CHAPTER 3

MOS TRANSISTOR

The *MOS Field Effect Transistor* (MOSFET) is the fundamental building block of MOS and CMOS digital integrated circuits. Compared to the bipolar junction transistor (BJT), the MOS transistor occupies a relatively smaller silicon area, and its fabrication involves fewer processing steps. These technological advantages, together with the relative simplicity of MOSFET operation, have helped make the MOS transistor the most widely used switching device in LSI and VLSI circuits. In this chapter, we will examine the basic structure and the electrical behavior of nMOS (n-channel MOS), as well as pMOS (p-channel MOS) devices. The nMOS transistor is used as the primary switching device in virtually all digital circuit applications, whereas the pMOS transistor is used mostly in conjunction with the nMOS device in CMOS circuits. However, the basic operation principles of both nMOS and pMOS transistors are very similar to each other.

This chapter starts with a detailed investigation of the basic electrical and physical properties of *Metal Oxide Semiconductor* (MOS) systems, upon which the MOSFET structure is based. We will consider the effects of external bias conditions on charge distribution in the MOS system and on the conductance of free carriers. It will be shown that, in field effect devices, the current flow is controlled by externally applied electric fields, and that the operation depends only on the majority carrier flow between two device terminals. Next, the current-voltage characteristics of MOS transistors will be examined in detail, including physical limitations imposed by small device geometries and various second-order effects observed in MOSFETs. Note that these considerations will be particularly important for the overall performance of large-scale digital circuits built by using small-geometry MOSFET devices.

3.1. The Metal Oxide Semiconductor (MOS) Structure

We will start our investigation by considering the electrical behavior of the simple two-terminal MOS structure shown in Fig. 3.1. Note that the structure consists of three layers: The metal *gate electrode*, the insulating oxide (SiO_2) layer, and the p-type bulk semiconductor (Si), called the *substrate*. As such, the MOS structure forms a capacitor, with the gate and the substrate acting as the two terminals (plates) and the oxide layer as the dielectric. The thickness of the silicon dioxide layer is usually between 10 nm and 50 nm. The carrier concentration and its local distribution within the semiconductor substrate can now be manipulated by the external voltages applied to the gate and substrate terminals. A basic understanding of the bias conditions for establishing different carrier concentrations in the substrate will also provide valuable insight into the operating conditions of more complicated MOSFET structures.

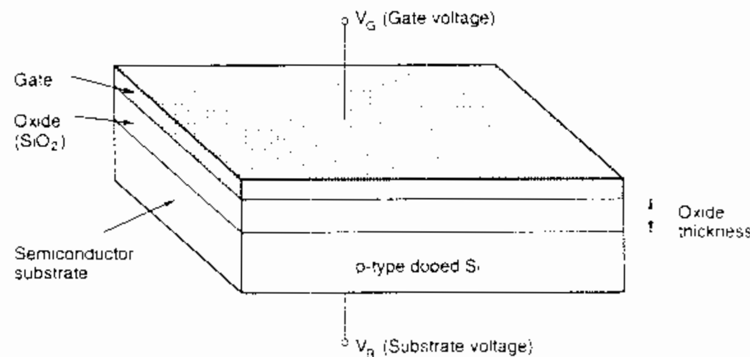

Figure 3.1. Two-terminal MOS structure.

Consider first the basic electrical properties of the semiconductor (Si) substrate, which acts as one of the electrodes of the MOS capacitor. The equilibrium concentrations of mobile carriers in a semiconductor always obey the *Mass Action Law* given by

$$n \cdot p = n_i^2 \tag{3.1}$$

Here, n and p denote the mobile carrier concentrations of electrons and holes, respectively, and n_i denotes the intrinsic carrier concentration of silicon, which is a function of the temperature T. At room temperature, i.e., $T = 300$ K, n_i is approximately equal to 1.45 x 10^{10} cm^{-3}. Assuming that the substrate is uniformly doped with an acceptor (e.g., Boron) concentration N_A, the equilibrium electron and hole concentrations in the p-type substrate are approximated by

$$n_{po} \cong \frac{n_i^2}{N_A} \qquad p_{po} \cong N_A \tag{3.2}$$

The doping concentration N_A is typically on the order of 10^{15} to 10^{16} cm^{-3}; thus, it is much greater than the intrinsic carrier concentration n_i. Note that the bulk electron and hole concentrations given in (3.2) are valid in the regions farther away from the *surface*, where the semiconductor substrate and the oxide layer meet. The conditions on the surface, however, are far more significant for the electrical behavior and the operation of the MOS system, and we will discuss these conditions in more detail.

The energy band diagram of the p-type substrate is shown in Fig. 3.2. The band-gap between the *conduction band* and the *valence band* for silicon is approximately 1.1 eV. The location of the *equilibrium Fermi level* E_F within the band-gap is determined by the doping type and the doping concentration in the silicon substrate. The *Fermi potential* ϕ_F, which is a function of temperature and doping, denotes the difference between the intrinsic Fermi level E_i, and the Fermi level E_F.

$$\phi_F = \frac{E_F - E_i}{q} \tag{3.3}$$

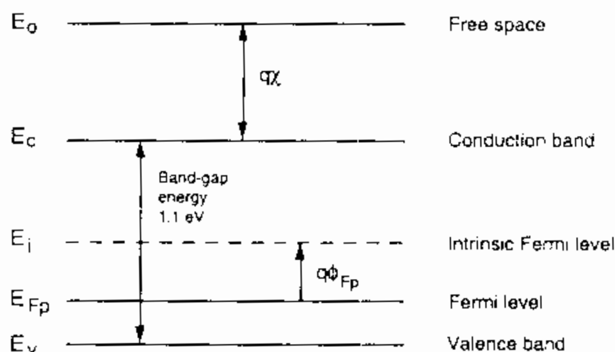

Figure 3.2. Energy band diagram of a p-type silicon substrate.

