces. This effect, schematically shown in Figure 11.9, is called *surface scattering*. The surface scattering effect reduces mobility. If there is a positive fixed oxide charge near the oxide-semiconductor interface, the mobility will be further reduced due to the additional coulomb interaction.
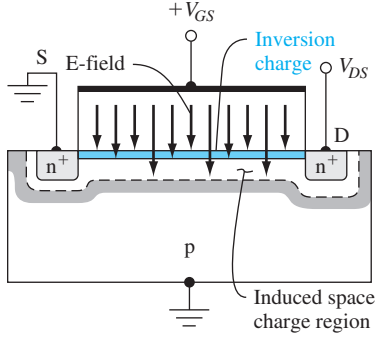
Figure 11.8 | Vertical electric field in an n-channel MOSFET.
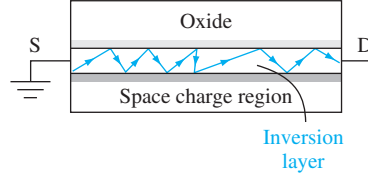
**Figure 11.9** | Schematic of carrier surface scattering effects.

The relationship between the inversion charge mobility and the transverse electric field is usually measured experimentally. An effective transverse electric field can be defined as

$$E_{eff} = \frac{1}{\epsilon_s} \left(|Q'_{SD}(\max)| + \frac{1}{2} Q'_n \right) \tag{11.14}$$

The effective inversion charge mobility can be determined from the channel conductance as a function of gate voltage. Figure 11.10 shows the effective electron mobility at $T = 300$ K for different doping levels and different oxide thicknesses. The effective mobility is only a function of the electric field at the inversion layer and is independent of oxide thickness. The effective mobility may be represented by

$$\mu_{eff} = \mu_0 \left(\frac{E_{eff}}{E_0} \right)^{-1/3} \tag{11.15}$$

where μ_0 and E_0 are constants determined from experimental results.
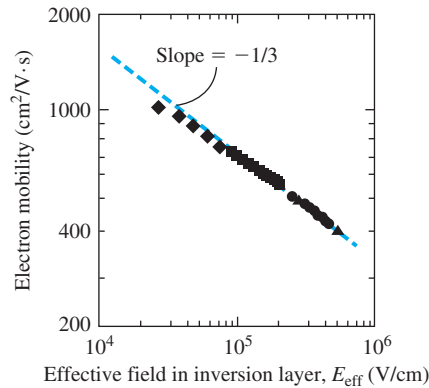
Figure 11.10 | Measured inversion layer electron mobility versus electric field at the inversion layer.
(From Yang [25].)

The effective inversion charge mobility is a strong function of temperature because of lattice scattering. As the temperature is reduced, the mobility increases.

EXAMPLE 11.2

Objective: Calculate the effective electric field at threshold for a given semiconductor doping concentration.

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

■ **Solution**

From the results of Chapter 10, we can calculate

$$\phi_{fp} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259) \ln\left(\frac{3 \times 10^{16}}{1.5 \times 10^{10}}\right) = 0.376 \text{ V}$$

and

$$x_{dT} = \left\{\frac{4\epsilon_s \phi_{fp}}{eN_a}\right\}^{1/2} = \left\{\frac{4(11.7)(8.85 \times 10^{-14})(0.376)}{(1.6 \times 10^{-19})(3 \times 10^{16})}\right\}^{1/2}$$

which is $x_{dT} = 0.18$ μm. Then

$$\left|Q'_{SD}(\text{max})\right| = eN_a x_{dT} = 8.64 \times 10^{-8} \text{ C/cm}^2$$

At the threshold inversion point, we may assume that $Q'_n = 0$, so the effective electric field from Equation (11.14) is found as

$$E_{\text{eff}} = \frac{1}{\epsilon_s}\left|Q'_{SD}(\text{max})\right| = \frac{8.64 \times 10^{-8}}{(11.7)(8.85 \times 10^{-14})} = 8.34 \times 10^4 \text{ V/cm}$$

■ **Comment**

We can see, from Figure 11.10, that this value of effective transverse electric field at the surface is sufficient for the effective inversion charge mobility to be significantly less than the bulk semiconductor value.

■ **EXERCISE PROBLEM**

Ex 11.2 Determine (using Figure 11.10) the effective inversion layer electron mobility for a surface electric field of $E_{eff} = 2 \times 10^5$ V/cm.

(Ans. $\mu_n \cong 550$ cm^2/V-s)

The effective mobility is a function of gate voltage through the inversion charge density in Equation (11.14). As the gate voltage increases, the carrier mobility decreases even further.

11.1.4 Velocity Saturation

In the analysis of the long-channel MOSFET, we assume the mobility to be constant, which means that the drift velocity increases without limit as the electric field increases. In this ideal case, the carrier velocity increases until the ideal current is attained. However, we have seen that the carrier velocity saturates with increasing electric field. Velocity saturation will become more prominent in shorter-channel devices since the corresponding horizontal electric field is generally larger.

In the ideal *I–V* relationship, current saturation occurs when the inversion charge density becomes zero at the drain terminal, or when

$$V_{DS} = V_{DS}(\text{sat}) = V_{GS} - V_T \tag{11.16}$$

for the n-channel MOSFET. However, velocity saturation can change this saturation condition. Velocity saturation will occur when the horizontal electric field is approximately 10^4 V/cm. If $V_{DS} = 5$ V in a device with a channel length of $L = 1$ μm, the average electric field is 5×10^4 V/cm. Velocity saturation, then, is very likely to occur in short-channel devices.

The modified $I_D(\text{sat})$ characteristics are described approximately by

$$I_D(\text{sat}) = W C_{\text{ox}}(V_{GS} - V_T)v_{\text{sat}} \tag{11.17}$$

where v_{sat} is the saturation velocity (approximately 10^7 cm/s for electrons in bulk silicon) and C_{ox} is the gate oxide capacitance per cm². The saturation velocity will decrease somewhat with applied gate voltage because of the vertical electric field and surface scattering. Velocity saturation will yield an $I_D(\text{sat})$ value smaller than that predicted by the ideal relation, and it will yield a smaller $V_{DS}(\text{sat})$ value than predicted. The $I_D(\text{sat})$ current is also approximately linear with V_{GS}, instead of having the ideal square law dependence predicted previously.

There are several models of mobility versus electric field. One particular relation that is commonly used is

$$\mu = \frac{\mu_{\text{eff}}}{\left[1 + \left(\dfrac{\mu_{\text{eff}}\text{E}}{v_{\text{sat}}}\right)^2\right]^{1/2}} \tag{11.18}$$

Figure 11.11 shows a comparison of drain current versus drain-to-source voltage characteristics for constant mobility and for field-dependent mobility. The smaller values of $I_D(\text{sat})$ and the approximate linear dependence on V_{GS} may be noted for the field-dependent mobility curves.

The transconductance is found from

$$g_{ms} = \frac{\partial I_D(\text{sat})}{\partial V_{GS}} = W C_{\text{ox}}\, v_{\text{sat}} \tag{11.19}$$

which is now independent of V_{GS} and V_{DS} when velocity saturation occurs. The drain current is saturated by the velocity saturation effect, which leads to a constant transconductance.

When velocity saturation occurs, the cutoff frequency is given by

$$f_T = \frac{g_m}{2\pi C_G} = \frac{W C_{\text{ox}} v_{\text{sat}}}{2\pi (C_{\text{ox}} WL)} = \frac{v_{\text{sat}}}{2\pi L} \tag{11.20}$$

where the parasitic capacitances are assumed to be negligible.

11.1.5 Ballistic Transport

Scattering events in a semiconductor limit the velocity of carriers to an average drift velocity as discussed in Chapter 5. The average drift velocity is a function of the
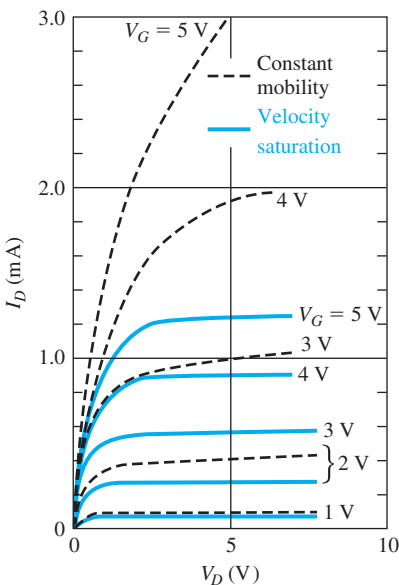
Figure 11.11 | Comparison of I_D versus V_D characteristics for constant mobility (dashed curves) and for field-dependent mobility and velocity saturation effects (solid curves).
(From Sze and Ng [22].)

mean time between collisions or the mean distance between scattering events. In the long-channel device, the channel length L is much longer than the mean distance between collisions l, so that an average carrier drift velocity exists. As the MOSFET channel length is reduced, the mean distance between collisions l may become comparable to L so that the previous analysis may not be valid. If the channel length is further reduced so that $L < l$, then a large fraction of carriers could travel from the source to the drain without experiencing a scattering event. This motion of carriers is called *ballistic transport*.

Ballistic transport means that carriers travel faster than the average drift velocity or the saturation velocity, and this effect can lead to very fast devices. Ballistic transport occurs in submicron ($L < 1$ μm) devices. As the MOSFET technology continues to shrink the channel length toward the 0.1 μm value, the ballistic transport phenomenon will become more important.

TEST YOUR UNDERSTANDING

TYU 11.1   Consider a MOSFET biased in the subthreshold region with $V_D \gg kT/e$. For the ideal relationship given, what change in gate-to-source voltage produces a factor of 10 change in drain current?

(Ans. $\Delta V_{GS} = 59.64$ mV)

TYU 11.2 Consider an NMOS transistor with the following parameters: $\mu_n = 1000$ cm²/V-s, $C_{ox} = 10^{-8}$ F/cm², $W = 10$ μm, $L = 1$ μm, $V_T = 0.4$ V, and $v_{sat} = 5 \times 10^6$ cm/s. Plot on the same graph $I_D(\text{sat})$ versus V_{GS} over the range $0 \leq V_{GS} \leq 4$ V for the case (a) of an ideal transistor (Equation (10.45a)) and (b) when velocity saturation occurs (Equation (11.17)).

