

CHAPTER 4

MOS Field-Effect Transistors (MOSFETs)

Introduction	235	4.8	The MOSFET Internal Capacitances and High-Frequency Model	320
4.1 Device Structure and Physical Operation	236	4.9	Frequency Response of the CS Amplifier	326
4.2 Current-Voltage Characteristics	248	4.10	The CMOS Digital Logic Inverter	336
4.3 MOSFET Circuits at DC	262	4.11	The Depletion-Type MOSFET	346
4.4 The MOSFET as an Amplifier and as a Switch	270	4.12	The SPICE MOSFET Model and Simulation Example	351
4.5 Biasing in MOS Amplifier Circuits	280		Summary	359
4.6 Small-Signal Operation and Models	287		Problems	360
4.7 Single-Stage MOS Amplifiers	299			

INTRODUCTION

Having studied the junction diode, which is the most basic two-terminal semiconductor device, we now turn our attention to three-terminal semiconductor devices. Three-terminal devices are far more useful than two-terminal ones because they can be used in a multitude of applications, ranging from signal amplification to digital logic and memory. The basic principle involved is the use of the voltage between two terminals to control the current flowing in the third terminal. In this way a three-terminal device can be used to realize a controlled source, which as we have learned in Chapter 1 is the basis for amplifier design. Also, in the extreme, the control signal can be used to cause the current in the third terminal to change from zero to a large value, thus allowing the device to act as a switch. As we also

learned in Chapter 1, the switch is the basis for the realization of the logic inverter, the basic element of digital circuits.

There are two major types of three-terminal semiconductor device: the metal-oxide-semiconductor field-effect transistor (MOSFET), which is studied in this chapter, and the bipolar junction transistor (BJT), which we shall study in Chapter 5. Although each of the two transistor types offers unique features and areas of application, the MOSFET has become by far the most widely used electronic device, especially in the design of integrated circuits (ICs), which are circuits fabricated on a single silicon chip.

Compared to BJTs, MOSFETs can be made quite small (i.e., requiring a small area on the silicon IC chip), and their manufacturing process is relatively simple (see Appendix A). Also, their operation requires comparatively little power. Furthermore, circuit designers have found ingenious ways to implement digital and analog functions utilizing MOSFETs almost exclusively (i.e., with very few or no resistors). All of these properties have made it possible to pack large numbers of MOSFETs (>200 million!) on a single IC chip to implement very sophisticated, very-large-scale-integrated (VLSI) circuits such as those for memory and microprocessors. Analog circuits such as amplifiers and filters are also implemented in MOS technology, albeit in smaller less-dense chips. Also, both analog and digital functions are increasingly being implemented on the same IC chip, in what is known as mixed-signal design.

The objective of this chapter is to develop in the reader a high degree of familiarity with the MOSFET: its physical structure and operation, terminal characteristics, circuit models, and basic circuit applications, both as an amplifier and a digital logic inverter. Although discrete MOS transistors exist, and the material studied in this chapter will enable the reader to design discrete MOS circuits, our study of the MOSFET is strongly influenced by the fact that most of its applications are in integrated-circuit design. The design of IC analog and digital MOS circuits occupies a large proportion of the remainder of this book.



4.1 DEVICE STRUCTURE AND PHYSICAL OPERATION

The enhancement-type MOSFET is the most widely used field-effect transistor. In this section, we shall study its structure and physical operation. This will lead to the current-voltage characteristics of the device, studied in the next section.

4.1.1 Device Structure

Figure 4.1, shows the physical structure of the *n*-channel enhancement-type MOSFET. The meaning of the names “enhancement” and “*n*-channel” will become apparent shortly. The transistor is fabricated on a *p*-type substrate, which is a single-crystal silicon wafer that provides physical support for the device (and for the entire circuit in the case of an integrated circuit). Two heavily doped *n*-type regions, indicated in the figure as the *n*⁺ **source**¹ and the *n*⁺ **drain** regions, are created in the substrate. A thin layer of silicon dioxide (SiO_2) of thickness t_{ox} (typically 2–50 nm),² which is an excellent electrical insulator, is grown on the surface of the substrate, covering the area between the source and drain regions. Metal is deposited on top of the oxide layer to form the **gate electrode** of the device. Metal contacts are also made to the source region, the drain region, and the substrate, also known as the

¹ The notation *n*⁺ indicates heavily doped *n*-type silicon. Conversely, *n*[−] is used to denote lightly doped *n*-type silicon. Similar notation applies for *p*-type silicon.