For a p-type semiconductor, the Fermi potential can be approximated by

$$\phi_{Fp} = \frac{kT}{q} \ln \frac{n_i}{N_A} \tag{3.4}$$

whereas for an n-type semiconductor (doped with a donor concentration N_D), the Fermi potential is given by

$$\phi_{Fn} = \frac{kT}{q} \ln \frac{N_D}{n_i} \tag{3.5}$$

Here, k denotes the Boltzmann constant and q denotes the unit (electron) charge. Note that the definitions given in (3.4) and (3.5) result in a positive Fermi potential for n-type

material, and a negative Fermi potential for p-type material. We will use this convention throughout the text. The *electron affinity* of silicon, which is the potential difference between the conduction band level and the vacuum (free-space) level, is denoted by $q\chi$ in Fig. 3.2. The energy required for an electron to move from the Fermi level into free space is called the *work function* $q\Phi_S$, and is given by

$$q\Phi_S = q\chi + \left(E_c - E_F\right) \tag{3.6}$$

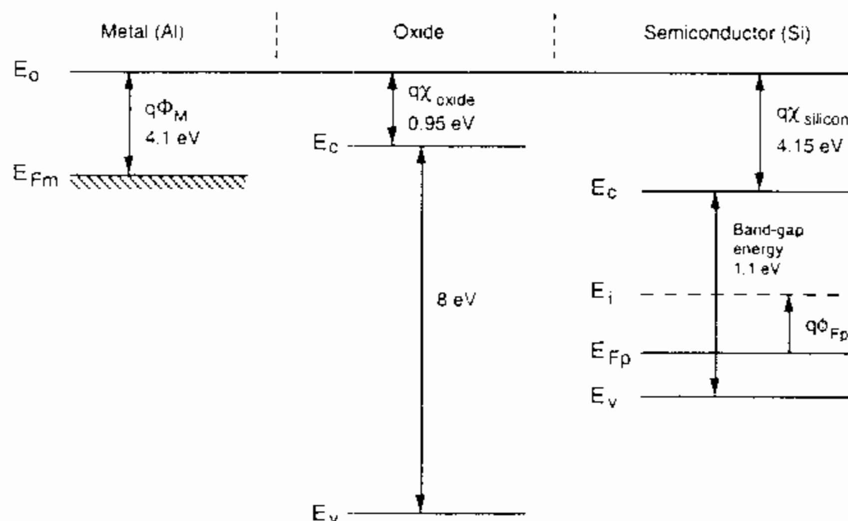

Figure 3.3. Energy band diagrams of the components that make up the MOS system.

The insulating silicon dioxide layer between the silicon substrate and the gate has a large band-gap of about 8 eV and an electron affinity of about 0.95 eV. On the other hand, the work function $q\Phi_M$ of an aluminum gate is about 4.1 eV. Figure 3.3 shows the energy band diagrams of metal, oxide, and semiconductor layers in a MOS system as three separate components.

Now consider that the three components of the ideal MOS system are brought into physical contact. The Fermi levels of all three materials must line up, as they form the MOS capacitor shown in Fig. 3.1. Because of the work-function difference between the metal and the semiconductor, a voltage drop occurs across the MOS system. Part of this built-in voltage drop occurs across the insulating oxide layer. The rest of the voltage drop (potential difference) occurs at the silicon surface next to the silicon-oxide interface, forcing the energy bands of silicon to bend in this region. The resulting combined energy band diagram of the MOS system is shown in Fig. 3.4. Notice that the equilibrium Fermi levels of the semiconductor (Si) substrate and the metal gate are at the same potential. The bulk Fermi level is not significantly affected by the band bending, whereas the surface Fermi level moves closer to the intrinsic Fermi (mid-gap) level. The Fermi potential at the surface, also called *surface potential* ϕ_s, is smaller in magnitude than the bulk Fermi potential ϕ_F.

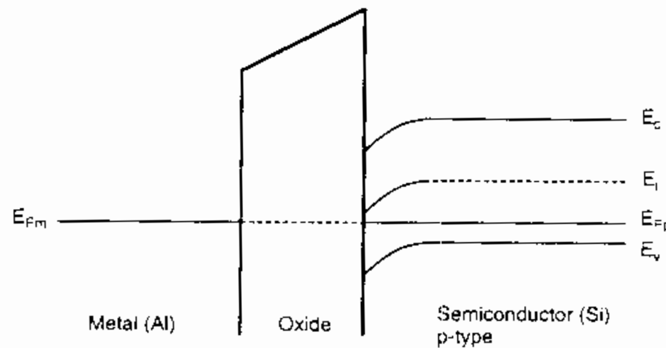

Figure 3.4. Energy band diagram of the combined MOS system.

Example 3.1.

Consider the MOS structure that consists of a p-type doped silicon substrate, a silicon dioxide layer, and a metal (aluminum) gate. The equilibrium Fermi potential of the doped silicon substrate is given as $q\phi_{Fp} = 0.2$ eV. Using the electron affinity for silicon and the work function for aluminum given in Fig. 3.3, calculate the built-in potential difference across the MOS system. Assume that the MOS system contains no other charges in the oxide or on the silicon-oxide interface.