[Ans. (a) $I_D(\text{sat}) = 50(V_{GS} - 0.4)^2$ μA; (b) $I_D(\text{sat}) = 50(V_{GS} - 0.4)$ μA]

11.2 | MOSFET SCALING

As we noted in the previous chapter, the frequency response of MOSFETs increases as the channel length decreases. The driving force in CMOS technology evolution in the last couple of decades has been reduced channel lengths. Channel lengths of 0.13 μm or less are now the norm. One question that must be considered is what other device parameters must be scaled as the channel length is scaled down.
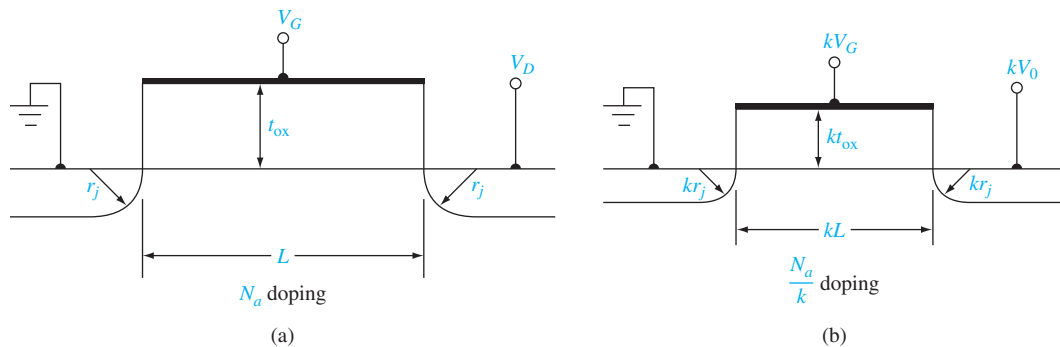
11.2.1 Constant-Field Scaling

The principle of constant-field scaling is that device dimensions and device voltages be scaled such that electric fields (both horizontal and vertical) remain essentially constant. To ensure that the reliability of the scaled device is not compromised, the electric fields in the scaled device must not increase.

Figure 11.12a shows the cross section and parameters of an original NMOS device and Figure 11.12b shows the scaled device, where the scaling parameter is k. Typically, $k \approx 0.7$ per generation of a given technology.

As seen in the figure, the channel length is scaled from L to kL. To maintain a constant horizontal electric field, the drain voltage must also be scaled from V_D to kV_D. The maximum gate voltage will also be scaled from V_G to kV_G so that the gate and drain voltages remain compatible. To maintain a constant vertical electric field, the oxide thickness then must also be scaled from t_{ox} to kt_{ox}.

The maximum depletion width at the drain terminal, for a one-sided pn junction, is

$$x_D = \sqrt{\frac{2\epsilon(V_{bi} + V_D)}{eN_a}} \qquad (11.21)$$

Figure 11.12 | Cross section of (a) original NMOS transistor and (b) scaled NMOS transistor.

Table 11.1 | Summary of constant-field device scaling

	Device and circuit parameters	Scaling factor ($k < 1$)
Scaled parameters	Device dimensions (L, t_{ox}, W, x_j)	k
	Doping concentration (N_a, N_d)	$1/k$
	Voltages	k
Effect on device parameters	Electric field	1
	Carrier velocity	1
	Depletion widths	k
	Capacitance ($C = \epsilon A/t$)	k
	Drift current	k
Effect on circuit parameters	Device density	$1/k^2$
	Power density	1
	Power dissipation per device ($P = IV$)	k^2
	Circuit delay time ($\approx CV/I$)	k
	Power−delay product ($P\tau$)	k^3

Source: Taur and Ning [23].

Since the channel length is being reduced, the depletion widths also need to be reduced. If the substrate doping concentration is increased by the factor ($1/k$), then the depletion width is reduced by approximately the factor k since V_D is reduced by k.

The drain current per channel width, for the transistor biased in the saturation region, can be written as

$$\frac{I_D}{W} = \frac{\mu_n \epsilon_{ox}}{2t_{ox}L} (V_G - V_T)^2 \rightarrow \frac{\mu_n \epsilon_{ox}}{2(kt_{ox})(kL)} (kV_G - V_T)^2 \approx \text{constant} \qquad (11.22)$$

The drift current per channel width remains essentially a constant, so if the channel width is reduced by k, then the drain current is also reduced by k. The area of the device, $A \approx WL$, is then reduced by k^2 and the power, $P = IV$, is also reduced by k^2. The power density in the chip remains unchanged.

Table 11.1 summarizes the device scaling and the effect on circuit parameters. Keep in mind that the width and length of interconnect lines are also assumed to be reduced by the same scaling factor.

11.2.2   Threshold Voltage—First Approximation

In constant-field scaling, the device voltages are reduced by the scaling factor k. It would seem appropriate that the threshold voltage should also be scaled by the same factor. The threshold voltage, for a uniformly doped substrate, can be written as

$$V_T = V_{FB} + 2\phi_{fp} + \frac{\sqrt{2\epsilon e N_a(2\phi_{fp})}}{C_{ox}} \qquad (11.23)$$

The first two terms in Equation (11.23) are functions of material parameters that do not scale and are only very slight functions of doping concentration. The last term is approximately proportional to $\sqrt{k}$, so the threshold voltage does not scale directly with the scaling factor k.

The effect of short channels on the threshold voltage is discussed further in section 11.3 of this chapter.

11.2.3 Generalized Scaling

In constant-field scaling, the applied voltages are scaled with the same scaling factor k as the device dimensions. However, in actual technology evolution, voltages have not been reduced with the same scaling factor. There has been reluctance, for example, to change standardized power supply levels that have been used in circuits earlier. In addition, other factors that do not scale, such as threshold voltage and subthreshold currents, have made the reduction in applied voltages less desirable. As a consequence, electric fields in MOS devices have tended to increase as device dimensions shrinked.

Consequences of increased electric fields are reduced reliability and increased power density. As the power density increases, the device temperature may increase. Increased temperature may affect the device reliability. As the oxide thickness is reduced and the electric field is increased, gate oxides are closer to breakdown and oxide integrity may be more difficult to maintain. In addition, direct tunneling of carriers through the oxide may be more likely to occur. Increased electric fields may also increase the chances of hot-electron effects, which are discussed later in this chapter. Reducing device dimensions, then, can introduce challenging problems that must be solved.


TEST YOUR UNDERSTANDING

TYU 11.3 An NMOS transistor has the following parameters: $L = 1\ \mu$m, $W = 10\ \mu$m, t_{ox} $= 250$ Å, $N_a = 5 \times 10^{15}$ cm^{-3}, and applied voltages of 3 V. If the device is to be scaled using constant-field scaling, determine the new device parameters for a scaling factor of $k = 0.7$. (Ans. $L = 0.7\ \mu$m, $W = 7\ \mu$m, $t_{ox} = 175$ Å, $N_a = 7.14 \times 10^{15}$ cm^{-3}, and applied voltages of 2.1 V.)

11.3 | THRESHOLD VOLTAGE MODIFICATIONS

We have derived the ideal MOSFET relations in the previous chapter, including expressions for threshold voltage and for the current–voltage characteristics. We now consider some of the nonideal effects including channel length modulation. Additional effects on threshold voltage occur as the devices shrink in size. A reduction in channel length increases the transconductance and frequency response of the MOSFET, and a reduction in channel width increases the packing density in an integrated circuit. A reduction in either or both the channel length and channel width can affect the threshold voltage.

11.3.1 Short-Channel Effects

For the ideal MOSFET, we have derived the threshold voltage using the concept of charge neutrality in which the sum of charges in the metal oxide inversion layer and
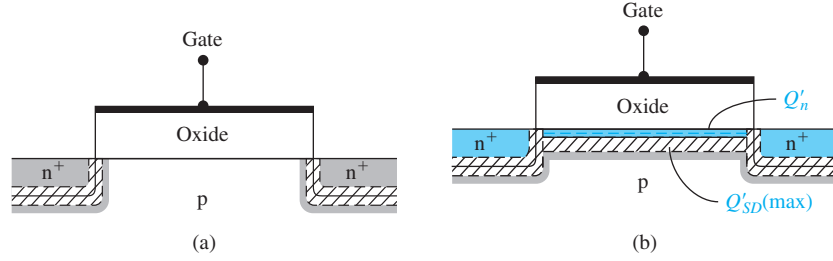
Figure 11.13 | Cross section of a long n-channel MOSFET (a) at flat band and (b) at inversion.

semiconductor space charge region is zero. We assumed that the gate area was the same as the active area in the semiconductor. Using this assumption, we have considered only equivalent surface charge densities and neglected any effects on threshold voltage that may occur due to source and drain space charge regions that extend into the active channel region.

Figure 11.13a shows the cross section of a long n-channel MOSFET at flat band, with zero source and drain voltage applied. The space charge regions at the source and drain extend into the channel region but occupy only a small fraction of the entire channel region. The gate voltage, then, will control essentially all of the space charge induced in the channel region at inversion as shown in Figure 11.13b.

As the channel length decreases, the fraction of charge in the channel region controlled by the gate decreases. This effect can be seen in Figure 11.14 for the flat-band condition. As the drain voltage increases, the reverse-biased space charge region at the drain extends further into the channel area and the gate controls even less bulk charge. The amount of charge in the channel region, $Q'_{SD}(\text{max})$, controlled by the gate, affects the threshold voltage and can be seen from Equation (11.24)

$$V_{TN} = \left(|Q'_{SD}(\text{max})| - Q'_{ss} \right)\left(\frac{t_{\text{ox}}}{\epsilon_{\text{ox}}}\right) + \phi_{ms} + 2\phi_{fp} \qquad (11.24)$$

We can quantitatively determine the short-channel effects on the threshold voltage by considering the parameters shown in Figure 11.15. The source and drain junctions are characterized by a diffused junction depth r_j. We will assume that the lateral diffusion distance under the gate is the same as the vertical diffusion distance. This assumption is a reasonably good approximation for diffused junctions but becomes less accurate for ion implanted junctions. We will initially consider the case when the source, drain, and body contacts are all at ground potential.

The basic assumption in this analysis is that the bulk charge in the trapezoidal region under the gate is controlled by the gate. The potential difference across the bulk space charge region is $2\phi_{fp}$ at the threshold inversion point, and the built-in potential barrier height of the source and drain junctions is also approximately $2\phi_{fp}$, implying that the three space charge widths are essentially equal. We can then write

$$x_s \approx x_d \approx x_{dT} \equiv x_{dT} \qquad (11.25)$$
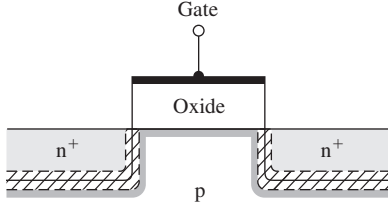
Figure 11.14 | Cross section of a short n-channel MOSFET at flat band.
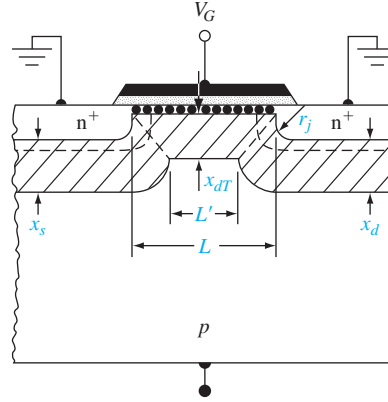
Figure 11.15 | Charge sharing in the short-channel threshold voltage model. *(From Yau [26].)*

Using the geometrical approximation, the average bulk charge per unit area Q'_B in the trapezoid is

$$|Q'_B| \cdot L = eN_a x_{dT} \left(\frac{L + L'}{2} \right) \tag{11.26}$$

From the geometry, we can show that

$$\frac{L + L'}{2L} = \left[1 - \frac{r_j}{L} \left(\sqrt{1 + \frac{2x_{dT}}{r_j}} - 1 \right) \right] \tag{11.27}$$

Then

$$|Q'_B| = eN_a x_{dT} \left[1 - \frac{r_j}{L} \left(\sqrt{1 + \frac{2x_{dT}}{r_j}} - 1 \right) \right] \tag{11.28}$$

Equation (11.28) is now used in place of $|Q'_{SD}(\text{max})|$ in the expression for the threshold voltage.