² A nanometer (nm) is 10^{-9} m or 0.001 μm . A micrometer (μm), or micron, is 10^{-6} m. Sometimes the oxide thickness is expressed in angstroms. An angstrom (\AA) is 10^{-10} m, or 10^{-10} m.

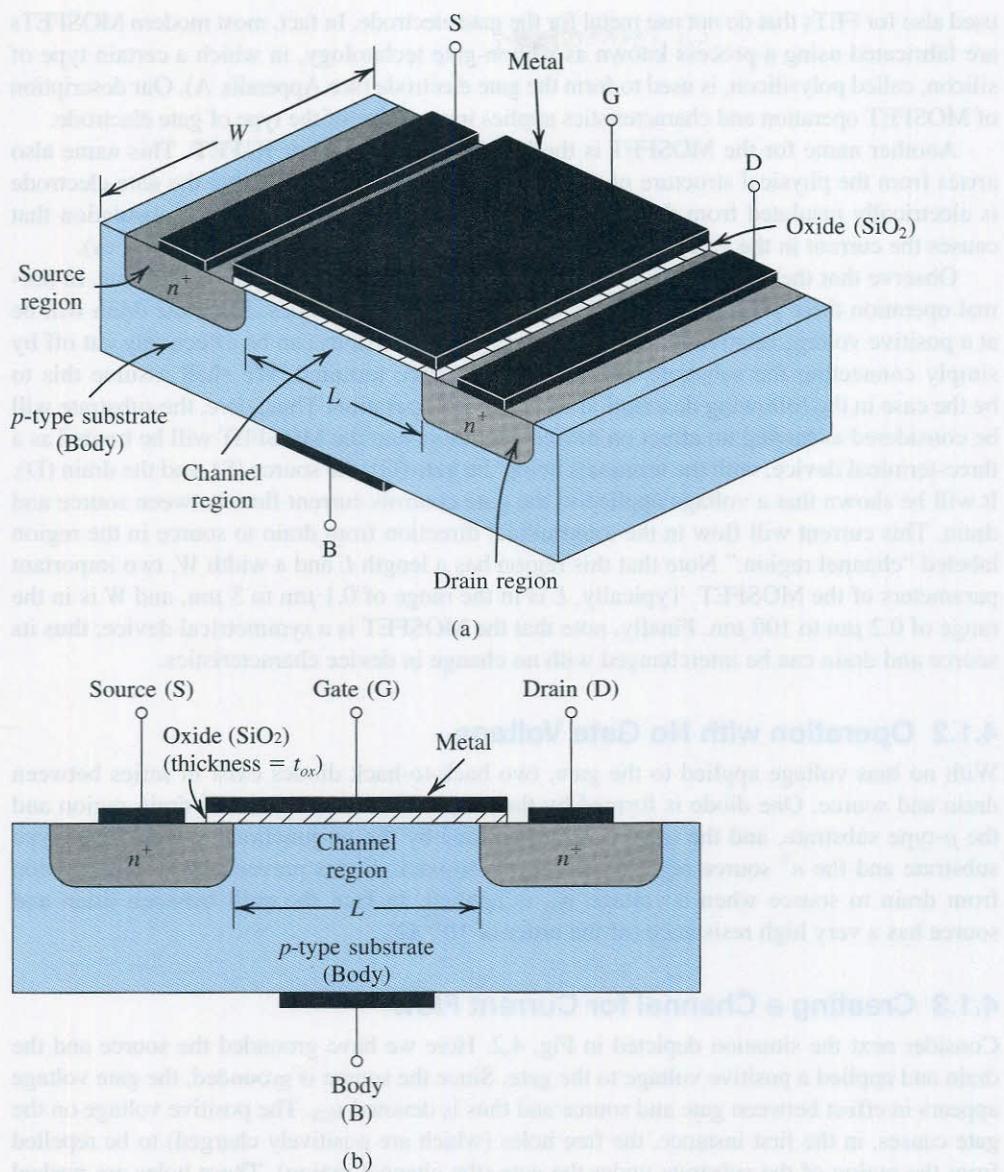


FIGURE 4.1 Physical structure of the enhancement-type NMOS transistor: (a) perspective view; (b) cross-section. Typically $L = 0.1$ to $3 \mu\text{m}$, $W = 0.2$ to $100 \mu\text{m}$, and the thickness of the oxide layer (t_{ox}) is in the range of 2 to 50 nm.

body.³ Thus four terminals are brought out: the gate terminal (G), the source terminal (S), the drain terminal (D), and the substrate or body terminal (B).