First, we have to calculate the work function for the doped silicon, which is given by (3.6). Since the electron affinity of silicon is 4.15 eV, the work function $q\Phi_S$ is found as

$$q\Phi_S = 4.15 \text{ eV} + 0.75 \text{ eV} = 4.9 \text{ eV}$$

Now calculate the work function difference between the silicon substrate and the aluminum gate. Note that the work function of aluminum is given as 4.1 eV in Fig. 3.3. Thus, the built-in potential difference across this MOS system is

$$q\Phi_M - q\Phi_S = 4.1 \text{ eV} - 4.9 \text{ eV} = -0.8 \text{ eV}$$

If a voltage corresponding to this potential difference is applied externally between the gate and the substrate, the bending of the energy bands near the surface can be compensated, i.e., the energy bands become "flat." Thus, the voltage defined by

$$V_{FB} = \Phi_M - \Phi_S$$

is called the *flat-band* voltage.

3.2. The MOS System under External Bias

We now turn our attention to the electrical behavior of the MOS structure under externally applied bias voltages. Assume that the substrate voltage is set at $V_B = 0$, and let the gate voltage be the controlling parameter. Depending on the polarity and the magnitude of V_G, three different operating regions can be observed for the MOS system: *accumulation*, *depletion*, and *inversion*.

If a negative voltage V_G is applied to the gate electrode, the holes in the p-type substrate are attracted to the semiconductor-oxide interface. The majority carrier concentration near the surface becomes larger than the equilibrium hole concentration in the substrate; hence, this condition is called carrier *accumulation* on the surface (Fig. 3.5). Note that in this case, the oxide electric field is directed towards the gate electrode. The negative surface potential also causes the energy bands to bend upward near the surface. While the hole density near the surface increases as a result of the applied negative gate bias, the electron (minority carrier) concentration decreases as the negatively charged electrons are pushed deeper into the substrate.

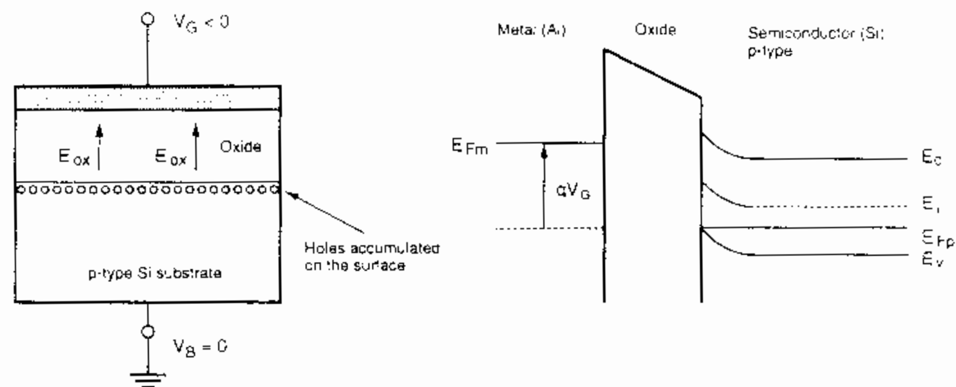

Figure 3.5. The cross-sectional view and the energy band diagram of the MOS structure operating in accumulation region.

Now consider the next case in which a small positive gate bias V_G is applied to the gate electrode. Since the substrate bias is zero, the oxide electric field will be directed towards the substrate in this case. The positive surface potential causes the energy bands to bend downward near the surface, as shown in Fig. 3.6. The majority carriers, i.e., the holes in the substrate, will be repelled back into the substrate as a result of the positive gate bias, and these holes will leave negatively charged fixed acceptor ions behind. Thus, a *depletion* region is created near the surface. Note that under this bias condition, the region near the semiconductor-oxide interface is nearly devoid of all mobile carriers.

The thickness x_d of this depletion region on the surface can easily be found as a function of the surface potential ϕ_s. Assume that the mobile hole charge in a thin horizontal layer parallel to the surface is

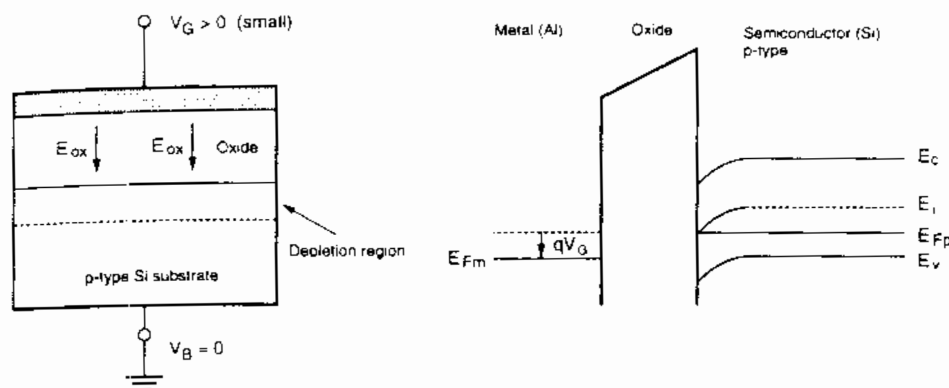

Figure 3.6. The cross-sectional view and the energy band diagram of the MOS structure operating in depletion mode, under small gate bias.

$$dQ = -q \cdot N_A \cdot dx \tag{3.7}$$

The *change* in surface potential required to displace this charge sheet dQ by a distance x_d away from the surface can be found by using the Poisson equation.