Since $|Q'_{SD}(\text{max})| = eN_a x_{dT}$, we can find ΔV_T as

$$\Delta V_T = -\frac{eN_a x_{dT}}{C_{\text{ox}}} \left[\frac{r_j}{L} \left(\sqrt{1 + \frac{2x_{dT}}{r_j}} - 1 \right) \right] \tag{11.29}$$

where

$$\Delta V_T = V_{T(\text{short channel})} - V_{T(\text{long channel})} \tag{11.30}$$

As the channel length decreases, the threshold voltage shifts in the negative direction so that an n-channel MOSFET shifts toward depletion mode.

Objective: Calculate the threshold voltage shift due to short-channel effects.

EXAMPLE 11.3

Consider an n-channel MOSFET with $N_a = 3 \times 10^{16}$ cm^{-3}, $L = 1.0$ μm, $r_j = 0.3$ μm, and $t_{ox} = 20$ nm $= 200$ Å.

■ **Solution**

We can find

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

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

and

$$x_{dT} = \left[\frac{4\epsilon_s \phi_{fp}}{eN_a}\right]^{1/2} = \left[\frac{4(11.7)(8.85 \times 10^{-14})(0.3758)}{(1.6 \times 10^{-19})(3 \times 10^{16})}\right]^{1/2}$$

$$= 0.18 \times 10^{-4} \text{ cm} = 0.18 \ \mu\text{m}$$

The shift in threshold voltage is now found as

$$\Delta V_T = -\frac{eN_a x_{dT}}{C_{ox}}\left[\frac{r_j}{L}\left(\sqrt{1 + \frac{2x_{dT}}{r_j}} - 1\right)\right]$$

$$= -\frac{(1.6 \times 10^{-19})(3 \times 10^{16})(0.18 \times 10^{-4})}{1.726 \times 10^{-7}}\left[\frac{0.3}{1.0}\left(\sqrt{1 + \frac{2(0.18)}{0.3}} - 1\right)\right]$$

or

$$\Delta V_T = -0.0726 \text{ V}$$

■ **Comment**

If the design value of the threshold voltage of an n-channel MOSFET is to be $V_T = 0.35$ V, for example, a shift of $\Delta V_T = -0.0726$ V due to short-channel effects is significant and needs to be taken into account.

■ **EXERCISE PROBLEM**

Ex 11.3   Repeat Example 11.3 if the device parameters are $N_a = 10^{16}$ cm^{-3}, $L = 0.75$ μm, $r_j = 0.25$ μm, and $t_{ox} = 12$ nm $= 120$ Å.

(Ans. $\Delta V_T = -0.0469$ V)

The effect of short channels becomes more pronounced as the channel length is reduced further.

The shift in threshold voltage with channel length for an n-channel MOSFET is shown in Figure 11.16. As the substrate doping increases, the initial threshold voltage increases, as we have seen in the previous chapter, and the short-channel threshold shift also becomes larger. The short-channel effects on threshold voltage do not become significant until the channel length becomes less than approximately 2 μm. The threshold voltage shift also becomes smaller as the diffusion depth r_j becomes smaller so that very shallow junctions reduce the threshold voltage dependence on channel length.

Equation (11.29) is derived assuming that the source, channel, and drain space charge widths are all equal. If we now apply a drain voltage, the space charge width at the drain terminal widens, which makes L' smaller, and the amount of bulk charge controlled by the gate voltage decreases. This effect makes the threshold voltage a function of drain voltage. As the drain voltage increases, the threshold voltage of an n-channel MOSFET decreases. The threshold voltage versus channel length is
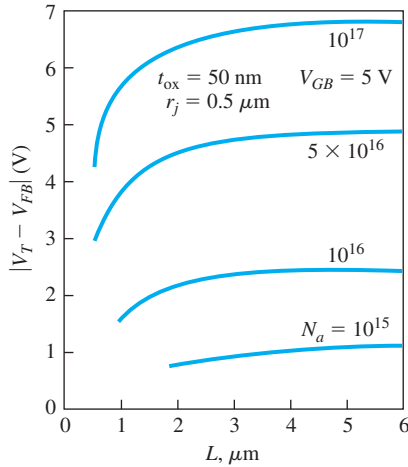
**Figure 11.16** | Threshold voltage versus channel length for various substrate dopings. (*From Yau [26].*)
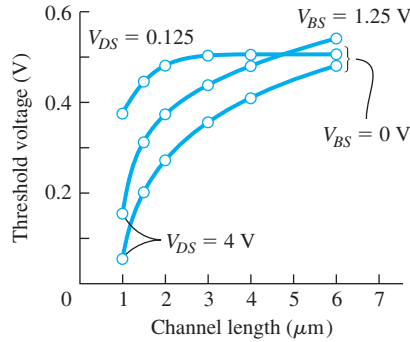
Figure 11.17 | Threshold voltage versus channel length for two values of drain-to-source and body-to-source voltage. (*From Yang [25].*)

plotted in Figure 11.17 for two values of drain-to-source voltage and two values of body-to-source voltage.

11.3.2 Narrow-Channel Effects

Figure 11.18 shows the cross section along the channel width of an n-channel MOSFET biased at inversion. The current is perpendicular to the channel width through the inversion charge. We may note in the figure that there is an additional space charge region at each end of the channel width. This additional charge is controlled by the gate voltage but is not included in the derivation of the ideal threshold voltage relation. The threshold voltage expression must be modified to include this additional charge.
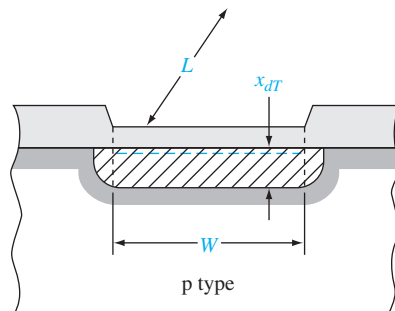
Figure 11.18 | Cross section of an n-channel MOSFET showing the depletion region along the width of the device.

If we neglect short-channel effects, the gate-controlled bulk charge can be written as

$$Q_B = Q_{B0} + \Delta Q_B \tag{11.31}$$

where Q_B is the total bulk charge, Q_{B0} is the ideal bulk charge, and ΔQ_B is the additional bulk charge at the ends of the channel width. For a uniformly doped p-type semiconductor biased at the threshold inversion point, we may write

$$|Q_{B0}| = eN_a WLx_{dT} \tag{11.32}$$

and

$$\Delta Q_B = eN_a Lx_{dT}(\xi x_{dT}) \tag{11.33}$$

where ξ is a fitting parameter that accounts for the lateral space charge width. The lateral space charge width may not be the same as the vertical width x_{dT} due to the thicker field oxide at the ends, and/or due to the nonuniform semiconductor doping created by an ion implantation. If the ends were a semicircle, then $\xi = \pi/2$.

We may now write

$$|Q_B| = |Q_{B0}| + |\Delta Q_B| = eN_a\, WL\, x_{dT} + eN_a\, Lx_{dT}\,(\xi x_{dT})$$

$$= eN_a WLx_{dT}\left(1 + \frac{\xi x_{dT}}{W}\right) \tag{11.34}$$

The effect of the end space charge regions becomes significant as the width W decreases and the factor $(\xi\, x_{dT})$ becomes a significant fraction of the width W.

The change in threshold voltage due to the additional space charge is

$$\boxed{\Delta V_T = \frac{eN_a x_{dT}}{C_{\text{ox}}}\left(\frac{\xi x_{dT}}{W}\right)} \tag{11.35}$$

The shift in threshold voltage due to a narrow channel is in the positive direction for the n-channel MOSFET. As the width W becomes smaller, the shift in threshold voltage becomes larger.

EXAMPLE 11.4

Objective: Design the channel width that will limit the threshold voltage shift because of narrow channel effects to a specified value.

Consider a silicon n-channel MOSFET with $N_a = 3 \times 10^{16}$ cm^{-3} and $t_{ox} = 20$ nm $= 200$ Å. Let $\xi = \pi/2$. Assume that the threshold voltage shift is to be limited to $\Delta V_T = 0.2$ V.

■ **Solution**
We find

$$C_{ox} = 1.726 \times 10^{-7} \text{ F/cm}^2 \qquad \text{and} \qquad x_{dT} = 0.18\ \mu\text{m}$$

From Equation (11.35), we can express the channel width as

$$W = \frac{eN_a\left(\xi x_{dT}^2\right)}{C_{ox}\,(\Delta V_T)} = \frac{(1.6 \times 10^{-19})(3 \times 10^{16})\left(\frac{\pi}{2}\right)(0.18 \times 10^{-4})^2}{(1.726 \times 10^{-7})(0.2)}$$

$$= 7.08 \times 10^{-5} \text{ cm}$$

or

$$W = 0.708 \ \mu\text{m}$$

■ Comment
We may note that the threshold voltage shift of $\Delta V_T = 0.2$ V occurs at a channel width of $W = 0.708 \ \mu$m, which is approximately four times larger than the induced space charge width x_{dT}.

### ■ EXERCISE PROBLEM
Ex 11.4 Repeat Example 11.4 for $N_a = 10^{16}$ cm^{-3} and $t_{ox} = 8$ nm $= 80$ Å. Determine the channel width such that the threshold voltage shift is limited to $\Delta V_T = 0.1$ V.

(Ans. $W = 0.524 \ \mu$m)

Figure 11.19 shows the threshold voltage as a function of channel width. We can again note that the threshold voltage shift begins to become apparent for channel widths that are large compared to the induced space charge width.