At this point it should be clear that the name of the device (metal-oxide-semiconductor FET) is derived from its physical structure. The name, however, has become a general one and is

³ In Fig. 4.1, the contact to the body is shown on the bottom of the device. This will prove helpful later in explaining a phenomenon known as the “body effect.” It is important to note, however, that in actual ICs, contact to the body is made at a location on the top of the device.



used also for FETs that do not use metal for the gate electrode. In fact, most modern MOSFETs are fabricated using a process known as silicon-gate technology, in which a certain type of silicon, called polysilicon, is used to form the gate electrode (see Appendix A). Our description of MOSFET operation and characteristics applies irrespective of the type of gate electrode.

Another name for the MOSFET is the **insulated-gate FET** or **IGFET**. This name also arises from the physical structure of the device, emphasizing the fact that the gate electrode is electrically insulated from the device body (by the oxide layer). It is this insulation that causes the current in the gate terminal to be extremely small (of the order of 10^{-15} A).

Observe that the substrate forms *pn* junctions with the source and drain regions. In normal operation these *pn* junctions are kept reverse-biased at all times. Since the drain will be at a positive voltage relative to the source, the two *pn* junctions can be effectively cut off by simply connecting the substrate terminal to the source terminal. We shall assume this to be the case in the following description of MOSFET operation. Thus, here, the substrate will be considered as having no effect on device operation, and the MOSFET will be treated as a three-terminal device, with the terminals being the gate (G), the source (S), and the drain (D). It will be shown that a voltage applied to the gate controls current flow between source and drain. This current will flow in the longitudinal direction from drain to source in the region labeled “channel region.” Note that this region has a length L and a width W , two important parameters of the MOSFET. Typically, L is in the range of $0.1 \mu\text{m}$ to $3 \mu\text{m}$, and W is in the range of $0.2 \mu\text{m}$ to $100 \mu\text{m}$. Finally, note that the MOSFET is a symmetrical device; thus its source and drain can be interchanged with no change in device characteristics.

4.1.2 Operation with No Gate Voltage

With no bias voltage applied to the gate, two back-to-back diodes exist in series between drain and source. One diode is formed by the *pn* junction between the n^+ drain region and the *p*-type substrate, and the other diode is formed by the *pn* junction between the *p*-type substrate and the n^+ source region. These back-to-back diodes prevent current conduction from drain to source when a voltage v_{DS} is applied. In fact, the path between drain and source has a very high resistance (of the order of $10^{12} \Omega$).

4.1.3 Creating a Channel for Current Flow

Consider next the situation depicted in Fig. 4.2. Here we have grounded the source and the drain and applied a positive voltage to the gate. Since the source is grounded, the gate voltage appears in effect between gate and source and thus is denoted v_{GS} . The positive voltage on the gate causes, in the first instance, the free holes (which are positively charged) to be repelled from the region of the substrate under the gate (the channel region). These holes are pushed downward into the substrate, leaving behind a carrier-depletion region. The depletion region is populated by the bound negative charge associated with the acceptor atoms. These charges are “uncovered” because the neutralizing holes have been pushed downward into the substrate.

As well, the positive gate voltage attracts electrons from the n^+ source and drain regions (where they are in abundance) into the channel region. When a sufficient number of electrons accumulate near the surface of the substrate under the gate, an *n* region is in effect created, connecting the source and drain regions, as indicated in Fig. 4.2. Now if a voltage is applied between drain and source, current flows through this induced *n* region, carried by the mobile electrons. The *induced n* region thus forms a **channel** for current flow from drain to source and is aptly called so. Correspondingly, the MOSFET of Fig. 4.2 is called an ***n*-channel MOSFET** or, alternatively, an **NMOS transistor**. Note that an *n*-channel MOSFET is formed in a *p*-type substrate: The channel is created by *inverting* the substrate surface from *p* type to *n* type. Hence the induced channel is also called an **inversion layer**.

