
Chapter 6

Field-Effect Transistors

OBJECTIVES

1. Derive I-V characteristics of JFETs, MESFETs, and HEMTs
2. Determine MOS C-V behavior and threshold voltage; study leakage characteristics of gate dielectrics
3. Understand MOSFET band diagrams, and linear and saturation behavior
4. Understand “effective” channel mobility, body effect, and subthreshold slopes
5. Analyze second-order effects—DIBL, GIDL, charge sharing, and V_T roll-off

The modern era of semiconductor electronics was ushered in by the invention of the bipolar transistor in 1948 by Bardeen, Brattain, and Shockley at the Bell Telephone Laboratories. This device, along with its field-effect counterpart, has had an enormous impact on virtually every area of modern life. In this chapter we will learn about the operation, applications, and fabrication of the field-effect transistor (FET).

The field-effect transistor comes in several forms. In a *junction FET* (called a *JFET*) the control (*gate*) voltage varies the depletion width of a reverse-biased p-n junction. A similar device results if the junction is replaced by a Schottky barrier (*metal-semiconductor FET*, called a *MESFET*). Alternatively, the metal gate electrode may be separated from the semiconductor by an insulator (*metal-insulator-semiconductor FET*, called a *MISFET*). A common special case of this type uses an oxide layer as the insulator (*MOSFET*).

In Chapter 5 we found that two dominant features of p-n junctions are the injection of minority carriers with forward bias and a variation of the depletion width W with reverse bias. These two p-n junction properties are used in two important types of transistors. The *bipolar junction transistor (BJT)* discussed in Chapter 7 uses the injection of minority carriers across a forward-biased junction, and the junction field-effect transistor discussed in this chapter depends on control of a junction depletion width under reverse bias. The FET is a majority carrier device, and is therefore often called a *unipolar* transistor. The BJT, on the other hand, operates by the injection and collection of *minority* carriers. Since the action of both electrons and holes is important in this device, it is called a *bipolar* transistor. Like its bipolar counterpart, the FET is a three-terminal device in which the current through two terminals is

controlled at the third. Unlike the BJT, however, field-effect devices are controlled by a voltage at the third terminal rather than by a current.

The history of BJTs and FETs is rather interesting. It was the FET that was proposed first in 1930 by Lilienfeld, but he never got it to work because he did not fully appreciate the role of surface defects or surface states. In the process of trying to demonstrate experimentally such a field-effect transistor, Bardeen and Brattain somewhat serendipitously invented the first bipolar transistor, the Ge point contact transistor. This major breakthrough was rapidly followed by Shockley's extension of the concept to the BJT. It was only much later, after the problem of surface states was resolved by growing an oxide insulator on Si, that the first MOSFET was demonstrated in 1960 by Kahng and Atalla. Although the BJT reigned supreme in the early days of semiconductor integrated electronics, it has been gradually supplanted in most applications by the Si MOSFET. The main reason is, unlike BJTs, the various types of FET are characterized by a high *input impedance*, since the control voltage is applied to a reverse-biased junction or Schottky barrier, or across an insulator. These devices are particularly well suited for controlled switching between a conducting state and a nonconducting state, and are therefore useful in digital circuits. They are also suitable for integration of many devices on a single chip, as we shall see in Chapter 9. In fact, millions of MOS transistors are commonly used together in semiconductor memory devices and microprocessors.

6.1 TRANSISTOR OPERATION

We begin this section with a general discussion of amplification and switching, the basic circuit functions performed by transistors. The transistor is a three-terminal device with the important feature that the current through two terminals can be controlled by small changes we make in the current or voltage at the third terminal. This control feature allows us to amplify small a-c signals or to switch the device from an *on* state to an *off* state and back. These two operations, *amplification* and *switching*, are the basis of a host of electronic functions. This section provides a brief introduction to these operations, as a foundation for understanding both bipolar and field-effect transistors.

6.1.1 The Load Line

Consider a two-terminal device that has a nonlinear I - V characteristic, as in Fig. 6-1. We might determine this curve experimentally by measuring the current for various applied voltages, or by using an oscilloscope called a *curve tracer*, which varies I and V repetitively and displays the resulting curve. When such a device is biased with the simple battery-resistor combination shown in the figure, steady state values of I_D and V_D are attained. To find these values we begin by writing a loop equation around the circuit:¹

$$E = i_D R + v_D \quad (6-1)$$

¹We use i_D to symbolize the total current i_D for the d-c value, and i_d for the a-c component. A similar scheme is used for other currents and voltages.

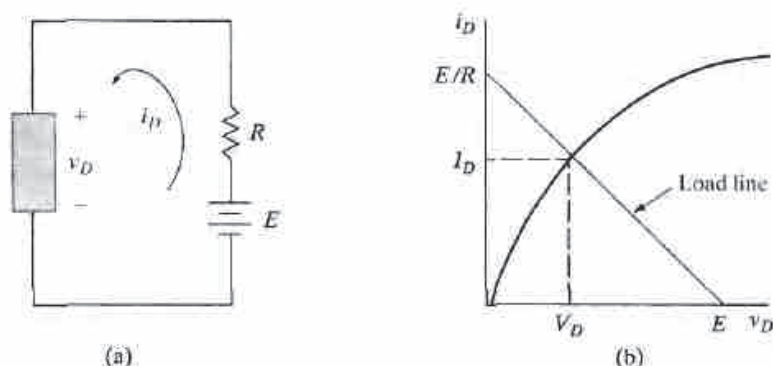


Figure 6-1
A two-terminal
nonlinear device:
(a) biasing
circuit; (b) I - V
characteristic and
load line.

This gives us one equation describing the circuit, but it contains two unknowns (i_D and v_D). Fortunately, we have another equation of the form $i_D = f(v_D)$ in the curve of Fig. 6-1b, giving us two equations with two unknowns. The steady state current and voltage are found by a simultaneous solution of these two equations. However, since one equation is analytical and the other is graphical, we must first put them into the same form. It is easy to make the linear equation (6-1) graphical, so we plot it on Fig. 6-1b to find the simultaneous solution. The end points of the line described by Eq. (6-1) are at E when $i_D = 0$ and at E/R when $v_D = 0$. The two graphs cross at $v_D = V_D$ and $i_D = I_D$, the steady state values of current and voltage for the device with this biasing circuit.

Now let's add a third terminal which somehow controls the I - V characteristic of the device. For example, assume that the device current-voltage curve can be moved up the current axis by increasing the control voltage as in Fig. 6-2b. This results in a family of i_D - v_D curves, depending upon the choice of v_G . We can still write the loop equation (6-1) and draw it on the set of

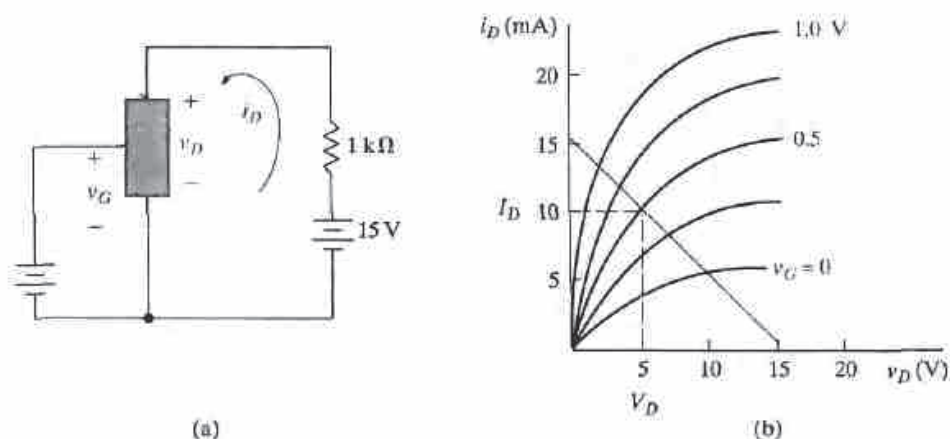


Figure 6-2

A three-terminal nonlinear device that can be controlled by the voltage at the third terminal v_G : (a) biasing circuit; (b) I - V characteristic and load line. If $V_G = 0.5$ V, the d-c values of I_D and V_D are as shown by the dashed lines.

curves, but now the simultaneous solution depends on the value of v_G . In the example of Fig. 6-2, V_G is 0.5 V and the d-c values of I_D and V_D are found at the intersection to be 10 mA and 5 V, respectively. Whatever the value of the control voltage v_G at the third terminal, values of I_D and V_D are obtained from points along the line representing Eq. (6-1). This is called the *load line*.

6.1.2 Amplification and Switching

If an a-c source is added to the control voltage, we can achieve large variations in i_D by making small changes in v_G . For example, as v_G varies about its d-c value by 0.25 V in Fig. 6-2, v_d varies about its d-c value V_D by 2 V. Thus the amplification of the a-c signal is $2/0.25 = 8$. If the curves for equal changes in v_G are equally spaced on the i_D axis, a faithful amplified version of the small control signal can be obtained. This type of voltage-controlled amplification is typical of field-effect transistors. For bipolar transistors, a small control current is used to achieve large changes in the device current, achieving current amplification.

Another important circuit function of transistors is the controlled switching of the device off and on. In the example of Fig. 6-2, we can switch from the bottom of the load line ($i_D = 0$) to almost the top ($i_D \approx E/R$) by appropriate changes in v_G . This type of switching with control at a third terminal is particularly useful in digital circuits.

6.2 THE JUNCTION FET

In a *junction FET (JFET)* the voltage-variable depletion region width of a junction is used to control the effective cross-sectional area of a conducting channel. In the device of Fig. 6-3, the current I_D flows through an n-type channel between two p^+ regions. A reverse bias between these p^+ regions and the channel causes the depletion regions to intrude into the n material, and therefore the effective width of the channel can be restricted. Since the resistivity of the channel region is fixed by its doping, the channel resistance varies with changes in the effective cross-sectional area. By analogy, the variable depletion regions serve as the two doors of a gate, which open and close on the conducting channel.

In Fig. 6-3 electrons in the n-type channel drift from left to right, opposite to current flow. The end of the channel from which electrons flow is called the *source*, and the end toward which they flow is called the *drain*. The p^+ regions are called *gates*. If the channel were p-type, holes would flow from the source to the drain, in the same direction as the current flow, and the gate regions would be n^+ . It is common practice to connect the two gate regions electrically; therefore, the voltage V_G refers to the potential from each gate region G to the source S . Since the conductivity of the heavily doped p^+ regions is high, we can assume that the potential is uniform throughout each gate. In the lightly doped channel material, however, the potential varies with position (Fig. 6-3b). If the channel of Fig. 6-3 is considered as a distributed resistor carrying a current I_D

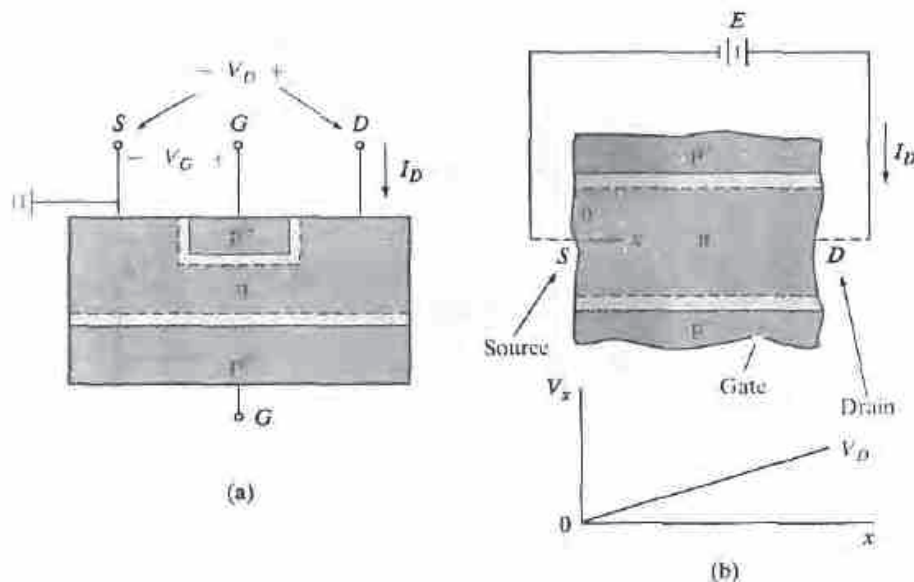


Figure 6-3

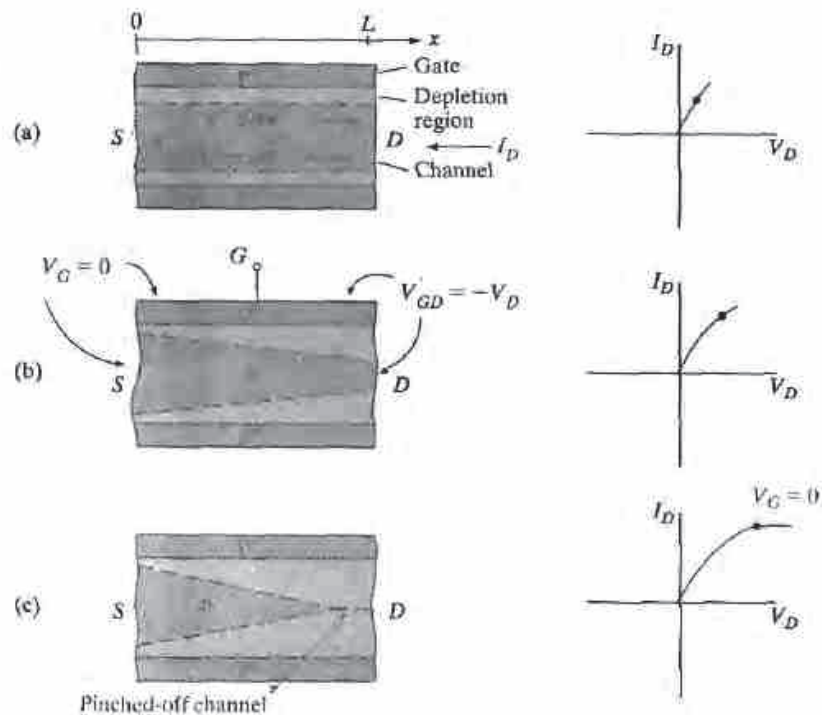
Simplified cross-sectional view of a junction FET: (a) transistor geometry; (b) detail of the channel and voltage variation along the channel with $V_G = 0$ and small I_D .

it is clear that the voltage from the drain end of the channel D to the source electrode S must be greater than the voltage from a point near the source end to S . For low values of current we can assume a linear variation of voltage V_x in the channel, varying from V_D at the drain end to zero at the source end (Fig. 6-3b).

6.2.1 Pinch-off and Saturation

In Figure 6-4 we consider the channel in a simplified way by neglecting voltage drops between the source and drain electrodes and the respective ends of the channel. For example, we assume that the potential at the drain end of the channel is the same as the potential at the electrode D . This is a good approximation if the source and drain regions are relatively large, so that there is little resistance between the ends of the channel and the electrodes. In Fig. 6-4 the gates are short circuited to the source ($V_G = 0$), such that the potential at $x = 0$ is the same as the potential everywhere in the gate regions. For very small currents, the widths of the depletion regions are close to the equilibrium values (Fig. 6-4a). As the current I_D is increased, however, it becomes important that V_x is large near the drain end and small near the source end of the channel. Since the reverse bias across each point in the gate-to-channel junction is simply V_x when V_G is zero, we can estimate the shape of the depletion regions as in Fig. 6-4b. The reverse bias is relatively large near the drain

Figure 6-4
Depletion regions
in the channel of
a JFET with zero
gate bias for sev-
eral values of V_D :
(a) linear range;
(b) near pinch-off;
(c) beyond pinch-
off.



($V_{GD} = -V_D$) and decreases toward zero near the source. As a result, the depletion region intrudes into the channel near the drain, and the effective channel area is constricted.

Since the resistance of the constricted channel is higher, the I - V plot for the channel begins to depart from the straight line that was valid at low current levels. As the voltage V_D and current I_D are increased still further, the channel region near the drain becomes more constricted by the depletion regions and the channel resistance continues to increase. As V_D is increased, there must be some bias voltage at which the depletion regions meet near the drain and essentially *pinch off* the channel (Fig. 6-4c). When this happens, the current I_D cannot increase significantly with further increase in V_D . Beyond pinch-off the current is *saturated* approximately at its value at pinch-off.² Once electrons from the channel enter the electric field of the depletion region, they are swept through and ultimately flow to the positive drain contact. After the current saturates beyond pinch-off, the differential channel resistance dV_D/dI_D becomes very high. To a good approximation, we can

²Saturation is used by device engineers in more different contexts than any other word is used. We have discussed velocity saturation, the reverse saturation current of a junction, and, now, the saturation of FET characteristics. In Chapter 7, we will discuss the saturation of a BJT. The student has probably also reached saturation by now in trying to absorb these various meanings!

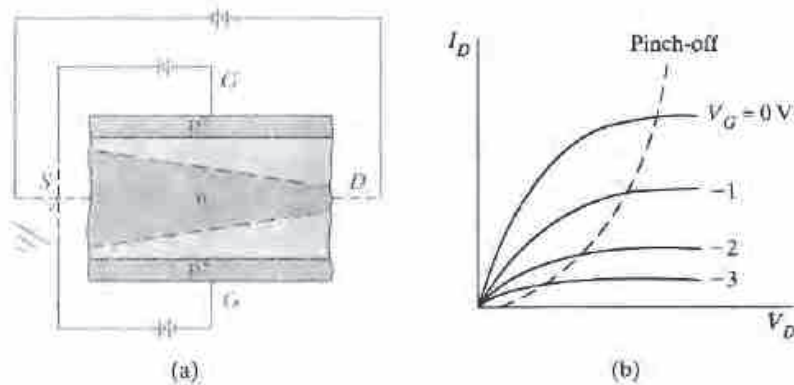


Figure 6-5
Effects of a negative gate bias: (a) increase of depletion region widths with V_G negative; (b) family of current-voltage curves for the channels as V_G is varied.

calculate the current at the critical pinch-off voltage and assume there is no further increase in current as V_D is increased.

6.2.2 Gate Control

The effect of a negative gate bias $-V_G$ is to increase the resistance of the channel and induce pinch-off at a lower value of current (Fig. 6-5). Since the depletion regions are larger with V_G negative, the effective channel width is smaller and its resistance is higher in the low-current range of the characteristic. Therefore, the slopes of the I_D vs. V_D curves below pinch-off become smaller as the gate voltage is made more negative (Fig. 6-5b). The pinch-off condition is reached at a lower drain-to-source voltage, and the saturation current is lower than for the case of zero gate bias. As V_G is varied, a family of curves is obtained for the I - V characteristic of the channel, as in Fig. 6-5b.

Beyond the pinch-off voltage the drain current I_D is controlled by V_G . By varying the gate bias we can obtain amplification of an a-c signal. Since the input control voltage V_G appears across the reverse-biased gate junctions, the input impedance of the device is high.

We can calculate the pinch-off voltage rather simply by representing the channel in the approximate form of Fig. 6-6. If the channel is symmetrical and the effects of the gates are the same in each half of the channel region, we can restrict our attention to the channel half-width $h(x)$, measured from the center line ($y = L$). The metallurgical half-width of the channel (i.e., neglecting the depletion region) is a . We can find the pinch-off voltage by calculating the reverse bias between the n channel and the p^+ gate at the drain end of the channel ($x = L$). For simplicity we shall assume that the channel width at the drain decreases uniformly as the reverse bias increases to pinch-off. If the reverse bias between the gate and the drain is $-V_{GD}$, the width of the depletion region at $x = L$ can be found from Eq. (5-57):

$$W(x = L) = \left[\frac{2\epsilon(-V_{GD})}{qN_d} \right]^{1/2} \quad (V_{GD} \text{ negative}) \quad (6-2)$$

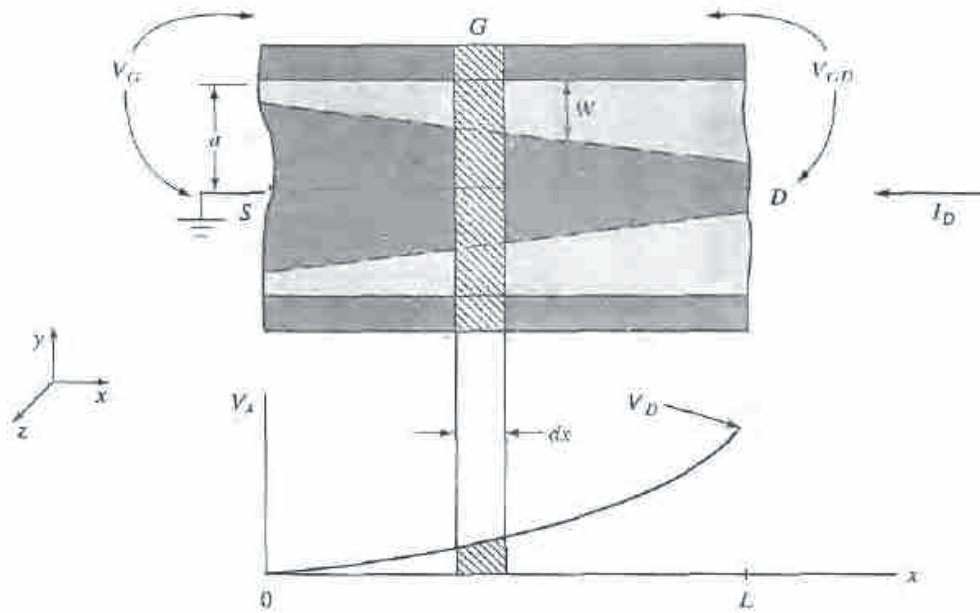


Figure 6-6

Simplified diagram of the channel with definitions of dimensions and differential volume for calculations.

In this expression we assume the equilibrium contact potential V_0 is negligible compared with V_{GD} and the depletion region extends primarily into the channel for the p^+n junction. Including V_0 is left for Prob. 6.1.

Pinch-off occurs at the drain end of the channel when

$$h(x=L) = a - W(x=L) = 0 \quad (6-3)$$

that is, when $W(x=L) = a$. If we define the value of $-V_{GD}$ at pinch-off as V_p , we have

$$\left[\frac{2\epsilon V_p}{qN_d} \right]^{1/2} = a$$

$$\boxed{V_p = \frac{qa^2N_d}{2\epsilon}} \quad (6-4)$$

The pinch-off voltage V_p is a positive number; its relation to V_D and V_G is

$$V_p = -V_{GD}(\text{pinch-off}) = -V_G + V_D \quad (6-5)$$

where V_G is zero or negative for proper device operation. A forward bias on the gate would cause hole injection from the p^+ regions into the channel, eliminating the field-effect control of the device. From Eq. (6-5) it is clear that

pinch-off results from a combination of gate-to-source voltage and drain-to-source voltage. Pinch-off is reached at a lower value of V_D (and therefore a lower I_D) when a negative gate bias is applied, in agreement with Fig. 6-5b.

6.2.3 Current-Voltage Characteristics

Calculation of the exact channel current is complicated, although the mathematics is relatively straightforward below pinch-off. The approach we shall take is to find the expression for I_D just at pinch-off, and then assume the saturation current beyond pinch-off remains fairly constant at this value.

In the coordinate system defined in Fig. 6-6, the center of the channel at the source end is taken as the origin. The length of the channel in the x -direction is L , and the depth of channel in the z -direction is Z . We shall call the resistivity of the n -type channel material ρ (valid only in the neutral n material, outside the depletion regions). If we consider the differential volume of neutral channel material $Z2h(x)dx$, the resistance of the volume element is $\rho dx/Z2h(x)$ [see Eq. (3-44)]. Since the current does not change with distance along the channel, I_D is related to the differential voltage change in the element dV_x by the conductance of the element:

$$I_D = \frac{Z2h(x)}{\rho} \frac{dV_x}{dx} \quad (6-6)$$

The term $2h(x)$ is the channel width at x .

The half-width of the channel at point x depends on the local reverse bias between gate and channel $-V_{Gx}$:

$$h(x) = a - W(x) = a - \left[\frac{2\epsilon(-V_{Gx})}{qN_d} \right]^{1/2} = a \left[1 - \left(\frac{V_x - V_G}{V_P} \right)^{1/2} \right] \quad (6-7)$$

since $V_{Gx} = V_G - V_x$ and $V_P = qa^2N_d/2\epsilon$. Implicit in Eq. (6-7) is the assumption that the expression for $W(x)$ can be obtained by a simple extension of Eq. (6-2) to point x in the channel. This is called the *gradual channel approximation*; it is valid if $h(x)$ does not vary abruptly in any element dx .

The voltage V_{Gx} will be negative since the gate voltage V_G is chosen zero or negative for proper operation. Substituting Eq. (6-7) into Eq. (6-6), we have

$$\frac{2Za}{\rho} \left[1 - \left(\frac{V_x - V_G}{V_P} \right)^{1/2} \right] dV_x = I_D dx \quad (6-8)$$

We can solve this equation to obtain

$$I_D = G_0 V_P \left[\frac{V_D}{V_P} + \frac{2}{3} \left(\frac{-V_G}{V_P} \right)^{3/2} - \frac{2}{3} \left(\frac{V_D - V_G}{V_P} \right)^{3/2} \right] \quad (6-9)$$

where V_G is negative and $G_0 \equiv 2aZ/pL$ is the conductance of the channel for negligible $W(x)$, i.e., with no gate voltage and low values of I_D . This equation is valid only up to pinch-off, where $V_D - V_G = V_p$. If we assume the saturation current remains essentially constant at its value at pinch-off, we have

$$\begin{aligned} I_D(\text{sat.}) &= G_0 V_p \left[\frac{V_D}{V_p} + \frac{2}{3} \left(-\frac{V_G}{V_p} \right)^{3/2} - \frac{2}{3} \right] \\ &= G_0 V_p \left[\frac{V_G}{V_p} + \frac{2}{3} \left(-\frac{V_G}{V_p} \right)^{3/2} + \frac{1}{3} \right] \end{aligned} \quad (6-10)$$

where

$$\frac{V_D}{V_p} = 1 + \frac{V_G}{V_p}$$

The resulting family of I - V curves for the channel agrees with the results we predicted qualitatively (Fig. 6-5b). The saturation current is greatest when V_G is zero and becomes smaller as V_G is made negative.

We can represent the device biased in the saturation region by an equivalent circuit where changes in drain current are related to gate voltage changes by

$$g_m(\text{sat.}) = \frac{\partial I_D(\text{sat.})}{\partial V_G} = G_0 \left[1 - \left(-\frac{V_G}{V_p} \right)^{1/2} \right] \quad (6-11)$$

The quantity g_m is the *mutual transconductance*, with units (A/V) called siemens (S), sometimes called mhos. As a figure of merit for FET devices it is common to describe the transconductance per unit channel width Z . This quantity g_m/Z is usually given in units of millisiemens per millimeter.

It is found experimentally that a square-law characteristic closely approximates the drain current in saturation:

$$I_D(\text{sat.}) \approx I_{DSS} \left(1 + \frac{V_G}{V_p} \right)^2, \quad (V_G \text{ negative}) \quad (6-12)$$

where I_{DSS} is the saturated drain current with $V_G = 0$.

The appearance of a constant value of channel resistivity (in the G_0 term) in Eqs. (6-9)–(6-11) implies that the electron mobility is constant. As mentioned in Section 3.4.4, electron velocity saturation at high fields may make this assumption invalid. This is particularly likely for very short channels, where even moderate drain voltage can result in a high field along the channel. Another departure from the ideal model results from the fact that the effective channel length decreases as the drain voltage is increased beyond pinch-off, as Fig. 6-4c suggests. In short-channel devices this effect can cause I_D to increase beyond pinch-off, since L appears in the denominator

of Eq. (6-10), in G_0 . Therefore, the assumption of constant saturation current is not valid for very short-channel devices.

The depletion of the channel discussed above for a JFET can be accomplished by the use of a reverse-biased Schottky barrier instead of a p-n junction. The resulting device is called a MESFET, indicating that a metal-semiconductor junction is used. This device is useful in high-speed digital or microwave circuits, where the simplicity of Schottky barriers allows fabrication to close geometrical tolerances. There are particular speed advantages for MESFET devices in III-V compounds such as GaAs or InP, which have higher mobilities and carrier drift velocities than Si.

6.3 THE METAL- SEMICONDUCTOR FET

6.3.1 The GaAs MESFET

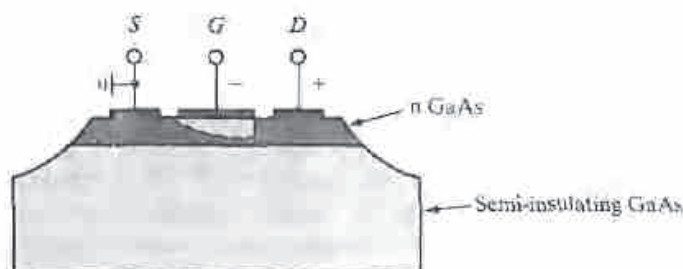
Figure 6-7 shows schematically a simple MESFET in GaAs. The substrate is undoped or doped with chromium, which has an energy level near the center of the GaAs band gap. In either case the Fermi level is near the center of the gap, resulting in very high resistivity material ($\sim 10^8 \Omega\text{-cm}$), generally called *semi-insulating* GaAs. On this nonconducting substrate a thin layer of lightly-doped n-type GaAs is grown epitaxially, to form the channel region of the FET.³ The photolithographic processing consists of defining patterns in the metal layers for source and drain ohmic contacts (e.g., Au-Ge) and for the Schottky barrier gate (e.g., Al). By reverse biasing the Schottky gate, the channel can be depleted to the semi-insulating substrate, and the resulting I - V characteristics are similar to the JFET device.

By using GaAs instead of Si, a higher electron mobility is available (see Appendix III), and furthermore GaAs can be operated at higher temperatures (and therefore higher power levels). Since no diffusions are involved in Fig. 6-7, close geometrical tolerances can be achieved and the MESFET can be made very small. Gate lengths $L \leq 0.25 \mu\text{m}$ are common in these devices. This is important at high frequencies, since drift time and capacitances must be kept to a minimum.

It is possible to avoid the epitaxial growth of the n-type layer and the etched isolation in Fig. 6-7 by using ion implantation. Starting with a semi-insulating GaAs substrate, a thin n-type layer at the surface of each transistor region can be formed by implanting Si or a column VI donor impurity such as Se. This implantation requires an anneal to remove the radiation damage, but the epitaxial growth step is eliminated. In either the fully implanted device or the epitaxial device of Fig. 6-7, the source and drain contacts may be improved by further n^+ implantation in these regions. Because

³In many cases, a high-resistivity GaAs epitaxial layer (called a *buffer layer*) is grown between the two layers shown in Fig. 6-7.

Figure 6-7
GaAs MESFET formed on an n-type GaAs layer grown epitaxially on a semi-insulating substrate. Common metals for the Schottky gate in GaAs are Al or alloys of Ti, W, and Au. The ohmic source and drain contacts may be an alloy of Au and Ge. In this example the device is isolated from others on the same chip by etching through the n region to the semi-insulating substrate.



of the relative simplicity of implanted GaAs MESFETs and the isolation between devices provided by the semi-insulating substrate, these structures are commonly used in GaAs integrated circuits.