$$d\phi_s = -x \cdot \frac{dQ}{\varepsilon_{Si}} = \frac{q \cdot N_A \cdot x}{\varepsilon_{Si}} dx \tag{3.8}$$

Integrating (3.7) along the vertical dimension (perpendicular to the surface) yields

$$\int_{\phi_F}^{\phi_s} d\phi_s = \int_0^{x_d} \frac{q \cdot N_A \cdot x}{\varepsilon_{Si}} \, dx \tag{3.9}$$

$$\phi_s - \phi_F = \frac{q \cdot N_A \cdot x_d^2}{2\,\varepsilon_{Si}} \tag{3.10}$$

Thus, the depth of the depletion region is

$$x_d = \sqrt{\frac{2\,\varepsilon_{Si} \cdot |\phi_s - \phi_F|}{q \cdot N_A}} \tag{3.11}$$

and the depletion region charge density, which consists solely of fixed acceptor ions in this region, is given by the following expression

$$Q = -q \cdot N_A \cdot x_d = -\sqrt{2q \cdot N_A \cdot \varepsilon_{Si} \cdot |\phi_s - \phi_F|} \qquad (3.12)$$

The amount of this depletion region charge plays a very important role in the analysis of threshold voltage, as we will examine shortly.

To complete our qualitative overview of different bias conditions and their effects upon the MOS system, consider next a further increase in the positive gate bias. As a result of the increasing surface potential, the downward bending of the energy bands will increase as well. Eventually, the mid-gap energy level E_i becomes smaller than the Fermi level E_{Fp} on the surface, which means that the substrate semiconductor in this region becomes n-type. Within this thin layer, the electron density is larger than the majority hole density, since the positive gate potential attracts additional minority carriers (electrons) from the bulk substrate to the surface (Fig. 3.7). The n-type region created near the surface by the positive gate bias is called the *inversion layer*, and this condition is called *surface inversion*. It will be seen that the thin inversion layer on the surface with a large mobile electron concentration can be utilized for conducting current between two terminals of the MOS transistor.

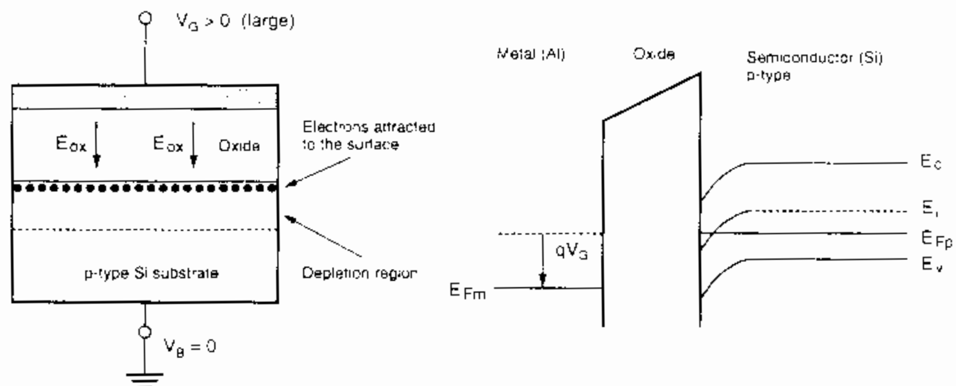

Figure 3.7. The cross-sectional view and the energy band diagram of the MOS structure in surface inversion, under larger gate bias voltage.

As a practical definition, the surface is said to be *inverted* when the density of mobile electrons on the surface becomes equal to the density of holes in the bulk (p-type) substrate. This condition requires that the surface potential has the same magnitude, but the reverse polarity, as the bulk Fermi potential ϕ_F. Once the surface is inverted, any further increase in the gate voltage leads to an increase of mobile electron concentration on the surface, but not to an increase of the depletion depth. Thus, the depletion region depth achieved at the onset of surface inversion is also equal to the maximum depletion depth, x_{dm}, which remains constant for higher gate voltages. Using the inversion condition $\phi_s = -\phi_F$, the maximum depletion region depth at the onset of surface inversion can be found from (3.11) as follows:

$$x_{dm} = \sqrt{\frac{2 \cdot \varepsilon_{Si} \cdot |2\phi_F|}{q \cdot N_A}} \tag{3.13}$$

The creation of a conducting surface inversion layer through externally applied gate bias is an essential phenomenon for current conduction in MOS transistors. In the following section, we will examine the structure and the operation of the MOS Field Effect Transistor (MOSFET).

3.3. Structure and Operation of MOS Transistor (MOSFET)

The basic structure of an n-channel MOSFET is shown in Fig. 3.8. This four-terminal device consists of a p-type substrate, in which two n^+ diffusion regions, the drain and the source, are formed. The surface of the substrate region between the drain and the source is covered with a thin oxide layer, and the metal (or polysilicon) gate is deposited on top of this gate dielectric. The midsection of the device can easily be recognized as the basic MOS structure which was examined in the previous sections. The two n^+ regions will be the current-conducting terminals of this device. Note that the device structure is completely symmetrical with respect to the drain and source regions; the different roles of these two regions will be defined only in conjunction with the applied terminal voltages and the direction of the current flow.