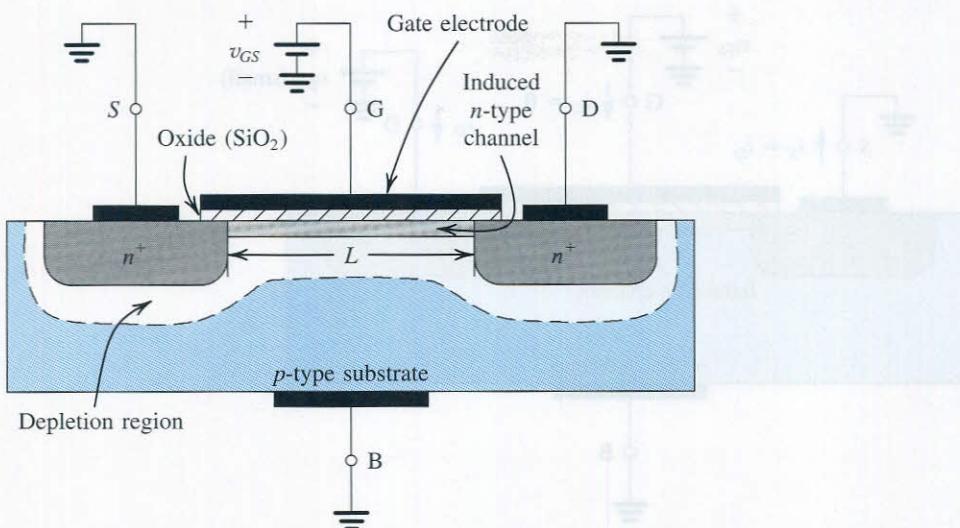


FIGURE 4.2 The enhancement-type NMOS transistor with a positive voltage applied to the gate. An *n* channel is induced at the top of the substrate beneath the gate.

The value of v_{GS} at which a sufficient number of mobile electrons accumulate in the channel region to form a conducting channel is called the **threshold voltage** and is denoted V_t .⁴ Obviously, V_t for an *n*-channel FET is positive. The value of V_t is controlled during device fabrication and typically lies in the range of 0.5 V to 1.0 V.

The gate and the channel region of the MOSFET form a parallel-plate capacitor, with the oxide layer acting as the capacitor dielectric. The positive gate voltage causes positive charge to accumulate on the top plate of the capacitor (the gate electrode). The corresponding negative charge on the bottom plate is formed by the electrons in the induced channel. An electric field thus develops in the vertical direction. It is this field that controls the amount of charge in the channel, and thus it determines the channel conductivity and, in turn, the current that will flow through the channel when a voltage v_{DS} is applied.

4.1.4 Applying a Small v_{DS}

Having induced a channel, we now apply a positive voltage v_{DS} between drain and source, as shown in Fig. 4.3. We first consider the case where v_{DS} is small (i.e., 50 mV or so). The voltage v_{DS} causes a current i_D to flow through the induced *n* channel. Current is carried by free electrons traveling from source to drain (hence the names source and drain). By convention, the direction of current flow is opposite to that of the flow of negative charge. Thus the current in the channel, i_D , will be from drain to source, as indicated in Fig. 4.3. The magnitude of i_D depends on the density of electrons in the channel, which in turn depends on the magnitude of v_{GS} . Specifically, for $v_{GS} = V_t$ the channel is just induced and the current conducted is still negligibly small. As v_{GS} exceeds V_t , more electrons are attracted into the channel. We may visualize the increase in charge carriers in the channel as an increase in the channel depth. The result is a channel of increased conductance or, equivalently, reduced resistance. In fact, the conductance of the channel is proportional to the **excess gate voltage** ($v_{GS} - V_t$), also

⁴ Some texts use V_T to denote the threshold voltage. We use V_t to avoid confusion with the thermal voltage V_T .

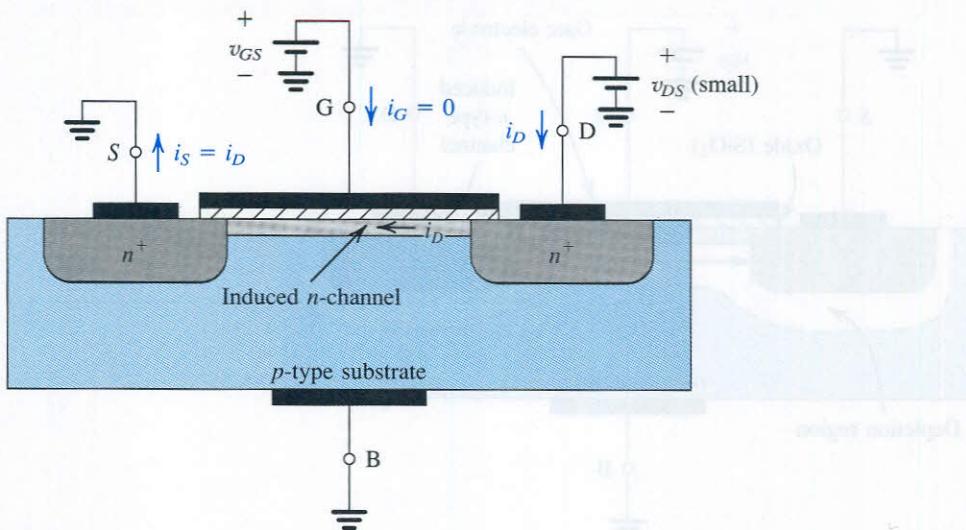


FIGURE 4.3 An NMOS transistor with $v_{GS} > V_t$ and with a small v_{DS} applied. The device acts as a resistance whose value is determined by v_{GS} . Specifically, the channel conductance is proportional to $v_{GS} - V_t$, and thus i_D is proportional to $(v_{GS} - V_t)v_{DS}$. Note that the depletion region is not shown (for simplicity).

known as the **effective voltage** or the **overdrive voltage**. It follows that the current i_D will be proportional to $v_{GS} - V_t$ and, of course, to the voltage v_{DS} that causes i_D to flow.