6.3.2 The High Electron Mobility Transistor (HEMT)

Since the metal-semiconductor field effect transistor (MESFET) is compatible with the use of III-V compounds, it is possible to exploit the band gap engineering available with heterojunctions in these materials. In order to maintain high transconductance in a MESFET, the channel conductivity must be as high as possible. Obviously, the conductivity can be increased by increasing the doping in the channel and thus the carrier concentration. However, increased doping also causes increased scattering by the ionized impurities, which leads to a degradation of mobility (see Fig. 3-23). What is needed is a way of creating a high electron concentration in the channel of a MESFET by some means other than doping. A clever approach to this requirement is to grow a thin undoped well (e.g., GaAs) bounded by wider band gap, doped barriers (e.g., AlGaAs). This configuration, called *modulation doping*, results in conductive GaAs when electrons from the doped AlGaAs barriers fall into the well and become trapped there, as shown in Fig. 6-8a. Since the donors are in the AlGaAs rather than the GaAs, there is no impurity scattering of electrons in the well. If a MESFET is constructed with the channel along the GaAs well (perpendicular to the page in Fig. 6-8), we can take advantage of this reduced scattering and resulting higher mobility. The effect is especially strong at low temperature where lattice (phonon) scattering is also low. This device is called a *modulation doped field-effect transistor (MODFET)* and is also called a *high electron mobility transistor (HEMT)*.

In Fig. 6-8a we have left out the band-bending expected at the AlGaAs/GaAs interfaces. Based upon the discussion in Section 5.8, we expect the electrons to accumulate at the corners of the well due to band-bending at the heterojunction. In fact, only one heterojunction is required to trap electrons, as shown in Fig. 6-8b. Generally, the donors in the AlGaAs layer are purposely separated from the interface by $\sim 100 \text{ \AA}$. Using this configuration, we can achieve a high electron concentration in the channel while retaining high mobility, since the GaAs channel region is spatially separated from the ionized impurities which provide the free carriers.

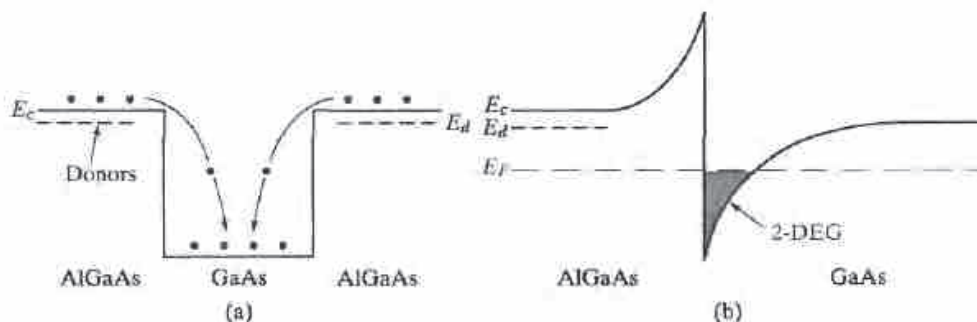


Figure 6-8

(a) Simplified view of modulation doping, showing only the conduction band. Electrons in the donor-doped AlGaAs fall into the GaAs potential well and become trapped. As a result, the undoped GaAs becomes n-type, without the scattering by ionized donors which is typical of bulk n-type material. (b) Use of a single AlGaAs/GaAs heterojunction to trap electrons in the undoped GaAs. The thin sheet of charge due to free electrons at the interface forms a two-dimensional electron gas (2-DEG), which can be exploited in HEMT devices.

In Fig. 6-8b, mobile electrons generated by the donors in the AlGaAs diffuse into the small band gap GaAs layer, and they are prevented from returning to the AlGaAs by the potential barrier at the AlGaAs/GaAs interface. The electrons in the (almost) triangular well form a two-dimensional electron gas (sometimes abbreviated 2-DEG). Sheet carrier densities as high as 10^{12} cm^{-2} can be obtained at a single interface such as that shown in Fig. 6-8b. Ionized impurity scattering is greatly reduced simply by separating the electrons from the donors. Also, screening effects due to the extremely high density of the two-dimensional electron gas can reduce ionized impurity scattering further. In properly designed structures, the electron transport approaches that of bulk GaAs with no impurities, so that mobility is limited by lattice scattering. As a result, mobilities above $250,000 \text{ cm}^2/\text{V-s}$ at 77 K and $2,000,000 \text{ cm}^2/\text{V-s}$ at 4 K can be achieved.

The advantages of a HEMT are its ability to locate a large electron density ($\sim 10^{12} \text{ cm}^{-2}$) in a very thin layer ($< 100 \text{ \AA}$ thick) very close to the gate while simultaneously eliminating ionized impurity scattering. The AlGaAs layer in a HEMT is fully depleted under normal operating conditions, and since the electrons are confined to the heterojunction, device behavior closely resembles that of a MOSFET. The advantages of the HEMT over the Si MOSFET are the higher mobility and maximum electron velocity in GaAs compared with Si, and the smoother interfaces possible with an AlGaAs/GaAs heterojunction compared with the Si/SiO₂ interface. The high performance of the HEMT translates into an extremely high cutoff frequency, and devices with fast access times.

Although we have discussed the HEMT in terms of the AlGaAs/GaAs heterojunction, other materials are also promising, such as the InGaAsP/InP

system. A motivation for avoiding $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is the presence of a deep-level defect called the DX center for $x > 0.2$, which traps electrons and impairs the HEMT operation. Since very thin layers are involved, materials with slight lattice mismatch can be grown to form *pseudomorphic* HEMTs. An example of such a system is the use of a thin layer of InGaAs grown pseudomorphically on GaAs, followed by AlGaAs. An advantage of this system is that a useful band discontinuity can be achieved using AlGaAs of low enough Al composition to avoid the DX center problem.

The HEMT, or MODFET, is also referred to as a *two-dimensional electron gas FET* (2-DEG FET, or *TEGFET*) to emphasize the fact that conduction along the channel occurs in a thin sheet of charge. The device has also been called a *separately doped FET* (SEDFET), to emphasize the fact that the doping occurs in a separate region from the channel.

6.3.3 Short Channel Effects

As mentioned in Section 6.2.3, a variety of modifications to the simple theory of JFET and MESFET operation must be made when the channel length is small (typically $< 1 \mu\text{m}$). In the past, these short-channel effects would be considered unusual, but now it is common to encounter FET devices in which these effects dominate the I - V characteristics. For example, high-field effects occur when 1 V appears across a channel length of $1 \mu\text{m}$ (10^{-4} cm), giving an electric field of 10 kV/cm .

A simple piecewise-linear approximation to the velocity-field curve assumes a constant mobility (linear) dependence up to some critical field $\mathcal{E}_c$ and a constant saturation velocity v_s for higher fields. For Si a better approximation is

$$v_d = \frac{\mu \mathcal{E}}{1 + \mu \mathcal{E}/v_s} \quad (6-13)$$

where μ is the low-field mobility. These two approximations are shown in Fig. 6-9a. If we assume that the electrons passing through the channel drift with a constant saturation velocity v_s , the current takes a simple form

$$I_D = qn v_s A = qN_d V_s Z h \quad (6-14)$$

where h is a slow function of V_G . In this case the saturated current follows the velocity saturation, and does not require a true pinch-off in the sense of depletion regions meeting at some point in the channel. In the saturated velocity case, the transconductance g_m is essentially constant, in contrast with the constant mobility case described by Eq. (6-11). As shown in Fig. 6-9b, the $I_D - V_D$ curves are more evenly spaced if constant saturation velocity dominates, compared with the V_G -dependent spacing shown in Fig. 6-5b for the long-channel constant-mobility case.

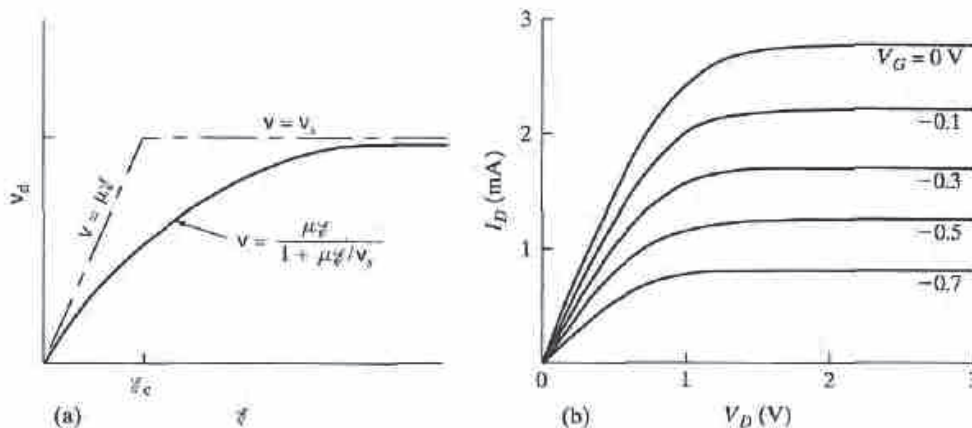


Figure 6-9

Effects of electron velocity saturation at high electric fields: (a) approximations to the saturation of drift velocity with increasing field; (b) drain current-voltage characteristics for the saturated velocity case, showing almost equally spaced curves with increasing gate voltage.

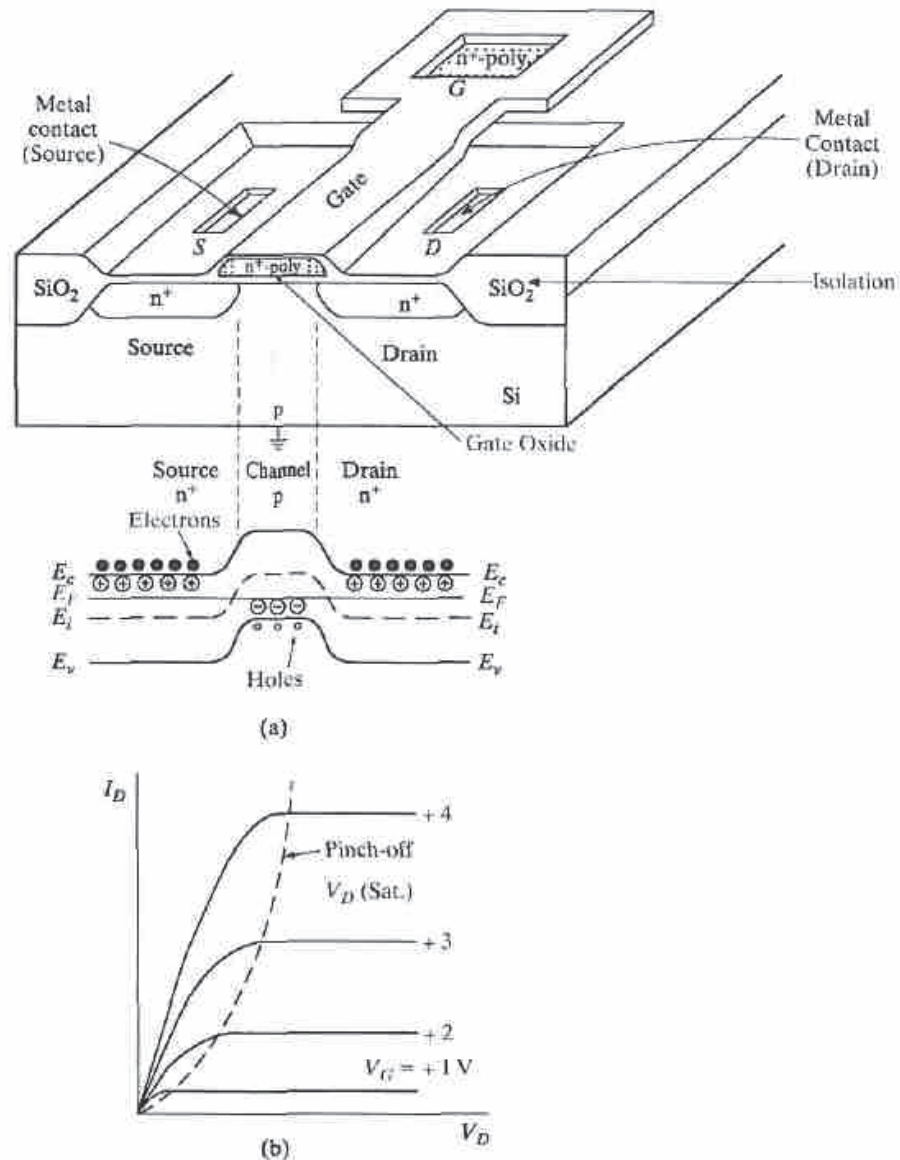
Most devices operate with characteristics intermediate between the constant mobility and the constant velocity regimes. Depending on the details of the field distribution, it is possible to divide up the channel into regions dominated by the two extreme cases, or to use an approximation such as Eq. (6-13).

Another important short-channel effect, described in Section 6.2.3, is the reduction in effective channel length after pinch-off as the drain voltage is increased. This effect is not significant in long-channel devices, since the change in L due to intrusion of the depletion region is a minor fraction of the total channel length. In short-channel devices, however, the effective channel length can be substantially shortened, leading to a slope in the saturated I - V characteristic that is analogous to the Early (base-width narrowing) effect in bipolar transistors discussed in Section 7.7.2.

One of the most widely used electronic devices, particularly in digital integrated circuits, is the *metal-insulator-semiconductor (MIS) transistor*. In this device the channel current is controlled by a voltage applied at a gate electrode that is isolated from the channel by an insulator. The resulting device may be referred to generically as an insulated-gate field-effect transistor (IGFET). However, since most such devices are made using silicon for the semiconductor, SiO_2 for the insulator, and metal or heavily doped polysilicon for the gate electrode, the term *MOS field-effect transistor (MOSFET)* is commonly used.

6.4 THE METAL- INSULATOR- SEMICONDUCTOR FET

Figure 6-10
An enhancement-type n-channel MOSFET: (a) isometric view of device and equilibrium band diagram along channel; (b) drain current-voltage output characteristics as a function of gate voltage.



6.4.1 Basic Operation and Fabrication

The basic MOS transistor is illustrated in Fig. 6-10a for the case of an enhancement-mode n-channel device formed on a p-type Si substrate. The n⁺ source and drain regions are diffused or implanted into a relatively lightly doped p-type substrate, and a thin oxide layer separates the conducting gate from the Si surface. No current flows from drain to source without a conducting n channel between them. This can be understood clearly by

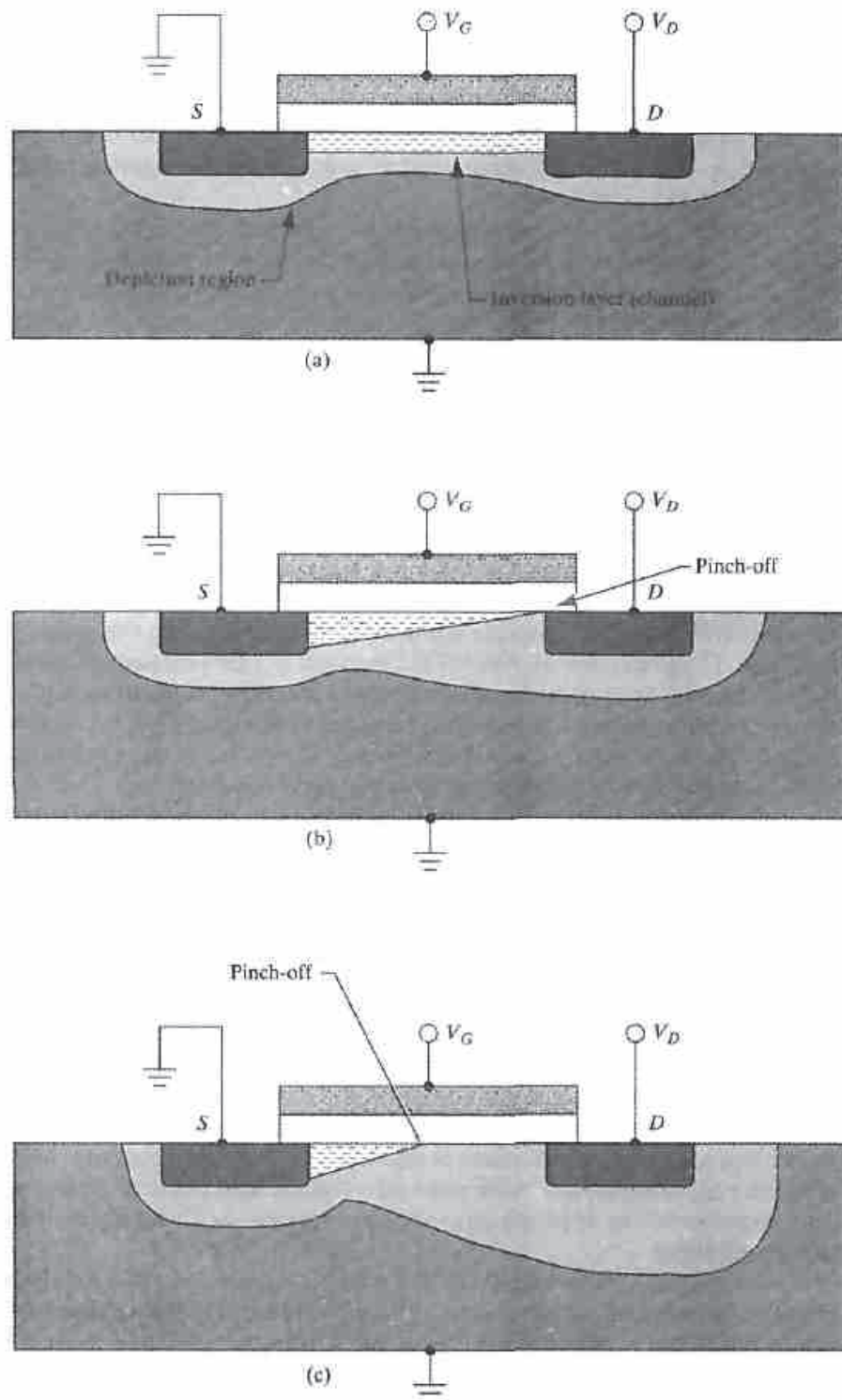
looking at the band diagram of the MOSFET in equilibrium along the channel (Fig. 6-10a). The Fermi level is flat in equilibrium. The conduction band is close to the Fermi level in the n^+ source/drain, while the valence band is closer to the Fermi level in the p-type material. Hence, there is a potential barrier for an electron to go from the source to the drain, corresponding to the built-in potential of the back-to-back p-n junctions between the source and drain.

When a positive voltage is applied to the gate relative to the substrate (which is connected to the source in this case), positive charges are in effect deposited on the gate metal. In response, negative charges are induced in the underlying Si, by the formation of a depletion region and a thin surface region containing mobile electrons. These induced electrons form the channel of the FET, and allow current to flow from drain to source. As Fig. 6-10b suggests, the effect of the gate voltage is to vary the conductance of this induced channel for low drain-to-source voltage, analogous to the JFET case. Since electrons are electrostatically induced in the p-type channel region, the channel becomes less p-type, and therefore the valence band moves down, farther away from the Fermi level. This obviously reduces the barrier for electrons between the source, the channel, and the drain. If the barrier is reduced sufficiently by applying a gate voltage in excess of what is known as the *threshold voltage*, V_T , there is significant current flow from the source to the drain. Thus, one view of a MOSFET is that it is a gate-controlled potential barrier. It is very important to have high-quality, low-leakage p-n junctions in order to ensure a low off-state leakage in the MOSFET. For a given value of V_G there will be some drain voltage V_D for which the current becomes saturated, after which it remains essentially constant.

The threshold voltage V_T is the minimum gate voltage required to induce the channel. In general, the positive gate voltage of an n-channel device (such as that shown in Fig. 6-11) must be larger than some value V_T before a conducting channel is induced. Similarly, a p-channel device (made on an n-type substrate with p-type source and drain implants or diffusions) requires a gate voltage more negative than some threshold value to induce the required positive charge (mobile holes) in the channel. There are exceptions to this general rule, however, as we shall see. For example, some n-channel devices have a channel already with zero gate voltage, and in fact a negative gate voltage is required to turn the device off. Such a "normally on" device is called a *depletion-mode* transistor, since gate voltage is used to deplete a channel which exists at equilibrium. The more common MOS transistor is "normally off" with zero gate voltage, and operates in the *enhancement mode* by applying gate voltage large enough to induce a conducting channel.

An alternative view of a MOSFET is that it is a gate-controlled resistor. If the (positive) gate voltage exceeds the threshold voltage in an n-channel device, electrons are induced in the p-type substrate. Since this channel is connected to the n^+ source and drain regions, the structure looks electrically like an induced n-type resistor. As the gate voltage increases, more electron

Figure 6-11
n-channel
MOSFET cross-
sections under dif-
ferent operating
conditions: (a) lin-
ear region for
 $V_G > V_T$ and
 $V_D < (V_G - V_T)$;
(b) onset of
saturation
at pinch-off,
 $V_G > V_T$ and
 $V_D = (V_G - V_T)$;
(c) strong satura-
tion, $V_G > V_T$ and
 $V_D > (V_G - V_T)$.



charge is induced in the channel and, therefore, the channel becomes more conducting. The drain current initially increases linearly with the drain bias (the *linear regime*) (Fig. 6-10b). As more drain current flows in the channel, however, there is more ohmic voltage drop along the channel such that the channel potential varies from zero near the grounded source to whatever the applied drain potential is near the drain end of the channel. Hence, the voltage difference between the gate and the channel reduces from V_G near the source to $(V_G - V_D)$ near the drain end. Once the drain bias is increased to the point that $(V_G - V_D) = V_T$, threshold is barely maintained near the drain end, and the channel is said to be pinched off. Increasing the drain bias beyond this point ($V_D(\text{sat})$) causes the point at which the channel gets pinched off to move more and more into the channel, closer to the source end (Fig. 6-11c). Electrons in the channel are pulled into the pinch-off region and travel at the saturation drift velocity because of the very high longitudinal electric field along the channel. Now, the drain current is said to be in the *saturation region* because it does not increase with drain bias significantly (Fig. 6-10b). Actually, there is a slight increase of drain current with drain bias due to various effects such as channel length modulation and drain-induced barrier lowering (DIBL) that will be discussed in Section 6.5.10.

The MOS transistor is particularly useful in digital circuits, in which it is switched from the "off" state (no conducting channel) to the "on" state. The control of drain current is obtained at a gate electrode which is insulated from the source and drain by the oxide. Thus the d-c input impedance of an MOS circuit can be very large.

Both n-channel and p-channel MOS transistors are in common usage. The n-channel type illustrated in Fig. 6-10 is generally preferred because it takes advantage of the fact that the electron mobility in Si is larger than the mobility of holes. In much of the discussion to follow we will use the n-channel (p-type substrate) example, although the p-channel case will be kept in mind also.

Let us give a very simplified description of how such an n-channel MOSFET can be fabricated. A much more detailed discussion is given in Section 9.3.1. An ultra-thin ($\sim 5\text{--}10\text{ nm}$) dry thermal silicon dioxide is grown on the p-type substrate. This serves as the gate insulator between the conducting gate and the channel. We immediately cover it with LPCVD of polysilicon, which is doped very heavily n^+ using P diffusion in order to make it behave electrically like a metal electrode. The doped polysilicon layer is then patterned to form the gates, and etched anisotropically by RIE to achieve vertical walls (Section 5.1.7). The gate itself is used as an implant mask for an n^+ implant which forms the source/drain junctions abutted to the gate edges, but is blocked from the channel region. Such a scheme is called a *self-aligned process* because we did not have to use a separate lithography step for the source/drain formation. Self-alignment is simple and is very useful because we thereby guarantee that there will be some overlap of the gate with the source/drain but not too much overlap. The advantages of this are discussed in Section 6.5.8. The implanted dopants must be annealed for reasons discussed

in Section 5.1.4. Finally, the MOSFETs have to be properly interconnected according to the circuit layout, using metallization. This involves LPCVD of an oxide dielectric, etching contact holes by RIE, sputter depositing a suitable metal such as Al, patterning and etching it.

As shown in Fig. 6-10a, the MOSFET is surrounded on all sides by a thick SiO_2 layer. This layer provides critical electrical isolation between adjacent transistors on an integrated circuit. We shall see in Section 9.3.1 that such *isolation* or *field* regions can be formed in several ways, such as *LOCAL Oxidation of Silicon (LOCOS)*. Briefly, it involves depositing a LPCVD Si_3N_4 layer over the entire substrate before the fabrication of the MOSFETs, patterning and etching it so that it is removed only in the isolation regions, but not in the *active* regions where the MOSFETs will be formed subsequently. A boron *channel stop* implant is then done in the isolation regions. Exploiting the useful property of Si_3N_4 that it blocks thermal oxidation, a thick LOCOS oxide is selectively grown by wet oxidation only in the isolation regions. The reason why a thick field oxide layer and a boron channel stop implant leads to electrical isolation is discussed in Section 6.5.5.

6.4.2 The Ideal MOS Capacitor

The surface effects that arise in an apparently simple MOS structure are actually quite complicated. Although many of these effects are beyond the scope of this discussion, we will be able to identify those which control typical MOS transistor operation. We begin by considering an uncomplicated idealized case, and then include effects encountered in real surfaces in the next section.

Some important definitions are made in the energy band diagram of Fig. 6-12a. The work function characteristic of the metal (see Section 2.2.1) can be defined in terms of the energy required to move an electron from the Fermi level to outside the metal. In MOS work it is more convenient to use a *modified work function* $q\Phi_m$ for the metal-oxide interface. The energy $q\Phi_m$ is measured from the metal Fermi level to the conduction band of the oxide.⁴ Similarly, $q\Phi_s$ is the modified work function at the semiconductor-oxide interface. In this idealized case we assume that $\Phi_m = \Phi_s$, so there is no difference in the two work functions. Another quantity that will be useful in later discussions is $q\Phi_F$, which measures the position of the Fermi level below the intrinsic level E_i for the semiconductor. This quantity indicates how strongly p-type the semiconductor is [see Eq. (3-25)].

The MOS structure of Fig. 6-12a is essentially a capacitor in which one plate is a semiconductor. If we apply a negative voltage between the metal and the semiconductor (Fig. 6-12b), we effectively deposit a negative charge on the metal. In response, we expect an equal net positive charge to accumulate

⁴On the MOS band diagrams of this section, we show a break in the electron energy scale leading to the insulator conduction band, since the band gap of SiO_2 (or other typical insulators) is much greater than that of the Si.

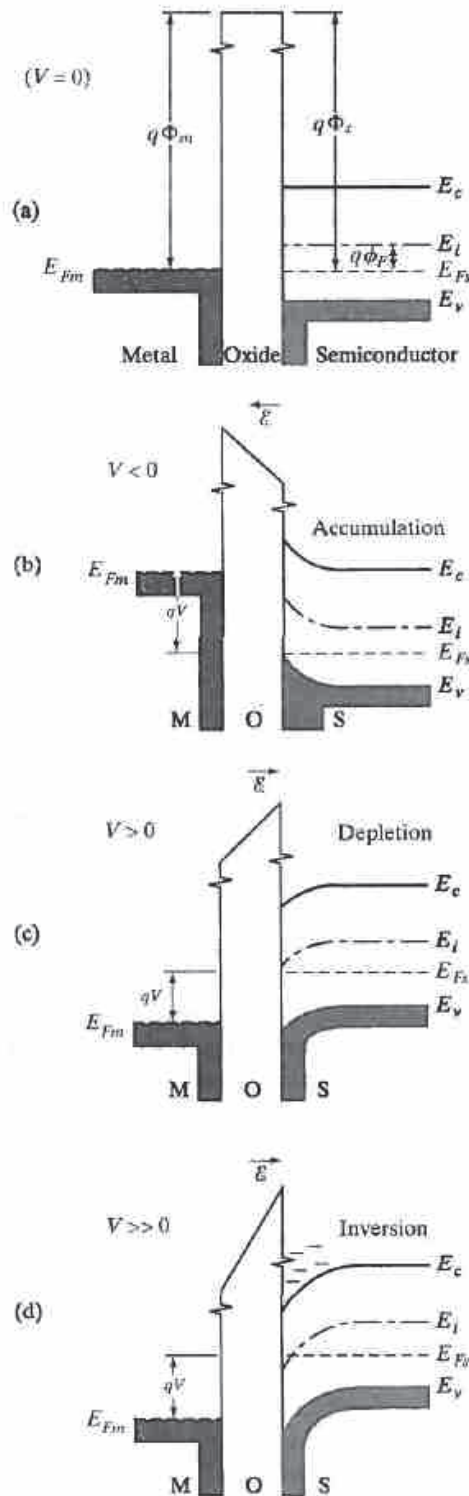


Figure 6-12
Band diagram for the ideal MOS structure at: (a) equilibrium; (b) negative voltage causes hole accumulation in the p-type semiconductor; (c) positive voltage depletes holes from the semiconductor surface; (d) a larger positive voltage causes inversion—an "n-type" layer at the semiconductor surface.

at the surface of the semiconductor. In the case of a p-type substrate this occurs by *hole accumulation* at the semiconductor-oxide interface.

Since the applied negative voltage *depresses* the electrostatic potential of the metal relative to the semiconductor, the electron energies are *raised* in the metal relative to the semiconductor.⁵ As a result, the Fermi level for the metal E_{Fm} lies above its equilibrium position by qV , where V is the applied voltage.

Since Φ_m and Φ_s do not change with applied voltage, moving E_{Fm} up in energy relative to E_{Fs} causes a tilt in the oxide conduction band. We expect such a tilt since an electric field causes a gradient in E_i (and similarly in E_v and E_c) as described in Section 4.4.2:

$$\mathcal{E}(x) = \frac{1}{q} \frac{dE_i}{dx} \quad (\text{see 4-26})$$

The energy bands of the semiconductor bend near the interface to accommodate the accumulation of holes. Since

$$p = n_i e^{(E_i - E_F)/kT} \quad (\text{see 3-25})$$

it is clear that an increase in hole concentration implies an increase in $E_i - E_F$ at the semiconductor surface.

Since no current passes through the MOS structure, there can be no variation in the Fermi level within the semiconductor. Therefore, if $E_i - E_F$ is to increase, it must occur by E_i moving up in energy near the surface. The result is a bending of the semiconductor bands near the interface. We notice in Fig. 6-12b that the Fermi level near the interface lies closer to the valence band, indicating a larger hole concentration than that arising from the doping of the p-type semiconductor.

In Fig. 6-12c we apply a positive voltage from the metal to the semiconductor. This raises the potential of the metal, lowering the metal Fermi level by qV relative to its equilibrium position. As a result, the oxide conduction band is again tilted. We notice that the slope of this band, obtained by simply moving the metal side down relative to the semiconductor side, is in the proper direction for the applied field, according to Eq. (4-26).

The positive voltage deposits positive charge on the metal and calls for a corresponding net negative charge at the surface of the semiconductor. Such a negative charge in p-type material arises from *depletion* of holes from the region near the surface, leaving behind uncompensated ionized acceptors. This is analogous to the depletion region at a p-n junction discussed in Section 5.2.3. In the depleted region the hole concentration decreases, moving E_i closer to E_{Fs} , and bending the bands down near the semiconductor surface.

If we continue to increase the positive voltage, the bands at the semiconductor surface bend down more strongly. In fact, a sufficiently large voltage can

⁵Recall that an electrostatic potential diagram is drawn for positive test charges, in contrast to an electron energy diagram, which is drawn for negative charges.

bend E_i below E_F (Fig. 6-12d). This is a particularly interesting case, since $E_F \gg E_i$ implies a large electron concentration in the conduction band.

The region near the semiconductor surface in this case has conduction properties typical of n-type material, with an electron concentration given by Eq. (3-25a). This n-type surface layer is formed not by doping, but instead by inversion of the originally p-type semiconductor due to the applied voltage. This inverted layer, separated from the underlying p-type material by a depletion region, is the key to MOS transistor operation.

We should take a closer look at the inversion region, since it becomes the conducting channel in the FET. In Fig. 6-13 we define a potential ϕ at any point x , measured relative to the equilibrium position of E_i . The energy $q\phi$ tells us the extent of band-bending at x , and $q\phi_s$ represents the band-bending at the surface. We notice that $\phi_s = 0$ is the *flat band* condition for this ideal MOS case (i.e., the bands look like Fig. 6-12a). When $\phi_s < 0$, the bands bend up at the surface, and we have hole accumulation (Fig. 6-12b). Similarly, when $\phi_s > 0$, we have depletion (Fig. 6-12c). Finally, when ϕ_s is positive and larger than ϕ_F , the bands at the surface are bent down such that $E_i(x=0)$ lies below E_F , and inversion is obtained (Fig. 6-12d).