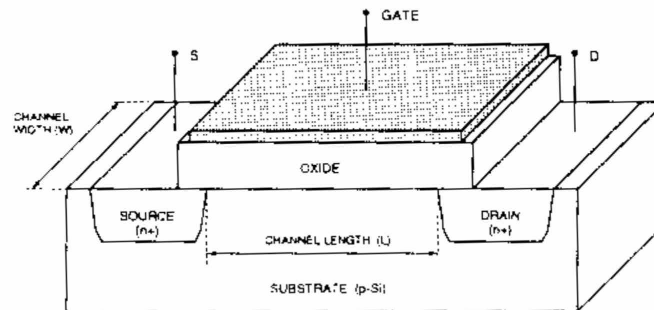

Figure 3.8. The physical structure of an n-channel enhancement-type MOSFET.

A conducting *channel* will eventually be formed through applied gate voltage in the section of the device between the drain and the source diffusion regions. The distance between the drain and source diffusion regions is the *channel length* L, and the lateral extent of the channel (perpendicular to the length dimension) is the *channel width* W. Both the channel length and the channel width are important parameters which can be used to control some of the electrical properties of the MOSFET. The thickness of the oxide layer covering the channel region, t_{ox}, is also an important parameter.

A MOS transistor which has no conducting channel region at zero gate bias is called an *enhancement-type* (or *enhancement-mode*) MOSFET. If a conducting channel already exists at zero gate bias, on the other hand, the device is called a *depletion-type* (or

depletion-mode) MOSFET. In a MOSFET with p-type substrate and with n^+ source and drain regions, the channel region to be formed on the surface is n-type. Thus, such a device with p-type substrate is called an *n-channel MOSFET*. In a MOSFET with n-type substrate and with p^+ source and drain regions, on the other hand, the channel is p-type and the device is called a *p-channel MOSFET*.

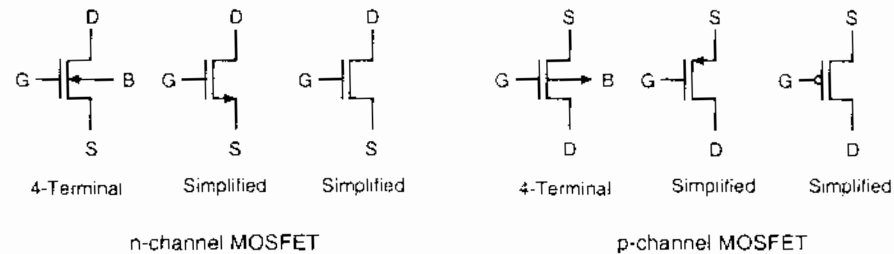

Figure 3.9. Circuit symbols for n-channel and p-channel enhancement-type MOSFETs.

The abbreviations used for the device terminals are: G for the gate, D for the drain, S for the source, and B for the substrate (or body). In an n-channel MOSFET, the source is defined as the n^+ region which has a *lower* potential than the other n^+ region, the drain. By convention, all terminal voltages of the device are defined with respect to the source potential. Thus, the gate-to-source voltage is denoted by V_{GS}, the drain-to-source voltage is denoted by V_{DS}, and the substrate-to-source voltage is denoted by V_{BS}. Circuit symbols for both n-channel and p-channel enhancement-type MOSFETs are shown in Fig. 3.9. While the four-terminal symbolic representation shows all external terminals of the device, the simple three-terminal representation will also be used extensively. Note that in the simple MOSFET circuit symbol, the small arrow always marks the source terminal.

Consider first the n-channel enhancement-type MOSFET shown in Fig. 3.8. The simple operation principle of this device is: *control the current conduction between the source and the drain, using the electric field generated by the gate voltage as a control variable*. Since the current flow in the channel is also controlled by the drain-to-source voltage and by the substrate voltage, the current can be considered a function of these external terminal voltages. We will examine in detail the functional relationships between the channel current (also called the *drain current*) and the terminal voltages. In order to start current flow between the source and the drain regions, however, we have to form a conducting channel first.

The simplest bias condition that can be applied to the n-channel enhancement-type MOSFET is shown in Fig. 3.10. The source, the drain, and the substrate terminals are all connected to ground. A positive gate-to-source voltage V_{GS} is then applied to the gate in order to create the conducting channel underneath the gate. With this bias arrangement, the channel region between the source and the drain diffusions behaves exactly the same as for the simple MOS structure we examined in Section 3.2. For small gate voltage levels, the majority carriers (holes) are repelled back into the substrate, and the surface of the p-type substrate is depleted. Since the surface is devoid of any mobile carriers, current conduction between the source and the drain is not possible.

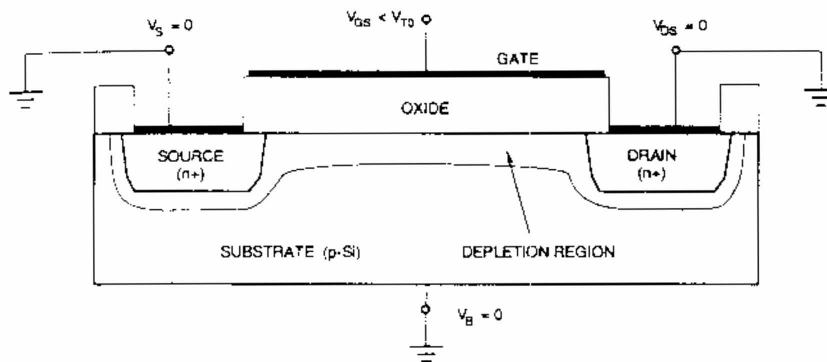

Figure 3.10. Formation of a depletion region in an n-channel enhancement-type MOSFET.