Figure 4.4 shows a sketch of i_D versus v_{DS} for various values of v_{GS} . We observe that the MOSFET is operating as a linear resistance whose value is controlled by v_{GS} . The resistance is infinite for $v_{GS} \leq V_t$, and its value decreases as v_{GS} exceeds V_t .

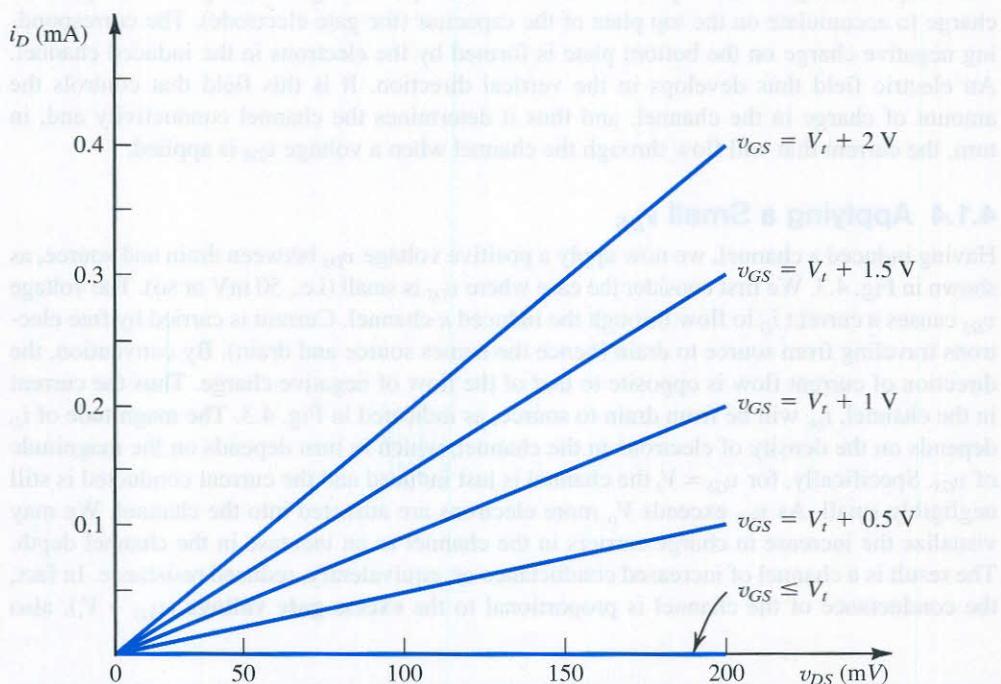


FIGURE 4.4 The i_D-v_{DS} characteristics of the MOSFET in Fig. 4.3 when the voltage applied between drain and source, v_{DS} , is kept small. The device operates as a linear resistor whose value is controlled by v_{GS} .

The description above indicates that for the MOSFET to conduct, a channel has to be induced. Then, increasing v_{GS} above the threshold voltage V_t enhances the channel, hence the names **enhancement-mode operation** and **enhancement-type MOSFET**. Finally, we note that the current that leaves the source terminal (i_S) is equal to the current that enters the drain terminal (i_D), and the gate current $i_G = 0$.

EXERCISE

- 4.1** From the description above of the operation of the MOSFET for small v_{DS} , we note that i_D is proportional to $(v_{GS} - V_t)v_{DS}$. Find the constant of proportionality for the particular device whose characteristics are depicted in Fig. 4.4. Also, give the range of drain-to-source resistances corresponding to an overdrive voltage, $v_{GS} - V_t$, of 0.5 V to 2 V.