While it is true that the surface is inverted whenever ϕ_s is larger than ϕ_F , a practical criterion is needed to tell us whether a true n-type conducting channel exists at the surface. The best criterion for *strong inversion* is that the surface should be as strongly n-type as the substrate is p-type. That is, E_i should lie as far below E_F at the surface as it is above E_F far from the surface. This condition occurs when

$$\phi_{s(\text{inv.})} = 2\phi_F = 2 \frac{kT}{q} \ln \frac{N_a}{n_i} \quad (6-15)$$

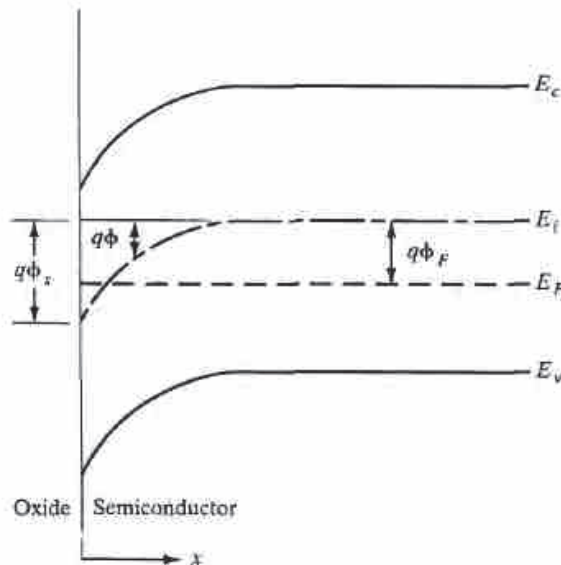


Figure 6-13
Bending of the semiconductor bands at the onset of strong inversion: the surface potential ϕ_s is twice the value of ϕ_F in the neutral p material.

A surface potential of ϕ_F is required to bend the bands down to the intrinsic condition at the surface ($E_i = E_F$), and E_i must then be depressed another $q\phi_F$ at the surface to obtain the condition we call strong inversion.

The electron and hole concentrations are related to the potential $\phi(x)$ defined in Fig. 6-13. Since the equilibrium electron concentration is

$$n_0 = n_i e^{(E_F - E_i)/kT} = n_i e^{-q\phi_F/kT} \quad (6-16)$$

we can easily relate the electron concentration at any x to this value:

$$n = n_i e^{-q(\phi_F - \phi)/kT} = n_0 e^{q\phi/kT} \quad (6-17)$$

and similarly for holes:

$$p_0 = n_i e^{q\phi_F/kT} \quad (6-18a)$$

$$p = p_0 e^{-q\phi/kT} \quad (6-18b)$$

at any x . We could combine these equations with Poisson's equation (6-19) and the usual charge density expression (6-20) to solve for $\phi(x)$:

$$\frac{\partial^2 \phi}{\partial x^2} = -\frac{\rho(x)}{\epsilon_s} \quad (6-19)$$

$$\rho(x) = q(N_d^+ - N_a^- + p - n) \quad (6-20)$$

Let us solve this equation to determine the total integrated charge per unit area, Q_s , as a function of the surface potential, ϕ_s . Substituting Eqs. (6-16), (6-17), and (6-18) for the electron and hole concentrations in Eqs. (6-19) and (6-20), we get

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial \phi}{\partial x} \right) = -\frac{q}{\epsilon_s} \left[p_0 \left(e^{-\frac{q\phi}{kT}} - 1 \right) - n_0 \left(e^{\frac{q\phi}{kT}} - 1 \right) \right] \quad (6-21)$$

It should be kept in mind that

$$\frac{-\partial \phi}{\partial x}$$

is the electric field, $\mathcal{E}$, at a depth x .

Integrating Eq. (6-21) from the bulk (where the bands are flat, the electric fields are zero, and the carrier concentrations are determined solely by the doping), towards the surface, we get

$$\int_0^{\phi_s} \left(\frac{\partial \phi}{\partial x} \right) d \left(\frac{\partial \phi}{\partial x} \right) = -\frac{q}{\epsilon_s} \int_0^{\phi_s} \left[p_0 \left(e^{-\frac{q\phi}{kT}} - 1 \right) - n_0 \left(e^{\frac{q\phi}{kT}} - 1 \right) \right] d\phi \quad (6-22)$$

After integration, we then get

$$\mathcal{E}^2 = \left(\frac{2kT}{\epsilon_s} \right) \left[\left(e^{-\frac{q\phi}{kT}} + \frac{q\phi}{kT} - 1 \right) + \frac{n_0}{p_0} \left(e^{\frac{q\phi}{kT}} - \frac{q\phi}{kT} - 1 \right) \right] \quad (6-23)$$

A particularly important case is at the surface ($x = 0$) where the surface perpendicular electric field, $\mathcal{E}_s$, becomes

$$\mathcal{E}_s = \frac{\sqrt{2kT}}{qL_D} \left[\left(e^{-\frac{q\phi_s}{kT}} + \frac{q\phi_s}{kT} - 1 \right) + \frac{n_0}{p_0} \left(e^{\frac{q\phi_s}{kT}} - \frac{q\phi_s}{kT} - 1 \right) \right]^{\frac{1}{2}} \quad (6-24)$$

where we have introduced a new term, the *Debye screening length*,

$$L_D = \sqrt{\frac{\epsilon_s kT}{q^2 p_0}} \quad (6-25)$$

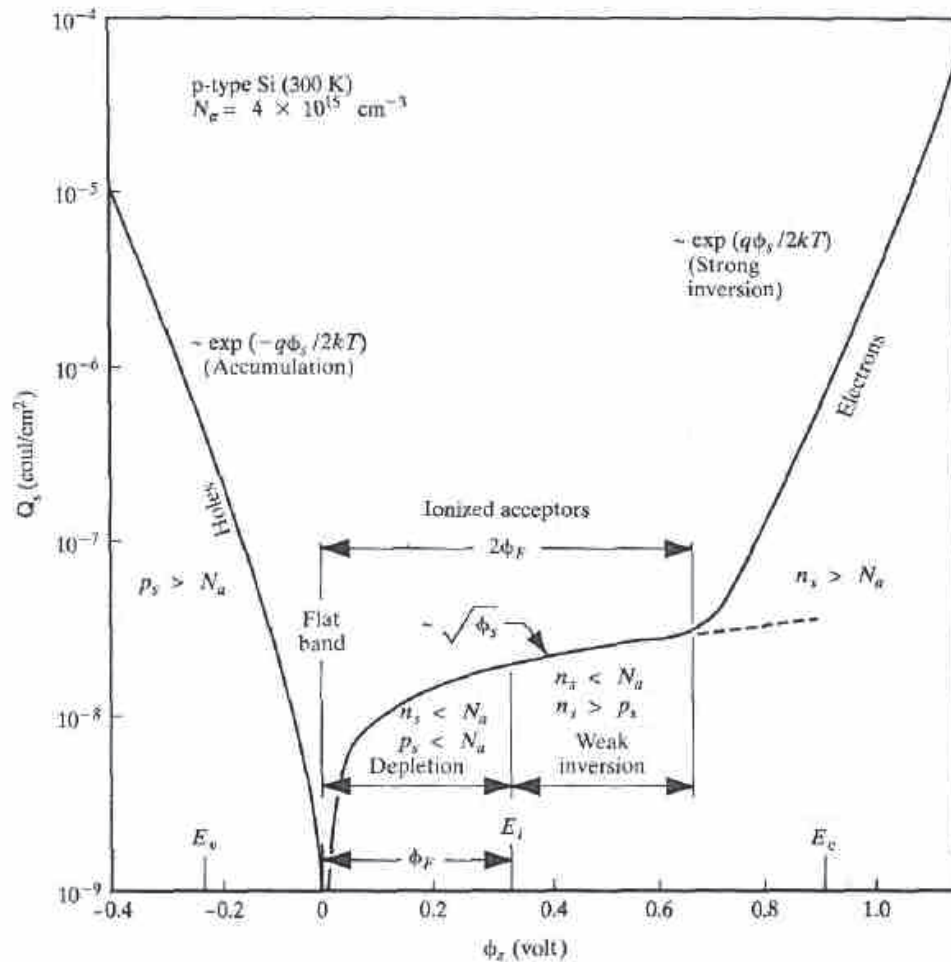
The Debye length is a very important concept in semiconductors. It gives us an idea of the distance scale in which charge imbalances are screened or smeared out. For example, if we think of inserting a positively charged sphere in an n-type semiconductor we know that the mobile electrons will crowd around the sphere. If we move away from the sphere by several Debye lengths, the positively charged sphere and the negative electron cloud will look like a neutral entity. Not surprisingly, L_D depends inversely on doping because the higher the carrier concentration, the more easily screening takes place. For n-type material we should use n_0 in Eq. (6-25).

By using Gauss's law at the surface, we can relate the integrated space charge per unit area to the electric displacement, keeping in mind that the electric field or displacement deep in the substrate is zero.

$$Q_s = -\epsilon_s \mathcal{E}_s \quad (6-26)$$

The space charge density per unit area Q_s in Eq. (6-26) is plotted as a function of the surface potential ϕ_s in Fig. 6-14. We see from Eq. (6-24) and Fig. 6-14 that when the surface potential is zero (flat band conditions), the net space charge is zero. This is because the fixed dopant charges are cancelled by the mobile carrier charges at flat band. When the surface potential is negative, it attracts and forms an accumulation layer of the majority carrier holes at the surface. The first term in Eq. (6-24) is the dominant one, and the accumulation space charge increases very strongly (exponentially) with negative surface potential. It is easy to see why by looking at Eq. (6-18), which gives the surface hole concentration in a p-type semiconductor as a function of surface potential. Since the band-bending decreases as a function of depth, the integrated accumulation charge involves averaging over depth and introduces a factor of 2 in the exponent. Mathematically, this is due to the square root in Eq. (6-24). It must be noted that since this charge is due to the mobile majority carriers (holes in this case), the charge piles up near the oxide-silicon interface, since typical accumulation layer thicknesses are ~ 20 nm. Also, because of the exponential dependence of accumulation charge on surface potential, the band-bending is generally small or is said to be *pinned* to nearly zero.

On the other hand, for a positive surface potential, we see from Eq. (6-24) that initially the second (linear) term is the dominant one. Although

**Figure 6-14**

Variation of space-charge density in the semiconductor as a function of the surface potential ϕ_s for p-type silicon with $N_a = 4 \times 10^{15} \text{ cm}^{-3}$ at room temperature. p_s and n_s are the hole and electron concentrations at the surface, ϕ_F is the potential difference between the Fermi level and the intrinsic level of the bulk. [Garrett and Brattain, Phys. Rev., 99, 376 (1955).]

the exponential term $\exp(q\phi_s/kT)$ is very large, it is multiplied by the ratio of the minority to majority carrier concentration which is very small, and is initially negligible. Hence, the space charge for small positive surface potentials increases as $\sim \sqrt{\phi_s}$, as shown in Fig. 6-14. As discussed in detail later in this section, this corresponds to the depletion region charge due to the exposed, fixed immobile dopants (acceptors in this case). The depletion width typically extends over several hundred nm. At some point, the band-bending is twice the Fermi potential ϕ_F , which is enough for the onset of strong inversion. Now, the exponential term $\exp(q\phi_s(\text{inv.})/kT)$ multiplied by the minority carrier concentration n_0 is equal to the majority carrier concentration

p_0 . Hence, for band bending beyond this point, it becomes the dominant term. As in the case of accumulation, the mobile inversion charge now increases very strongly with bias, as indicated by Eq. (6-17) and shown in Fig. 6-14. The typical inversion layer thicknesses are ~ 5 nm, and the surface potential now is essentially pinned at $2\phi_F$.

It may be pointed out that in accumulation, and especially in inversion, the carriers are confined in the x -direction in narrow, essentially triangular potential wells, causing quantum mechanical particle-in-a-box states or subbands, similar to those discussed in Chapter 2. However, the carriers are free in the other directions (parallel to the oxide-silicon interface). This leads to a two-dimensional electron gas (2DEG) or hole gas, with a "staircase" constant density of states, as discussed in Appendix IV. The detailed analysis of these effects is, unfortunately, beyond the scope of our discussion here.

The charge distribution, electric field, and electrostatic potential for the inverted surface are sketched in Fig. 6-15. For simplicity we use the depletion approximation of Chapter 5 in this figure, assuming complete depletion for $0 < x < W$, and neutral material for $x > W$. In this approximation the charge per unit area⁶ due to uncompensated acceptors in the depletion region is $-qN_aW$. The positive charge Q_m on the metal is balanced by the negative charge Q_s in the semiconductor, which is the depletion layer charge plus the charge due to the inversion region Q_n :

$$Q_m = -Q_s = qN_aW - Q_n \quad (6-27)$$

The width of the inversion region is exaggerated in Fig. 6-15 for illustrative purposes. Actually, the width of this region is generally less than 100 Å. Thus we have neglected it in sketching the electric field and potential distribution. In the potential distribution diagram we see that an applied voltage V appears partially across the insulator (V_i) and partially across the depletion region of the semiconductor (ϕ_s):

$$V = V_i + \phi_s \quad (6-28)$$

The voltage across the insulator is obviously related to the charge on either side, divided by the capacitance:

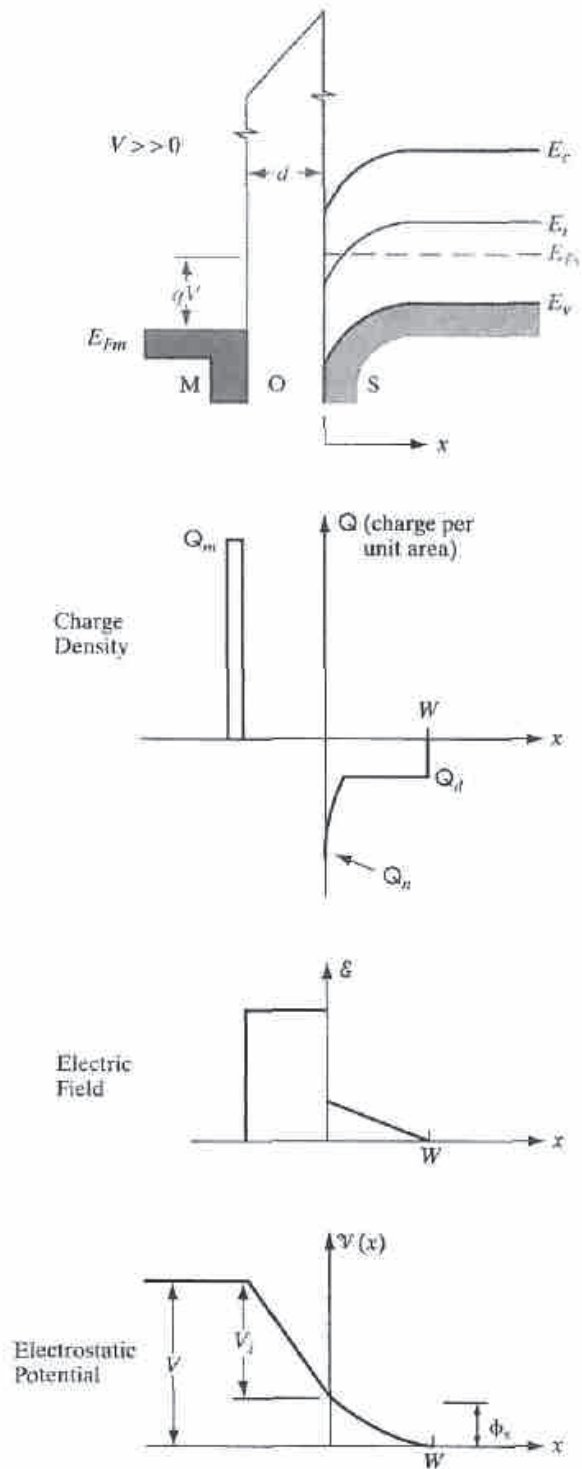
$$V_i = \frac{-Q_s d}{\epsilon_i} = \frac{-Q_s}{C_i} \quad (6-29)$$

where ϵ_i is permittivity of the insulator and C_i is the insulator capacitance per unit area. The charge Q_s will be negative for the n channel, giving a positive V_i .

Using the depletion approximation, we can solve for W as a function of ϕ_s (Prob. 6.7). The result is the same as would be obtained for an n^+-p junction in Chapter 5, for which the depletion region extends almost entirely into the p region:

⁶In this chapter, we will use charge per unit area (Q) and capacitance per unit area (C) to avoid repeating A throughout the discussion.

Figure 6-15
Approximate distributions of charge, electric field, and electrostatic potential in the ideal MOS capacitor in inversion. The relative width of the inverted region is exaggerated for illustrative purposes, but is neglected in the field and potential diagrams.



$$W = \left[\frac{2\epsilon_s \phi_s}{qN_a} \right]^{1/2} \quad (6-30)$$

This depletion region grows with increased voltage across the capacitor until strong inversion is reached. After that, further increases in voltage result in stronger inversion rather than in more depletion. Thus the maximum value of the depletion width is

$$W_m = \left[\frac{2\epsilon_s \phi_s(\text{inv.})}{qN_a} \right]^{1/2} = 2 \left[\frac{\epsilon_s kT \ln(N_a/n_i)}{q^2 N_a} \right]^{1/2} \quad (6-31)$$

using Eq. (6-15). We know the quantities in this expression, so W_m can be calculated.

The charge per unit area in the depletion region Q_d at strong inversion is⁷

$$Q_d = -qN_a W_m = -2(\epsilon_s q N_a \phi_s)^{1/2} \quad (6-32)$$

The applied voltage must be large enough to create this depletion charge plus the surface potential $\phi_s(\text{inv.})$. The *threshold* voltage required for strong inversion, using Eqs. (6-15), (6-28), and (6-29), is

$$V_T = -\frac{Q_d}{C_i} + 2\phi_F \quad (\text{ideal case}) \quad (6-33)$$

This assumes the negative charge at the semiconductor surface Q_s at inversion is mostly due to the depletion charge Q_d . The threshold voltage represents the minimum voltage required to achieve strong inversion, and is an extremely important quantity for MOS transistors. We will see in the next section that other terms must be added to this expression for real MOS structures.

The capacitance-voltage characteristics of this ideal MOS structure (Fig. 6-16) vary depending on whether the semiconductor surface is in accumulation, depletion, or inversion.

Since the capacitance for MOSFETs is voltage dependent, we must use the more general expression in Eq. (5-55) for the voltage-dependent semiconductor capacitance,

$$C_s = \frac{dQ}{dV} = \frac{dQ_s}{d\phi_s} \quad (6-34)$$

Actually, if one looks at the electrical equivalent circuit of a MOS capacitor or MOSFET, it is the series combination of a fixed, voltage-independent gate oxide (insulator) capacitance, and a voltage-dependent semiconductor capacitance [defined according to Eq. (6-34)], such that the overall MOS capacitance becomes voltage dependent. The semiconductor capacitance itself can be determined from the slope of the Q_s versus ϕ_s plot (Fig. 6-14). It is

⁷In the p-channel (n-type substrate) case, for which ϕ_F is negative, we use $Q_d = +qN_a W_m = 2(\epsilon_s q N_a |\phi_F|)^{1/2}$.

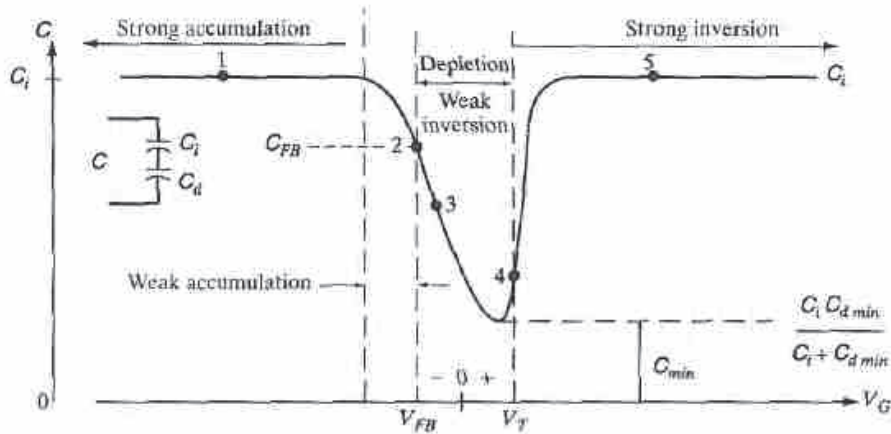


Figure 6-16

Capacitance-voltage relation for an n-channel (p-substrate) MOS capacitor. The dashed curve for $V > V_T$ is observed at high measurement frequencies. The flat band voltage V_{FB} will be discussed in Section 6.4.3. When the semiconductor is in depletion, the semiconductor capacitance C_s is denoted as C_d .

clear that the semiconductor capacitance in accumulation is very high because the slope is so steep; i.e., the accumulation charge changes a lot with surface potential. Hence, the series capacitance in accumulation is basically the insulator capacitance, C_i . Since, for negative voltage, holes are accumulated at the surface (Fig. 6-12b), the MOS structure appears almost like a parallel-plate capacitor, dominated by the insulator properties $C_i = \epsilon/d$ (point 1 in Fig. 6-16). As the voltage becomes less negative, the semiconductor surface is depleted. Thus a depletion-layer capacitance C_d is added in series with C_i :

$$C_i = \frac{\epsilon_s}{W} \quad (6-35)$$

where ϵ_s is the semiconductor permittivity and W is the width of the depletion layer from Eq. (6-30). The total capacitance is

$$C = \frac{C_i C_d}{C_i + C_d} \quad (6-36)$$

The capacitance decreases as W grows from flatband (point 2), past weak inversion (point 3), until finally strong inversion is reached at V_T (point 4). In the depletion region, the small signal semiconductor capacitance is given by the same formula [Eq. (6-34)] which gives the variation of the (depletion) space charge with surface potential. Since the charge increases as $\sim \sqrt{\phi_s}$, the depletion capacitance will obviously decrease as $1/\sqrt{\phi_s}$, exactly as for the depletion capacitance of a p-n junction [see Eq. (5-63)].

After inversion is reached, the small signal capacitance depends on whether the measurements are made at high (typically ~ 1 MHz) or low (typically ~ 1 – 100 Hz) frequency, where “high” and “low” are with respect to the generation-recombination rate of the minority carriers in the inversion layer. If the gate voltage is varied rapidly, the charge in the inversion layer cannot change in response, and thus does not contribute to the small signal a-c capacitance. Hence, the semiconductor capacitance is at a minimum, corresponding to a maximum depletion width.

On the other hand, if the gate bias is changed slowly, there is time for minority carriers to be generated in the bulk, drift across the depletion region to the inversion layer, or go back to the substrate and recombine. Now, the semiconductor capacitance, using the same Eq. (6-34), is very large because we saw in Fig. 6-14 that the inversion charge increases exponentially with ϕ_s . Hence, the low frequency MOS series capacitance in strong inversion is basically C_i once again (point 5).

What is the frequency dependence of the capacitance in accumulation (Fig. 6-12a)? We get a very high capacitance both at low and high frequencies because the majority carriers in the accumulation layer can respond much faster than minority carriers. While minority carriers respond on the time scale of generation-recombination times (typically hundreds of microseconds in Si), majority carriers respond on the time scale of the dielectric relaxation time, $\tau_D = \rho\epsilon$, where ρ is the resistivity and ϵ is the permittivity. τ_D is analogous to the RC time constant of a system, and is small for the majority carriers ($\sim 10^{-13}$ s). As an interesting aside, it may be pointed out that in inversion, although the high-frequency capacitance for MOS capacitors is low, it is high ($= C_i$) for MOSFETs because now the inversion charge can flow in readily and very fast ($\sim \tau_D$) from the source/drain regions rather than having to be created by generation-recombination in the bulk.

6.4.3 Effects of Real Surfaces

When MOS devices are made using typical materials (e.g., n^+ polysilicon-SiO₂-Si), departures from the ideal case described in the previous section can strongly affect V_T and other properties. First, there is a work function difference between the doped polysilicon gate and substrate, which depends on the substrate doping. Here, the heavily doped polysilicon acts as a metal electrode. Second, there are inevitably charges at the Si-SiO₂ interface and within the oxide which must be taken into account.

Work Function Difference. We expect Φ_s to vary depending on the doping of the semiconductor. Figure 6-17 illustrates the work function potential difference $\Phi_{ms} = \Phi_m - \Phi_s$ for n^+ polysilicon on Si as the doping is varied. We note that Φ_{ms} is always negative for this case, and is most negative for heavily doped p-type Si (i.e., for E_F close to the valence band).

If we try to construct an equilibrium diagram with Φ_{ms} negative (Fig. 6-18a), we find that in aligning E_F we must include a tilt in the oxide conduction band

Figure 6-17
Variation of the metal-semiconductor work function potential difference Φ_m with substrate doping concentration, for n^+ poly-Si.

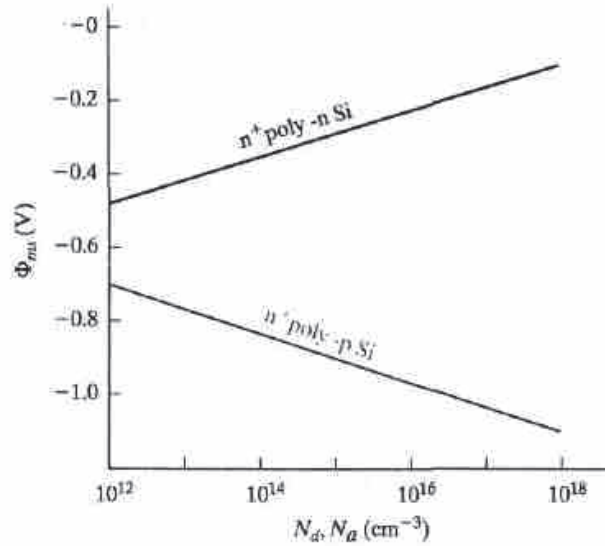
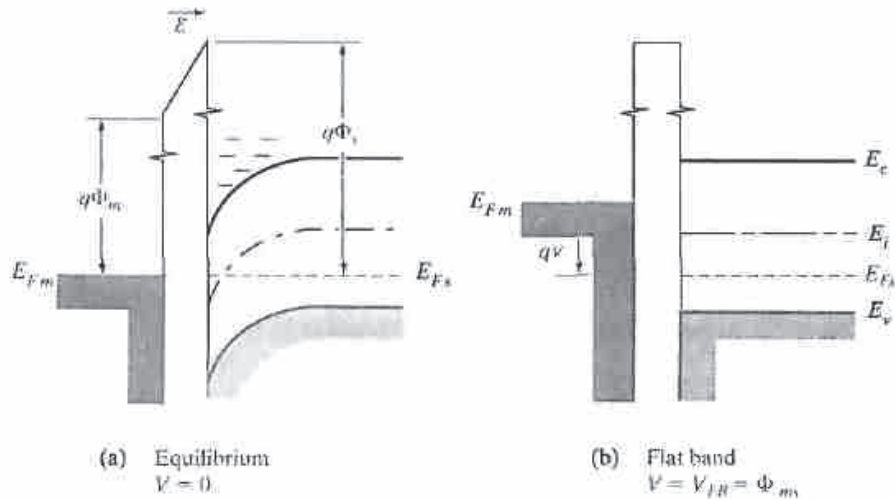


Figure 6-18
Effect of a negative work function difference ($\Phi_m < 0$):
(a) band bending and formation of negative charge at the semiconductor surface;
(b) achievement of the flat band condition by application of a negative voltage.



(implying an electric field). Thus the metal is positively charged and the semiconductor surface is negatively charged at equilibrium, to accommodate the work function difference. As a result, the bands bend down near the semiconductor surface. In fact, if Φ_m is sufficiently negative, an inversion region can exist with no external voltage applied. To obtain the *flat band* condition pictured in Fig. 6-18b, we must apply a negative voltage to the metal ($V_{FB} = \Phi_m$).

Interface Charge. In addition to the work function difference, the equilibrium MOS structure is affected by charges in the insulator and at the semiconductor-oxide interface (Fig. 6-19). For example, alkali metal ions (particularly Na^+) can be incorporated inadvertently in the oxide during

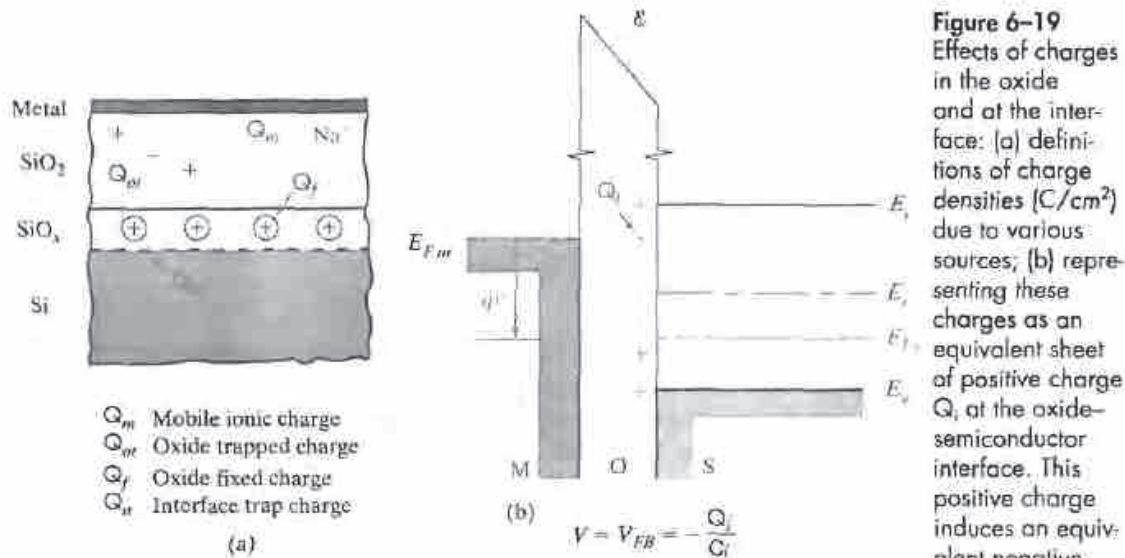


Figure 6-19
Effects of charges in the oxide and at the interface: (a) definitions of charge densities (C/cm^2) due to various sources; (b) representing these charges as an equivalent sheet of positive charge Q_f at the oxide-semiconductor interface. This positive charge induces an equivalent negative charge in the semiconductor, which requires a negative gate voltage to achieve the flat band condition.

growth or subsequent processing steps. Since sodium is a common contaminant, it is necessary to use extremely clean chemicals, water, gases, and processing environment to minimize its effect on dielectric layers. Sodium ions introduce positive charges (Q_m) in the oxide, which in turn induce negative charges in the semiconductor. The effect of such positive ionic charges in the oxide depends upon the number of ions involved and their distance from the semiconductor surface (Prob. 6.12). The negative charge induced in the semiconductor is greater if the Na^+ ions are near the interface than if they are farther away. The effect of this ionic charge on threshold voltage is complicated by the fact that Na^+ ions are relatively mobile in SiO_2 , particularly at elevated temperatures, and can thus drift in an applied electric field. Obviously, a device with V_T dependent on its past history of voltage bias is unacceptable. Fortunately, Na contamination of the oxide can be reduced to tolerable levels by proper care in processing. The oxide also contains trapped charges (Q_{ot}) due to imperfections in the SiO_2 .