Figure 11.20a, b shows qualitatively the threshold voltage shifts due to short-channel and narrow-channel effects, respectively, in n-channel MOSFETs. The narrow-channel device produces a larger threshold voltage; the short-channel device produces a smaller threshold voltage. For devices exhibiting both short-channel and narrow-channel effects, the two models need to be combined into a three-dimensional volume approximation of the space charge region controlled by the gate.
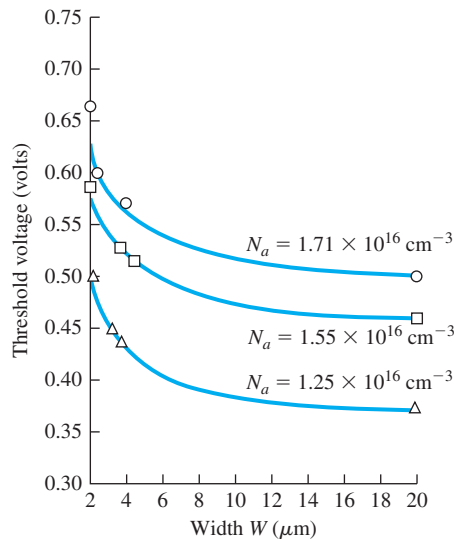
Figure 11.19 | Threshold voltage versus channel width (solid curves, theoretical; points, experimental).
(*From Akers and Sanchez [1].*)
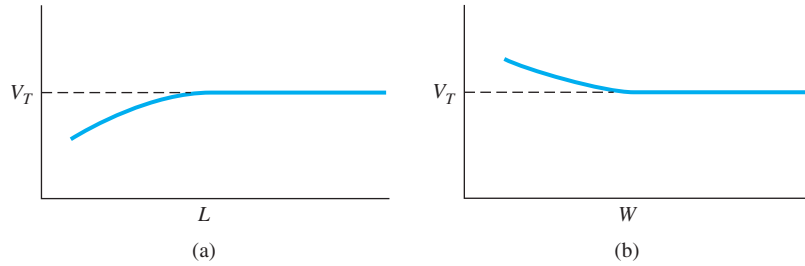
Figure 11.20 | Qualitative variation of threshold voltage (a) with channel length and (b) with channel width.

11.4 | ADDITIONAL ELECTRICAL CHARACTERISTICS

There is a tremendous volume of information on MOSFETs that cannot be included in an introductory text on semiconductor physics and devices. However, two additional topics should be included here: breakdown voltage and threshold adjustment by ion implantation.
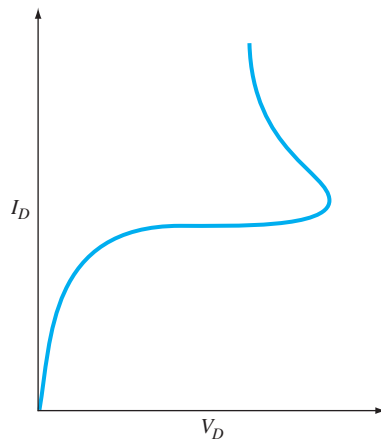
11.4.1 Breakdown Voltage

Several voltage breakdown mechanisms in the MOSFET must be considered, including voltage breakdown across the oxide as well as the various voltage breakdown mechanisms in the semiconductor junctions.

Oxide Breakdown   We have assumed that the oxide is a perfect insulator. However, if the electric field in the oxide becomes large enough, breakdown can occur, which can lead to a catastrophic failure. In silicon dioxide, the electric field at breakdown is on the order of 6×10^6 V/cm. This breakdown field is larger than that in silicon, but the gate oxides are also quite thin. A gate voltage of approximately 30 V would produce breakdown in an oxide with a thickness of 500 Å. However, a safety margin of a factor of 3 is common, so that the maximum safe gate voltage with $t_{\text{ox}} = 500$ Å would be 10 V. A safety margin is necessary since there may be defects in the oxide that lower the breakdown field. Oxide breakdown is normally not a serious problem except in power devices and ultrathin oxide devices. Other oxide degradation problems are discussed later in this chapter.

Avalanche Breakdown   Avalanche breakdown may occur by impact ionization in the space charge region near the drain terminal. We have considered avalanche breakdown in pn junctions in Chapter 7. In an ideal planar one-sided pn junction, breakdown is a function primarily of the doping concentration in the low-doped region of the junction. For the MOSFET, the low-doped region corresponds to the semiconductor substrate. If a p-type substrate doping is $N_a = 3 \times 10^{16}$ cm^{-3}, for example, the pn junction breakdown voltage would be approximately 25 V for a planar junction. However, the n$^+$ drain contact may be a fairly shallow diffused region with
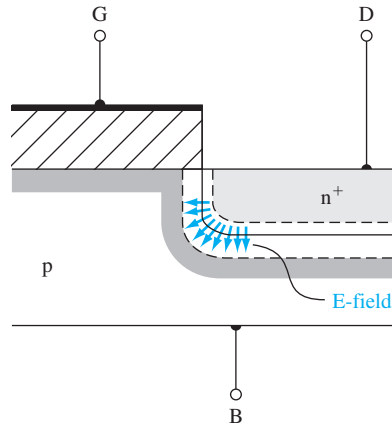
Figure 11.21 | Curvature effect on the electric field in the drain junction.

Figure 11.22 | Current–voltage characteristic showing the snapback breakdown effect.

a large curvature. The electric field in the depletion region tends to be concentrated at the curvature, which lowers the breakdown voltage. This curvature effect is shown in Figure 11.21.

Near Avalanche and Snapback Breakdown[1]   Another breakdown mechanism results in the S-shaped breakdown curve shown in Figure 11.22. This breakdown process is due to second order effects and can be explained with the aid of Figure 11.23. The n-channel enhancement-mode MOSFET geometry in Figure 11.23a shows the

---

[1]This section may be postponed until after the bipolar transistor is considered in Chapter 12.
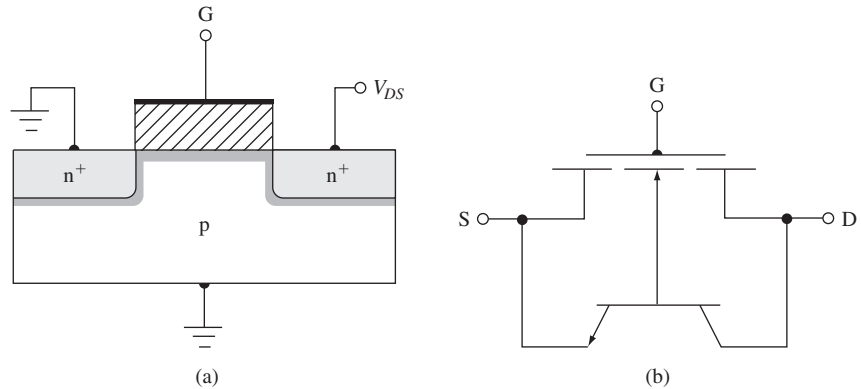
Figure 11.23 | (a) Cross section of the n-channel MOSFET. (b) Equivalent circuit including the parasitic bipolar transistor.

n-type source and drain contacts along with the p-type substrate. The source and body are at ground potential. The n(source)-p(substrate)-n(drain) structure also forms a parasitic bipolar transistor. The equivalent circuit is shown in Figure 11.23b.

Figure 11.24a shows the device when avalanche breakdown is just beginning in the space charge region near the drain. We tend to think of the avalanche breakdown suddenly occurring at a particular voltage. However, avalanche breakdown is a gradual process that starts at low current levels and for electric fields somewhat below the breakdown field. The electrons generated by the avalanche process flow into the drain and contribute to the drain current. The avalanche-generated holes generally flow through the substrate to the body terminal. Since the substrate has a nonzero resistance, a voltage drop is produced as shown. This potential difference drives the source-to-substrate pn junction into forward bias near the source terminal. The source is heavily doped n-type; thus, a large number of electrons can be injected from the source contact into the substrate under forward bias. This process becomes
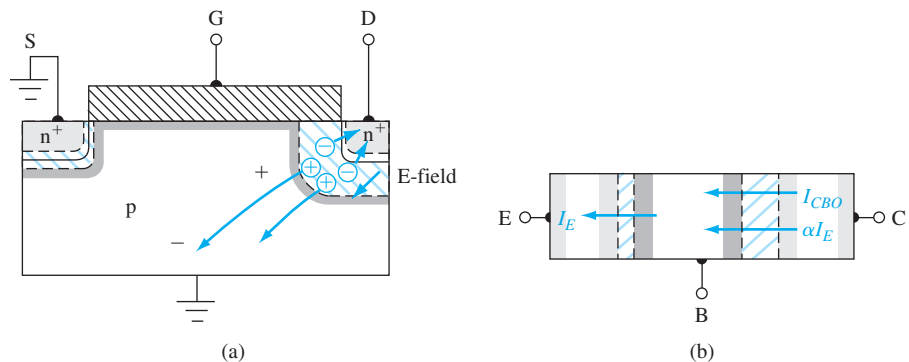
Figure 11.24 | (a) Substrate current–induced voltage drop caused by avalanche multiplication at the drain. (b) Currents in the parasitic bipolar transistor.

severe as the voltage drop in the substrate approaches 0.6 to 0.7 V. A fraction of the injected electrons diffuses across the parasitic base region into the reverse-biased drain space charge region where they also add to the drain current.

The avalanche breakdown process is a function of not only the electric field but the number of carriers involved. The rate of avalanche breakdown increases as the number of carriers in the drain space charge region increases. We now have a regenerative or positive feedback mechanism. Avalanche breakdown near the drain terminal produces the substrate current, which produces the forward-biased source-substrate pn junction voltage. The forward-biased junction injects carriers that can diffuse back to the drain and increase the avalanche process. The positive feedback produces an unstable system.

The snapback or negative resistance portion of the curve shown in Figure 11.22 can now be explained by using the parasitic bipolar transistor. The potential of the base of the bipolar transistor near the emitter (source) is almost floating, since this voltage is determined primarily by the avalanche-generated substrate current rather than an externally applied voltage.

For the open-base bipolar transistor shown in Figure 11.24, we can write

$$I_C = \alpha I_E + I_{CB0} \tag{11.36}$$

where α is the common base current gain and I_{CB0} is the base-collector leakage current. For an open base, $I_C = I_E$, so Equation (11.36) becomes

$$I_C = \alpha I_C + I_{CB0} \tag{11.37}$$

At breakdown, the current in the B–C junction is multiplied by the multiplication factor M, so we have

$$I_C = M(\alpha I_C + I_{CB0}) \tag{11.38}$$

Solving for I_C we obtain

$$I_C = \frac{M I_{CB0}}{1 - \alpha M} \tag{11.39}$$

Breakdown is defined as the condition that produces $I_C \rightarrow \infty$. For a single reverse-biased pn junction, $M \rightarrow \infty$ at breakdown. However, from Equation (11.39), breakdown is now defined to be the condition when $\alpha M \rightarrow 1$ or, for the open-base condition, breakdown occurs when $M \rightarrow 1/\alpha$, which is a much lower multiplication factor than for the simple pn junction.

An empirical relation for the multiplication factor is usually written as

$$M = \frac{1}{1 - (V_{CE}/V_{BD})^m} \tag{11.40}$$

where m is an empirical constant in the range of 3 to 6 and V_{BD} is the junction breakdown voltage.

The common base current gain factor α is a strong function of collector current for small values of collector current. This effect will be discussed in Chapter 12 on bipolar transistors. At low currents, the recombination current in the B-E junction is a significant fraction of the total current so that the common base current gain is small.

As the collector current increases, the value of α increases. As avalanche breakdown begins and I_C is small, particular values of M and V_{CE} are required to produce the condition of $\alpha M = 1$. As the collector current increases, α increases; therefore, smaller values of M and V_{CE} are required to produce the avalanche breakdown condition. The snapback, or negative resistance, breakdown characteristic is then produced.