Ans. 1 mA/V^2 ; $2 \text{ k}\Omega$ to $0.5 \text{ k}\Omega$

4.1.5 Operation as v_{DS} Is Increased

We next consider the situation as v_{DS} is increased. For this purpose let v_{GS} be held constant at a value greater than V_t . Refer to Fig. 4.5, and note that v_{DS} appears as a voltage drop across the length of the channel. That is, as we travel along the channel from source to drain, the voltage (measured relative to the source) increases from 0 to v_{DS} . Thus the voltage between the gate and points along the channel decreases from v_{GS} at the source end to $v_{GS} - v_{DS}$ at the drain end. Since the channel depth depends on this voltage, we find that the channel is no longer of uniform depth; rather, the channel will take the tapered form shown in Fig. 4.5, being deepest at the source end and shallowest at the drain end. As v_{DS} is increased, the channel becomes more tapered and its resistance increases correspondingly. Thus the i_D-v_{DS} curve does not continue as a straight line but bends as shown in Fig. 4.6. Eventually, when v_{DS} is increased to the value that reduces the voltage between gate and

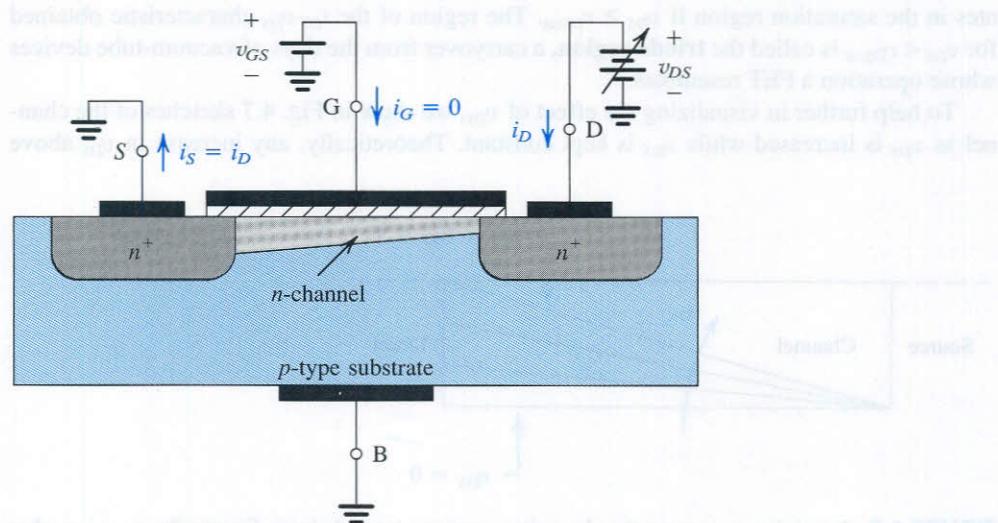


FIGURE 4.5 Operation of the enhancement NMOS transistor as v_{DS} is increased. The induced channel acquires a tapered shape, and its resistance increases as v_{DS} is increased. Here, v_{GS} is kept constant at a value $> V_t$.

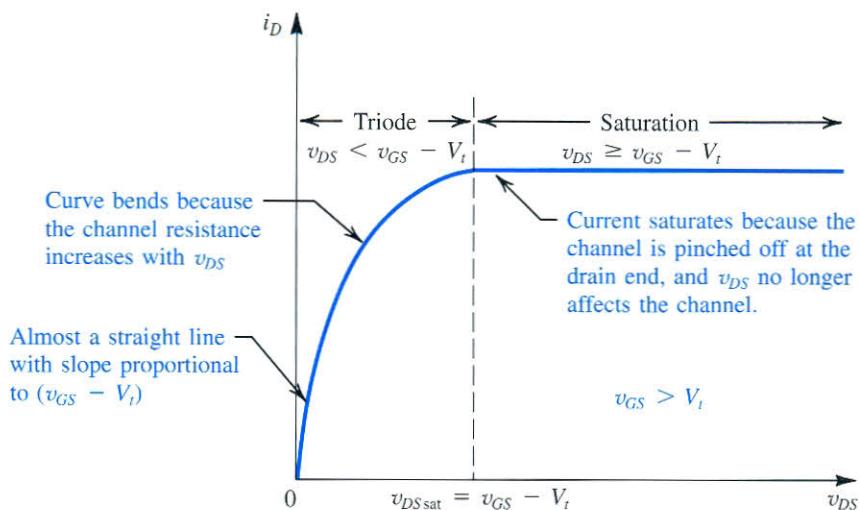


FIGURE 4.6 The drain current i_D versus the drain-to-source voltage v_{DS} for an enhancement-type NMOS transistor operated with $v_{GS} > V_t$.