In addition to oxide charges, a set of positive charges arises from *interface states* at the Si-SiO₂ interface. These charges, which we will call Q_{it} , result from the sudden termination of the semiconductor crystal lattice at the oxide interface. Near the interface is a transition layer (SiO_x) containing fixed charges (Q_f). As oxidation takes place in forming the SiO_2 layer, Si is removed from the surface and reacts with the oxygen. When the oxidation is stopped, some ionic Si is left near the interface. These ions, along with uncompleted Si bonds at the surface, result in a sheet of positive charge Q_f near the interface. This charge depends on oxidation rate and subsequent heat treatment, and also on crystal orientation. For carefully treated Si-SiO₂ interfaces, typical charge densities due to Q_{it} and Q_f are about 10^{10} charges/cm² for samples with [100] surfaces. The interface charge density is about a factor

of ten higher on [111] surfaces. That is why MOS devices are generally made on {100} Si.

For simplicity we will include the various oxide and interface charges in an *effective* positive charge at the interface Q_i (C/cm^2). The effect of this charge is to induce an equivalent negative charge in the semiconductor. Thus an additional component must be added to the flat band voltage:

$$V_{FB} = \Phi_{ms} - \frac{Q_i}{C_i} \quad (6-37)$$

Since the difference in work function and the positive interface charge both tend to bend the bands down at the semiconductor surface, a negative voltage must be applied to the metal relative to the semiconductor to achieve the flat band condition of Fig. 6-19b.

6.4.4 Threshold Voltage

The voltage required to achieve flat band should be added to the threshold voltage equation (6-33) obtained for the ideal MOS structure (for which we assumed a zero flat band voltage):

$$V_T = \Phi_{ms} - \frac{Q_i}{C_i} - \frac{Q_d}{C_i} + 2\phi_F \quad (6-38)$$

Thus the voltage required to create strong inversion must be large enough to first achieve the flat band condition (Φ_{ms} and Q_i/C_i terms), then accommodate the charge in the depletion region (Q_d/C_i), and finally to induce the inverted region ($2\phi_F$). This equation accounts for the dominant threshold voltage effects in typical MOS devices. It can be used for both n-type and p-type substrates⁸ if appropriate signs are included for each term (Fig. 6-20). Typically Φ_{ms} is negative, although its value varies as in Fig. 6-17. The interface charge is positive, so the contribution of the $-Q_i/C_i$ term is negative for either substrate type. On the other hand, the charge in the depletion region is negative for ionized acceptors (p-type substrate, n-channel device) and is positive for ionized donors (n-type substrate, p channel). Also, the term ϕ_F , which is defined as $(E_i - E_F)/q$ in the neutral substrate, can be positive or negative, depending on the conductivity type of the substrate. Considering the signs in Fig. 6-20, we see that all four terms give negative contributions in the p-channel case. Thus we expect negative threshold voltages for typical p-channel devices. On the other hand, n-channel devices may have either positive or negative threshold voltages, depending on the relative values of terms in Eq. (6-38).

⁸It is important to remember that n-channel devices are made on p-type substrates and that p-channel devices have n-type substrates.

$$V_T = \left| \begin{array}{c|c|c|c|c} \Phi_{ms} & -\frac{Q_f}{C_i} & -\frac{Q_d}{C_i} & +2\phi_F \\ \hline (-) & (-) & \begin{array}{l} (+) \text{ n-channel} \\ (-) \text{ p-channel} \end{array} & \begin{array}{l} (+) \text{ n-channel} \\ (-) \text{ p-channel} \end{array} \end{array} \right|$$

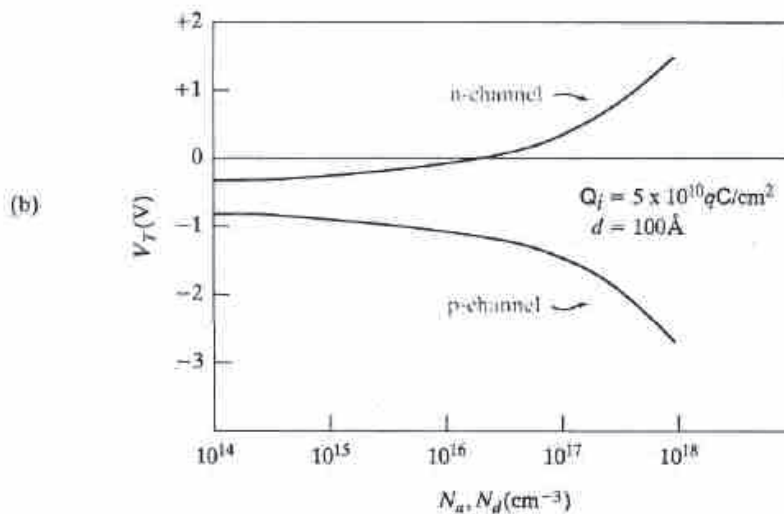


Figure 6-20 Influence of materials parameters on threshold voltage: (a) the threshold voltage equation indicating signs of the various contributions; (b) variation of V_T with substrate doping for n-channel and p-channel n^+ poly-SiO₂-Si devices.

All terms in Eq. (6-38) except Q_f/C_i depend on the doping in the substrate. The terms Φ_{ms} and ϕ_F have relatively small variations as E_F is moved up or down by the doping. Large changes can occur in Q_d , which varies with the square root of the doping impurity concentration as in Eq. (6-32). We illustrate the variation of threshold voltage with substrate doping in Fig. 6-20. As expected from Eq. (6-38), V_T is always negative for the p-channel case. In the n-channel case, the negative flat band voltage terms can dominate for lightly doped p-type substrates, resulting in a negative threshold voltage. However, for more heavily doped substrates, the increasing contribution of N_a to the Q_d term dominates, and V_T becomes positive.

We should pause here and consider what positive or negative V_T means for the two cases. In a p-channel device we expect to apply a negative voltage from metal to semiconductor in order to induce the positive charges in the channel. In this case a negative threshold voltage means that the negative voltage we apply must be larger than V_T in order to achieve strong inversion. In the n-channel case we expect to apply a positive voltage to the metal to induce the channel. Thus a positive value for V_T means the applied voltage must be larger than this threshold value to obtain strong inversion and a conducting n channel. On the other hand, a negative V_T in this case means

that a channel exists at $V = 0$ due to the Φ_{ms} and Q_i effects (Figs. 6-18 and 6-19), and we must apply a negative voltage V_T to turn the device off. Since lightly doped substrates are desirable to maintain a high breakdown voltage for the drain junction, Fig. 6-20 suggests that V_T will be negative for n-channel devices made by standard processing. This tendency for the formation of depletion mode (normally on) n-channel transistors is a problem which must be dealt with by special fabrication methods to be described in Section 6.5.5.

EXAMPLE 6-1

An n⁺-polysilicon-gate n-channel MOS transistor is made on a p-type Si substrate with $N_a = 5 \times 10^{15} \text{ cm}^{-3}$. The SiO_2 thickness is 100 Å in the gate region, and the effective interface charge Q_i is $4 \times 10^{10} \text{ qC/cm}^2$. Find C_i and $C_{\min}$ on the C - V characteristics, and find W_m , V_{FB} , and V_T .

SOLUTION

$$\phi_F = \frac{kT}{q} \ln \frac{N_a}{n_i} = 0.0259 \ln \frac{5 \times 10^{15}}{1.5 \times 10^{10}} = 0.329 \text{ eV}$$

$$W_m = 2 \left[\frac{\epsilon_s \phi_F}{q N_a} \right]^{1/2} = 2 \left[\frac{11.8 \times 8.85 \times 10^{-14} \times 0.329}{1.6 \times 10^{-19} \times 5 \times 10^{15}} \right]^{1/2}$$

$$= 4.15 \times 10^{-5} \text{ cm} = \mathbf{0.415 \mu\text{m}}$$

From Fig. 6-17, $\Phi_{ms} \approx -0.95 \text{ V}$, and we have

$$Q_i = 4 \times 10^{10} \times 1.6 \times 10^{-19} = 6.4 \times 10^{-9} \text{ C/cm}^2$$

$$C_i = \frac{\epsilon_i}{d} = \frac{3.9 \times 8.85 \times 10^{-14}}{0.1 \times 10^{-5}} = 3.45 \times 10^{-7} \text{ F/cm}^2$$

$$V_{FB} = \Phi_{ms} - Q_i/C_i = -0.95 - 6.4 \times 10^{-9}/3.45 \times 10^{-7} = \mathbf{-0.969 \text{ V}}$$

$$Q_d = -q N_a W_m = -1.6 \times 10^{-19} \times 5 \times 10^{15} \times 4.15 \times 10^{-5}$$

$$= -3.32 \times 10^{-8} \text{ C/cm}^2$$

$$V_T = V_{FB} - \frac{Q_d}{C_i} + 2\phi_F = -0.969 + \frac{3.32 \times 10^{-8}}{3.45 \times 10^{-7}} + 0.658 = \mathbf{-0.215 \text{ V}}$$

$$C_d = \frac{\epsilon_s}{W_m} = \frac{11.8 \times 8.85 \times 10^{-14}}{4.15 \times 10^{-5}} = 2.5 \times 10^{-8} \text{ F/cm}$$

$$C_{\min} = \frac{C_i C_d}{C_i + C_d} = \frac{3.45 \times 10^{-7} \times 2.5 \times 10^{-8}}{3.45 \times 10^{-7} + 2.5 \times 10^{-8}} = \mathbf{2.33 \times 10^{-8} \text{ F/cm}^2}$$

6.4.5 MOS Capacitance-Voltage Analysis

Let us see how the various parameters of a MOS device such as insulator thickness, substrate doping, and V_T can be determined from the C - V characteristics (Fig. 6-21). First, the shape of the C - V curve depends upon the type of substrate

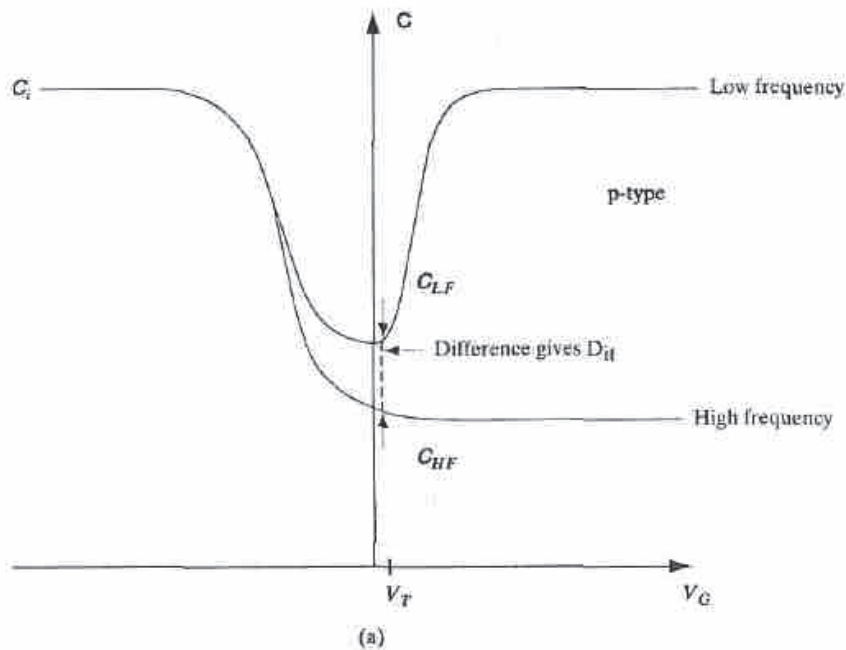
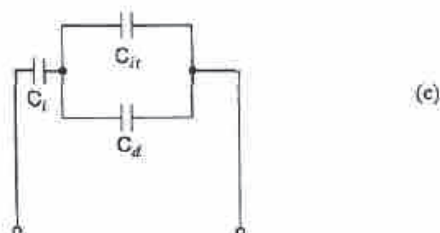
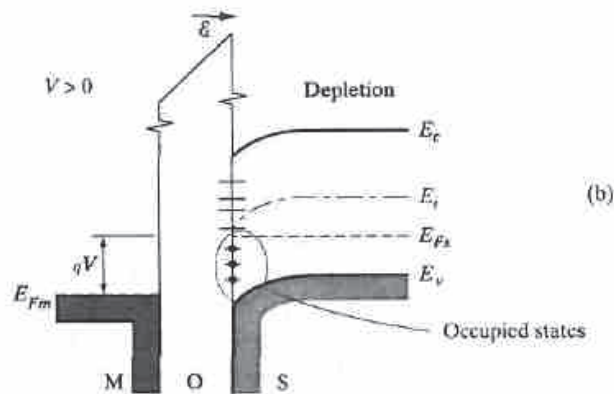


Figure 6-21
Fast interface state determination: (a) High-frequency and low-frequency C - V curves showing impact of fast interface states; (b) energy levels in the band gap due to fast interface states; (c) equivalent circuit of MOS structure showing capacitance components due to gate oxide (C_i), depletion layer in the channel (C_d), and fast interface states (C_{it}).



doping. If the high-frequency capacitance is large for negative gate biases and small for positive biases, it is a p-type substrate, and vice versa. From the low-frequency C - V curve for p-type material, as the gate bias is made more positive (or less negative), the capacitance goes down slowly in depletion and then rises rapidly in inversion. As a result, the low-frequency C - V is not quite symmetric in shape. For n-type substrates, the C - V curves would be the mirror image of Fig. 6-21.

The capacitance $C_i = \epsilon/d$ in accumulation or strong inversion (at low frequencies) gives us the insulator thickness, d . The minimum MOS capacitance, $C_{\min}$, is the series combination of C_i and the minimum depletion capacitance, $C_{d\min} = \epsilon_s/W_m$, corresponding to the maximum depletion width. We can in principle use the measurement of $C_{\min}$ to determine the substrate doping. However, from Eq. (6-31) we see that the dependence of W_m on N_a is complicated, and we get a transcendental equation which can only be solved numerically. Actually, an approximate, iterative solution exists which gives us N_a in terms of the minimum depletion capacitance, $C_{d\min}$.

$$N_a = 10^{[30.388 + 1.683 \log C_{d\min} - 0.03177(\log C_{d\min})^2]} \quad (6-39)$$

where $C_{d\min}$ is in F/cm².

Once the substrate doping is obtained, we can determine the flat band capacitance from it. It can be shown that the semiconductor capacitance at flat band C_{FB} (point 2 in Fig. 6-16) is determined from the Debye length capacitance

$$C_{\text{debye}} = \frac{\epsilon_s}{L_D} \quad (6-40)$$

where the Debye length is dependent on doping as described in Eq. (6-25). The overall MOS flat band capacitance, C_{FB} , is the series combination of C_{debye} and C_i . We can thus determine V_{FB} corresponding to the C_{FB} . Once C_i , V_{FB} , and substrate doping are obtained, all terms in the V_T expression (Eq. 6-38) are known. Interestingly, the threshold voltage V_T does not correspond to exactly the minimum of the C - V characteristics, $C_{\min}$, but a slightly higher capacitance marked as point 4 in Fig. 6-16. In fact, it corresponds to the series combination of C_i and $2C_{d\min}$, rather than the series combination of C_i and $C_{d\min}$. The reason for this is that when we change the gate bias around strong inversion, the change of charge in the semiconductor is the sum of the change in depletion charge and the mobile inversion charge, where the two are equal in magnitude at the onset of strong inversion.

We can also determine MOS parameters such as the *fast interface state* density, D_{it} , and mobile ion charges, Q_m , from C - V measurements (Figs. 6-21 and 6-22). The term *fast interface state* refers to the fact that these defects can change their charge state relatively fast in response to changes of the gate bias. As the surface potential in a MOS device is varied, the fast interface

states or traps in the band gap can move above or below the Fermi level in response to the bias, because their positions relative to the band edges are fixed (Fig. 6-21b). Keeping in mind the property of the Fermi-Dirac distribution that energy levels below the Fermi level have a high probability of occupancy by electrons, while levels above the Fermi level tend to be empty, we see that a fast interface state moving above the Fermi level would tend to give up its trapped electron to the semiconductor (or equivalently capture a hole). Conversely, the same fast interface state below the Fermi level captures an electron (or gives up a hole). It obviously makes sense to talk in terms of electrons or holes, depending on which is the majority carrier in the semiconductor. Since charge storage results in capacitance, the fast interface states give rise to a capacitance which is in parallel with the depletion capacitance in the channel (and hence is additive), and this combination is in series with the insulator capacitance C_i . The fast interface states can keep pace with low frequency variations of the gate bias (~ 1 – 1000 Hz), but not at extremely high frequencies (~ 1 MHz). So the fast interface states contribute to the low frequency capacitance C_{LF} , but not the high frequency capacitance C_{HF} . Clearly, from the difference between the two, we ought to be able to compute the fast interface state density. Although we will not do the detailed derivation here, it can be shown that

$$D_{it} = \frac{1}{q} \left(\frac{C_i C_{LF}}{C_i - C_{LF}} - \frac{C_i C_{HF}}{C_i - C_{HF}} \right) \text{ cm}^{-2} \text{ eV}^{-1} \quad (6-41)$$

While the fast interface states can respond quickly to voltage changes, the fixed oxide charges Q_f , as the name implies, do not change their charge state regardless of the gate bias or surface potential. As mentioned above, the effect of these charges on the flat band and threshold voltage depends not only on the number of charges but also their location relative to the oxide-silicon interface (Fig. 6-22). Hence, we must take a weighted sum of these charges, counting charges closer to the oxide-silicon interface more heavily than those that are farther away. This position dependence is the basis of what is called the *bias-temperature stress test* for measuring the mobile ion content, Q_m . We heat up the MOS device to ~ 200 – 300°C (to make the ions more mobile) and apply a positive gate bias to generate a field of ~ 1 MV/cm within the oxide. After cooling the capacitor to room temperature, the C - V characteristics are measured. We have seen how V_{FB} can be determined from the C - V curve, using Eq. (6-40) and $C_i V_{FB}$ is also given by Eq. (6-37). The positive bias repels positive mobile ions such as Na^+ to the oxide-silicon interface so that they contribute fully to a flat band voltage we can call V_{FB}^+ . Next, the capacitor is heated up again, a negative bias is applied so that the ions drift to the gate electrode, and another C - V measurement is made. Now, the mobile ions are too far away to affect the semiconductor band-bending, but induce an equal and opposite charge on the gate electrode. From the resulting C - V , the new flat band, V_{FB}^- , is determined. From

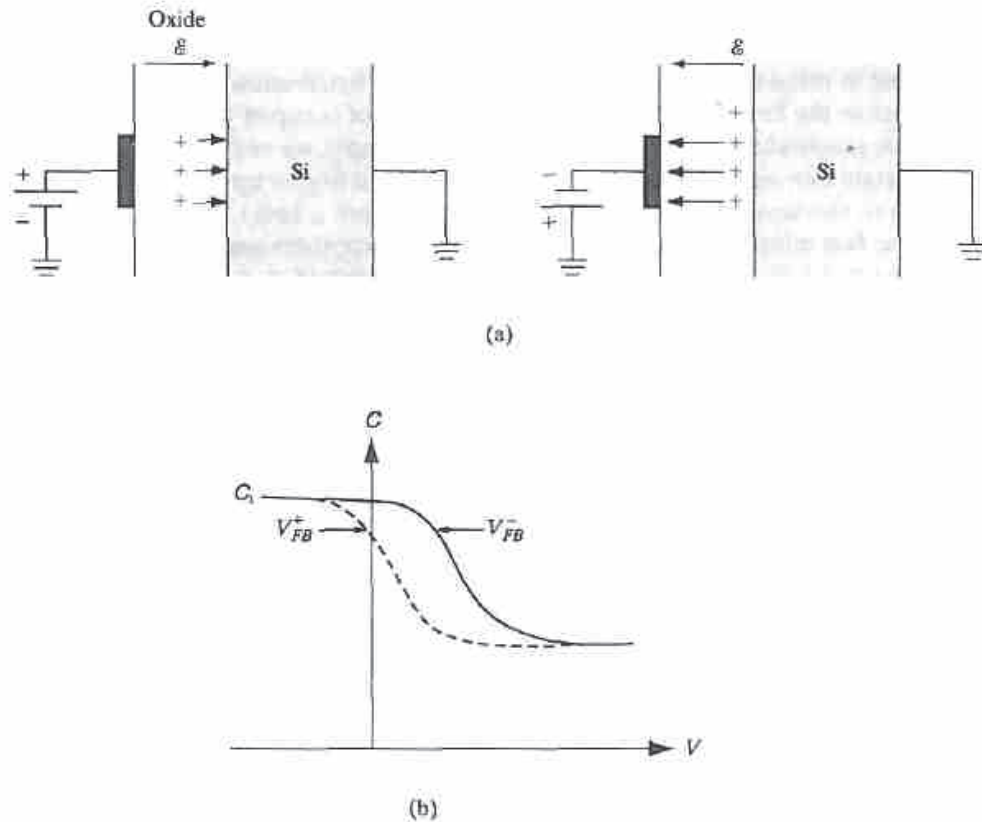


Figure 6-22

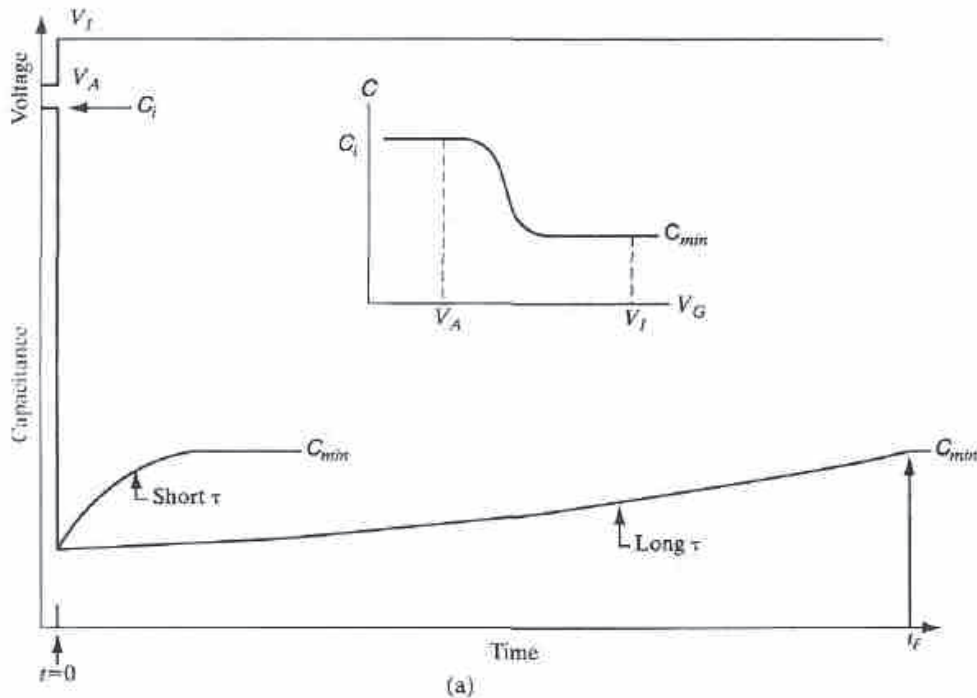
Mobile ion determination: (a) Movement of mobile ions due to positive and negative bias-temperature stress; (b) C - V characteristics under positive (dashed line) and negative (solid line) bias-temperature.

the difference of the two flat band voltages, we can determine the mobile ion content using

$$Q_m = C_i(V_{FB}^- - V_{FB}^+) \quad (6-42)$$

6.4.6 Time-Dependent Capacitance Measurements

During C - V measurements, if the gate bias is varied rapidly from accumulation to inversion, the depletion width can momentarily become greater than the theoretical maximum for gate biases beyond V_T . This phenomenon is known as *deep depletion*, and causes the MOS capacitance to drop below the theoretical minimum, C_{min} , for a transient period. After a time period characteristic of the minority carrier lifetime, which determines

**Figure 6-23**

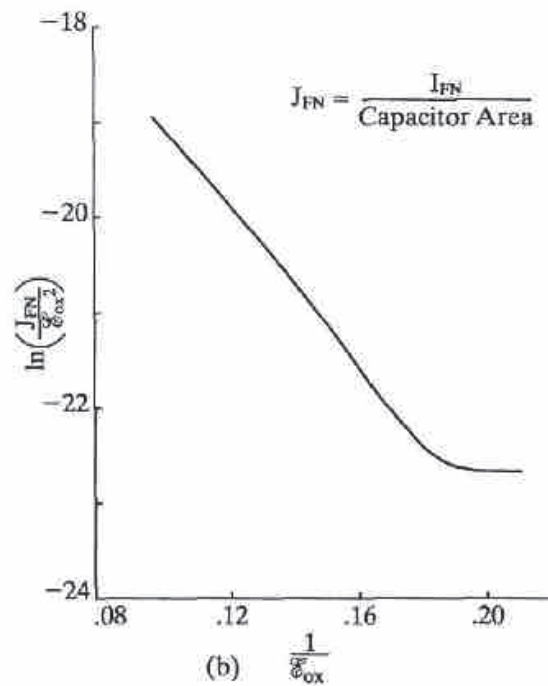
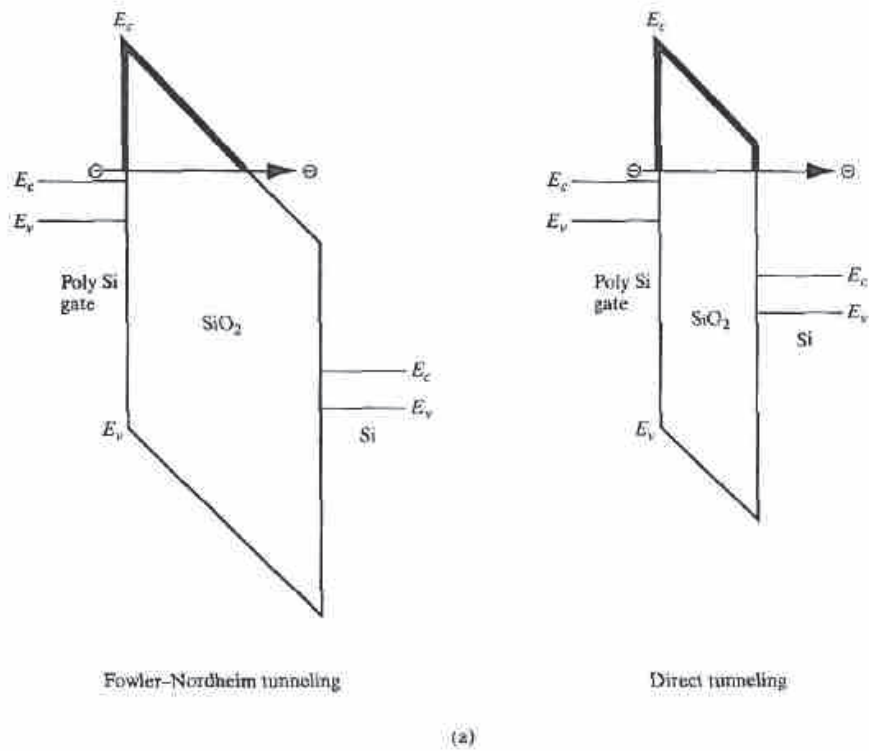
Time-dependent MOS capacitance (C_{HH}) due to the application of a step voltage V_A (which puts the capacitor in accumulation) to V_I (which puts the capacitor in inversion).

the rate of generation of the minority carriers in the MOS device, the depletion width collapses back to the theoretical maximum and the capacitance recovers to C_{min} (Fig. 6-23). This capacitance transient, $C-t$, forms the basis of a powerful technique to measure the lifetime, known as the Zerbst technique.

6.4.7 Current-Voltage Characteristics of MOS Gate Oxides

An ideal gate insulator does not conduct any current, but for real insulators there can be some leakage current which varies with the voltage or electric field across the gate oxide. By looking at the band diagram of the MOS system perpendicular to the oxide-silicon interface (Fig. 6-24), we see that for electrons in the conduction band, there is a barrier, ΔE_c ($= 3.1$ eV). Although electrons with energy less than this barrier cannot go through the oxide classically, it was discussed in Chapter 2 that quantum mechanically electrons can tunnel through a barrier, especially if the barrier thickness is sufficiently small. The detailed calculation of the *Fowler-Nordheim* tunneling current for electrons going from the Si conduction band to the conduction band of SiO_2 , and then having the electrons "hop" along in the oxide to the

Figure 6-24
Current-voltage characteristics of gate oxides: (a) Fowler-Nordheim and direct tunneling through thin gate oxides; (b) plot of Fowler-Nordheim tunneling leakage current as a function of electric field across the oxide.



gate electrode, involves solving the Schrödinger equation for the electron wave function. The Fowler-Nordheim tunneling current I_{FN} can be expressed as a function of the electric field in the gate oxide:

$$I_{FN} \propto \epsilon_{ox}^2 \exp\left(\frac{-B}{\epsilon_{ox}}\right) \quad (6-43)$$

where B is a constant depending on m_n^* and the barrier height.

As gate oxides are made thinner in successive generations of MOSFETs, the tunneling barrier in the gate oxide becomes so thin that the electrons in the conduction band of Si can tunnel through the gate oxide and emerge in the gate, without having to go via the conduction band of the gate oxide. This is known as *direct tunneling* rather than Fowler-Nordheim tunneling. The overall physics is similar, but some of the details are different. For instance, Fowler-Nordheim tunneling involves a triangular barrier, while direct tunneling is through a trapezoidal barrier (Fig. 6-24a). Such tunneling currents are becoming a major problem in modern devices because the useful feature of high input impedance for MOS devices is degraded.

It is necessary to increase the gate capacitance in successive generations of MOSFETs in order to increase the drain current, as discussed later in Section 6.5. One can achieve higher C_i by using insulators with a dielectric constant higher than SiO_2 instead of reducing the gate oxide thickness, d , as can be seen from Eq. (6-29). The advantage of not reducing d too much is that it keeps the tunneling barrier wide, and the gate oxide field low, thereby keeping the gate tunneling leakage current low (Eq. 6-43). Such high dielectric constant insulators such as hafnium dioxide, HfO_2 , are also known as high- k gate dielectrics because k is also sometimes used as a symbol for dielectric constant.