Only a fraction of the injected electrons from the forward-biased source-substrate junction are collected by the drain terminal. A more exact calculation of the snapback characteristic would necessarily involve taking into account this fraction; thus, the simple model would need to be modified. However, the above discussion qualitatively describes the snapback effect. The snapback effect can be minimized by using a heavily doped substrate that will prevent any significant voltage drop from being developed. A thin epitaxial p-type layer with the proper doping concentration to produce the required threshold voltage can be grown on a heavily doped substrate.

Near Punch-Through Effects   Punch-through is the condition at which the drain-to-substrate space charge region extends completely across the channel region to the source-to-substrate space charge region. In this situation, the barrier between the source and drain is completely eliminated and a very large drain current would exist.

However, the drain current will begin to increase rapidly before the actual punch-through condition is reached. This characteristic is referred to as the near punch-through condition, also known as **D**rain-**I**nduced **B**arrier **L**owering (DIBL). Figure 11.25a shows the ideal energy-band diagram from source to drain for a long n-channel MOSFET for the case when $V_{GS} < V_T$ and when the drain-to-source voltage is relatively small. The large potential barriers prevent significant current between the drain and source. Figure 11.25b shows the energy-band diagram when a relatively large drain voltage V_{DS2} is applied. The space charge region near the drain terminal is beginning to interact with the source space charge region and the potential barrier is being lowered. Since the current is an exponential function of barrier height, the current will increase very rapidly with drain voltage once this near punch-through
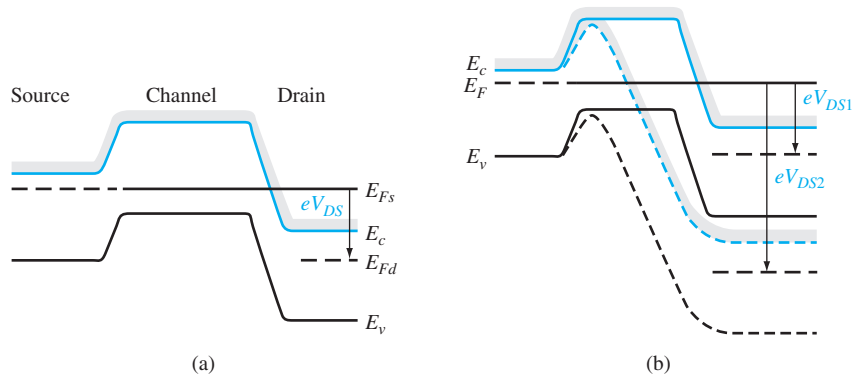
Figure 11.25 | (a) Equipotential plot along the surface of a long-channel MOSFET. (b) Equipotential plot along the surface of a short-channel MOSFET before and after punch-through.
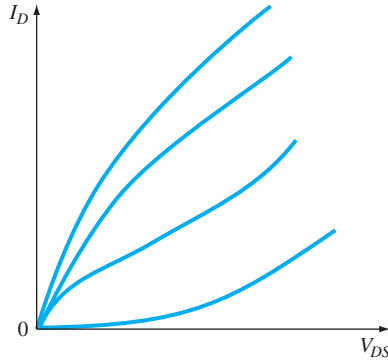
Figure 11.26 | Typical *I–V* character-
istics of a MOSFET exhibiting punch-
through effects.

condition has been reached. Figure 11.26 shows some typical characteristics of a
short-channel device with a near punch-through condition.

EXAMPLE 11.5

Objective: Calculate the theoretical punch-through voltage assuming the abrupt junction
approximation.

Consider an n-channel MOSFET with source and drain doping concentrations of $N_d = 10^{19}$ cm^{-3} and a channel region doping of $N_a = 10^{16}$ cm^{-3}. Assume a channel length of $L = 1.2\ \mu$m, and assume the source and body are at ground potential.

■ **Solution**

The pn junction built-in potential barrier is given by

$$V_{bi} = V_t \ln\left(\frac{N_a N_d}{n_i^2}\right) = (0.0259)\ln\left[\frac{(10^{16})(10^{19})}{(1.5 \times 10^{10})^2}\right] = 0.874\text{ V}$$

The zero-biased source–substrate pn junction width is

$$x_{d0} = \left[\frac{2\epsilon_s V_{bi}}{eN_a}\right]^{1/2} = \left[\frac{2(11.7)(8.85 \times 10^{-14})(0.874)}{(1.6 \times 10^{-19})(10^{16})}\right]^{1/2} = 0.336\ \mu\text{m}$$

The reverse-biased drain-substrate pn junction width is given by

$$x_d = \left[\frac{2\epsilon_s(V_{bi} + V_{DS})}{eN_a}\right]^{1/2}$$

At punch-through, we will have

$$x_{d0} + x_d = L \qquad \text{or} \qquad 0.336 + x_d = 1.2$$

which gives $x_d = 0.864\ \mu$m at the punch-through condition. We can then find

$$V_{bi} + V_{DS} = \frac{x_d^2 eN_a}{2\epsilon_s} = \frac{(0.864 \times 10^{-4})^2(1.6 \times 10^{-19})(10^{16})}{2(11.7)(8.85 \times 10^{-14})}$$

$$= 5.77\text{ V}$$

The punch-through voltage is then found as

$$V_{DS} = 5.77 - 0.874 = 4.9 \text{ V}$$

■ Comment

As the two space charge regions approach punch-through, the abrupt junction approximation is no longer a good assumption. The drain current will begin to increase rapidly before the theoretical punch-through voltage is reached.

■ EXERCISE PROBLEM

Ex 11.5   Repeat Example 11.5 for a substrate doping concentration of $N_a = 3 \times 10^{16} \text{ cm}^{-3}$ and a channel length of $L = 0.8 \; \mu$m.

(Ans. $V_{DS} = 7.52$ V)

For a doping of 10^{16} cm^{-3}, the two space charge regions will begin to interact when the abrupt depletion layers are approximately 0.25 μm apart. The drain voltage at which this near punch-through condition, also known as drain-induced barrier lowering, occurs is significantly less than the ideal punch-through voltage such as calculated in Example 11.5 (see Problem 11.33).

## *11.4.2   The Lightly Doped Drain Transistor

The junction breakdown voltage is a function of the maximum electric field. As the channel length becomes smaller, the bias voltages may not be scaled down accordingly, so the junction electric fields become larger. As the electric field increases, near avalanche breakdown and near punch-through effects become more serious. In addition, as device geometries are scaled down, the parasitic bipolar device becomes more dominant and breakdown effects are enhanced.

One approach that reduces these breakdown effects is to alter the doping profile of the drain contact. The **L**ightly **D**oped **D**rain (LDD) design and doping profiles are shown in Figure 11.27a, the conventional MOSFET and doping profiles are shown in Figure 11.27b for comparison. By introducing the lightly doped region, the peak electric field in the space charge region is reduced and the breakdown effects are minimized. The peak electric field at the drain junction is a function of the semiconductor doping as well as the curvature of the n^+ drain region. Figure 11.28 shows the physical geometries of a conventional n^+ drain contact and an LDD structure superimposed on the same plot. The magnitude of the electric field at the oxide–semiconductor interface in the LDD structure is less than in the conventional structure. The electric field in the conventional device peaks approximately at the metallurgical junction and drops quickly to zero in the drain because no field can exist in the highly conductive n^+ region. On the other hand, the electric field in the LDD device extends across the n-region before dropping to zero at the drain. This effect minimizes breakdown and the hot electron effects, which we discuss in Section 11.5.3.

Two disadvantages of the LDD device are an increase in both fabrication complexity and drain resistance. The added processing steps, however, produce a device with significant improvements in performance. The cross section of the LDD device
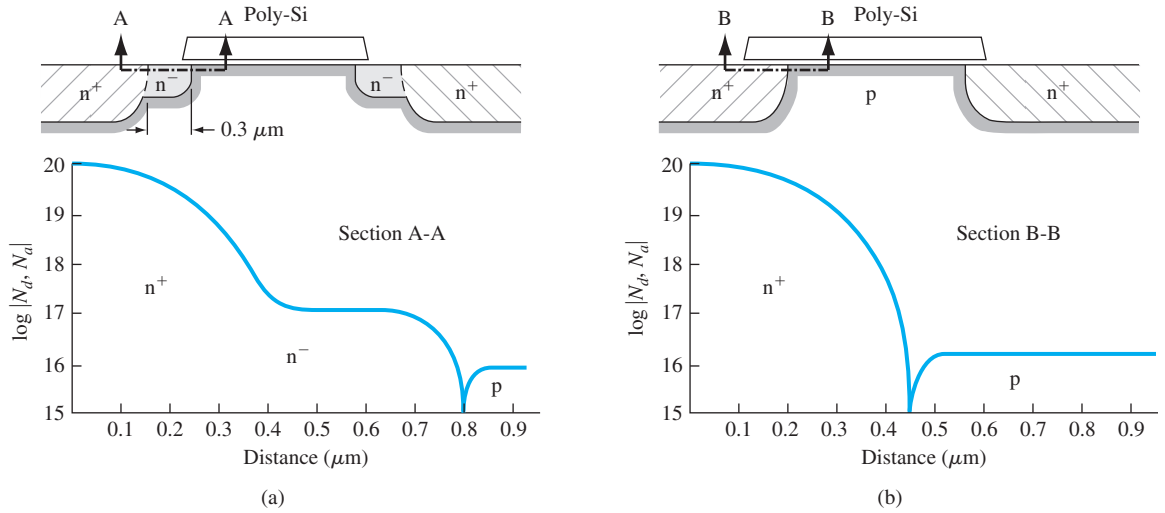
Figure 11.27 | (a) The lightly doped drain (LDD) structure. (b) Conventional structure.
(*From Ogura et al. [12].*)
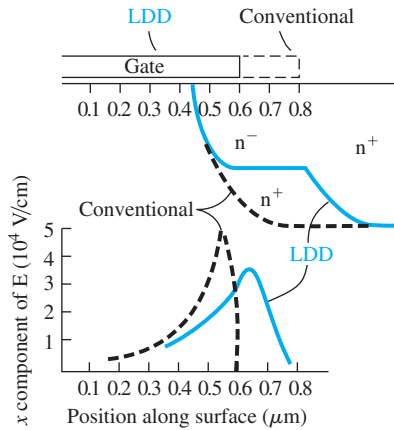
Figure 11.28 | Magnitude of the electric field at the Si–SiO$_2$ interface as a function of distance; $V_{DS} = 10$ V. $V_{SB} = 2$ V, $V_{GS} = V_T$.
(*From Ogura et al. [12].*)

shown in Figure 11.27 indicates a lightly doped n region at the source terminal also. The inclusion of this region does not improve device performance, but does reduce the fabrication complexity as much as possible. The added series resistances increase power dissipation in the device; this must be taken into account in high-power devices.