channel at the drain end to V_t —that is, $v_{GD} = V_t$ or $v_{GS} - v_{DS} = V_t$ or $v_{DS} = v_{GS} - V_t$ —the channel depth at the drain end decreases to almost zero, and the channel is said to be **pinched off**. Increasing v_{DS} beyond this value has little effect (theoretically, no effect) on the channel shape, and the current through the channel remains constant at the value reached for $v_{DS} = v_{GS} - V_t$. The drain current thus **saturates** at this value, and the MOSFET is said to have entered the **saturation region** of operation. The voltage v_{DS} at which saturation occurs is denoted v_{DSsat} ,

$$v_{DSsat} = v_{GS} - V_t \quad (4.1)$$

Obviously, for every value of $v_{GS} \geq V_t$, there is a corresponding value of v_{DSsat} . The device operates in the saturation region if $v_{DS} \geq v_{DSsat}$. The region of the i_D-v_{DS} characteristic obtained for $v_{DS} < v_{DSsat}$ is called the **triode region**, a carryover from the days of vacuum-tube devices whose operation a FET resembles.

To help further in visualizing the effect of v_{DS} , we show in Fig. 4.7 sketches of the channel as v_{DS} is increased while v_{GS} is kept constant. Theoretically, any increase in v_{DS} above

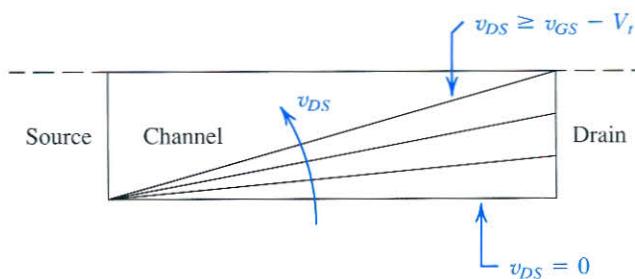


FIGURE 4.7 Increasing v_{DS} causes the channel to acquire a tapered shape. Eventually, as v_{DS} reaches $v_{GS} - V_t$, the channel is pinched off at the drain end. Increasing v_{DS} above $v_{GS} - V_t$ has little effect (theoretically, no effect) on the channel's shape.

v_{DSsat} (which is equal to $v_{GS} - V_t$) has no effect on the channel shape and simply appears across the depletion region surrounding the channel and the n^+ drain region.

4.1.6 Derivation of the i_D-v_{DS} Relationship

The description of physical operation presented above can be used to derive an expression for the i_D-v_{DS} relationship depicted in Fig. 4.6. Toward that end, assume that a voltage v_{GS} is applied between gate and source with $v_{GS} > V_t$ to induce a channel. Also, assume that a voltage v_{DS} is applied between drain and source. First, we shall consider operation in the triode region, for which the channel must be continuous and thus v_{GD} must be greater than V_t , or, equivalently, $v_{DS} < v_{GS} - V_t$. In this case the channel will have the tapered shape shown in Fig. 4.8.

The reader will recall that in the MOSFET, the gate and the channel region form a parallel-plate capacitor for which the oxide layer serves as a dielectric. If the capacitance per unit gate area is denoted C_{ox} and the thickness of the oxide layer is t_{ox} , then

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \quad (4.2)$$

where ϵ_{ox} is the permittivity of the silicon oxide,

$$\epsilon_{ox} = 3.9\epsilon_0 = 3.9 \times 8.854 \times 10^{-12} = 3.45 \times 10^{-11} \text{ F/m}$$

The oxide thickness t_{ox} is determined by the process technology used to fabricate the MOSFET. As an example, for $t_{ox} = 10 \text{ nm}$, $C_{ox} = 3.45 \times 10^{-3} \text{ F/m}^2$, or $3.45 \text{ fF}/\mu\text{m}^2$ as it is usually expressed.

Now refer to Fig. 4.8 and consider the infinitesimal strip of the gate at distance x from the source. The capacitance of this strip is $C_{ox}Wdx$. To find the charge stored on this infinitesimal strip of the gate capacitance, we multiply the capacitance by the *effective voltage*