Prolonged charge transport through gate oxides can ultimately cause catastrophic electrical breakdown of the oxides. This is known as *Time-Dependent Dielectric Breakdown (TDDB)*. One of the popular models that explains this degradation involves electrons tunneling into the conduction band of the gate oxide from the negative electrode (cathode), then gaining energy from the electric field, thus becoming "hot" electrons in the gate oxide. If they gain sufficient energy, they can cause impact ionization within the oxide and create electron-hole pairs. The electrons are accelerated toward the (positive) Si substrate, while the holes travel toward the gate. However, electron and hole mobilities are extremely small in SiO_2 . Hole mobilities are particularly low ($\sim 0.01 \text{ cm}^2/\text{Vs}$). Hence, there is a great propensity for these impact-generated holes to be trapped at defect sites within the oxide, near the cathode. The resulting band diagram (Fig. 6-25) is altered by this sheet of trapped positive charge, which causes the internal electric field between this point and the gate to increase. A similar distortion of the electric field near the Si anode is created by the trapped impact-generated electrons. However, the steepest slope in Fig. 6-25, and therefore the highest field, is near the gate. As a result, the

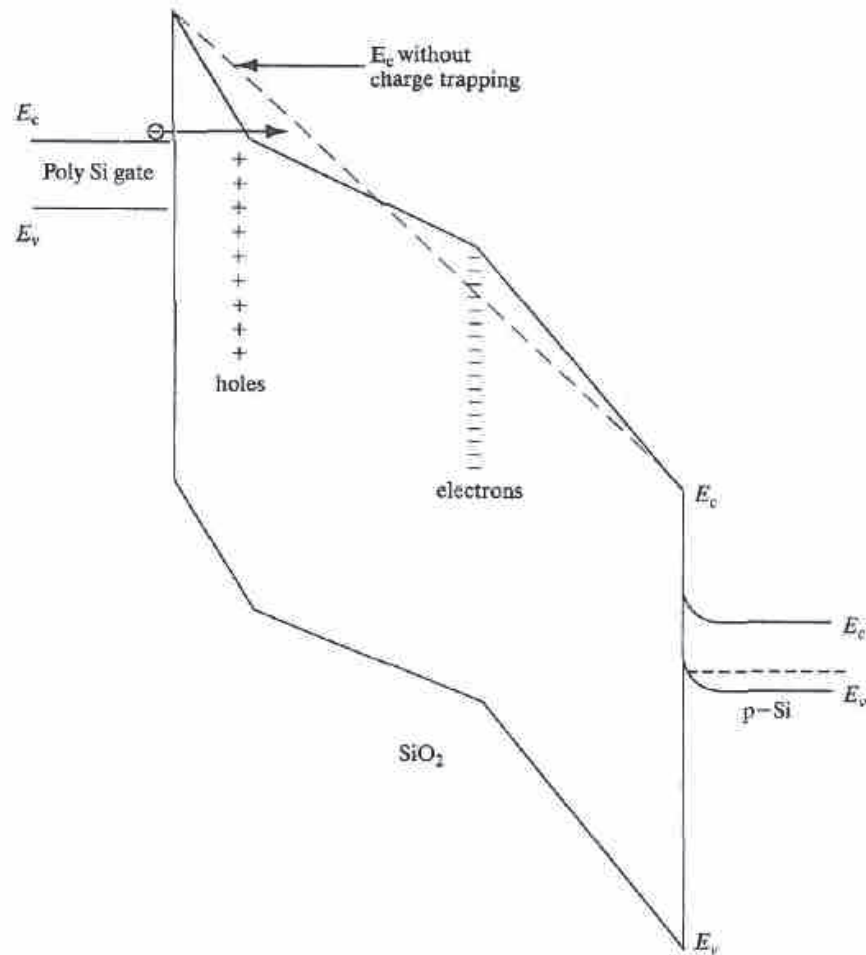


Figure 6-25

Time-dependent dielectric breakdown of oxides: Band diagram of a MOS device showing the band edges in the polysilicon gate, oxide, and Si substrate. Trapped holes and electrons in the oxide distort the band edges, and increase the electric field in the oxide near the gate. The tunneling barrier width is seen to be less than if there were no charge trapping (dashed line).

barrier for electron tunneling from the gate into the oxide is reduced. More electrons can tunnel into the oxide, and cause more impact ionization. We get a positive feedback effect that can lead to a runaway TDDB process.

6.5 THE MOS FIELD- EFFECT TRANSISTOR

The MOS transistor is also called a surface field-effect transistor, since it depends on control of current through a thin channel at the surface of the semiconductor (Fig. 6-10). When an inversion region is formed under the gate, current can flow from drain to source (for an n-channel device). In this section

we analyze the conductance of this channel and find the $I_D - V_D$ characteristics as a function of gate voltage V_G . As in the JFET case, we will find these characteristics below saturation and then assume I_D remains essentially constant above saturation.

6.5.1 Output Characteristics

The applied gate voltage V_G is accounted for by Eq. (6-28) plus the voltage required to achieve flat band:

$$V_G = V_{FB} - \frac{Q_s}{C_i} + \phi_s \quad (6-44)$$

The induced charge Q_s in the semiconductor is composed of mobile charge Q_n and fixed charge in the depletion region Q_d . Substituting $Q_n + Q_d$ for Q_s , we can solve for the mobile charge:

$$Q_n = -C_i \left[V_G - \left(V_{FB} + \phi_s - \frac{Q_d}{C_i} \right) \right] \quad (6-45)$$

At threshold the term in brackets can be written $V_G - V_T$ from Eq. (6-38).

With a voltage V_D applied, there is a voltage rise V_x from the source to each point x in the channel. Thus the potential $\phi_s(x)$ is that required to achieve strong inversion ($2\phi_F$) plus the voltage V_x :

$$Q_n = -C_i \left[V_G - V_{FB} - 2\phi_F - V_x - \frac{1}{C_i} \sqrt{2q\epsilon_s N_a (2\phi_F + V_x)} \right] \quad (6-46)$$

If we neglect the variation of $Q_d(x)$ with bias V_x , Eq. (6-46) can be simplified to

$$Q_n(x) = -C_i (V_G - V_T - V_x) \quad (6-47)$$

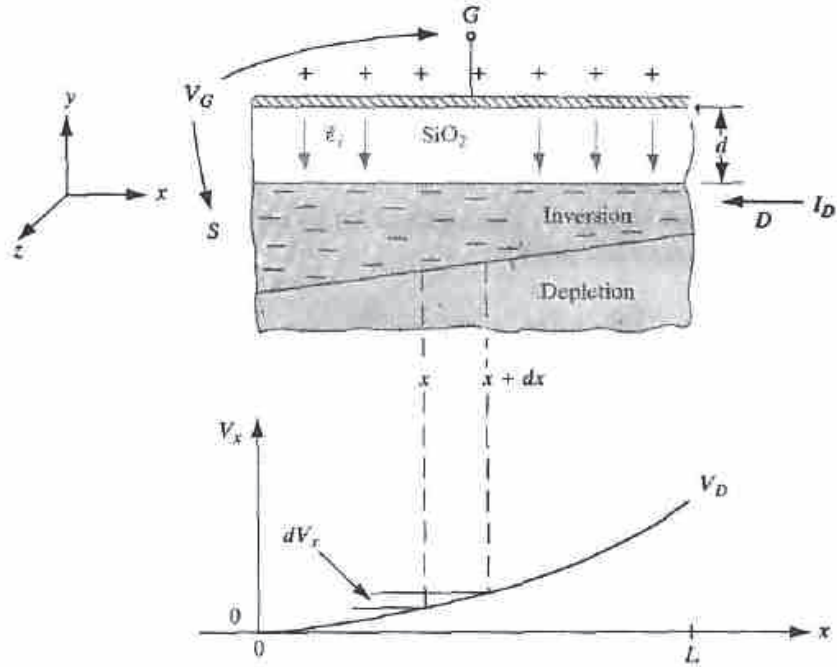
This equation describes the mobile charge in the channel at point x (Fig. 6-26). The conductance of the differential element dx is $\bar{\mu}_n Q_n(x) Z / dx$, where Z is the width of the channel and $\bar{\mu}_n$ is a *surface* electron mobility (indicating the mobility in a thin region near the surface is not the same as in the bulk material). At point x we have

$$I_D dx = \bar{\mu}_n Z |Q_n(x)| dV_x \quad (6-48)$$

Integrating from source to drain,

$$\begin{aligned} \int_0^L I_D dx &= \bar{\mu}_n Z C_i \int_0^{V_D} (V_G - V_T - V_x) dV_x \\ I_D &= \frac{\bar{\mu}_n Z C_i}{L} \left[(V_G - V_T) V_D - \frac{1}{2} V_D^2 \right] \end{aligned} \quad (6-49)$$

Figure 6-26
Schematic view of
the n-channel re-
gion of a MOS
transistor under
bias below pinch-
off, and the varia-
tion of voltage V_x
along the conduct-
ing channel.



where

$$\frac{\bar{\mu}_n Z C_i}{L} = k_N$$

determines the conductance and transconductance of the *n*-channel MOSFET [see Eqs. (6-51) and (6-54)].

In this analysis the depletion charge Q_d in the threshold voltage V_T is simply the value with no drain current. This is an approximation, since $Q_d(x)$ varies considerably when V_D is applied, to reflect the variation in V_x (see Fig. 6-26b). However, Eq. (6-49) is a fairly accurate description of drain current for low values of V_D , and is often used in approximate design calculations because of its simplicity. A more accurate and general expression is obtained by including the variation of $Q_d(x)$. Performing the integration of Eq. (6-48) using Eq. (6-46) for $Q_n(x)$, one obtains

$$I_D = \frac{\bar{\mu}_n Z C_i}{L} \times \left\{ (V_G - V_{FB} - 2\phi_F - \frac{1}{2}V_D)V_D - \frac{2}{3} \frac{\sqrt{2\epsilon_n q N_a}}{C_i} [(V_D + 2\phi_F)^{3/2} - (2\phi_F)^{3/2}] \right\} \quad (6-50)$$

The drain characteristics that result from these questions are shown in Fig. 6-10c. If the gate voltage is above threshold ($V_G > V_T$), the drain current

is described by Eq. (6-50) or approximately by Eq. (6-49) for low V_D . Initially the channel appears as an essentially linear resistor, dependent on V_G . The conductance of the channel in this linear region can be obtained from Eq. (6-49) with $V_D \ll (V_G - V_T)$:

$$g = \frac{\partial I_D}{\partial V_D} \approx \frac{Z}{L} \bar{\mu}_n C_i (V_G - V_T) \quad (6-51)$$

where $V_G > V_T$ for a channel to exist.

As the drain voltage is increased, the voltage across the oxide decreases near the drain, and Q_n becomes smaller there. As a result the channel becomes pinched off at the drain end, and the current saturates. The saturation condition is approximately given by

$$V_D(\text{sat.}) \approx V_G - V_T \quad (6-52)$$

The drain current at saturation remains essentially constant for larger values of drain voltage. Substituting Eq. (6-52) into Eq. (6-49), we obtain

$$I_D(\text{sat.}) \approx \frac{1}{2} \bar{\mu}_n C_i \frac{Z}{L} (V_G - V_T)^2 = \frac{Z}{2L} \bar{\mu}_n C_i V_D^2(\text{sat.}) \quad (6-53)$$

for the approximate value of drain current at saturation.

The transconductance in the saturation range can be obtained approximately by differentiating Eq. (6-53) with respect to the gate voltage:

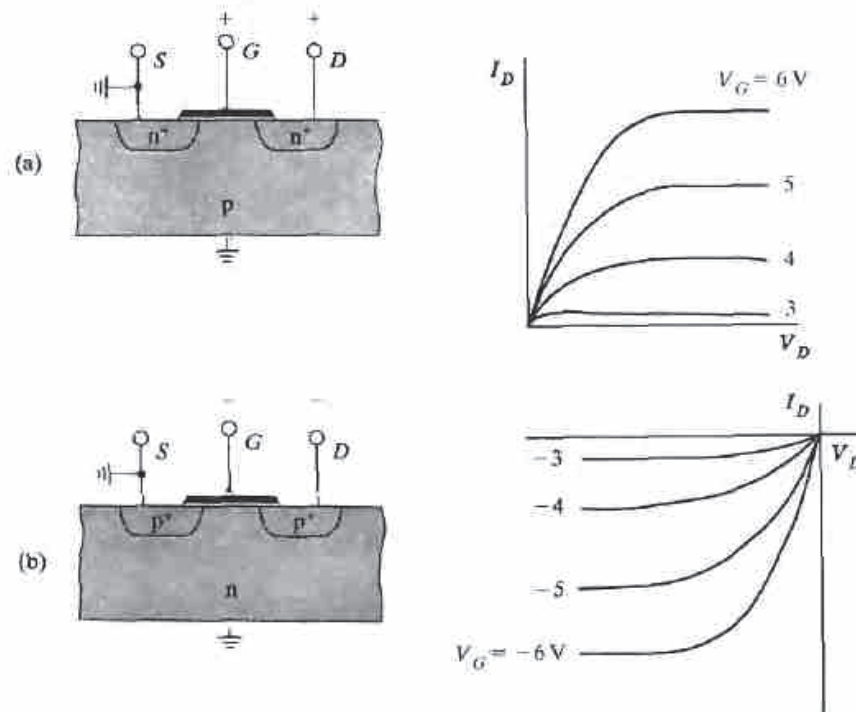
$$g_m(\text{sat.}) = \frac{\partial I_D(\text{sat.})}{\partial V_G} \approx \frac{Z}{L} \bar{\mu}_n C_i (V_G - V_T) \quad (6-54)$$

The derivations presented here are based on the n-channel device. For the p-channel enhancement transistor the voltages V_D , V_G , and V_T are negative, and current flows from source to drain (Fig. 6-27).

6.5.2 Transfer Characteristics

The output characteristics plot the drain current as a function of the drain bias, with gate bias as a parameter (Fig. 6-27). On the other hand, the *transfer* characteristics plot the output drain current as a function of the input gate bias, for fixed drain bias (Fig. 6-28a). Clearly, in the linear region, I_D versus V_G should be a straight line from Eq. (6-49). The intercept of this line on the V_G axis is the linear region threshold voltage, V_T (lin.) and the slope (divided by the applied V_D) gives us the linear value of k_N , $k_N(\text{lin.})$, of the n-channel MOSFET. If we look at actual data, however, we see that while the characteristics are approximately linear at low gate bias, at high gate biases the drain current increases sub-linearly. The transconductance, g_m (lin.), in the linear region can be obtained by differentiating the right-hand side of Eq. (6-49) with respect to gate bias.

Figure 6-27
Drain current-voltage characteristics for enhancement transistors: (a) for n-channel V_D , V_G , V_T , and I_D are positive; (b) for p-channel all these quantities are negative.



The g_m (lin.) is plotted as a function of V_G in Fig. 6-28b. It may be noted that the transconductance is zero below V_T because there is little drain current. It goes through a maximum at the point of inflection of the I_D - V_G curve, and then decreases. This decrease is due to two factors that will be discussed in Sections 6.5.3 and 6.5.8: degradation of the effective channel mobility as a function of increasing transverse electric field across the gate oxide, and source/drain series resistance.

For the transfer characteristics in the saturation region, since Eq. (6-53) shows a quadratic dependence of I_D on V_G , we get a linear behavior by plotting not the drain current, but rather the square root of I_D , as a function of V_G (Fig. 6-29). In this case the intercept gives us the threshold voltage in the saturation region, $V_{T(\text{sat.})}$. We shall see in Section 6.5.10 that due to effects such as drain-induced barrier lowering (DIBL), for short channel length MOSFETs the $V_{T(\text{sat.})}$ can be lower than $V_{T(\text{lin.})}$, while the long channel values are similar. Similarly, the slope of the transfer characteristics can be used to determine the value of k_N in the saturation region, $k_{N(\text{sat.})}$ for the n-channel MOSFET, which can be different from $k_{N(\text{lin.})}$ for short channel devices.

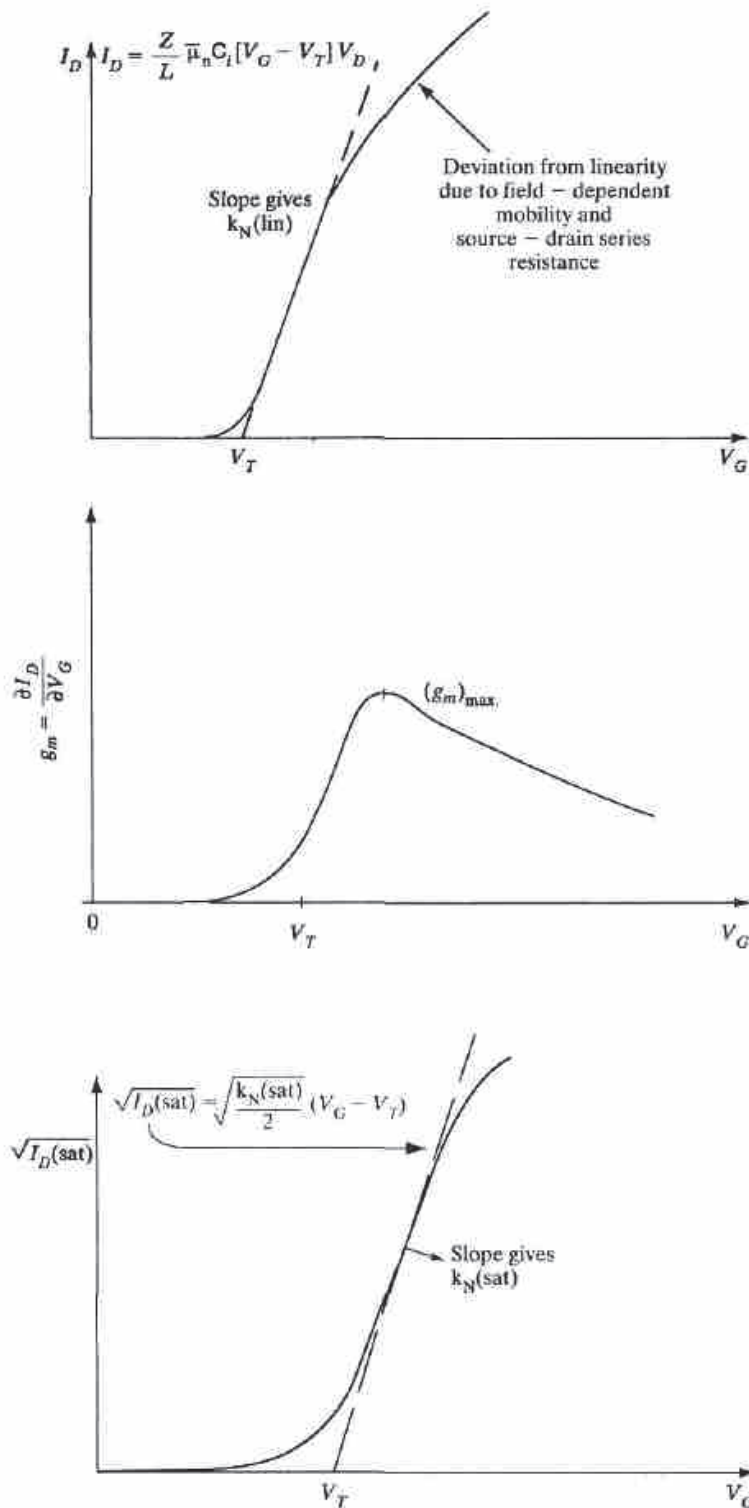


Figure 6-28
Linear region transfer characteristics: (a) plot of drain current versus gate voltage for MOSFETs in the linear region; (b) transconductance as a function of gate bias.

Figure 6-29
Saturation region transfer characteristics: plot of square root of the drain current versus gate voltage for MOSFETs.

EXAMPLE 6-2

For an n-channel MOSFET with a gate oxide thickness of 10 nm, $V_T = 0.6$ V, and $Z = 25$ μm , $L = 1$ μm . Calculate the drain current at $V_G = 5$ V and $V_D = 0.1$ V. Repeat for $V_G = 3$ V and $V_D = 5$ V. Discuss what happens for $V_D = 7$ V. Assume an electron channel mobility of $\bar{\mu}_n = 200$ $\text{cm}^2/\text{V}\cdot\text{s}$.

SOLUTION

$$C_i = \frac{\epsilon_i}{d} = \frac{(3.9)(8.85 \times 10^{-14})}{10^{-6}} = 3.45 \times 10^{-7} \text{ F/cm}^2$$

For $V_G = 5$ V and it follows that $V_D = 0.1$ V, and $V_T = 0.6$ V, $V_D < (V_G - V_T)$, we are in the linear region. Thus,

$$\begin{aligned} I_D &= \frac{Z}{L} \bar{\mu}_n C_i \left[(V_G - V_T) V_D - \frac{1}{2} V_D^2 \right] \\ &= \frac{25}{1} (200) (3.45 \times 10^{-7}) \left[(5 - 0.6) \times 0.1 - \frac{1}{2} (0.1)^2 \right] = 7.51 \times 10^{-4} \text{ A} \end{aligned}$$

For $V_G = 3$ V, $V_D = 5$ V, $V_{D(\text{sat.})} = V_G - V_T = 3 - 0.6 = 2.4$ V

$$\begin{aligned} I_D &= \frac{Z}{L} \bar{\mu}_n C_i \left[(V_G - V_T) V_{D(\text{sat.})} - \frac{1}{2} V_{D(\text{sat.})}^2 \right] \\ &= \frac{25}{1} (200) (3.45 \times 10^{-7}) \left[(2.4)^2 - \frac{1}{2} (2.4)^2 \right] = 4.97 \times 10^{-3} \text{ A} \end{aligned}$$

For $V_D = 7$ V, I_D will not increase, because we are in the saturation region.

6.5.3 Mobility Models

The mobility of carriers in the channel of a MOSFET is lower than in bulk semiconductors because there are additional scattering mechanisms. Since carriers in the channel are very close to the semiconductor-oxide interface, they are scattered by surface roughness and by coulombic interaction with fixed charges in the gate oxide. When the carriers travel in the inversion layer from the source to the drain, they encounter microscopic roughness on an atomistic scale at the oxide-silicon interface and undergo scattering because, as discussed in Section 3.4.1, any deviation from a perfectly periodic crystal potential results in scattering. This mobility degradation increases with the gate bias because a higher gate bias draws the carriers closer to the oxide-silicon interface, where they are more influenced by the interfacial roughness.

It is very interesting to note that if we plot the effective carrier mobility in the MOSFET as a function of the average transverse electric field in the middle of the inversion layer, we get what is known as a “universal”

mobility degradation curve for any MOSFET, which is independent of the technology or device structural parameters such as oxide thickness and channel doping (Fig. 6-30). We can apply Gauss's law to the region marked by the colored box in Fig. 6-31, which encloses all the depletion charge and half of the inversion charge in the channel. We see that the average transverse field in the middle of the inversion region is given by

$$\mathcal{E}_{eff} = \frac{1}{\epsilon_s} \left(Q_d + \frac{1}{2} Q_n \right) \quad (6-55a)$$

While this model works quite well for electrons, for reasons that are not clearly understood at present, it has to be modified slightly for holes in the sense that the average transverse field must now be defined as

$$\mathcal{E}_{eff} = \frac{1}{\epsilon_s} \left(Q_d + \frac{1}{3} Q_n \right) \quad (6-55b)$$

This degradation of mobility with gate bias is often compactly described by writing the drain current expression as

$$I_D = \frac{\bar{\mu}_n Z C_i}{L \{1 + \theta(V_G - V_T)\}} \left[(V_G - V_T)V_D - \frac{1}{2} V_D^2 \right] \quad (6-56)$$

where θ is called the *mobility degradation parameter*. Because of the additional $(V_G - V_T)$ term in the denominator, the drain current increases sub-linearly with gate bias for high gate voltages.

In addition to this dependence of the channel mobility on gate bias or transverse electric field, there is also a strong dependence on drain bias or the

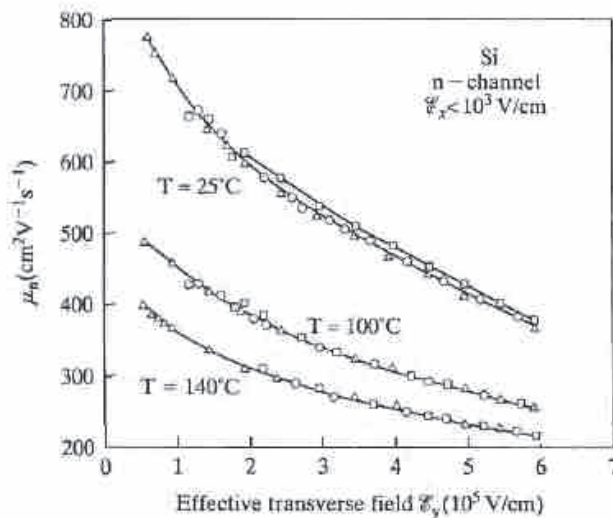


Figure 6-30
Inversion layer
electron mobility
versus effective
transverse field, at
various tempera-
tures. The trian-
gles, circles, and
squares refer to
different MOSFETs
with different gate
oxide thicknesses
and channel dop-
ings. (After Sabnis
and Clemens,
IEEE IEDM,
1979.)

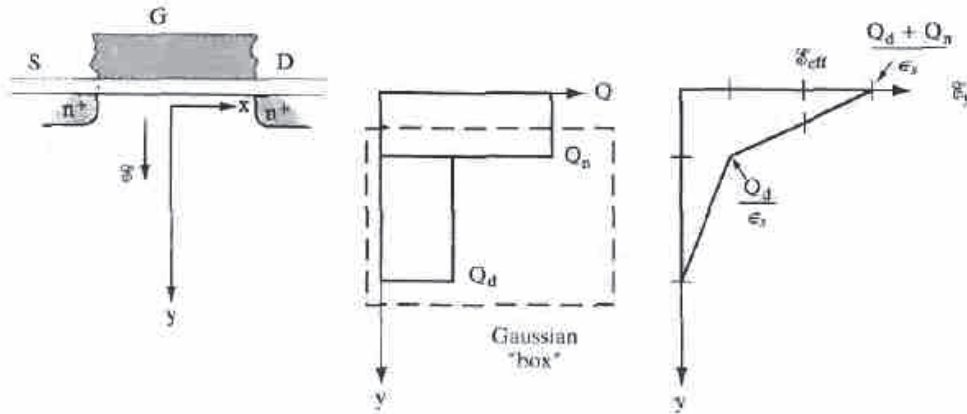


Figure 6-31

Determination of effective transverse field. Idealized charge distribution and transverse electric field in the inversion layer and depletion layer, as a function of depth in the channel of a MOSFET. The region to which we apply Gauss's law is shown in color.

longitudinal electric field. As shown in Fig. 3-24, the carrier drift velocity increases linearly with electric field (ohmic behavior) until the field reaches $\mathcal{E}_{\text{sat}}$; in other words, the mobility is constant up to $\mathcal{E}_{\text{sat}}$. After this, the velocity saturates at v_{sat} , and it can no longer be described in terms of mobility. These effects can be described as:

$$v = \mu \mathcal{E} \text{ for } \mathcal{E} < \mathcal{E}_{\text{sat}} \quad (6-57)$$

$$\text{and } v = v_{\text{sat}} \text{ for } \mathcal{E} > \mathcal{E}_{\text{sat}} \quad (6-58)$$

The maximum longitudinal electric field near the drain end of the channel is approximately given by the voltage drop along the pinch-off region ($V_D - V_D(\text{sat.})$), divided by the length of the pinch-off region, ΔL :

$$\mathcal{E}_{\text{max}} = \left(\frac{V_D - V_D(\text{sat.})}{\Delta L} \right) \quad (6-59)$$

From a two-dimensional solution of the Poisson equation near the drain end, one can show that the pinch-off region ΔL shown in Fig. 6-11c is approximately equal to $\sqrt{(3dx_j)}$, where d is the gate oxide thickness and x_j is the source/drain junction depth. The factor of 3 is due to the ratio of the dielectric constant for Si to that of SiO_2 .

6.5.4 Short Channel MOSFET I-V Characteristics

In short channel devices, the analysis has to be somewhat modified. As mentioned in the previous section, the effective channel mobility decreases with increasing transverse electric field perpendicular to the gate oxide (i.e., the

gate bias). Furthermore, for very high longitudinal electric fields in the pinch-off region, the carrier velocity saturates (Fig. 3-24). For short channel lengths, the carriers travel at the saturation velocity over most of the channel. In that case, the drain current is given by the width times the channel charge per unit area times the saturation velocity:

$$I_D(\text{sat.}) \approx ZC_i(V_G - V_T)v_s \quad (6-60)$$

As a result, the saturation drain current does not increase quadratically with $(V_G - V_T)$ as shown in Eq. (6-53), but rather shows a linear dependence (note the equal spacing of curves in Fig. 6-32). Due to the advances in Si device processing, particularly photolithography, MOSFETs used in modern integrated circuits tend to have short channels, and are commonly described by Eq. (6-60) rather than Eq. (6-53).

6.5.5 Control of Threshold Voltage

Since the threshold voltage determines the requirements for turning the MOS transistor on or off, it is very important to be able to adjust V_T in designing the

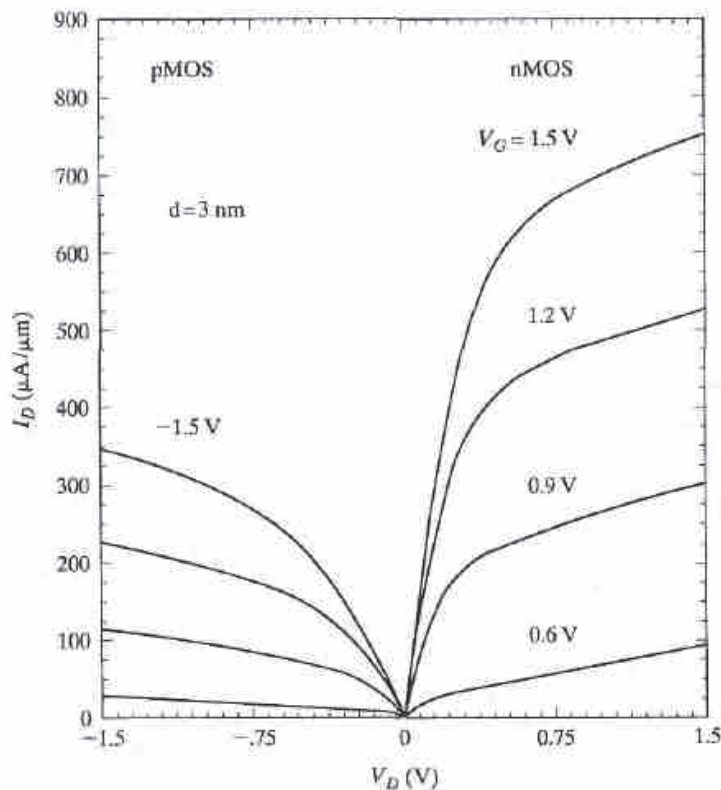


Figure 6-32
Experimental output characteristics of n-channel and p-channel MOSFETs with 0.1- μm channel lengths. The curves exhibit almost equal spacing, indicating a linear dependence of I_D on V_G , rather than a quadratic dependence. We also see that I_D is not constant but increases somewhat with V_D in the saturation region. The p-channel devices have lower currents because hole mobilities are lower than electron mobilities.

device. For example, if the transistor is to be used in a circuit driven by a 3-V battery, it is clear that a 4-V threshold voltage is unacceptable. Some applications require not only a low value of V_T , but also a precisely controlled value to match other devices in the circuit.

All of the terms in Eq. (6-38) can be controlled to some extent. The work function potential difference Φ_{ms} is determined by choice of the gate conductor material; Φ_F depends on the substrate doping; Q_i can be reduced by proper oxidation methods and by using Si grown in the (100) orientation; Q_d can be adjusted by doping of the substrate; and C_i depends on the thickness and dielectric constant of the insulator. We shall discuss here several methods of controlling these quantities in device fabrication.

Choice of Gate Electrode. Since V_T depends on Φ_{ms} , the choice of the gate electrode material (i.e., the gate electrode work function) has an impact on the threshold voltage. When MOSFETs were first made in the 1960s, they used Al gates. However, since Al has a low melting point, it precluded the use of a self-aligned source/drain technology because that required a high temperature source/drain implant anneal after the gate formation. Hence, Al was supplanted by n^+ doped LPCVD polysilicon *refractory* (high melting point) gates, where the Fermi level lines up with the conduction band edge in Si. While this works quite well for n-channel MOSFETs, we shall see in Section 9.3.1 that it can create problems for p-channel MOSFETs. Therefore, sometimes, a p^+ doped polysilicon gate is used for p-channel devices. Refractory metal gates with suitable work functions are also being researched as possible replacements for doped polysilicon. One attractive candidate is tungsten, whose work function is such that the Fermi level happens to lie near the midgap of Si.

Control of C_i . Since a low value of V_T and a high drive current is usually desired, a thin oxide layer is used in the gate region to increase $C_i = \epsilon_i/d$ in Eq. (6-38). From Fig. 6-20 we see that increasing C_i makes V_T less negative for p-channel devices and less positive for n-channel with $-Q_d > Q_i$. For practical considerations, the gate oxide thickness is generally 20-100 Å (2-10 nm) in modern devices having submicron gate length. An example of such a device is shown in Figure 6-33. The gate oxide, easily observable in this micrograph, is 40 Å thick. The interfacial layer between the crystalline silicon and the amorphous SiO_2 is also observable.

Although a low threshold voltage is desirable in the gate region of a transistor, a large value of V_T is needed between devices. For example, if a number of transistors are interconnected on a single Si chip, we do not want inversion layers to be formed inadvertently between devices (generally called the *field*). One way to avoid such parasitic channels is to increase V_T in the field by using a very thick oxide. Figure 6-34 illustrates a transistor with a gate oxide 10 nm thick and a field oxide of 0.5 μm .