Now assume that the gate-to-source voltage is further increased. As soon as the surface potential in the channel region reaches $-\phi_{Fp}$, surface inversion will be established, and a conducting n-type layer will form between the source and the drain diffusion regions (Fig. 3.11). This channel now provides an electrical connection between the two n^+ regions, and it allows current flow, as long as there is a potential difference between the source and the drain terminal voltages (Fig. 3.12). The bias conditions for the onset of surface inversion and for the creation of the conducting channel are therefore very significant for MOSFET operation.

The value of the gate-to-source voltage V_{GS} needed to cause surface inversion (to create the conducting channel) is called the *threshold voltage* V_{T0}. Any gate-

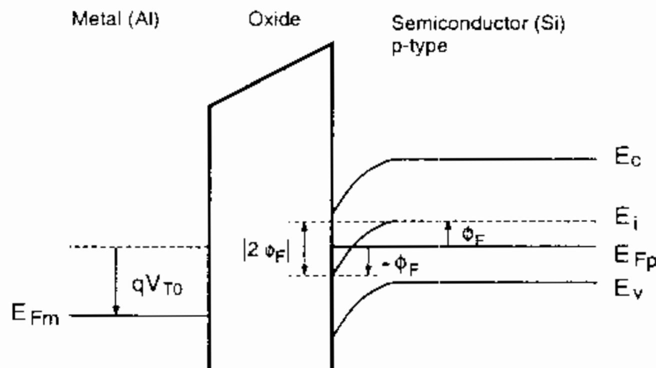

Figure 3.11. Band diagram of the MOS structure underneath the gate, at surface inversion. Notice the band bending by $|2\phi_F|$ at the surface.

to-source voltage smaller than V_{T0} is not sufficient to establish an inversion layer; thus, the MOSFET can conduct no current between its source and drain terminals unless $V_{GS} > V_{T0}$. For gate-to-source voltages larger than the threshold voltage, on the other hand,

a larger number of minority carriers (electrons) are attracted to the surface, which ultimately contribute to channel current conduction. Also note that increasing the gate-to-source voltage above and beyond the threshold voltage will not affect the surface potential and the depletion region depth. Both quantities will remain approximately constant and equal to their values attained at the onset of surface inversion.

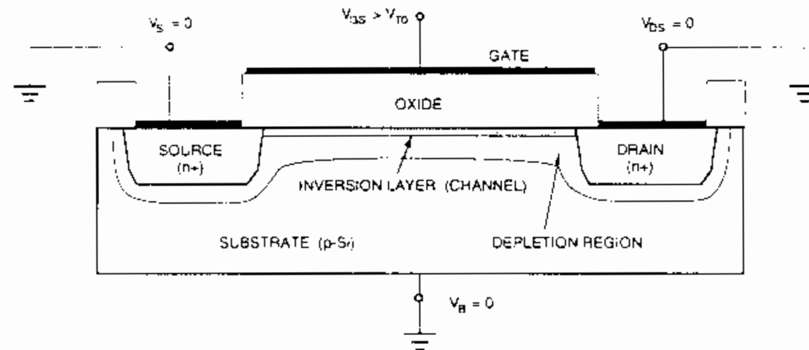

Figure 3.12. Formation of an inversion layer (channel) in an n-channel enhancement-type MOSFET.

The Threshold Voltage

In the following, physical parameters affecting the threshold voltage of a MOS structure will be examined by considering the various components of V_{T0}. For all practical purposes, we can identify four physical components of the threshold voltage: (i) the work function difference between the gate and the channel, (ii) the gate voltage component to change the surface potential, (iii) the gate voltage component to offset the depletion region charge, and (iv) the voltage component to offset the fixed charges in the gate oxide and in the silicon-oxide interface. The analysis will be carried out for an n-channel device, but the results are applicable to p-channel devices as well, with minor modifications.

The work function difference Φ_{GC} between the gate and the channel reflects the built-in potential of the MOS system, which consists of the p-type substrate, the thin silicon dioxide layer, and the gate electrode. Depending on the gate material, the work function difference is

$$\Phi_{GC} = \phi_F(substrate) - \phi_M \qquad \text{for metal gate} \qquad (3.14)$$

$$\Phi_{GC} = \phi_F(substrate) - \phi_F(gate) \qquad \text{for polysilicon gate} \qquad (3.15)$$

This first component of the threshold voltage accounts for part of the voltage drop across the MOS system that is built-in. Now, the externally applied gate voltage must be changed to achieve surface inversion, i.e., to change the surface potential by $-2\phi_F$. This will be the second component of the threshold voltage.

Another component of the applied gate voltage is necessary to offset the depletion region charge, which is due to the fixed acceptor ions located in the depletion region near the surface. We can calculate the depletion region charge density at surface inversion ($\phi_s = -\phi_F$) using (3.12).

$$Q_{B0} = -\sqrt{2q \cdot N_A \cdot \varepsilon_{Si} \cdot |-2\phi_F|} \tag{3.16}$$

Note that if the substrate (body) is biased at a different voltage level than the source, which is at ground potential (reference), then the depletion region charge density can be expressed as a function of the source-to-substrate voltage V_{SB}.

$$Q_B = -\sqrt{2q \cdot N_A \cdot \varepsilon_{Si} \cdot |-2\phi_F + V_{SB}|} \tag{3.17}$$

The component that offsets the depletion region charge is then equal to $-Q_B/C_{ox}$, where C_{ox} is the gate oxide capacitance per unit area.