### 11.4.3 Threshold Adjustment by Ion Implantation

Several factors, such as fixed oxide charge, metal–semiconductor work function difference, oxide thickness, and semiconductor doping, influence the threshold voltage. Although all of these parameters may be fixed in a particular design and fabrication process, the resulting threshold voltage may not be acceptable for all applications. Ion implantation can be used to change and adjust the substrate doping near the oxide–semiconductor surface to provide the desired threshold voltage. In addition, ion implantation is used for more than doping the channel. It is used extensively as a standard part of device fabrication; for example, it is used to form the source and drain regions of the transistor.

To change the doping and thereby change the threshold voltage, a precise, controlled number of either donor or acceptor ions are implanted into the semiconductor near the oxide surface. When an MOS device is biased in either depletion or inversion and when the implanted dopant atoms are within the induced space charge region, then the ionized dopant charge adds to (or subtracts from) the maximum space charge density, which controls the threshold voltage. An implant of acceptor ions into either a p- or n-type substrate will shift the threshold voltage to more positive values, while an implant of donor ions will shift the threshold voltage to more negative values. Ion implantation can be carried out to change a depletion-mode device to enhancement-mode or an enhancement-mode device to depletion-mode, which is an important application of this technology.

As a first approximation, assume that D_I acceptor atoms per cm^2 are implanted into a p-type substrate directly adjacent to the oxide–semiconductor interface as shown in Figure 11.29a. The shift in threshold voltage due to the implant is

$$\Delta V_T = + \frac{eD_I}{C_{ox}} \tag{11.41}$$

If donor atoms were implanted into the p-type substrate, the space charge density would be reduced; thus, the threshold voltage would shift in the negative voltage direction.
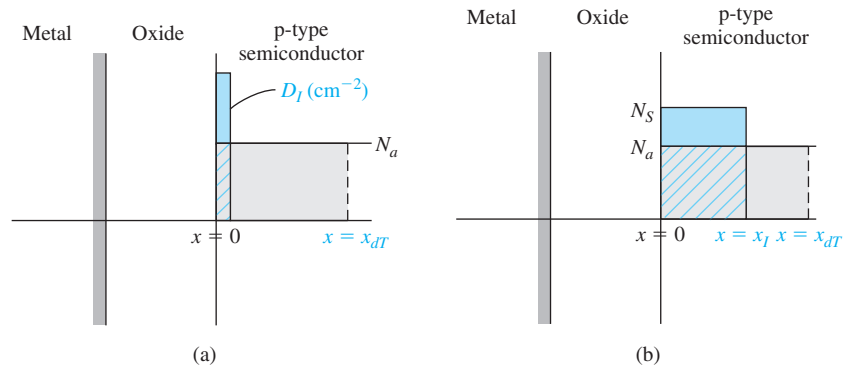
Figure 11.29 | (a) Ion-implanted profile approximated by a delta function. (b) Ion-implanted profile approximated by a step function in which the depth x_i is less than the space charge width x_{dT}.

A second type of implant approximation is the step junction, shown in Figure 11.29b. If the induced space charge width at the threshold inversion point is less than x_I, then the threshold voltage is determined on the basis of a semiconductor with a uniform doping concentration of N_s atoms per cm^3. On the other hand, if the induced space charge width is greater than x_I at the threshold inversion point, then a new expression for x_{dT} must be derived. We can apply Poisson's equation and show that the maximum induced space charge width after the step implant is

$$x_{dT} = \sqrt{\frac{2\epsilon_s}{eN_a} \left[2\phi_{fp} - \frac{ex_I^2}{2\epsilon_s} (N_s - N_a) \right]^{1/2}}$$ (11.42)

The threshold voltage after a step implant for the case when $x_{dT} > x_I$ can be written as

$$\boxed{V_T = V_{T0} + \frac{eD_I}{C_{ox}}}$$ (11.43)

where V_{T0} is the preimplant threshold voltage. The parameter D_I is given by

$$D_I = (N_s - N_a)x_I$$ (11.44)

which is the number per cm^2 of implanted ions. The preimplant threshold voltage is

$$V_{T0} = V_{FB0} + 2\phi_{fp0} + \frac{eN_a x_{dT0}}{C_{ox}}$$ (11.45)

where the subscript 0 indicates the preimplant values.

Objective: Design the ion implant dose required to adjust the threshold voltage to a specified value.

    EXAMPLE 11.6

Consider an n-channel silicon MOSFET with a doping of $N_a = 5 \times 10^{15}$ cm^{-3}, an oxide thickness of $t_{ox} = 18$ nm $= 180$ Å, and an initial flat-band voltage of $V_{FBO} = -1.25$ V. Determine the ion implantation dose such that a threshold voltage of $V_T = +0.4$ V is obtained.

■ **Solution**
We find that

$$\phi_{fpO} = V_t \ln\left(\frac{N_a}{n_i}\right) = (0.0259) \ln\left(\frac{5 \times 10^{15}}{1.5 \times 10^{10}}\right) = 0.3294 \text{ V}$$

$$x_{dTO} = \left[\frac{4\epsilon_s \phi_{fpO}}{eN_a}\right]^{1/2} = \left[\frac{4(11.7)(8.85 \times 10^{-14})(0.3294)}{(1.6 \times 10^{-19})(5 \times 10^{15})}\right]^{1/2}$$

$$= 0.4130 \times 10^{-4} \text{ cm}$$

and

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

The initial pre-implant threshold voltage is

$$V_{TO} = V_{FBO} + 2\phi_{fpO} + \frac{eN_a x_{dTO}}{C_{ox}}$$

$$= -1.25 + 2(0.3294) + \frac{(1.6 \times 10^{-19})(5 \times 10^{15})(0.4130 \times 10^{-4})}{1.9175 \times 10^{-7}}$$

$$= -0.419 \text{ V}$$

The threshold voltage after implant, from Equation (11.43), is

$$V_T = V_{TO} + \frac{eD_I}{C_{ox}}$$

so that

$$+0.40 = -0.419 + \frac{(1.6 \times 10^{-19})D_I}{1.9175 \times 10^{-7}}$$

which gives

$$D_I = 9.815 \times 10^{11} \text{ cm}^{-2}$$

If the uniform step implant extends to a depth of $x_I = 0.15$ μm, for example, then the equivalent acceptor concentration at the surface is

$$N_s - N_a = \frac{D_I}{x_I}$$

or

$$N_s - 5 \times 10^{15} = \frac{9.815 \times 10^{11}}{0.15 \times 10^{-4}}$$

so that

$$N_s = 7.04 \times 10^{16} \text{ cm}^{-3}$$

■ Comment
It is assumed in the above calculation that the induced space charge width in the channel region is greater than the ion implant dept x_I. The calculation satisfies our assumptions.

■ EXERCISE PROBLEM
Ex 11.6   A silicon MOSFET has the following parameters: $N_a = 10^{15}$ cm^{-3}, p$^+$ polysilicon gate with an initial flat-band voltage of $V_{FBO} = +0.95$ V, and an oxide thickness of $t_{ox} = 12$ nm $= 120$ Å. A final threshold voltage of $V_T = +0.40$ V is required. Assume the idealized delta function for the ion implant profile shown in Figure 11.29a. (*a*) What type of ion (acceptor or donor) should be implanted? (*b*) Determine the required ion dose D_I.
[Ans. (*a*) donor ions; (*b*) $D_I = 2.11 \times 10^{12}$ cm^{-2}]

The actual implant dose versus distance is neither a delta function nor a step function; it tends to be a Gaussian-type distribution. The threshold shift due to a nonuniform ion implant density may be defined as the shift in curves of N_{inv} versus V_G, where N_{inv} is the inversion carrier density per cm^2. This shift corresponds to an experimental shift of drain current versus V_G when the transistor is biased in the linear mode. The criteria of the threshold inversion point as $\phi_s = 2\phi_{fp}$ in the implanted devices have an uncertain meaning because of the nonuniform doping in the substrate. The determination of the threshold voltage becomes more complicated and will not be made here.

TEST YOUR UNDERSTANDING

TYU 11.4   Repeat Exercise Problem Ex 11.6 for a final threshold voltage of (*a*) $V_T = +0.25$ V and (*b*) $V_T = -0.25$ V.
[Ans. (*a*) $D_I = 8.02 \times 10^{11}$ cm^{-2}; (*b*) $D_I = 2.03 \times 10^{11}$ cm^{-2}]

*11.5 | RADIATION AND HOT-ELECTRON EFFECTS

We have considered the effects of fixed trapped oxide charge and interface state charge on the capacitance–voltage characteristics of MOS capacitors and on the MOSFET characteristics. These charges can exist because the oxide is essentially a perfect dielectric and a net charge density can exist in a dielectric material. Two processes that generate these charges are ionizing radiation and impact ionization in the drain region of a MOSFET operating near avalanche breakdown.

MOS devices are exposed to ionizing radiation, for example, in communication satellites orbiting through the Van Allen radiation belts. The ionizing radiation can produce additional fixed oxide charge and also additional interface states. In this short discussion of radiation effects in MOSFETs, we are concerned only with the permanent effects that occur in the device characteristics.

Another source can generate oxide charge and interface states: the hot electron effect. Electrons near the drain terminal of a MOSFET operating near avalanche breakdown can have energies that are much larger than the thermal-equilibrium value. These *hot electrons* have energies sufficient to penetrate the oxide–semiconductor barrier.

11.5.1 Radiation-Induced Oxide Charge

Gamma-rays or x-rays incident on semiconductor or oxide materials can interact with valence band electrons. The incident radiation photons can impart enough energy to a valence electron to elevate the electron into the conduction band; an empty state or hole is also produced in the valence band. This process generates electron–hole pairs. These newly generated electrons and holes can move through a material under the influence of an electric field.

Figure 11.30 shows the energy-band diagram of an MOS device with a p-type substrate and a positive gate voltage. The bandgap energy of silicon dioxide is
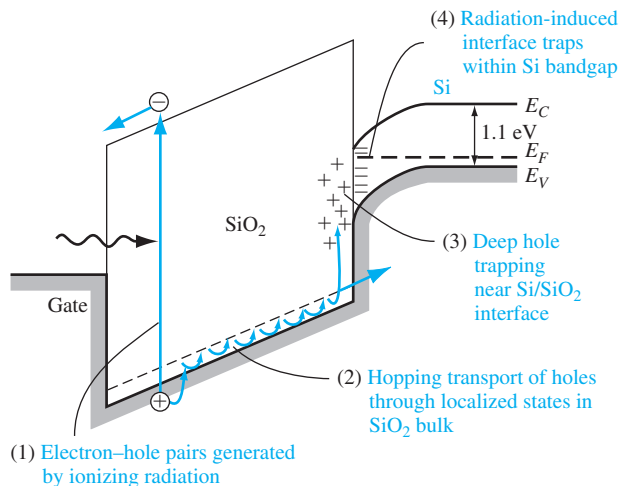
Figure 11.30 | Schematic of ionizing radiation–induced processes in an MOS capacitor with a positive gate bias. (*From Ma and Dressendorfer [7].*)

approximately 9 eV. The figure schematically shows the creation of an electron–hole pair in the oxide by ionizing radiation. The force on the radiation-induced electron is toward the gate and the force on the radiation-induced hole is toward the semiconductor. It has been found that generated electrons in the oxide are fairly mobile, with a mobility value on the order of 20 cm^2/V-s. At high electric fields, the electron velocity in the oxide also saturates at approximately 10^7 cm/s, so that the electron transit time for typical gate oxide thicknesses is on the order of a few picoseconds. For positive gate voltages, the vast majority of radiation-induced electrons flow out through the gate terminal; for this reason, in general these electrons do not play a significant role in the radiation response of MOS devices.