The value of C_i can also be controlled by varying ϵ_i . A SiO_2 layer which has some N incorporated in it, leading to the formation of a silicon

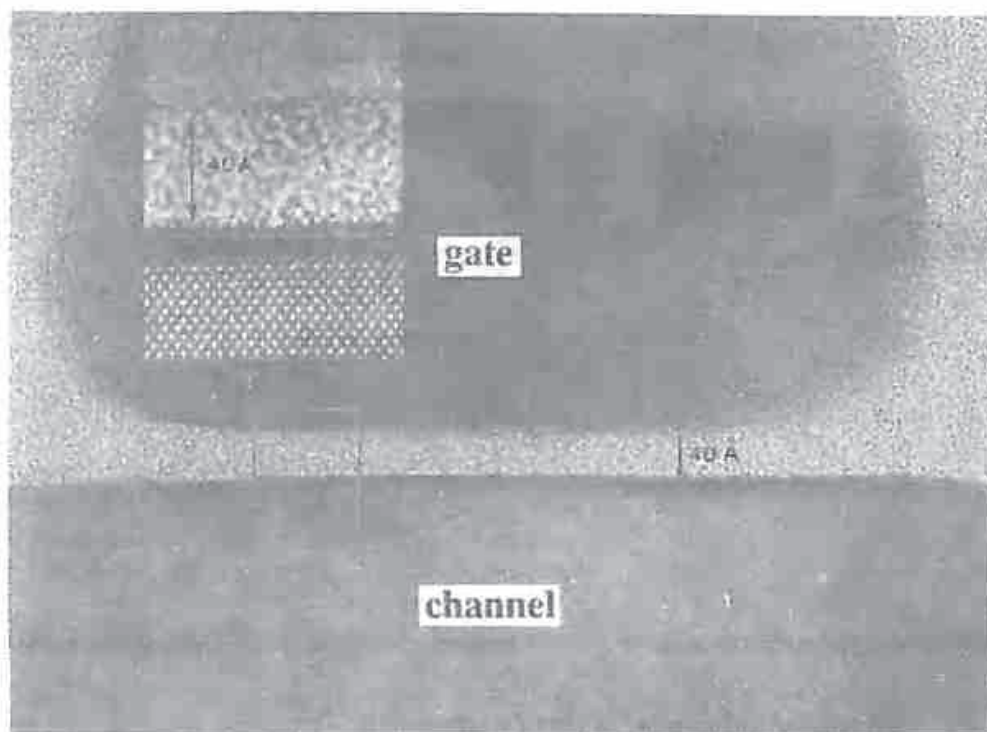


Figure 6-33

Cross section of a MOSFET. This high resolution transmission electron micrograph of a silicon Metal-Oxide Semiconductor Field Effect Transistor shows the silicon channel and metal gate separated by a thin (40 Å, 4 nm) silicon-dioxide insulator. The inset shows a magnified view of the three regions, in which individual rows of atoms in the crystalline silicon can be distinguished. (Photograph courtesy of AT&T Bell Laboratories.)

oxynitride, is often used. Such silicon oxynitrides have slightly higher ϵ_i and C_i than SiO_2 , with excellent interface properties. Other high dielectric constant materials such as Ta_2O_5 , ZrO_2 , and ferroelectrics (e.g., barium-strontium-titanate) are also being investigated as replacements for SiO_2 as the gate dielectric in MOSFETs in order to increase $C_i = \epsilon_i/d$ and, therefore, the drive current of the MOSFET. Generally speaking, we cannot use these high dielectric constant materials directly on the Si substrate; a very thin (~ 0.5 nm) interfacial SiO_2 layer is needed to achieve a low fast interface state density. It is clear from the expression for C_i that for these high dielectric constant materials, a physically thicker layer, d , can be used than for SiO_2 and still achieve a certain C_i . This is very useful for reducing the tunneling leakage current through the gate dielectric, discussed in Section 6.4.7. A physically thicker layer implies a wider tunneling barrier with a reduced tunneling probability.

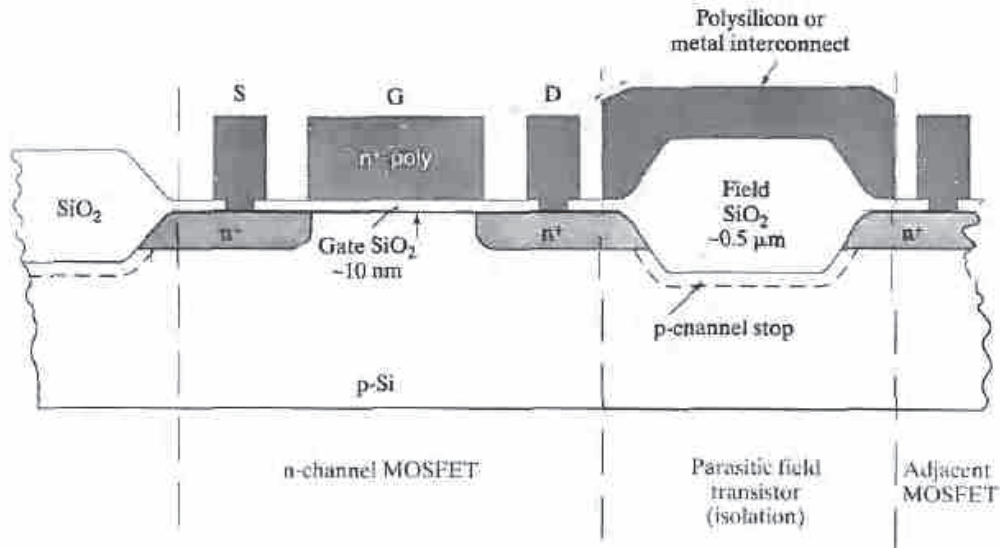


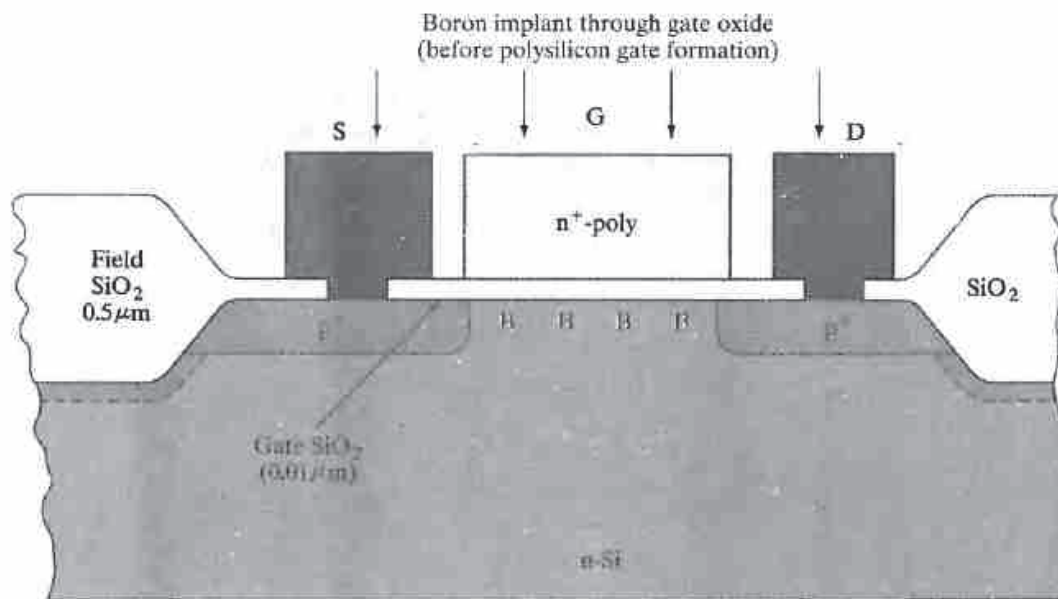
Figure 6-34

Thin oxide in the gate region and thick oxide in the field-between transistors for V_T control (not to scale).

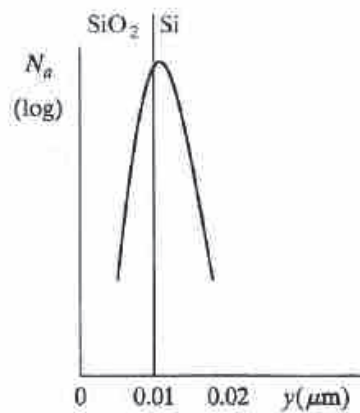
Threshold Adjustment by Ion Implantation. The most valuable tool for controlling threshold voltage is ion implantation (Section 5.1.4). Since very precise quantities of impurity can be introduced by this method, it is possible to maintain close control of V_T . For example, Fig. 6-35 illustrates a boron implantation through the gate oxide of a p-channel device such that the implanted peak occurs just below the Si surface. The negatively charged boron acceptors serve to reduce the effects of the positive depletion charge Q_d . As a result, V_T becomes less negative. Similarly, a shallow boron implant into the p-type substrate of an n-channel transistor can make V_T positive, as required for an enhancement device.

If the implantation is performed at higher energy, or into the bare Si instead of through an oxide layer, the impurity distribution lies deeper below the surface. In such cases the essentially gaussian impurity concentration profile cannot be approximated by a spike at the Si surface. Therefore, effects of distributed charge on the Q_d term of Eq. (6-38) must be considered. Calculations of the effects on V_T in this case are more complicated, and the shift of threshold voltage with implantation dose is often obtained empirically instead.

The implantation energy required for shallow V_T adjustment implants is low (50–100 keV), and relatively low doses are needed. A typical V_T adjustment requires only about 10 s of implantation for each wafer, and therefore this procedure is compatible with large-scale production requirements.



(a)



(b)

Figure 6-35

Adjustment of V_T in a p-channel transistor by boron implantation: (a) boron ions are implanted through the thin gate oxide but are absorbed within the thick oxide regions; (b) variation of implanted boron concentration in the gate region—here the peak of the boron distribution lies just below the Si surface.

EXAMPLE 6-3

For a p-channel transistor with a gate oxide thickness of 10 nm, calculate the boron ion dose F_B (B^+ ions/cm²) required to reduce V_T from -1.1 V to -0.5 V. Assume that the implanted acceptors form a sheet of negative charge just below the Si surface. If, instead of a shallow B implant, it was a much broader distribution, how would the V_T calculation change? Assuming a boron ion beam current of 10^{-5} A, and supposing that the area scanned by the ion beam is 650 cm², how long does this implant take?

SOLUTION

Since the B negative ions are assumed to form a sheet of charge just below the gate oxide, they affect the fixed oxide charge term in V_{FB} . Thus,

$$C_i = \frac{\epsilon_i}{d} = \frac{(3.9)(8.85 \times 10^{-14})}{10^{-6}} = 3.45 \times 10^{-7} \text{ F/cm}^2$$

$$-0.5 = -1.1 + \frac{qF_B}{C_i}$$

$$F_B = \frac{3.45 \times 10^{-7}}{1.6 \times 10^{-19}}(0.6) = 1.3 \times 10^{12} \text{ cm}^{-2}$$

If the B distribution is deeper, we cannot assume it to be a sheet charge in V_{FB} . If it is approximately *constant over the maximum depletion width*, we should instead change the substrate doping term in the expression for V_T . Otherwise, the calculation has to be done by numerically solving Poisson's equation in the depletion region, with a varying doping concentration corresponding to the B distribution, to obtain the voltage drop in the semiconductor.

For a beam current of 10 μ A scanned over a 650-cm² target area,

$$\frac{10^{-5}(\text{C/s})}{650 \text{ cm}^2} t(\text{s}) = 1.3 \times 10^{12} (\text{ions/cm}^2) \times 1.6 \times 10^{-19} (\text{C/ion})$$

The implant time is $t = 13.5$ s.

If the implantation is continued to higher doses, V_T can be moved past zero to the *depletion-mode* condition (Fig. 6-36). This capability provides considerable flexibility to the integrated-circuit designer, by allowing enhancement- and depletion-mode devices to be incorporated on the same chip. For example, a depletion-mode transistor can be used instead of a resistor as a load element for the enhancement device. Thus an array of MOS transistors can be fabricated in an IC layout, with some adjusted by implantation to have the desired enhancement mode V_T and others implanted to become depletion loads.

As mentioned, V_T control is important not only in the MOSFETs but also in the isolation or field regions. In addition to using a thick field oxide, we can do a *channel stop implant* (so called because it stops turning on a parasitic channel in the isolation regions) selectively in the isolation regions under the

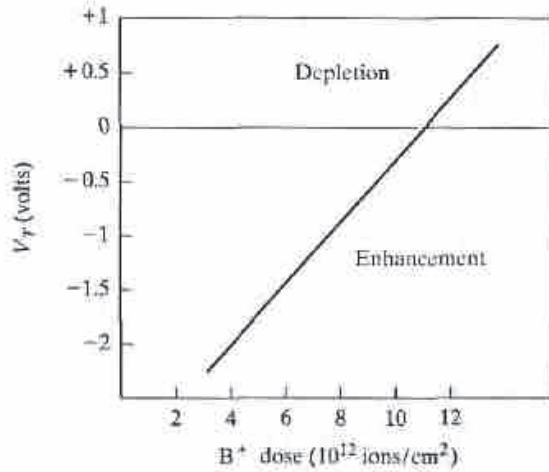


Figure 6-36
Typical variation of V_T for a p-channel device with increased implanted boron dose. The originally enhancement p-channel transistor becomes a depletion-mode device ($V_T > 0$) by sufficient B implantation.

field oxide (Fig. 6-34). Generally, a B channel stop implant is used for n-channel devices. (It must be noted that such an acceptor implant will raise the field thresholds for n-channel MOSFETs made in a p-substrate, but will *decrease* the field thresholds for p-MOSFETs made in an n-substrate).

6.5.6 Substrate Bias Effects

In the derivation of Eq. (6-49) for current along the channel, we assumed that the source S was connected to the substrate B (Fig. 6-27). In fact, it is possible to apply a voltage between S and B (Fig. 6-37). With a reverse bias between the substrate and the source (V_B negative for an n-channel device), the depletion region is widened and the threshold gate voltage required to achieve inversion must be increased to accommodate the larger Q_d . A simplified view of the result is that W is widened uniformly along the channel, so that Eq. (6-32) should be changed to

$$Q_d = -[2\epsilon_s q N_a (2\phi_F - V_B)]^{1/2} \quad (6-61)$$

The change in threshold voltage due to the substrate bias is

$$\Delta V_T = \frac{\sqrt{2\epsilon_s q N_a}}{C_i} [(2\phi_F - V_B)^{1/2} - (2\phi_F)^{1/2}] \quad (6-62)$$

If the substrate bias V_B is much larger than $2\phi_F$ (typically ~ 0.6 V), the threshold voltage is dominated by V_B and

$$\Delta V_T \approx \frac{\sqrt{2\epsilon_s q N_a}}{C_i} (-V_B)^{1/2} \quad (\text{n channel}) \quad (6-63)$$

where V_B will be negative for the n-channel case. As the substrate bias is increased, the threshold voltage becomes more positive. The effect of this bias becomes more dramatic as the substrate doping is increased, since ΔV_T is also proportional to $\sqrt{N_a}$. For a p-channel device the bulk-to-source voltage

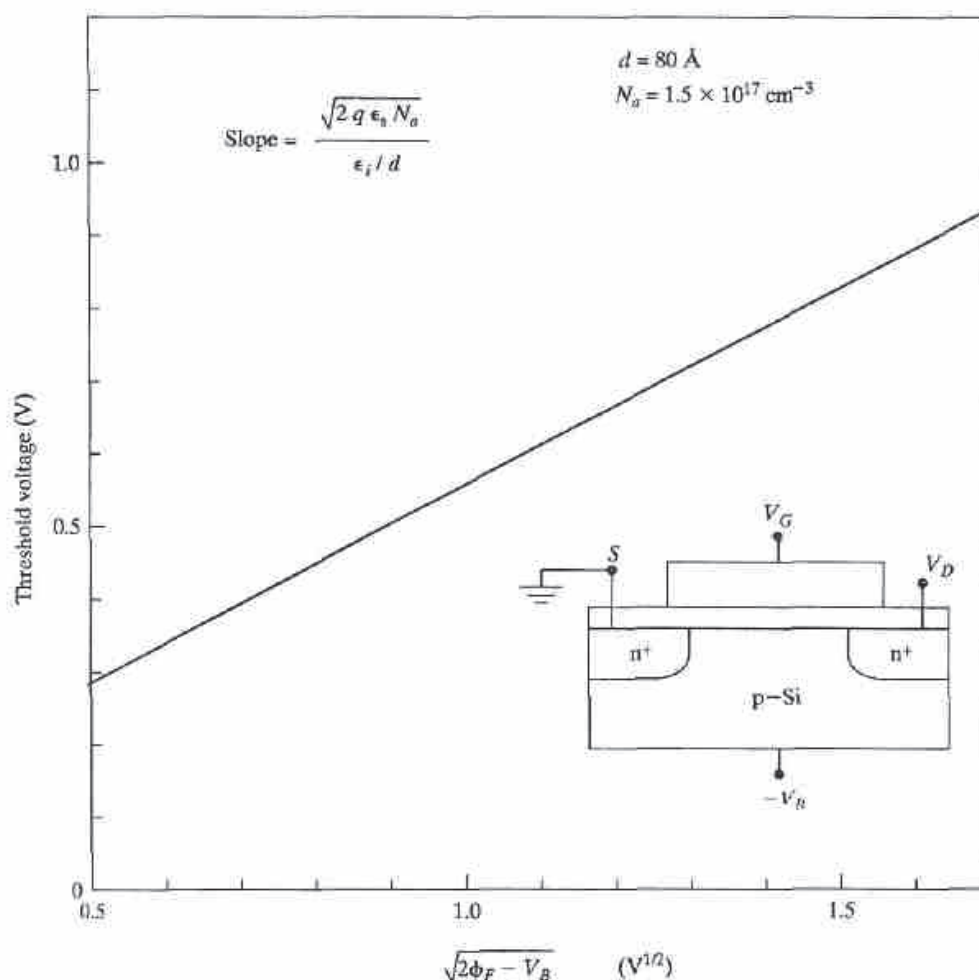


Figure 6-37

Threshold voltage dependence on substrate bias resulting from application of a voltage V_B from the substrate (i.e., bulk) to the source. For n channel, V_B must be zero or negative to avoid forward bias of the source junction. For p channel, V_B must be zero or positive.

V_B is positive to achieve a reverse bias, and the approximate change ΔV_T for $V_B \gg 2\phi_F$ is

$$\Delta V_T \approx - \frac{\sqrt{2\epsilon_s q N_d}}{C_i} V_B^{1/2} \quad (\text{p channel}) \quad (6-64)$$

Thus the p-channel threshold voltage becomes more negative with substrate bias.

The substrate bias effect (also called the *body effect*) increases V_T for either type of device. This effect can be used to raise the threshold voltage

of a marginally enhancement device ($V_T \approx 0$) to a somewhat larger and more manageable value. This can be an asset for n-channel devices particularly (see Fig. 6-20). The effect can present problems, however, in MOS integrated circuits for which it is impractical to connect each source region to the substrate. In these cases, possible V_T shifts due to the body effect must be taken into account in the circuit design.

6.5.7 Subthreshold Characteristics

If we look at the drain current expression (Eq. 6-53), it appears that the current abruptly goes to zero as soon as V_G is reduced to V_T . In reality, there is still some drain conduction below threshold, and this is known as *subthreshold conduction*. This current is due to weak inversion in the channel between flat band and threshold (for band-bending between zero and $2\phi_F$), which leads to a diffusion current from source to drain. The drain current in the subthreshold region is equal to

$$I_D = \mu(C_d + C_u) \frac{Z}{L} \left(\frac{kT}{q} \right)^2 \left(1 - e^{-\frac{qV_D}{kT}} \right) \left(e^{\frac{q(V_G - V_T)}{c kT}} \right) \quad (6-65)$$

where

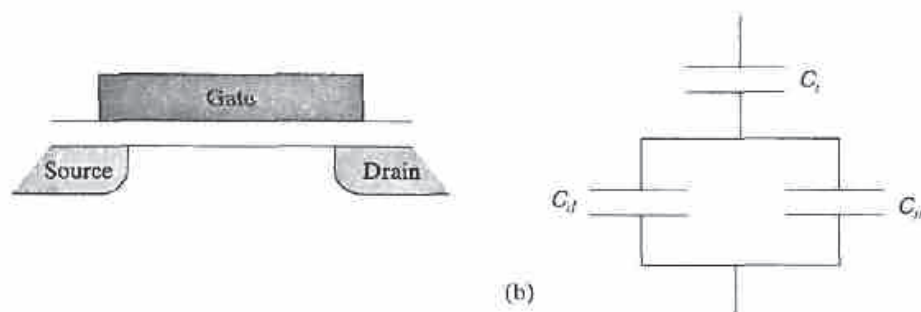
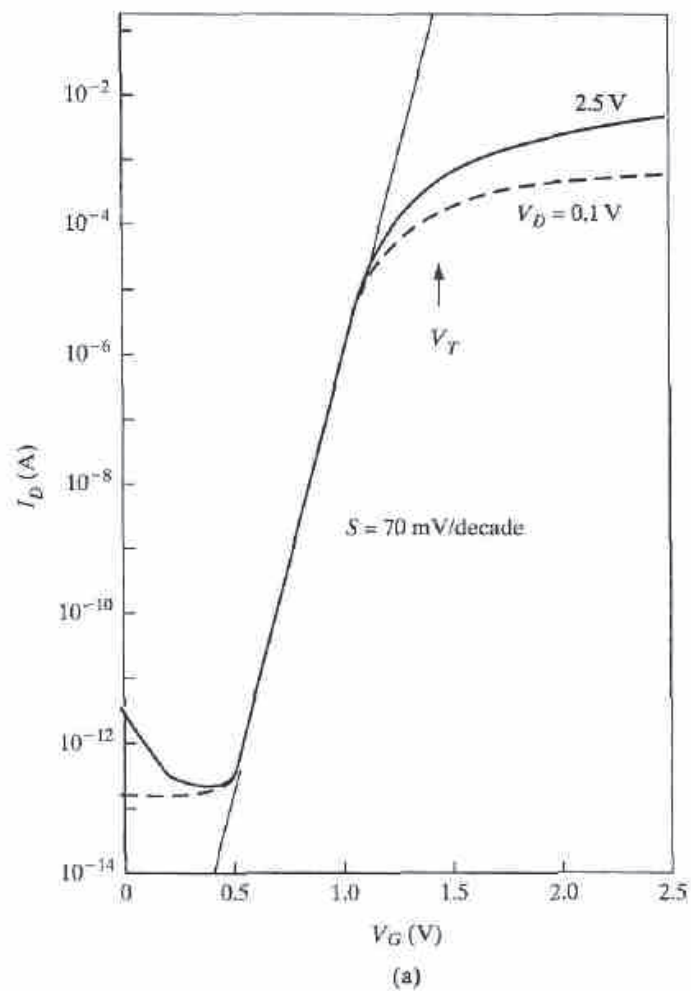
$$c_r = \left[1 + \frac{C_d + C_u}{C_i} \right]$$

It can be seen that I_D depends exponentially on gate bias, V_G . However, V_D has little influence once V_D exceeds a few kT/q . Obviously, if we plot $\ln I_D$ as a function of gate bias V_G , we should get a linear behavior in the subthreshold regime, as shown in Fig. 6-38a. The slope of this line (or more precisely the reciprocal of the slope) is known as the subthreshold slope, S , which has typical values of ~ 70 mV/decade at room temperature for state-of-the-art MOSFETs. This means that a change in the input V_G of 70 mV will change the output I_D by an order of magnitude. Clearly, the smaller the value of S , the better the transistor is as a switch. A small value of S means a small change in the input bias can modulate the output current considerably.

It can be shown that the expression for S is given by

$$S = \frac{dV_G}{d(\log I_D)} = \ln 10 \frac{dV_G}{d(\ln I_D)} = 2.3 \frac{kT}{q} \left[1 + \frac{C_d + C_u}{C_i} \right] \quad (6-66)$$

Here, the factor $\ln 10$ (≈ 2.3) is introduced to change from $\log_{10}$ to natural logarithm, $\ln$. This equation can be understood by looking at the electrical equivalent circuit of the MOSFET in terms of the capacitors (Fig. 6-38b). Between the gate and the substrate, we find the gate capacitance, C_i , in series with the parallel combination of the depletion capacitance in the channel, C_d , and the fast interface state capacitance, $C_u = qD_{it}$. The expression in brackets in Eq. (6-66) is simply the capacitor divider ratio which tells us what

**Figure 6-38**

Subthreshold conduction in MOSFETs: (a) Semilog plot of I_D versus V_G ; (b) equivalent circuit showing capacitor divider which determines subthreshold slope.

fraction of the applied gate bias, V_G , appears at the Si-SiO₂ interface as the surface potential. Ultimately, it is the surface potential that is responsible for modulating the barrier between source and drain, and therefore the drain current. Hence, S is a measure of the efficacy of the gate potential in modulating I_D . From Eq. (6-66), we observe that S is improved by reducing the gate oxide thickness, which is reasonable because if the gate electrode is closer to the channel, the gate control is obviously better. The value of S is higher for heavy channel doping (which increases the depletion capacitance) or if the silicon-oxide interface has many fast interface states.

For a very small gate voltage, the subthreshold current is reduced to the leakage current of the source/drain junctions. This determines the off-state leakage current, and therefore the standby power dissipation in many complementary MOS (CMOS) circuits involving both n-channel and p-channel MOSFETs. It also underlines the importance of having high quality source/drain junctions. From the subthreshold characteristics, it can be seen that if the V_t of a MOSFET is too low, it cannot be turned off fully at $V_G = 0$. Also, unavoidable statistical variations of V_T cause drastic variations of the subthreshold leakage current. On the other hand, if V_T is too high, one sacrifices drive current, which depends on the difference between the power supply voltage and V_T . For these reasons, the V_T of MOSFETs has historically been designed to be ~ 0.7 V. However, with the recent advent of various types of low-voltage, low-power portable electronics, there are new challenges in device and circuit design to optimize speed and power dissipation.

6.5.8 Equivalent Circuit for the MOSFET

When we attempt to draw an equivalent circuit of a MOSFET, we find that in addition to the intrinsic MOSFET itself, there are a variety of parasitic elements associated with it. An important addition to the gate capacitance is the so-called *Miller overlap capacitance* due to the overlap between the gate and the drain region (Fig. 6-39). This capacitance is particularly problematic because it represents a feedback path between the output drain terminal and the input gate terminal. One can measure the Miller capacitance at high frequency by holding the gate at ground ($V_G = 0$) so that an inversion layer is not formed in the channel. Thereby, most of the measured capacitance between gate and drain is due to the Miller capacitance, rather than the gate capacitance C_i . It is possible to minimize this capacitance by using a so-called *self-aligned gate*. In this process, the gate itself is used to mask the source/drain implants, thereby achieving alignment. Even in this design, however, there is still a certain amount of overlap because of the lateral straggle or spread of the implanted dopants underneath the gate, further exacerbated by the lateral diffusion which occurs during high temperature annealing. This spread of the source/drain junctions under the gate edge determines what is called the channel length reduction, ΔL_R (Fig. 6-40). Hence, we get the electrical or "effective" channel length, L_{eff} , in terms of the physical gate length, L as

$$L_{eff} = L - \Delta L_R \quad (6-67)$$

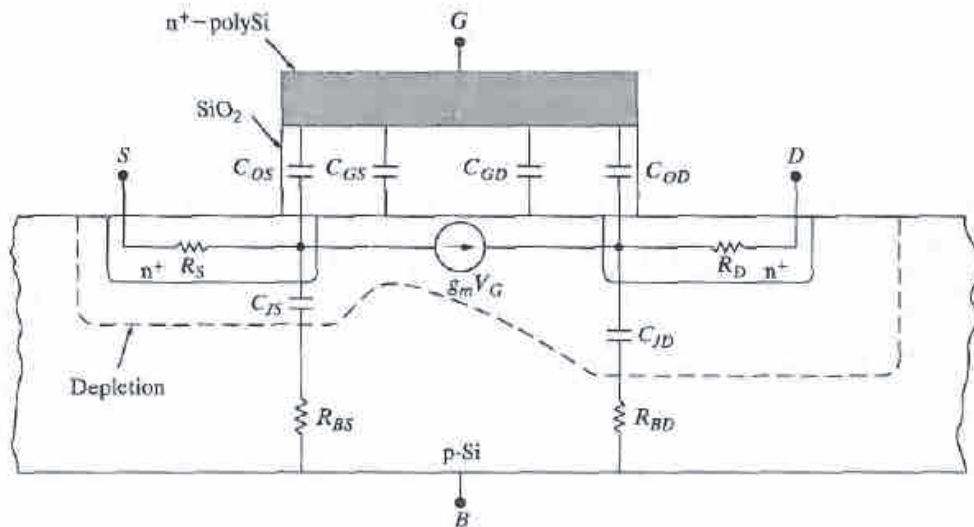


Figure 6-39

Equivalent circuit of a MOSFET, showing the passive capacitive and resistive components. The gate capacitance C_i is the sum of the distributed capacitances from the gate to the source end of the channel (C_{GS}) and the drain end (C_{GD}). In addition, we have an overlap capacitance (where the gate electrode overlaps the source/drain junctions) from the gate-to-source (C_{OS}) and gate-to-drain (C_{OD}). C_{OD} is also known as the Miller overlap capacitance. We also have p-n junction depletion capacitances associated with the source (C_{JS}) and drain (C_{JD}). The parasitic resistances include the source/drain series resistances (R_S and R_D), and the resistances in the substrate between the bulk contact and the source and drain (R_{BS} and R_{BD}). The drain current can be modeled as a (gate) voltage-controlled constant-current source.

There can also be a width reduction, ΔZ , which changes the effective width, Z_{eff} , from the physical width Z of the MOSFET. The width reduction results from the electrical isolation regions that are formed around all transistors, generally by LOCOS. The LOCOS isolation technique is discussed in Section 9.3.1.

Another very important parameter in the equivalent circuit is the source/drain series resistance, $R_{SD} = (R_S + R_D)$, because it degrades the drain current and transconductance. For a certain applied drain bias to the source/drain terminals, part of the applied voltage is “wasted” as an ohmic voltage drop across these resistances, depending on the drain current (or gate bias). Hence, the actual drain voltage applied to the intrinsic MOSFET itself is less; this causes I_D to increase sub-linearly with V_G .

We can determine R_{SD} , along with ΔL_R , from the overall resistance of the MOSFET in the linear region,

$$\left(\frac{V_D}{I_D} \right)$$

This corresponds to the intrinsic channel impedance R_{ch} , plus the source-drain resistance R_{SD} . Modifying Eq. (6-51) we get

$$\frac{V_D}{I_D} = R_{ch} + R_{SD} = \frac{L - \Delta L_R}{Z - \Delta Z} \frac{1}{\mu_n C_i (V_G - V_T)} + R_{SD} \quad (6-68)$$

We can measure V_D/I_D in the linear range for various MOSFETs having the same width, but different channel lengths, as a function of substrate bias. Varying the substrate bias changes the V_T through the body effect, and therefore the slope of the straight lines that result from plotting the overall resistance as a function of L . The lines pass through a point, having values which correspond to ΔL_R and R_{SD} , as shown in Fig. 6-40.