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \tag{3.18}$$

Finally, we must consider the influence of a nonideal physical phenomenon which we have neglected until now. There always exists a fixed positive charge density Q_{ox} at the interface between the gate oxide and the silicon substrate, due to impurities and/or lattice imperfections at the interface. The gate voltage component that is necessary to offset this positive charge at the interface is $-Q_{ox}/C_{ox}$. Now, we can combine all of these voltage components to find the threshold voltage. For zero substrate bias, the threshold voltage V_{T0} is expressed as follows:

$$V_{T0} = \Phi_{GC} - 2\phi_F - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \tag{3.19}$$

For nonzero substrate bias, on the other hand, the depletion charge density term must be modified to reflect the influence of V_{SB} upon that charge, resulting in the following generalized threshold voltage expression.

$$V_T = \Phi_{GC} - 2\phi_F - \frac{Q_B}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \tag{3.20}$$

The generalized form of the threshold voltage can also be written as

$$V_T = \Phi_{GC} - 2\phi_F - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} - \frac{Q_B - Q_{B0}}{C_{ox}} = V_{T0} - \frac{Q_B - Q_{B0}}{C_{ox}} \tag{3.21}$$

Note that in this case, the threshold voltage differs from V_{T0} only by an additive term. This substrate-bias term is a simple function of the material constants and of the source-to-substrate voltage V_{SB}.

$$\frac{Q_B - Q_{B0}}{C_{ox}} = -\frac{\sqrt{2q \cdot N_A \cdot \varepsilon_{Si}}}{C_{ox}} \cdot \left(\sqrt{\left| -2\phi_F + V_{SB} \right|} - \sqrt{\left| 2\phi_F \right|} \right) \tag{3.22}$$

Thus, the most general expression of the threshold voltage V_T can be found as follows:

$$V_T = V_{T0} + \gamma \cdot \left(\sqrt{\left| -2\phi_F + V_{SB} \right|} - \sqrt{\left| 2\phi_F \right|} \right) \tag{3.23}$$

where the parameter γ

$$\gamma = \frac{\sqrt{2q \cdot N_A \cdot \varepsilon_{Si}}}{C_{ox}} \tag{3.24}$$

is the *substrate-bias* (or *body-effect*) coefficient.

The threshold voltage expression given in (3.23) can be used both for n-channel and p-channel MOS transistors. One must be careful, however, since some of the terms and coefficients in this equation have different polarities for the n-channel (nMOS) case and for the p-channel (pMOS) case. The reason for this polarity difference is that the substrate semiconductor is p-type in an n-channel MOSFET and n-type in a p-channel MOSFET. Specifically,

- The substrate Fermi potential ϕ_F is *negative in nMOS, positive in pMOS.*
- The depletion region charge densities Q_{B0} and Q_B are *negative in nMOS, positive in pMOS.*
- The substrate bias coefficient γ is *positive in nMOS, negative in pMOS.*
- The substrate bias voltage V_{SB} is *positive in nMOS, negative in pMOS.*

Typically, the threshold voltage of an enhancement-type n-channel MOSFET is a positive quantity, whereas the threshold voltage of a p-channel MOSFET is negative.

Example 3.2.

Calculate the threshold voltage V_{T0} at $V_{SB} = 0$, for a polysilicon gate n-channel MOS transistor, with the following parameters: substrate doping density $N_A = 10^{16}$ cm^{-3}, polysilicon gate doping density $N_D = 2 \times 10^{20}$ cm^{-3}, gate oxide thickness $t_{ox} = 500$ Å, and oxide-interface fixed charge density $N_{ox} = 4 \times 10^{10}$ cm^{-2}.

First, calculate the Fermi potentials for the p-type substrate and for the n-type polysilicon gate:

$$\phi_F(substrate) = \frac{kT}{q} \ln\left(\frac{n_i}{N_A}\right) = 0.026 \text{ V} \cdot \ln\left(\frac{1.45 \cdot 10^{10}}{10^{16}}\right) = -0.35 \text{ V}$$

Since the doping density of the polysilicon gate is very high, the heavily doped n-type gate material is expected to be degenerate. Thus, we may assume that the Fermi potential of the polysilicon gate is approximately equal to the conduction band potential, i.e., $\phi_F(gate) = 0.55$ V. Now, calculate the work function difference between the gate and the channel:

$$\Phi_{GC} = \phi_F(substrate) - \phi_F(gate) = -0.35 \text{ V} - 0.55 \text{ V} = -0.90 \text{ V}$$

The depletion region charge density at $V_{SB} = 0$ is found as follows:

$$\begin{aligned} Q_{B0} = & -\sqrt{2 \cdot q \cdot N_A \cdot \varepsilon_{Si} \cdot |-2\phi_F(substrate)|} \\ & = -\sqrt{2 \cdot 1.6 \cdot 10^{-19} \cdot 10^{16} \cdot 11.7 \cdot 8.85 \cdot 10^{-14} |-2 \cdot 0.35|} = -4.82 \cdot 10^{-8} \text{C/cm}^2 \end{aligned}$$

The oxide-interface charge is :

$$Q_{ox} = q \cdot N_{ox} = 1.6 \cdot 10^{-19} \text{C} \times 4 \cdot 10^{10} \text{cm}^{-2} = 6.4 \cdot 10^{-9} \text{C/cm}^2$$

The gate oxide capacitance per unit area is calculated using the dielectric constant of silicon dioxide and the oxide thickness t_{ox}.