The generated holes, on the other hand, undergo a stochastic hopping transport process through the oxide (shown schematically in Figure 11.30). The hole transport process is dispersive in time and is a function of the electric field, temperature, and oxide thickness. The effective hole mobility in silicon dioxide is typically in the range of 10^{-4} to 10^{-11} cm^2/V-s; thus, holes are relatively immobile when compared with electrons.

When holes reach the silicon–silicon dioxide (Si–SiO_2) interface, a fraction are captured in trapping sites while the remainder flow into the silicon. A net positive radiation-induced charge is then trapped in the oxide due to these captured holes. This trapped charge can last from hours to years. As we have seen, a positive oxide charge causes a negative shift in threshold voltage.

The measured areal hole trap densities are in the range of 10^{12} to 10^{13} cm^{-2} depending upon oxide and device processing. In general, these traps are located within approximately 50 Å of the Si–SiO_2 interface. The hole trap is usually associated with a trivalent silicon defect that has an oxygen vacancy in the SiO_2 structure. The oxygen vacancies are located in a silicon-rich region near the Si–SiO_2 interface.

Since the threshold or flat-band voltage shift is a function of the amount of trapped charge, the voltage shift is a function of applied voltage across the oxide. Figure 11.31 shows the flat-band voltage shift of an MOS capacitor as a function of gate voltage applied during irradiation. For small values of gate voltage, some radiation-generated
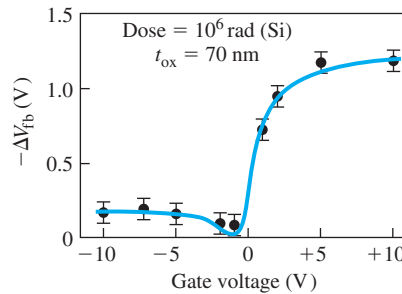
Figure 11.31 | Radiation-induced flat-band voltage shift in an MOS capacitor as a function of applied gate bias during irradiation.
(From Ma and Dressendorfer [7].)

holes and electrons recombine in the oxide. Hence, the amount of charge reaching the Si–SiO$_2$ interface and being trapped is less than for a large positive gate voltage, where essentially all radiation-generated holes reach the interface without recombining with electrons. If the fraction of generated holes that become trapped is relatively constant, then the voltage shift becomes independent of positive gate bias, as shown in the figure. For negative applied gate voltages, the radiation-induced holes move toward the gate terminal. There can be positive charge trapping in the oxide near the gate, but the effect of this trapped charge on the threshold voltage is small.

Objective: Calculate the threshold voltage shift due to radiation-induced oxide charge trapping.

EXAMPLE 11.7

Consider a MOS device with an oxide thickness of $t_{ox} = 25$ nm $= 250$ Å. Assume that a pulse of ionizing radiation creates 10^{18} electron–hole pairs per cm^3 in the oxide. Also assume that the electrons are swept out through the gate terminal with zero recombination, and that 20 percent of the generated holes are trapped at the oxide–semiconductor interface.

■ **Solution**

The areal density of holes generated in the oxide is

$$N_h = (10^{18})(250 \times 10^{-8}) = 2.5 \times 10^{12} \text{ cm}^{-2}$$

The equivalent trapped surface charge is then

$$Q'_{ss} = (2.5 \times 10^{12})(0.2)(1.6 \times 10^{-19}) = 8 \times 10^{-8} \text{ C/cm}^2$$

We find

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} = \frac{(3.9)(8.85 \times 10^{-14})}{250 \times 10^{-8}} = 1.381 \times 10^{-7} \text{ F/cm}^2$$

The threshold voltage shift is then

$$\Delta V_T = -\frac{Q'_{ss}}{C_{ox}} = -\frac{8 \times 10^{-8}}{1.381 \times 10^{-7}} = -0.579 \text{ V}$$

■ **Comment**

As we have seen previously, a positive fixed oxide charge shifts the threshold voltage in the negative voltage direction. The ionizing radiation may shift an enhancement-mode device into the depletion mode.

■ **EXERCISE PROBLEM**

Ex 11.7 Repeat Example 11.7 for a MOS device with an oxide thickness of (a) $t_{ox} = 12$ nm $= 120$ Å and (b) $t_{ox} = 8$ nm $= 80$ Å. (c) What can be said about the shift in threshold voltage as the oxide thickness decreases?

[Ans. (a) $\Delta V_T = -0.134$ V; (b) $\Delta V_T = -0.0593$ V; (c) less shift]

One failure mechanism, therefore, caused by the radiation-induced oxide charge in an n-channel MOSFET in an integrated circuit is a shift from enhancement mode to depletion mode. The device will be turned on rather than off at zero gate voltage; consequently, the circuit function may be disrupted or an excessive power supply current may be generated in the circuit.

The gate voltage in a p-channel MOSFET is normally negative with respect to the substrate. Radiation-generated holes in the oxide are forced to the gate–oxide interface. The trapped charge in this region has less effect on the threshold voltage, so threshold shifts in p-channel MOSFETs are normally smaller if the trap concentrations at the gate–oxide and oxide–semiconductor interfaces are of the same order of magnitude.

11.5.2 Radiation-Induced Interface States

We have considered the effect of interface states on the *C–V* characteristics of an MOS capacitor and on the MOSFET characteristics. The net charge in the interface states of an n-channel MOS device at the threshold inversion point is negative. This negative charge causes a shift in threshold voltage in the positive voltage direction, which is opposite to the shift due to the positive oxide charge. In addition, since the interface states can be charged, they are another source of coulomb interaction with the inversion charge carrier, which means that the inversion carrier mobility is a function of the interface state density through surface-scattering effects. Interface states, then, affect both threshold voltage and carrier mobility.

When MOS devices are exposed to ionizing radiation, additional interface states are generated at the Si–SiO$_2$ interface. The radiation-induced interface states tend to be donor states in the lower half of the bandgap and acceptor states in the upper half. Figure 11.32 shows the threshold voltage in an n-channel and a p-channel MOSFET as a function of ionizing radiation dose. We initially see the negative threshold voltage shift in both devices due to the radiation-induced positive oxide charge. The reversal in threshold shift at the higher dose levels is attributable to the creation of radiation-induced interface states that tend to compensate the radiation-induced positive oxide charge.

In our discussion of subthreshold conduction, we have mentioned that the slope of the ln I_D versus V_{GS} curves in the subthreshold region is a function of the density of interface states. Figure 11.33 shows the subthreshold current at several total ionizing
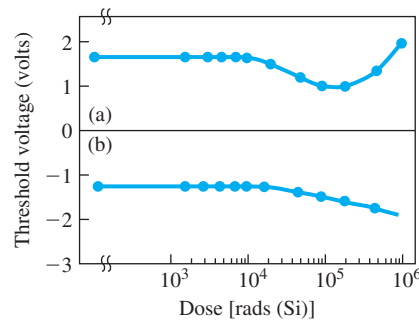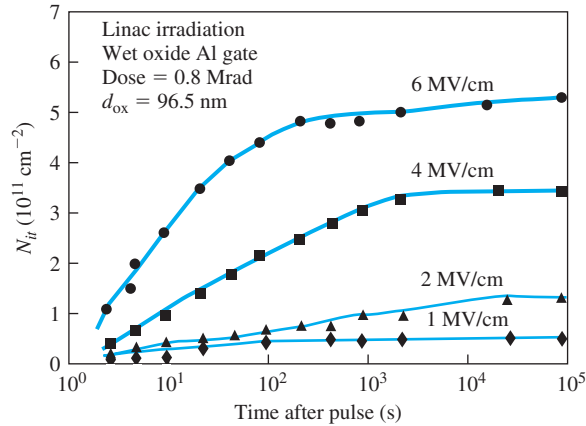

Figure 11.32 | Threshold voltage versus total ionizing radiation dose of (a) an n-channel MOSFET and (b) a p-channel MOSFET.
(From Ma and Dressendorfer [7].)
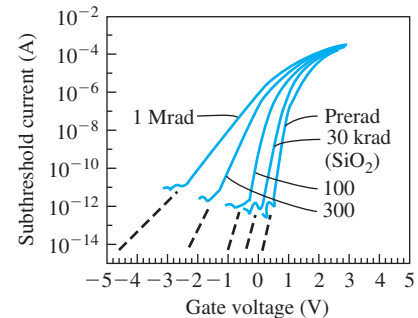
Figure 11.33 | Subthreshold current versus gate voltage of a MOSFET prior to irradiation and at four total radiation dose levels.
(From Ma and Dressendorfer [7].)

Figure 11.34 | Radiation-induced interface state density versus time after a pulse of ionizing radiation for several values of oxide electric field.
(From Ma and Dressendorfer [7].)

dose levels. The change in slope indicates that the density of interface states is increasing with total dose.

The buildup of radiation-induced interface states occurs over a relatively long time period and is a very strong function of the applied electric field in the oxide. Figure 11.34 shows the radiation-induced interface state density versus time for several applied fields. The final interface state density is reached between 100 to 10,000 seconds after a pulse of ionizing radiation. Almost all models for the generation of radiation-induced interface states depend on the transport or trapping of radiation-generated holes near the Si–SiO$_2$ interface. This transport and trapping process is time and field dependent, supporting the time and field dependence of the interface state buildup.

The sensitivity of the Si–SiO$_2$ interface to the buildup of radiation-induced interface states is a strong function of device processing. The interface state buildup in aluminum-gate MOSFETs tends to be smaller than in polysilicon-gate devices. This difference is probably more a result of variations between the two processing technologies than an inherent difference. Hydrogen appears to be important in the radiation-induced interface state buildup—hydrogen tends to passivate dangling silicon bonds at the interface, reducing the preradiation density of interface states. However, devices passivated with hydrogen appear to be more susceptible to the buildup of radiation-induced interface states. The silicon–hydrogen bond at the interface may be broken by the radiation process, which leaves a dangling silicon bond that acts like an interface state trap. These traps at the interface have been identified through electron spin resonance experiments.

Interface states may seriously affect the MOSFET characteristics, which in turn can affect MOSFET circuit performance. Radiation-induced interface states can cause shifts in threshold voltage, affecting circuit performance as we have discussed. A reduction in mobility can affect the speed and output drive capability of a circuit.