6.5.9 MOSFET Scaling and Hot Electron Effects

Much of the progress in semiconductor integrated circuit technology can be attributed to the ability to shrink or *scale* the devices. Scaling down MOSFETs has a multitude of benefits. From Table 6-1, we see the benefits of scaling in terms of the improvement of packing density, speed and power dissipation. A key concept in scaling, first due to Dennard at IBM, is that the various structural parameters of the MOSFET should be scaled in concert if the device is to keep functioning properly. In other words, if lateral dimensions such as the channel length and width are reduced by a factor of K , so should the vertical dimensions such as source/drain junction depths (x_j) and gate insulator thickness (Table 6-1). Scaling of depletion widths is achieved indirectly by scaling up doping concentrations. However, if we simply reduced the dimensions of the device and kept the power supply voltages the same, the internal electric fields in the device would increase. For ideal scaling, power supply voltages should also be reduced to keep the internal electric fields reasonably constant from one technology generation to the next. Unfortunately, in practice, power supply voltages are not scaled hand-in-hand with the device dimensions, partly because of other system-related constraints. The longitudinal electric fields in the pinch-off region, and the transverse electric fields across the gate oxide, increase with MOSFET scaling. A variety of problems then arise which are generically known as hot electron effects and short channel effects (Fig. 6-41).

When an electron travels from the source to the drain along the channel, it gains kinetic energy at the expense of electrostatic potential energy in the pinch-off region, and becomes a "hot" electron. At the conduction band edge, the electron has potential energy only; as it gains more kinetic energy, it moves higher up in the conduction band. A few of the electrons can become energetic enough to surmount the 3.1-eV potential barrier between the Si channel and the gate oxide (Fig. 6-25). Some of these injected hot electrons can go through the gate oxide and be collected as gate current, thereby reducing the input impedance. More importantly, some of these electrons can

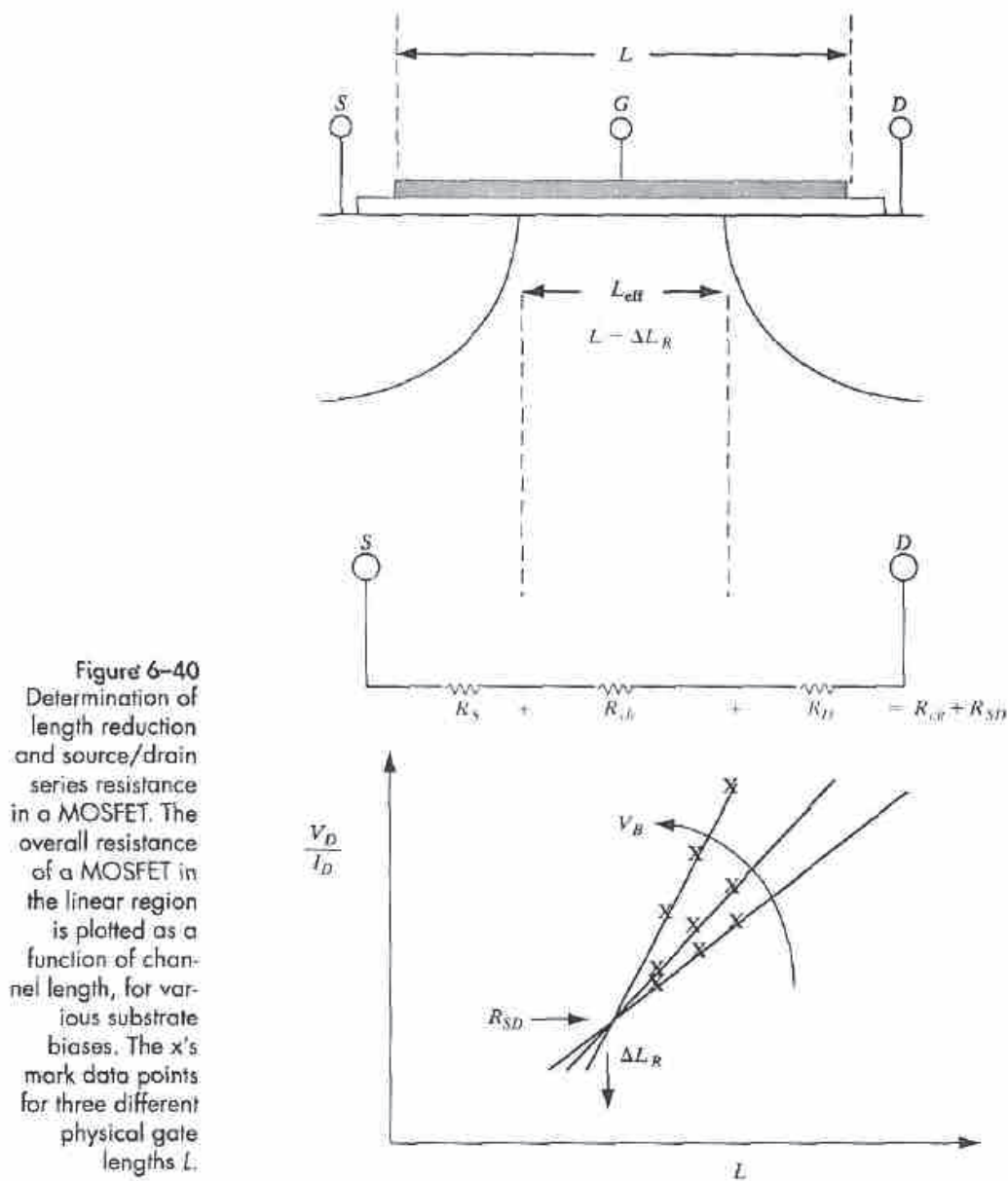


Figure 6-40
Determination of
length reduction
and source/drain
series resistance
in a MOSFET. The
overall resistance
of a MOSFET in
the linear region
is plotted as a
function of chan-
nel length, for var-
ious substrate
biases. The x's
mark data points
for three different
physical gate
lengths L .

be trapped in the gate oxide as fixed oxide charges. According to Eq. (6-37), this increases the flat band voltage, and therefore the V_T . In addition, these energetic hot carriers can rupture Si-H bonds that exist at the Si-SiO₂ interface, creating fast interface states that degrade MOSFET parameters such as transconductance and subthreshold slope, with stress. The results of such hot carrier degradation are shown in Fig. 6-42, where we see the increase of V_T and decrease of slope, and therefore transconductance, with stress. The solution to this problem is to use what is known as a *lightly doped drain* (LDD),

Table 6-1 Scaling rules for MOSFETs according to a constant factor K . The horizontal and vertical dimensions are scaled by the same factor. The voltages are also scaled to keep the internal electric fields more or less constant, and the hot carrier effects manageable.

Scaling factor	
Surface dimensions (L, Z)	$1/K$
Vertical dimensions (d, x_i)	$1/K$
Impurity Concentrations	K
Current, Voltages	$1/K$
Current Density	K
Capacitance (per unit area)	K
Transconductance	1
Circuit Delay Time	$1/K$
Power Dissipation	$1/K^2$
Power Density	1
Power-Delay Product	$1/K^3$

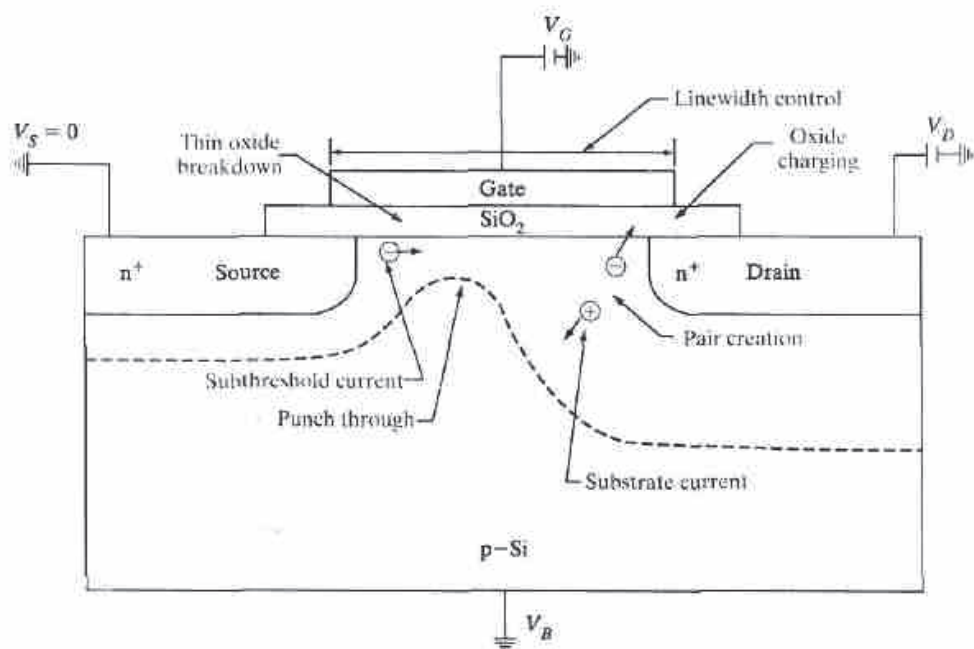


Figure 6-41

Short channel effects in MOSFETs. As MOSFETs are scaled down, potential problems due to short channel effects include hot carrier generation (electron-hole pair creation) in the pinch-off region, punch-through breakdown between source and drain, and thin gate oxide breakdown.

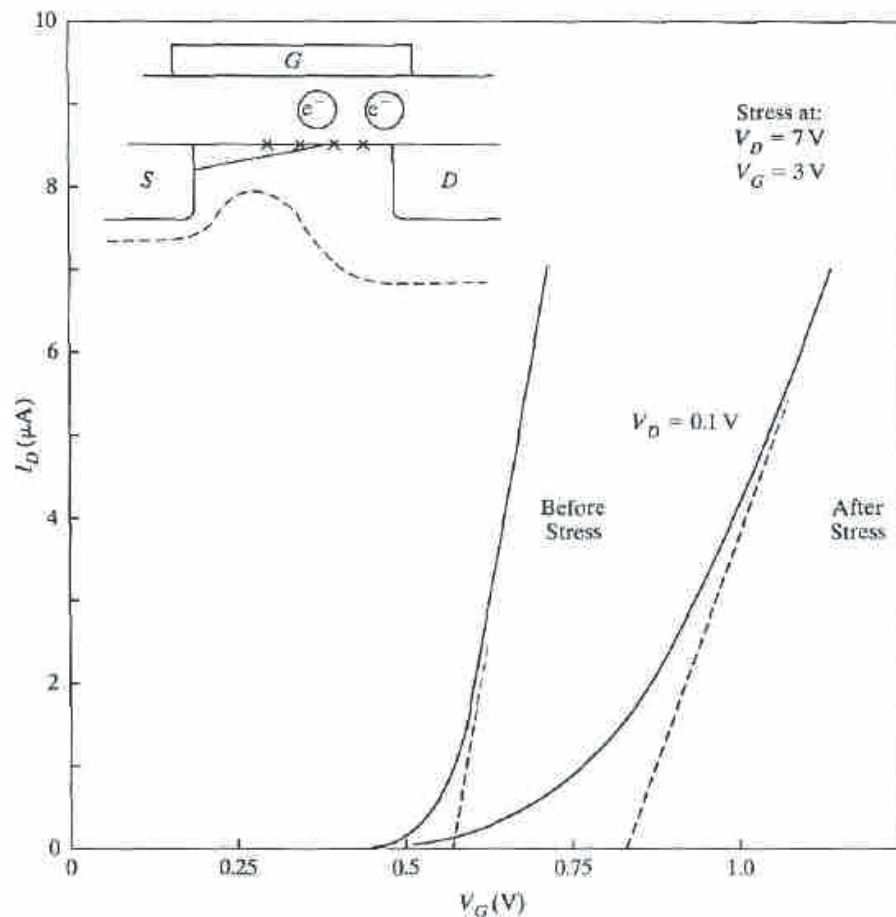


Figure 6-42

Hot carrier degradation in MOSFETs. The linear region transfer characteristics before and after hot carrier stress indicate an increase of V_T and decrease of transconductance (or channel mobility) due to hot electron damage. The damage can be due to hot electron injection into the gate oxide which increases the fixed oxide charge, and increasing fast interface state densities at the oxide-silicon interface (indicated by *x*).

As discussed in more detail in Section 9.3.1, by reducing the doping concentration in the source/drain, the depletion width at the reverse-biased drain-channel junction is increased and the electric field is reduced.

Hot carrier effects are less problematic for holes in p-channel MOSFETs than for electrons in n-channel devices for two reasons. The channel mobility of holes is approximately half that of electrons; hence, for the same electric field, there are fewer hot holes than hot electrons. Unfortunately, the lower hole mobility is also responsible for lower drive currents in p-channel than in n-channel. Also, the barrier for hole injection in the valence band between Si and SiO_2 is higher (5 eV) than for electrons in the conduction band (3.1 eV), as

shown in Fig. 6-25. Hence, while LDD is mandatory for n-channel, it is often not used for p-channel devices.

One "signature" for hot electron effects is substrate current (Fig. 6-43). As the electrons travel towards the drain and become hot, they can create secondary electron-hole pairs by impact ionization (Fig. 6-41). The secondary electrons are collected at the drain, and cause the drain current in saturation to increase with drain bias at high voltages, thereby leading to a decrease of the output impedance. The secondary holes are collected at the substrate as substrate current. This current can create circuit problems such as noise or latchup in CMOS circuits (Section 9.3.1). It can also be used as a monitor for hot electron effects. As shown in Fig. 6-43, substrate current initially increases with gate bias (for a fixed, high drain bias), goes through a peak and then

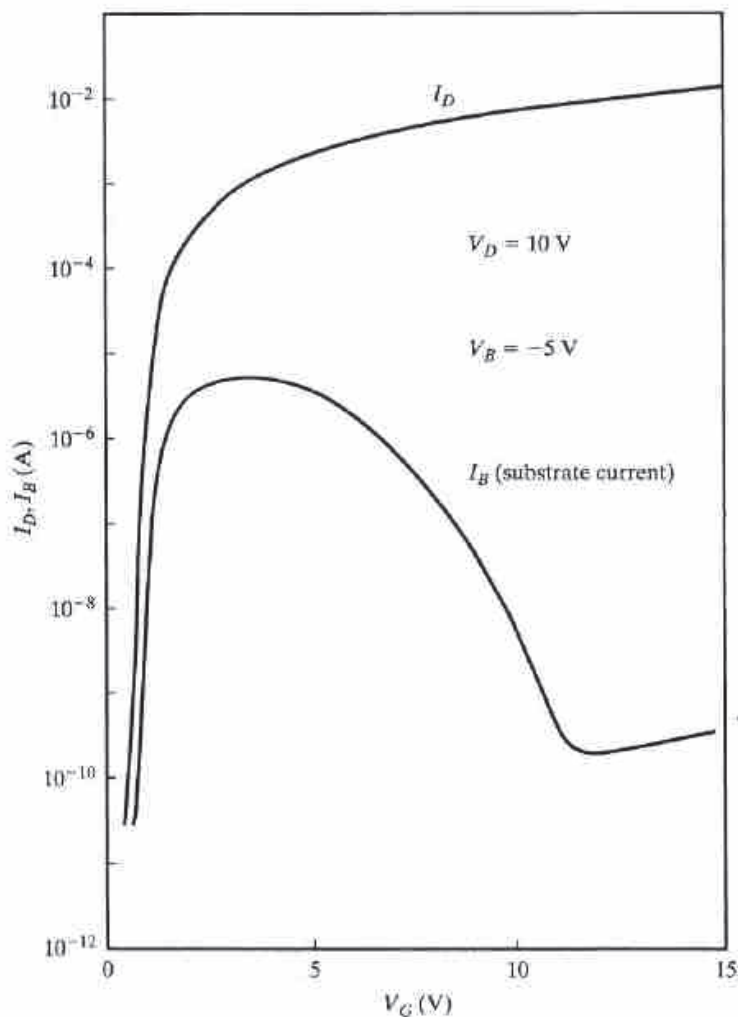


Figure 6-43
Substrate current in a MOSFET. The substrate current in n-channel MOSFETs due to impact-generated holes in the pinch-off region, as a function of gate bias. The substrate current initially increases with V_G because of the corresponding increase of I_D . However, for even higher V_G , the MOSFET goes from the saturation to the linear region, and the high electric fields in the pinch-off region decrease, causing less impact ionization. (After Kamata, et. al., *Jpn. J. Appl. Phys.*, 15 (1976), 1127.)

decreases. The reason for this behavior is that initially, as the gate bias increases, the drain current increases and thereby provides more primary carriers into the pinch-off region for impact ionization. However, for even higher gate bias, the MOSFET goes from the saturation region into the linear region when the fixed V_D drops below $V_{D(\text{sat.})} = (V_G - V_T)$. The longitudinal electric field in the pinch-off region drops, thereby reducing the impact ionization rates. Hot electron reliability studies are done under "worst case" conditions of peak substrate current. These are generally done under accelerated conditions of higher-than-normal operating voltages so that if there are any potential problems, they show up in a reasonable time period. The degradation data is then extrapolated to the actual operating conditions.

6.5.10 Drain-Induced Barrier Lowering

If small channel length MOSFETs are not scaled properly, and the source/drain junctions are too deep or the channel doping is too low, there can be unintended electrostatic interactions between the source and the drain known as *Drain-Induced Barrier Lowering (DIBL)*. This leads to punch-through leakage or breakdown between the source and the drain, and loss of gate control. The phenomenon can be understood from Fig. 6-44, where we have schematically plotted the surface potential along the channel for a long channel device and a short device. We see that as the drain bias is increased, the conduction band edge (which reflects the electron energies) in the drain is pulled down, and the drain-channel depletion width expands. For a long channel MOSFET, the drain bias does not affect the source-to-channel potential barrier, which corresponds to the built-in potential of the source-channel p-n junction. Hence, unless the gate bias is increased to lower this potential barrier, there is little drain current. On the other hand, for a short channel MOSFET, as the drain bias is raised and the conduction band edge in the drain is pulled down (with a concomitant increase of the drain depletion width), the source-channel potential barrier is lowered due to DIBL. This can be shown numerically by a solution of the two-dimensional Poisson equation in the channel region. Simplistically, the onset of DIBL is sometimes considered to correspond to the drain depletion region expanding and merging with the source depletion region, and causing punch-through breakdown between source and drain. However, it must be kept in mind that DIBL is ultimately caused by the lowering of the source-junction potential barrier below the built-in potential. Hence, if we get DIBL in a MOSFET for a grounded substrate, the problem can be mitigated by applying a substrate reverse bias, because that raises the potential barrier at the source end. This works in spite of the fact that the drain depletion region interacts even more with the source depletion region under such back bias. Once the source-channel barrier is lowered by DIBL, there can be significant drain leakage current, with the gate being unable to shut it off.

What are the solutions to this problem? The source/drain junctions must be made sufficiently shallow (i.e., scaled properly) as the channel lengths are reduced, to prevent DIBL. Secondly, the channel doping must be made

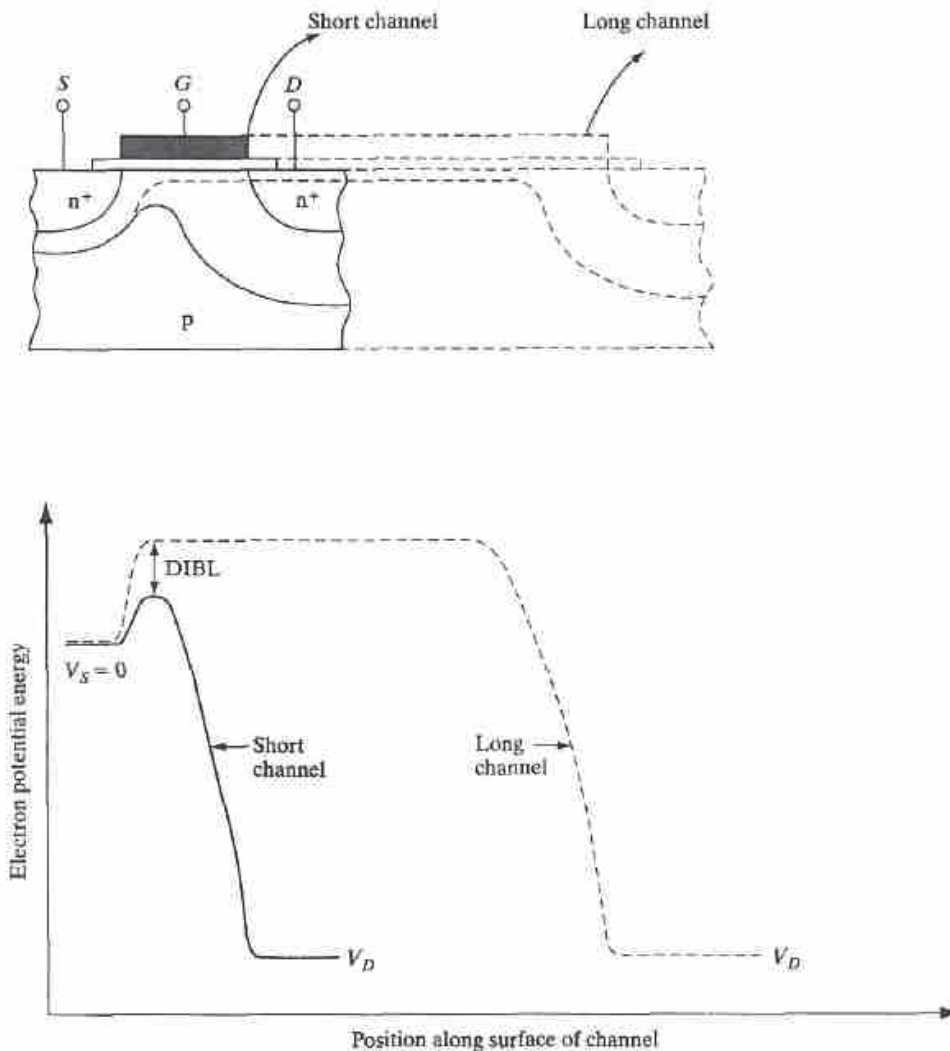


Figure 6-44

Drain-induced barrier lowering in MOSFETs. Cross sections and potential distribution along the channel for a long channel and short channel MOSFET.

sufficiently high to prevent the drain from being able to control the source junction. This is achieved by performing what is known as an *anti-punch-through* implant in the channel. Sometimes, instead of such an implant throughout the channel (which can have undesirable consequences such as raising the V_T or the body effect), a localized implant is done only near the source/drains. These are known as *halo* or *pocket* implants. The higher doping reduces the source/drain depletion widths and prevents their interaction.

For short channel MOSFETs, DIBL is related to the electrical modulation of the channel length in the pinch-off region, ΔL . Since the drain current

is inversely proportional to the electrical channel length, we get

$$I_D \propto \frac{1}{L - \Delta L} = \frac{1}{L} \left(1 + \frac{\Delta L}{L} \right) \quad (6-69)$$

for small pinch-off regions, ΔL . We assume that the fractional change in the channel length is proportional to the drain bias,

$$\frac{\Delta L}{L} = \lambda V_D \quad (6-70)$$

where λ is the *channel length modulation parameter*. Hence, in the saturation region, the expression for the drain current becomes

$$I_D = \frac{Z}{2L} \bar{\mu}_n C_A (V_G - V_T)^2 (1 + \lambda V_D) \quad (6-71)$$

This leads to a slope in the output characteristics, or a lowering of the output impedance (Fig. 6-32).

6.5.11 Short Channel Effect and Narrow Width Effect

If we plot the threshold voltage as a function of channel length in MOSFETs, we find that V_T decreases with L for very small geometries. This effect is called the *short channel effect (SCE)*, and is somewhat similar to DIBL. The mechanism is due to something called *charge sharing* between the source/drain and the gate (Fig. 6-45)⁹. From the equation for the threshold voltage (6-38), we notice that one of the terms is the depletion charge under the gate.

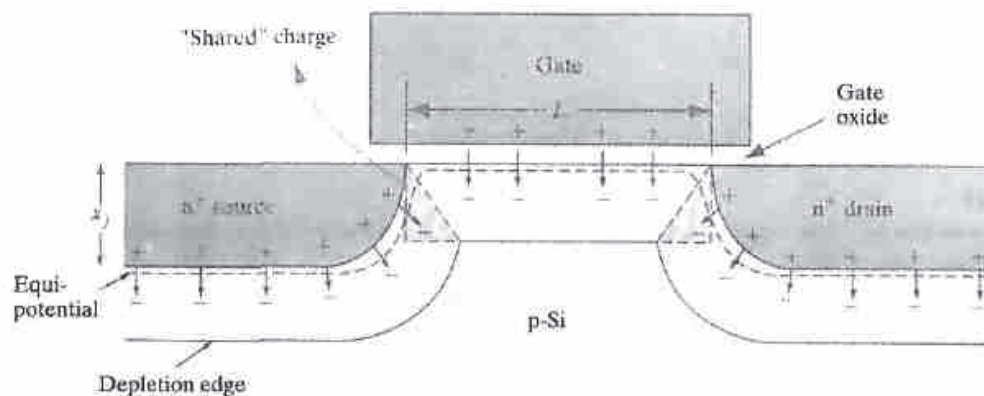


Figure 6-45

Short channel effect in a MOSFET. Cross-sectional view of MOSFET along the length showing depletion charge sharing (colored regions) between the gate, source and drain.

⁹L. Yau, "A simple theory to predict the threshold voltage of short-channel IGFETs," *Solid-State Electronics*, 17 (1974): 1059.

The equipotential lines in Fig. 6-45 designating the depletion regions curve around the contours of the source/drain junctions. Keeping in mind that the electric field lines are perpendicular to the equipotential contours, we see that the depletion charges that are physically underneath the gate in the approximately triangular regions near the source/drains have their field lines terminate not on the gate, but instead on the source/drains. Hence, electrically these depletion charges are "shared" with the source and drain regions and should not be counted in the V_T expression, Eq. (6-38). We can deal with this effect by replacing the original Q_d in the rectangular region underneath the gate by a lower Q_d in the trapezoidal region in Fig. 6-45. Clearly, for a long channel device, the triangular depletion charge regions near the source and drain are a very small fraction of the total depletion charge underneath the gate. However, as the channel lengths are reduced, the shared charge becomes a larger fraction of the total, and this results in a V_T roll-off as a function of L (Fig. 6-46). This is important because it is hard to control the channel lengths precisely in manufacturing. The channel length variations then lead to problems with V_T control.

In the last several years, another effect has been observed in n-channel MOSFETs with decreasing L . The V_T initially goes up before it goes down due to the short channel effect. This phenomenon has been dubbed the *reverse short channel effect (RSCE)*, and is due to interactions between Si point defects that are created during the source/drain implant and the B doping in the channel, causing the B to pile up near the source and drains, and thus raise the V_T .

Another related effect in MOSFETs is the *narrow width effect*, where the V_T goes up as the channel width Z is reduced for very narrow devices (Fig. 6-46). This can be understood from Fig. 6-47, where some of the depletion charges under the LOCOS isolation regions have field lines electrically terminating on the gate. Unlike the SCE, where the effective depletion charge is reduced due to charge sharing with the source/drain, here the depletion charge belonging to the gate is increased. The effect is not important for very wide devices, but becomes quite important as the widths are reduced below $1\mu\text{m}$.

6.5.12 Gate-Induced Drain Leakage

If we examine the subthreshold characteristics shown in Fig. 6-38, we find that as the gate voltage is reduced below V_T , the subthreshold current drops and

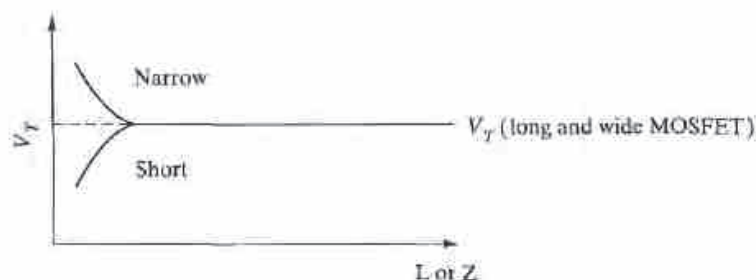


Figure 6-46
Roll-off of V_T with decreasing channel length, and increase of V_T with decreasing width.

Figure 6-47
Narrow width effect in a MOSFET. Cross-sectional view of MOSFET along the width, showing additional depletion charge (colored regions) underneath the field or the LOCOS isolation regions.

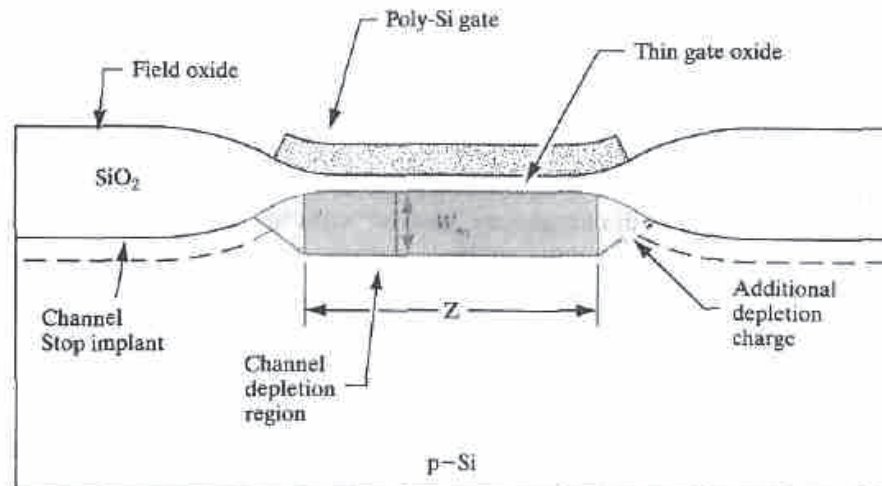
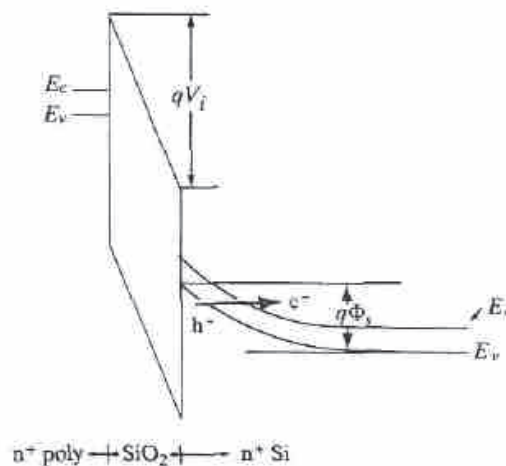
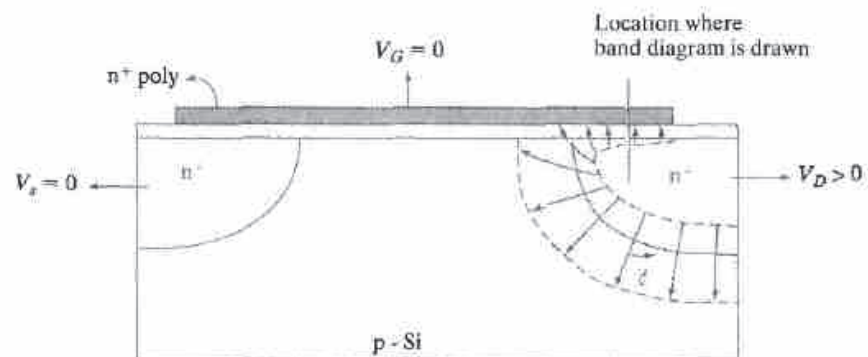


Figure 6-48
Gate-induced drain leakage in MOSFETs. The band diagram for the location shown in color is plotted as a function of depth in the gate-drain overlap region, indicating band-to-band tunneling and creation of electron-hole pairs in the drain region in the Si substrate.



then bottoms out at a level determined by the source/drain diode leakage. However, for even more negative gate biases we find that the off-state leakage current actually goes up as we try to turn off the MOSFET more for high V_{Di} ; this is known as *gate-induced drain leakage (GIDL)*. The same effect is seen at a fixed gate bias of near zero, for increasing drain bias. The reason for GIDL can be understood from Fig. 6-48, where we show the band diagram as a function of depth in the region where the gate overlaps the drain junction. As the gate is made more negative (or alternatively, for a fixed gate bias, the drain is made more positive), a depletion region forms in the n-type drain. Since the drain doping is high, the depletion widths tend to be narrow. If the band-bending is more than the band gap E_g across a narrow depletion region, the conditions are conducive to band-to-band tunneling in this region, thereby creating electron-hole pairs. The electrons then go to the drain as GIDL. It must be emphasized that this tunneling is not through the gate oxide (Section 6.4.7), but entirely in the Si drain region. For GIDL to occur, the drain doping level should be moderate ($\sim 10^{18} \text{ cm}^{-3}$). If it is much lower than this, the depletion widths and tunneling barriers are too wide. On the other hand, if the doping in the drain is very high, most of the voltage drops in the gate oxide, and the band-bending in the Si drain region drops below the value E_g . GIDL is an important factor in limiting the off-state leakage current in state-of-the-art MOSFETs.