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \frac{3.97 \cdot 8.85 \cdot 10^{-14} \text{ F/cm}}{500 \cdot 10^{-8} \text{ cm}} = 7.03 \cdot 10^{-8} \text{ F/cm}^2$$

Now, we can combine all components and calculate the threshold voltage.

$$\begin{aligned} V_{T0} & = \Phi_{GC} - 2\phi_F(substrate) - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \\ & = -0.90 - (-0.70) - (-0.69) - 0.09 = 0.40 \text{ V} \end{aligned}$$

In this simplified analysis, the doping concentrations of the source and the drain diffusion regions and the geometry (physical dimensions) of the channel region have no influence upon the threshold voltage V_{T0}.

Note that the exact value of the threshold voltage of an actual MOS transistor cannot be determined using (3.23) in most practical cases, due primarily to uncertainties and variations of the doping concentrations, the oxide thickness, and the fixed oxide-interface charge. The nominal value and the statistical range of the threshold voltage for any MOS process are ultimately determined by direct measurements, which will be described later in Section 3.4. In most MOS fabrication processes, the threshold voltage can be adjusted

by selective dopant ion implantation into the channel region of the MOSFET. For n-channel MOSFETs, the threshold voltage is *increased* (made more positive) by adding extra p-type impurities (acceptor ions). Alternatively, the threshold voltage of the n-channel MOSFET can be *decreased* (made more negative) by implanting n-type impurities (dopant ions) into the channel region.

The amount of change in the threshold voltage as a result of extra implants can be approximated as follows. Let the density of implanted impurities be represented by N_I [cm^{-2}]. Assume that all implanted ions are electrically active, i.e., each ion contributes to the depletion region charge. Then, the threshold voltage V_{T0} at zero substrate bias ($V_{SB} = 0$) will be shifted by an amount of qN_I/C_{ox}. This approximation obviously neglects the variation of the substrate Fermi level ϕ_F as the result of extra implants, but it nevertheless provides a fair estimate for the threshold voltage shift.

Exercise 3.1

Consider the following p-channel MOSFET process:

Substrate doping $N_D = 10^{15}$ cm^{-3}, polysilicon gate doping density $N_D = 10^{20}$ cm^{-3}, gate oxide thickness $t_{ox} = 650$ Å, and oxide-interface charge density $N_{ox} = 2 \times 10^{10}$ cm^{-2}. Use $\varepsilon_{Si} = 11.7\varepsilon_0$ and $\varepsilon_{ox} = 3.97\varepsilon_0$ for the dielectric coefficients of silicon and silicon-dioxide, respectively.

(a) Calculate the threshold voltage V_{T0}, for $V_{SB} = 0$.

(b) Determine the type and the amount of channel ion implantation which are necessary to achieve a threshold voltage of $V_{T0} = -2$ V.

Note that, using selective ion implantation into the channel, the threshold voltage of an n-channel MOSFET can also be made negative. This means that the resulting nMOS transistor will have a conducting channel at $V_{GS} = 0$, enabling current flow between its source and drain terminals as long as V_{GS} is larger than the negative threshold voltage. Such a device is called a *depletion-type* (or *normally-on*) n-channel MOSFET. We will see several practical applications for depletion-type nMOS transistors in the design of MOS digital circuits. Except for its negative threshold voltage, the depletion-type n-channel MOSFET exhibits the same electrical behavior as the enhancement-type n-channel MOSFET. Figure 3.13 shows the conventional circuit symbols used for depletion-type n-channel MOSFETs.

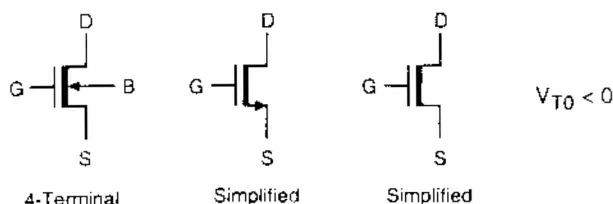

Figure 3.13. Circuit symbols for n-channel depletion-type MOSFETs.

Example 3.3.

Consider the n-channel MOSFET process given in Example 3.2. In several digital circuit applications, the condition $V_{SB} = 0$ cannot be guaranteed for all transistors. We will examine in this example how a nonzero source-to-substrate voltage V_{SB} affects the threshold voltage of the MOS transistor.

First, we must calculate the substrate-bias coefficient γ using the process parameters given in Example 3.2.

$$\gamma = \frac{\sqrt{2 \cdot q \cdot N_A \cdot \varepsilon_{Si}}}{C_{ox}} = \frac{\sqrt{2 \cdot 1.6 \cdot 10^{-19} \cdot 10^{16} \cdot 11.7 \cdot 8.85 \cdot 10^{-14}}}{7.03 \cdot 10^{-8}} = 0.82\ \mathrm{V}^{\frac{1}{2}}$$

Now we compute and plot the threshold voltage V_T as a function of the source-to-substrate voltage V_{SB}. The voltage V_{SB} will be assumed to vary between zero and 5 V.

$$V_T = V_{T0} + \gamma \cdot \left(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|2\phi_F|} \right) = 0.40 + 0.82 \cdot \left(\sqrt{0.7 + V_{SB}} - \sqrt{0.7} \right)$$

It is seen that the threshold voltage variation is about 1.3 V over this range, which could present serious design problems if neglected. We will see in the following chapters that the substrate-bias effect is unavoidable in most digital circuits and that the circuit designer usually must take appropriate measures to account for and/or to compensate for the threshold voltage variations.