11.5.3 Hot-Electron Charging Effects

We have considered breakdown voltage effects in a MOSFET. In particular, as the electric field in the drain junction space charge region increases, electron–hole pairs can be generated by impact ionization. The generated electrons tend to be swept to the drain and generated holes swept into the substrate in an n-channel MOSFET.

Some of the electrons generated in the space charge region are attracted to the oxide due to the electric field induced by a positive gate voltage; this effect is shown in Figure 11.35. These generated electrons have energies far greater than the thermal-equilibrium value and are called hot electrons. If the electrons have energies on the order of 1.5 eV, they may be able to tunnel into the oxide; or in some cases they may be able to overcome the silicon oxide potential barrier and produce a gate current, which may be in the range of femtoamperes (fA) (10^{-15} A) or perhaps picoamperes (pA) (10^{-12} A). A fraction of the electrons traveling through the oxide may be trapped, producing a net negative charge density in the oxide. The probability of electron trapping is usually less than that of hole trapping; but a hot electron–induced gate current may exist over a long period of time, therefore the negative charging effect may build up. The negative oxide charge trapping will cause a local positive shift in the threshold voltage.

The energetic electrons, as they cross the Si–SiO$_2$ interface, can generate additional interface states. The probable cause of interface state generation is due to the breaking up of silicon-hydrogen bonds—a dangling silicon bond is produced, which acts as an interface state. The charge trapping in interface states causes a shift in threshold voltage, additional surface scattering, and reduced mobility. The hot-electron charging effects are continuous processes, so the device degrades over a period of time. This degradation is obviously an undesirable effect and may tend to limit the useful life of the device. We have discussed the lightly doped drain (LDD) structure in Section 11.4.2. The maximum electric field is reduced in this device, decreasing the probability of impact ionization and hot-electron effects.
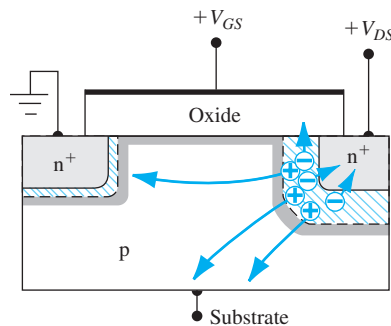
Figure 11.35 | Hot carrier generation, current components, and electron injection into the oxide.

11.6 | SUMMARY

- Advanced concepts of MOSFETs have been discussed in this chapter.
- Subthreshold conduction means that the drain current in a MOSFET is not zero even when the gate-to-source voltage is less than the threshold voltage. In this situation, the transistor is biased in the weak inversion mode and the drain current is dominated by the diffusion rather than the drift mechanism.
- The effective channel length decreases with an increase in drain voltage when the MOSFET is biased in the saturation region, since the depletion region at the drain extends into the channel. This effect is known as channel length modulation.
- The mobility of carriers in the inversion layer is not a constant. As the gate voltage increases, the transverse electric field at the oxide interface increases, causing additional surface scattering. The increased carrier scattering leads to a reduced mobility and a deviation from the ideal current–voltage relation.
- As the channel length decreases, the lateral electric field increases so that carriers flowing through the channel may reach their saturation velocity. In this case, the drain current becomes a linear function of gate-to-source voltage.
- The tendency in MOSFET design is to make devices smaller. The principle of constant-field scaling has been discussed.
- Modification in threshold voltage as device dimensions shrink has been discussed. Because of charge-sharing effects in the substrate, the threshold voltage decreases as channel length decreases and increases as channel width decreases.
- Several voltage breakdown mechanisms have been discussed. These include oxide breakdown, avalanche breakdown, near avalanche or snapback breakdown, and near punch-through effects. These mechanisms are all enhanced as device dimensions shrink. The lightly doped drain transistor tends to minimize the drain breakdown effects.
- Ion implantation can be used as essentially a final process step to change and adjust the substrate doping in the channel region to provide the desired threshold voltage. This process is referred to as threshold voltage adjustment by ion implantation and is used extensively in device fabrication.
- The effect of ionization radiation and hot-electron effects on MOSFET behavior have been briefly discussed.

GLOSSARY OF IMPORTANT TERMS

channel length modulation   The change in effective channel length with drain-to-source voltage when the MOSFET is biased in saturation.

drain-induced barrier lowering   The near punch-through condition in which the potential barrier between the source and channel region in an off transistor is lowered due to a large applied drain voltage.

hot electrons   Electrons with energies far greater than the thermal-equilibrium value caused by acceleration in high electric fields.

lightly doped drain (LDD)   A MOSFET with a lightly doped drain region adjacent to the channel to reduce voltage breakdown effects.

narrow-channel effects   The shift in threshold voltage as the channel width narrows.

near punch-through   The reduction in the potential barrier between source and substrate by the drain-to-substrate voltage, resulting in a rapid increase in drain current.

short-channel effects   The shift in threshold voltage as the channel length becomes smaller.

snapback   The negative resistance effect during breakdown in a MOSFET caused by the variable current gain in a parasitic bipolar transistor.

subthreshold conduction   The process of current conduction in a MOSFET when the transistor is biased below the threshold inversion point.

surface scattering   The process of electric field attraction and coulomb repulsion of carriers at the oxide–semiconductor interface as the carriers drift between source and drain.

threshold adjustment   The process of altering the threshold voltage by changing the semiconductor doping concentration through ion implantation.

CHECKPOINT

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

- Describe the concept and effects of subthreshold conduction.
- Discuss channel length modulation.
- Describe carrier mobility versus gate-to-source voltage and discuss the effects on the current–voltage characteristics of a MOSFET.
- Discuss the effect of velocity saturation on the current–voltage relationship of a MOSFET.
- Define what is meant by constant-field scaling in MOSFET device design, and discuss how device parameters change in constant-field scaling.
- Describe why the threshold voltage changes as the channel length decreases and as the channel width decreases.
- Describe the various voltage breakdown mechanisms in a MOSFET, such as oxide breakdown, avalanche breakdown, snapback breakdown, and near punch-through effects.
- Describe the advantages of the lightly doped drain transistor.
- Discuss the advantages and the process of threshold adjustment by ion implantation.

REVIEW QUESTIONS

1. What is subthreshold conduction? Sketch a drain current versus gate voltage plot that shows the subthreshold current for the transistor biased in the saturation region.
2. What is channel length modulation? Sketch an I–V curve that shows the channel length modulation effect.
3. Why, in general, is the mobility of carriers in the inversion layer not a constant with applied voltage?
4. What is velocity saturation and what is its effect on the I–V relation of a MOSFET?
5. What is constant-field scaling and what parameters in a MOSFET are changed in constant-field scaling?
6. Sketch the space charge region in the channel of a short-channel MOSFET and show the charge-sharing effect. Why does the threshold voltage decrease in a short-channel NMOS device?
7. Sketch the space charge region along the width of an NMOS device. Why does the threshold voltage increase as the channel width of the NMOS device decreases?
8. Sketch I_D versus V_D for an NMOS device, showing the snapback breakdown effect.
9. Sketch the energy bands of an NMOS device between source and drain, showing the near punch-through effect.

10. Sketch the cross section of a lightly doped drain transistor. What are the advantages and disadvantages of this design?

11. What type of ion should be implanted into a MOSFET to increase the threshold voltage? What type of ion should be implanted into a MOSFET to decrease the threshold voltage?

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.)

Section 11.1   Nonideal Effects

11.1 Assume that the subthreshold current of a MOSFET is given by

$$I_D = 10^{-15} \exp\left(\frac{V_{GS}}{(2.1)V_t}\right)$$

over the range $0 \le V_{GS} \le 1$ volt and where the factor 2.1 takes into account the effect of interface states. Assume that 10^6 identical transistors on a chip are all biased at the same V_{GS} and at $V_{DD} = 5$ V. (a) Calculate the total current that must be supplied to the chip at $V_{GS} = 0.5, 0.7,$ and 0.9 V. (b) Calculate the total power dissipated in the chip for the same V_{GS} values.

11.2 The subthreshold current in a MOSFET is given by $I_D = I_S \exp(V_{GS}/nV_t)$. Determine the change in applied V_{GS} for a factor of 10 increase in I_D for (a) $n = 1$, (b) $n = 1.5$, and (c) $n = 2.1$.

11.3 A silicon n-channel MOSFET has a doping concentration of $N_a = 2 \times 10^{16}$ cm^{-3} and a threshold voltage of $V_T = 0.40$ V. Determine the change in channel length, ΔL, for (a) $V_{GS} = 1.0$ V, $V_{DS} = 2.0$ V; (b) $V_{GS} = 1.0$, $V_{DS} = 4.0$ V; (c) $V_{GS} = 2.0$ V, $V_{DS} = 2.0$ V; and (d) $V_{GS} = 2.0$, $V_{DS} = 4.0$.

11.4 Consider the MOSFET described in Problem 11.3. (a) Determine the minimum channel length so that the incremental change ΔL is no more than 10 percent of the original length L for applied voltages of $V_{GS} = 2$ V and $V_{DS} = 3$ V. (b) Repeat part (a) for $V_{DS} = 5$ V.

11.5 A silicon MOSFET has parameters $N_a = 4 \times 10^{16}$ cm^{-3}, $t_{ox} = 12$ nm $= 120$ Å, $Q'_{ss} = 4 \times 10^{10}$ cm^{-2}, and $\phi_{ms} = -0.5$ V. The transistor is biased at $V_{GS} = 1.25$ V and $V_{SB} = 0$. (a) Calculate ΔL for (i) $\Delta V_{DS} = 1$ V, (ii) $\Delta V_{DS} = 2$ V, and (iii) $\Delta V_{DS} = 4$ V. (b) Determine the minimum channel length L such that $\Delta L/L = 0.12$ for $V_{GS} = 1.25$ V and $\Delta V_{DS} = 4$ V.

11.6 The parameters of a silicon MOSFET are $N_a = 3 \times 10^{16}$ cm^{-3}, $V_T = 0.40$ V, $k'_n = 50$ μA/V^2, $L = 0.80$ μm, and $W = 15$ μm. (a) Determine the current I'_D at $V_{GS} = 1.0$ V for (i) $V_{DS} = 2.0$ V and (ii) $V_{DS} = 4.0$ V. (b) Defining the output resistance as $r_o = (\Delta I'_D/\Delta V_{DS})^{-1}$, determine r_o for part (a). (c) Repeat parts (a) and (b) for $V_{GS} = 2.0$ V.

11.7 Consider an n-channel silicon MOSFET. The parameters are $k'_n = 75$ μA/V^2, $W/L = 10$, and $V_T = 0.35$ V. The applied drain-to-source voltage is $V_{DS} = 1.5$ V. (a) For $V_{GS} = 0.8$ V, find (i) the ideal drain current, (ii) the drain current if $\lambda = 0.02$ V^{-1}, and (iii) the output resistance for $\lambda = 0.02$ V^{-1}. (b) Repeat part (a) for $V_{GS} = 1.25$ V.