- 6.1 Three-terminal active devices capable of power gain are key to the switching and amplification of electronic signals. The *input impedance* is (desirably) high for voltage-controlled field-effect transistors and low for current-controlled bipolar junction transistors. The power gain of the a-c signal comes from the d-c power supply. It is also desirable to have a high output impedance and a linear dependence of output current with respect to input voltage (or current).
- 6.2 Field effect devices such as the JFETs (or MESFETs and HEMTs) use a reverse-biased p-n junction (or Schottky diode) as a *gate* to control the undepleted part of a *channel* and thereby modulate the current flow between a *source* (S) of carriers and a *drain* (D) of carriers.
- 6.3 The MOSFET uses an oxide-isolated gate (MOS capacitor) that has a very high input impedance to modulate the barrier between S and D and to change the conductivity of the channel.
- 6.4 An enhancement-mode NMOSFET has a p-type channel between the n⁺ source and the drain regions. At the *flat band conditions* (V_{FB}), there is no band-bending as a function of depth in the channel. A gate bias more negative than V_{FB} causes exponentially more holes to be accumulated near the surface with band-bending.
- 6.5 More positive gate voltages initially repel the majority carrier holes and create a *depletion* region. Then, above the *threshold voltage* (V_T), they attract minority carrier electrons near the surface and *invert* the surface, thereby forming a conductive channel between source and drain. PMOSFETs have the opposite polarities.

SUMMARY

- 6.6 The V_T can be increased by increasing the channel doping and the gate oxide thickness (lower C_i). It is also affected by V_{FB} , which depends on the gate work function and oxide charges, and increases with substrate reverse bias (*body effect*) as well.
- 6.7 The overall MOS capacitance is the series combination of a fixed gate capacitance C_i and the voltage-dependent semiconductor capacitance in the channel. The inversion charge per unit area is approximately $C_i(V_G - V_T)$.
- 6.8 For long channel MOSFETs, initially I_D increases *linearly* with V_D , and then it *saturates* once the inversion channel is *pinched off* near the drain end.
- 6.9 There is a quadratic dependence of I_D on V_G for long channel MOSFETs. For *short channel* devices that operate in the *velocity-saturated region* or *source-injection regime*, the dependence of I_D on V_G is approximately linear.
- 6.10 For short L , I_{DSat} often increases slightly with V_D due to various short channel effects, leading to a decrease of output impedance. Short channel effects include *drain-induced barrier lowering (DIBL)* and *charge sharing*, which leads to V_T roll-off with L .
- 6.11 There is some conduction below V_T , characterized by a subthreshold slope, that is a measure of the effectiveness of the MOSFET as a switch.
- 6.12 Conduction above V_T in the channel inversion region reflects an *effective channel mobility* lower than that in bulk due to *surface roughness scattering*, and increasingly so at higher V_G .
- 6.13 In highly scaled MOSFETs, there can be *gate-tunneling* leakage currents (mitigated by high- k dielectrics) and *hot electron* (and, to a lesser extent, hole) effects, mitigated by lower power supply voltages and lightly doped drains (*LDDs*). The resulting high-S/D parasitic series resistance is minimized by Self-Aligned silICIDES (*SALICIDE*), described in Chapter 9.

PROBLEMS

- 6.1 Assume the JFET shown in Fig. 6-6 is Si and has p^+ regions doped with 10^{18} acceptors/cm³ and a channel with 10^{16} donors/cm³. If the channel half-width a is 1 μm , compare V_P with V_D . What voltage V_{GD} is required to cause pinch-off when V_D is included? With $V_G = -3$ V, at what value of V_D does the current saturate?
- 6.2 If the ratio $Z/L = 10$ for the JFET of Prob. 6.1, and $\mu_n = 1000$ cm²/V-s, calculate $I_D(\text{sat.})$ for $V_G = 0, -2, -4$, and -6 V. Plot $I_D(\text{sat.})$ vs. $V_D(\text{sat.})$.
- 6.3 For the JFET of Prob. 6.2, plot I_D vs. V_D for the same four values of V_G . Terminate each plot at the point of saturation.
- 6.4 Use Eqs. (6-9) and (6-10) to calculate and plot $I_D(V_D, V_G)$ at 300 K for a Si JFET with $a = 1000$ Å, $N_d = 7 \times 10^{17}$ cm⁻³, $Z = 100$ μm , and $L = 5$ μm . Allow V_D to range from 0 to 5 V, and allow V_G to take on the values 0, -1, -2, -3, -4, and -5 V.
- 6.5 Use the field-dependent mobility expression to calculate and plot $I_D(V_D, V_G)$ for the Si JFET described in Prob. 6.4 for gate lengths of 0.25, 0.50, 1.0, 2.0,

and $5.0\text{ }\mu\text{m}$ and a gate voltage of 0 V . The field-dependent mobility model of Eq. (8-12) takes the form

$$\mu(\mathcal{E}) = \frac{v_d}{\mathcal{E}} = \frac{\mu_0}{1 + (\mu_0 \mathcal{E} / v_s)}$$

where μ_0 is the low-field mobility (as given in Fig. 3-23) and v_s is the saturation velocity (10^7 cm/s). Assume that $\mathcal{E} = V_D/L$, and assume that the drain current saturates when it reaches a maximum. Discuss the advantages of having a short gate length device.

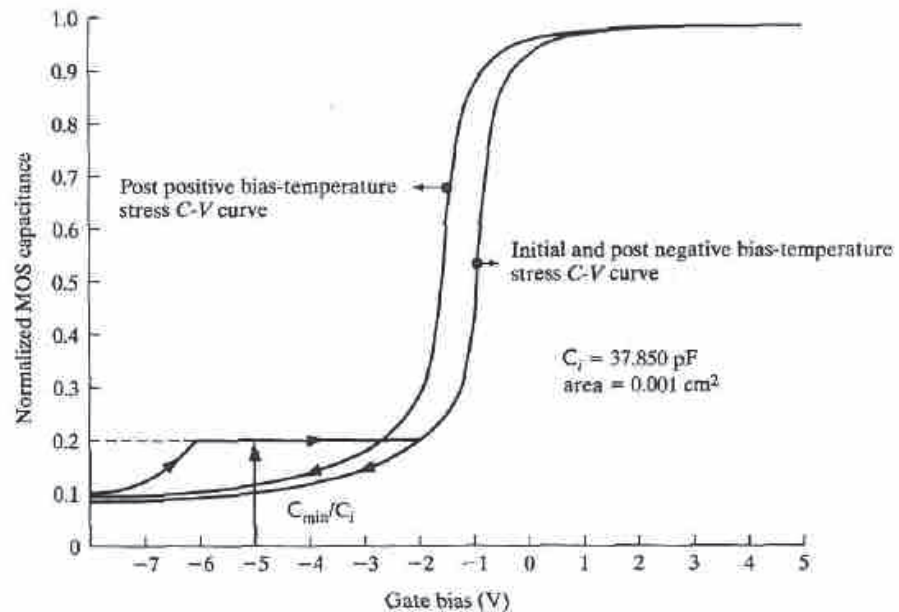
- 6.6 The current I_D varies almost linearly with V_D in a JFET for low values of V_D .
- Use the binomial expansion with $V_D/(-V_G) < 1$ to rewrite Eq. (6-9) as an approximation to this case.
 - Show that the expression for the channel conductance I_D/V_D in the linear range is the same as $g_m(\text{sat.})$ given by Eq. (6-11).
 - What value of gate voltage V_G turns the device off such that the channel conductance goes to zero?
- 6.7 Show that the width of the depletion region in Fig. 6-15 is given by Eq. (6-30). Assume the carriers are completely swept out within W , as was done in Section 5.2.3.
- 6.8 Find the maximum depletion width, minimum capacitance C_n , and threshold voltage for an *ideal* MOS capacitor with a 10-nm gate oxide (SiO_2) on p-type Si with $N_a = 10^{16}\text{ cm}^{-3}$. Next, include the effects of flat band voltage, assuming an n^+ polysilicon gate and fixed oxide charge of $5 \times 10^{10}\text{ q C/cm}^2$.
- 6.9 An Al-gate p-channel MOS transistor is made on an n-type Si substrate with $N_d = 5 \times 10^{17}\text{ cm}^{-3}$. The SiO_2 thickness is $100\text{ }\text{\AA}$ in the gate region, and the effective interface charge Q_i is $5 \times 10^{10}\text{ q C/cm}^2$. Find W_m , V_{FB} , and V_T . Sketch the C - V curve for this device and give important numbers for the scale.
- 6.10 Use Eq. (6-50) to calculate and plot $I_D(V_D, V_G)$ at 300 K for an n-channel Si MOSFET with an oxide thickness $d = 200\text{ }\text{\AA}$, a channel mobility $\bar{\mu}_n = 1000\text{ cm}^2/\text{V}\cdot\text{s}$, $Z = 100\text{ }\mu\text{m}$, $L = 5\text{ }\mu\text{m}$, and N_a of 10^{14} , 10^{15} , 10^{16} , and 10^{17} cm^{-3} . Allow V_D to range from 0 to 5 V and allow V_G to take on values of 0 , 1 , 2 , 3 , 4 , and 5 V . Assume that $Q_i = 5 \times 10^{11}\text{ q C/cm}^2$.
- 6.11 Calculate the V_T of a Si-P-channel MOS transistor for an n^+ -polysilicon gate with silicon oxide thickness $= 50\text{ }\text{\AA}$, $N_d = 1 \times 10^{18}\text{ cm}^{-3}$ and a fixed charge of $2 \times 10^{10}\text{ q C/cm}^2$. Is it an enhancement- or depletion-mode device? What B dose is required to change the V_T to 0 V ? Assume a shallow B implant.
- 6.12 (a) Find the voltage V_{FB} required to reduce to zero the negative charge induced at the semiconductor surface by a sheet of positive charge Q_{ox} located x' below the metal.

(b) In the case of an arbitrary distribution of charge $\rho(x')$ in the oxide, show that

$$V_{FB} = -\frac{1}{C_t} \int_0^d \frac{x'}{d} \rho(x') dx'$$

- 6.13 Redraw Figs. 6-12, 6-13, and 6-15 for the p-channel case (n-type substrate).
- 6.14 Calculate the V_T of an MOS capacitor where we deposit a high- k gate dielectric, HfO_2 , whose relative dielectric constant is 25, on a *novel* p-type semiconductor whose electron affinity is 4 eV, band gap is 1.5 eV, relative dielectric constant is 10, and intrinsic carrier concentration is 10^{12} cm^{-3} . The gate is made of a metal whose work function is 5 eV, gate oxide thickness is 100 Å, and N_A is 10^{18} cm^{-3} and that has a fixed oxide charge of $5 \times 10^{10} \text{ q C/cm}^2$. At V_T , what are the electron and hole concentrations at the oxide-semiconductor interface and deep in the substrate? Sketch a labeled band diagram normal to the surface at V_T and mark off the relevant values on the basis of the numbers given. Sketch the *low- and high-frequency* $C-V$ characteristics of this capacitor, and explain their differences. How would the characteristics change at *large negative gate bias* if we doubled the oxide thickness? How about if we doubled the substrate doping?
- 6.15 A Si MOS capacitor has the high frequency $C-V$ curve shown in Fig. P6-15 normalized to the capacitance in strong accumulation. Determine the oxide thickness and substrate doping assuming a gate-to-substrate work function difference of -0.35 V .

Figure P6-15



- 6.16 For the capacitor in Prob. 6.15, determine the initial flat band voltage.
- 6.17 For the capacitor in Prob. 6.15, determine the fixed oxide charge, Q_f , and the mobile ion content.
- 6.18 For a long channel n-MOSFET with $V_T = 1$ V, calculate the V_G required for an $I_D(\text{sat.})$ of 0.1 mA and $V_D(\text{sat.})$ of 5 V. Calculate the small-signal output conductance g and the transconductance $g_m(\text{sat.})$ at $V_D = 10$ V. Sketch the cross section of this MOSFET, and schematically show the inversion charge and depletion charge distributions for $V_D = 1$ V, 5 V, and 10 V. Recalculate the new I_D for $V_G - V_T = 3$ V and $V_D = 4$ V.
- 6.19 Calculate the V_T of a Si n-channel MOSFET for a gate-to-substrate work function difference $\phi_{ms} = -1.5$ eV, gate oxide thickness = 100 Å, $N_A = 10^{18} \text{ cm}^{-3}$, and fixed oxide charge of $5 \times 10^{10} \text{ q C/cm}^2$, for a substrate bias of -2.5 V. At V_T , what are the electron and hole concentrations at the oxide-Si interface and deep in the substrate? Sketch a labeled band diagram normal to the surface at V_T , showing the Fermi potential.
- 6.20 Calculate the V_T of a Si n-channel MOSFET for an n⁺-polysilicon gate with gate oxide thickness = 100 Å, $N_a = 10^{18} \text{ cm}^{-3}$, and a fixed oxide charge of $5 \times 10^{10} \text{ q C/cm}^2$.
- 6.21 For a MOSFET with $V_T = 1$ V and $Z = 50 \text{ } \mu\text{m}$, $L = 2 \text{ } \mu\text{m}$, calculate the drain current at $V_G = 5$ V, $V_D = 0.1$ V. Repeat for $V_G = 3$ V, $V_D = 5$ V. Assume an electron channel mobility $\bar{\mu}_n = 200 \text{ cm}^2/\text{V-s}$, and the substrate is connected to the source.
- 6.22 For the MOSFET characteristics shown in Fig. P6-22, calculate:

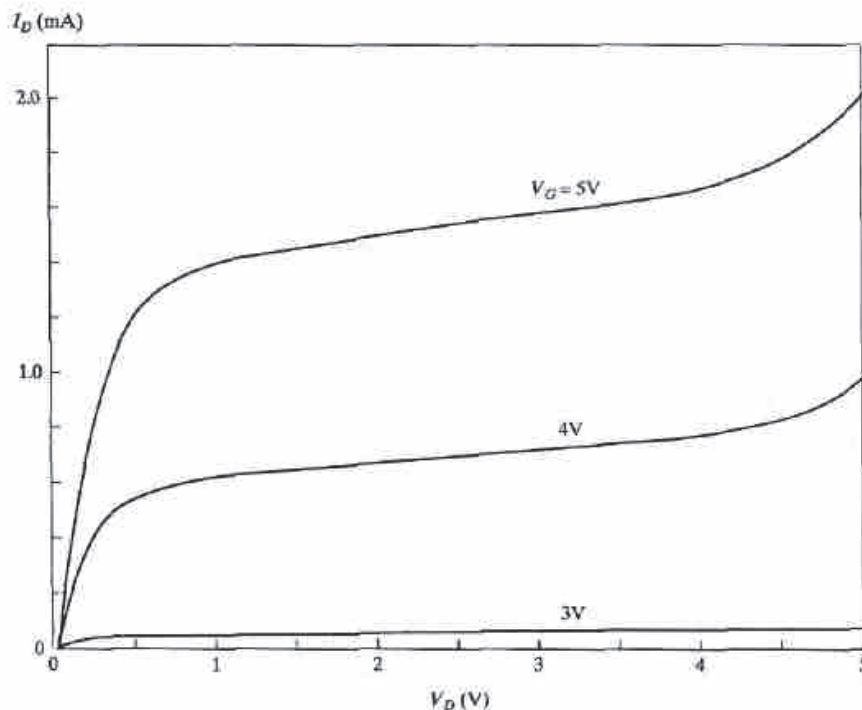


Figure P6-22

1. Linear V_T and k_N

2. Saturation V_T and k_N

Assume channel mobility, $\bar{\mu}_n = 500 \text{ cm}^2/\text{V}\cdot\text{s}$ and $V_{FH} = 0$.

- 6.23 For Prob. 6.22, calculate the gate oxide thickness and substrate doping, either graphically or iteratively.
- 6.24 Assume that the inversion layer in a Si MOSFET can be treated as a 2-D electron gas trapped in an infinite rectangular potential well of width $100 \text{ \AA}$. (In reality, it looks more like a triangular well.) Calculate the inversion charge per unit area assuming that the Fermi level lies midway between the second and third subbands. Assume $T = 77 \text{ K}$, and effective mass $= 0.2 m_0$. Assume also that the Fermi function can be treated as a rectangular function. Also sketch (E, k) for the first three subbands. Refer to Appendix IV.
- 6.25 An n-channel MOSFET with a $400\text{-\AA}$ gate oxide requires its V_T to be lowered by 2 V . Using a 50-keV implant of a singly charged species, and assuming that the implant distribution is peaked at the oxide-Si interface and can be regarded as a sheet charge at that interface, what implant parameters (species, energy, dose, and beam current) would you choose? The scan area is 200 cm^2 and the desired implant time is 20 s . Assume similar range statistics in oxide and Si.
- 6.26 Plot I_D vs. V_D with several values of V_G for a thin-oxide p-channel transistor with a $10\text{-}\mu\text{m}$ oxide and $V_T = -1.1 \text{ V}$. Assume that $I_D(\text{sat.})$ remains constant beyond pinch-off. Assume that $\bar{\mu}_p = 200 \text{ cm}^2/\text{V}\cdot\text{s}$, and $Z = 10L$.
- 6.27 A typical figure of merit for high-frequency operation of MOS transistors is the cutoff frequency $f_c = g_m/2\pi C_G L Z$, where the gate capacitance C_G is essentially C_i over most of the voltage range. Express f_c above pinch-off in terms of materials parameters and device dimensions, and calculate f_c for the transistor of Prob. 6.26, with $L = 1 \text{ }\mu\text{m}$.
- 6.28 When an MOS transistor is biased with $V_D > V_D(\text{sat.})$, the effective channel length is reduced by ΔL and the current I'_D is larger than $I_D(\text{sat.})$, as shown in Fig. 6-32. Assuming that the depleted region ΔL is described by an expression similar to Eq. (6-30) with $V_D - V_D(\text{sat.})$ for the voltage across ΔL , show that the conductance beyond saturation is

$$g'_D = \frac{\partial I'_D}{\partial V_D} = I_D(\text{sat.}) \frac{\partial}{\partial V_D} \left(\frac{L}{L - \Delta L} \right)$$

and find the expression for g'_D in terms of V_D .

- 6.29 From Fig. 6-44, it is clear that the depletion regions of the source and drain junctions can meet for short channels, a condition called *punch-through*. Assume that the source and drain regions of an n-channel Si MOSFET are doped with 10^{20} donors/cm³ and the $1\text{-}\mu\text{m}$ -long channel is doped with 10^{16} acceptors/cm³. If the source and substrate are grounded, what drain voltage will cause punch-through?

Very useful applets for understanding device operation are available at <http://jas.eng.buffalo.edu/>

Information about MOS devices used in integrated circuits can be found at <http://public.itrs.net/>

READING LIST

- Kahng, D. "A Historical Perspective on the Development of MOS Transistors and Related Devices," *IEEE Trans. Elec. Dev.*, ED-23 (1976): 655.
- Muller, R. S., and T. I. Kamins. *Device Electronics for Integrated Circuits*. New York: Wiley, 1986.
- Neamen, D. A. *Semiconductor Physics and Devices: Basic Principles*. Homewood, IL: Irwin, 2003.
- Pierret, R. F. *Field Effect Devices*. Reading, MA: Addison-Wesley, 1990.
- Sah, C. T. "Characteristics of the Metal-Oxide Semiconductor Transistors." *IEEE Trans. Elec. Dev.* ED-11 (1964): 324.
- Sah, C. T. "Evolution of the MOS Transistor—From Conception to VLSI." *Proceedings of the IEEE* 76 (October 1988): 1280–1326.
- Schroder, D. K. *Modular Series on Solid State Devices: Advanced MOS Devices*. Reading, MA: Addison-Wesley, 1987.
- Shockley, W., and G. Pearson. "Modulation of Conductance of Thin Films of Semiconductors by Surface Charges." *Phys. Rev.* 74 (1948): 232.
- Sze, S. M. *Physics of Semiconductor Devices*. New York: Wiley, 1981.
- Taur, Y., and T.H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge: Cambridge University Press, 1998.
- Tsividis, Y. *Operation and Modeling of the MOS Transistor*. Boston: McGraw-Hill College, 1998.

Question 1

SELF QUIZ

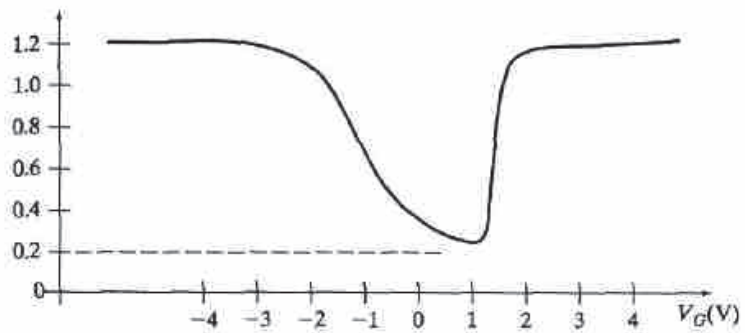
- What is the main distinction between an active device and a passive device?
- If a device has power gain, where is the higher energy of the a-c signal at the output coming from?
- What is the distinction between a current-controlled versus voltage-controlled three-terminal active device? Which is preferable?

Question 2

For parts (a) through (c) below, consider the following low-frequency gate capacitance (per unit area) vs. gate voltage characteristic for a metal gate n-channel MOSFET.

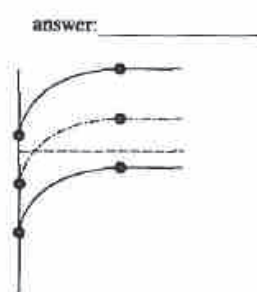
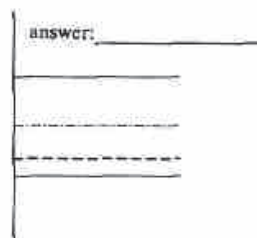
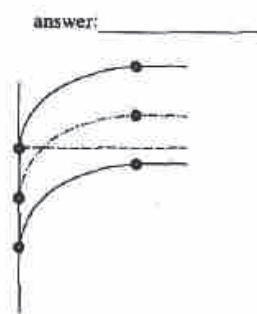
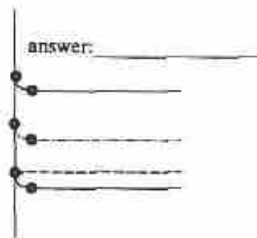
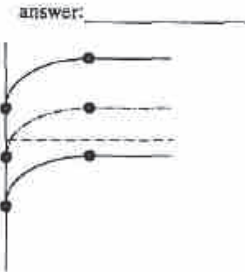
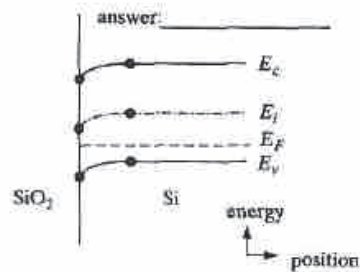
- On the below figure label the *approximate* regions or points of

▪ weak inversion (1)	▪ accumulation (4)
▪ flat band (2)	▪ threshold (5)
▪ strong inversion (3)	▪ depletion (6)
- What is the oxide capacitance per unit area? _____ $\mu\text{F}/\text{cm}^2$
- On the curve below, sketch the high-frequency curves for this MOSFET with grounded source/drain.



Question 3

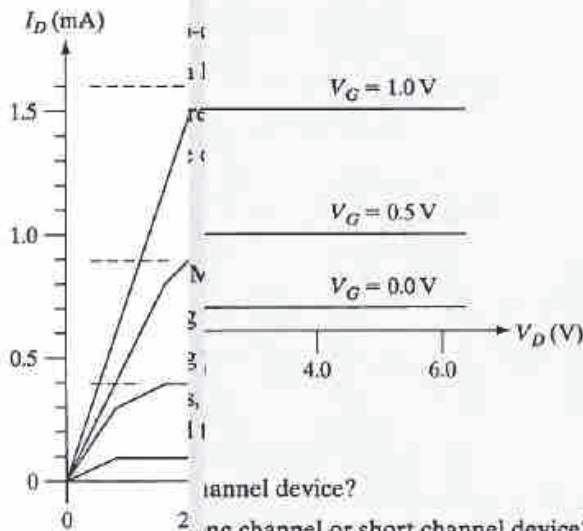
Label—in the “answer” space provided *above* the figure—the following MOS capacitor band diagrams as corresponding to accumulation, weak inversion, depletion, strong inversion, flat band or threshold. Use each possibility only once.



Field-Effect Transistor

Question 4

Consider the following MOSFET



- (a) Is this an n-channel or p-channel device?
 (b) Does this appear to be a long channel or short channel device?
 (c) What is the apparent threshold voltage V_T ?
 (d) Is this a depletion-mode or enhancement-mode MOSFET?

Question 5

Is the subthreshold slope of a MOSFET

- (a) Increasing or decreasing with increasing channel length?
 (b) Increasing or decreasing with increasing gate oxide thickness?

For each of these two methods, which the method can be used to

Question 6

Would increasing the device temperature (circle the correct answers):

- (a) the reverse saturation current I_{S0} of a p-n diode?
 (b) the subthreshold source-drain leakage current of MOSFET?

Question 7

Assuming no interface charge due to defects and/or traps, would decreasing the oxide thickness/increasing the oxide capacitance of an n-channel MOSFET *increase, decrease, or leave essentially unchanged* the following parameters (circle the correct answers):

(a) the flat band voltage, V_{FB} ?

increase / decrease / unchanged

(b) the threshold voltage, V_T ?

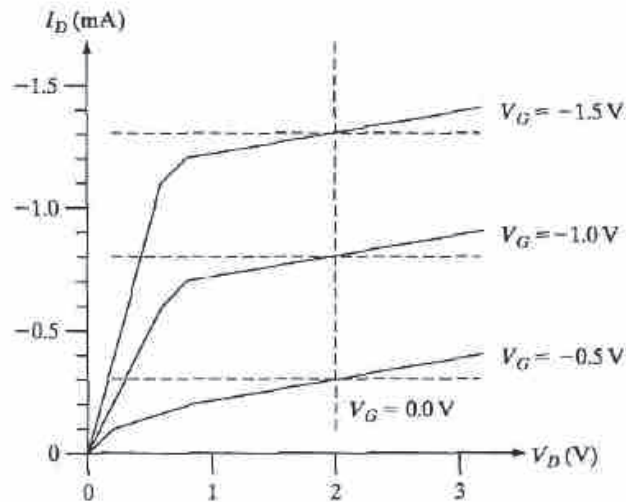
increase / decrease / unchanged

(c) the subthreshold slope?

increase / decrease / unchanged

Question 8

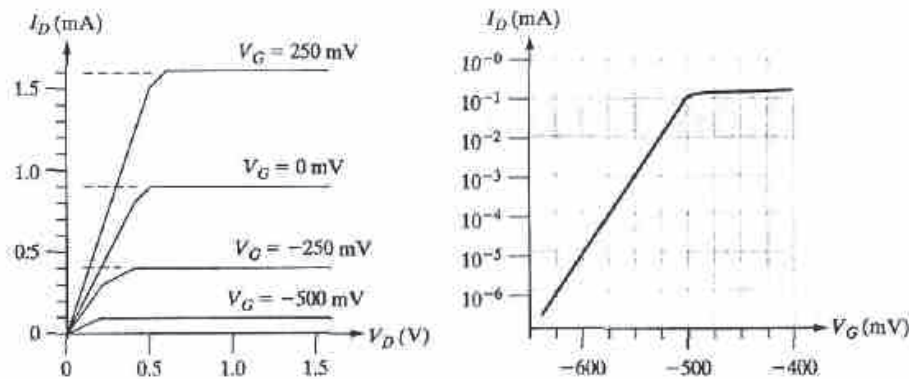
Consider the following MOSFET characteristic.



- What is the transconductance g_m at $V_D = -1.0$ V (in units of mhos)?
- What is the apparent threshold voltage at $V_D = -1.0$ V?
- Does this MOSFET appear to be a long channel or short channel device?
- Is this an n-channel or p-channel device?
- Is this a depletion-mode or enhancement-mode MOSFET?

Question 9

A senior in electrical and computer engineering in a device fabrication course presented the following characteristics as those of an n-channel MOSFET that he had fabricated and characterized at a temperature of 300 K.



- (a) If we were to believe this student, would this be *normally on/depletion mode* or a *normally off/enhancement mode* MOSFET? Circle one choice below.

normally off / normally on

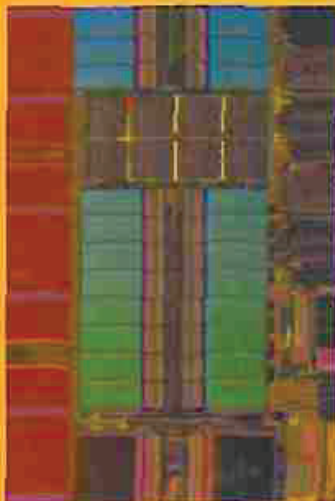
- (b) If we were to believe this student, would this MOSFET appear to be a *long channel* or *short channel* device? Circle one.

long channel / short channel

- (c) If we were to believe this student, what would be the *subthreshold slope/swing*, S , of this MOSFET?
- (d) This student, however, was subsequently expelled from the university for falsifying this data, leaving him with nothing to show for his years at school but huge college loan debts. What *clearly* physically unrealistic aspect of these MOSFET characteristics (besides less than smooth curves) should have drawn suspicion?

SOLID STATE ELECTRONIC DEVICES

SIXTH EDITION



BEN G. STREETMAN
SANJAY KUMAR BANERJEE

both of University of Texas at Austin

This classic introductory textbook in its sixth edition retains the two basic goals: (1) develop the basic semiconductor physics concepts to understand current and future devices and (2) provides a sound understanding of current semiconductor devices and technology so that their applications to electronic and optoelectronic circuits and systems can be appreciated.

The book is useful for advanced undergraduate and postgraduate students of electrical and electronics engineering and should also prove useful for practicing engineers and scientists, interested in updating their understanding of modern electronics.

Key Features

- The basics of semiconductor materials and conduction processes in solids. Incorporated to help students understand p-n junctions, bipolar and metal oxide semiconductor transistors, optoelectronic and other devices.
- Extensive discussion of circuit and other application examples to provide students with feedback about the practical relevance of the theory.
- Key equations for semiconductor devices are listed on the inside covers of the book, arranged thematically to help students digest concepts.
- Provides a summary of key concepts at end of each chapter, and chapter objectives at the beginning.
- Approximately 200 problems along with current references that extend concepts in the text.

New to this Edition

- Discussion of advanced MOS and bipolar processes and structures such as high-k gate dielectrics, polyemitter BJTs and graded SiGe base HBTs.
- Consolidated treatment of optoelectronic devices, discusses sources such as white LEDs for illumination, detectors such as SACM and waveguide photodiodes, and advances in solar cells.
- Updated discussion of device fabrication processes and CMOS integrated circuit technology provides a useful understanding of how semiconductor devices are made.
- Summary of key concepts at end of each chapter.
- Answers to Selected Problems in the back of the book.

Rs. 295.00

www.phindia.com