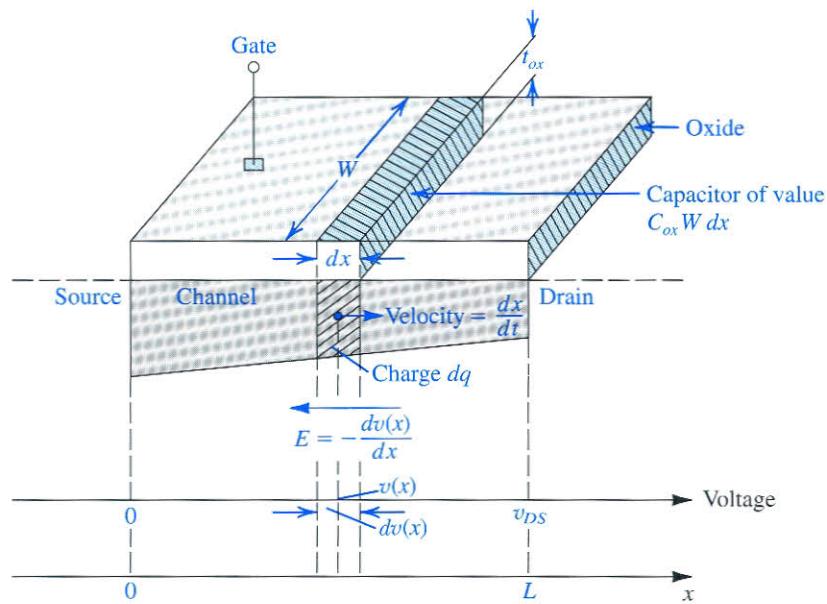


FIGURE 4.8 Derivation of the i_D-v_{DS} characteristic of the NMOS transistor.



between the gate and the channel at point x , where the effective voltage is the voltage that is responsible for inducing the channel at point x and is thus $[v_{GS} - v(x) - V_t]$ where $v(x)$ is the voltage in the channel at point x . It follows that the electron charge dq in the infinitesimal portion of the channel at point x is

$$dq = -C_{ox}(W dx)[v_{GS} - v(x) - V_t] \quad (4.3)$$

where the leading negative sign accounts for the fact that dq is a negative charge.

The voltage v_{DS} produces an electric field along the channel in the negative x direction. At point x this field can be expressed as

$$E(x) = -\frac{dv(x)}{dx}$$

The electric field $E(x)$ causes the electron charge dq to drift toward the drain with a velocity dx/dt ,

$$\frac{dx}{dt} = -\mu_n E(x) = \mu_n \frac{dv(x)}{dx} \quad (4.4)$$

where μ_n is the mobility of electrons in the channel (called surface mobility). It is a physical parameter whose value depends on the fabrication process technology. The resulting drift current i can be obtained as follows:

$$\begin{aligned} i &= \frac{dq}{dt} \\ &= \frac{dq}{dx} \frac{dx}{dt} \end{aligned}$$

Substituting for the charge-per-unit-length dq/dx from Eq. (4.3), and for the electron drift velocity dx/dt from Eq. (4.4), results in

$$i = -\mu_n C_{ox} W [v_{GS} - v(x) - V_t] \frac{dv(x)}{dx}$$

Although evaluated at a particular point in the channel, the current i must be constant at all points along the channel. Thus i must be equal to the source-to-drain current. Since we are interested in the drain-to-source current i_D , we can find it as

$$i_D = -i = \mu_n C_{ox} W [v_{GS} - v(x) - V_t] \frac{dv(x)}{dx}$$

which can be rearranged in the form

$$i_D dx = \mu_n C_{ox} W [v_{GS} - V_t - v(x)] dv(x)$$

Integrating both sides of this equation from $x = 0$ to $x = L$ and, correspondingly, for $v(0) = 0$ to $v(L) = v_{DS}$,

$$\int_0^L i_D dx = \int_0^{v_{DS}} \mu_n C_{ox} W [v_{GS} - V_t - v(x)] dv(x)$$

gives

$$i_D = (\mu_n C_{ox}) \left(\frac{W}{L} \right) \left[(v_{GS} - V_t) v_{DS} - \frac{1}{2} v_{DS}^2 \right] \quad (4.5)$$

This is the expression for the i_D - v_{DS} characteristic in the triode region. The value of the current at the edge of the triode region or, equivalently, at the beginning of the saturation region can be obtained by substituting $v_{DS} = v_{GS} - V_t$, resulting in

$$i_D = \frac{1}{2}(\mu_n C_{ox})\left(\frac{W}{L}\right)(v_{GS} - V_t)^2 \quad (4.6)$$

This is the expression for the i_D - v_{DS} characteristic in the saturation region; it simply gives the saturation value of i_D corresponding to the given v_{GS} . (Recall that in saturation i_D remains constant for a given v_{GS} as v_{DS} is varied.)

In the expressions in Eqs. (4.5) and (4.6), $\mu_n C_{ox}$ is a constant determined by the process technology used to fabricate the n -channel MOSFET. It is known as the **process transconductance parameter**, for as we shall see shortly, it determines the value of the MOSFET transconductance, is denoted k'_n , and has the dimensions of A/V²:

$$k'_n = \mu_n C_{ox} \quad (4.7)$$

Of course, the i_D - v_{DS} expressions in Eqs. (4.5) and (4.6) can be written in terms of k'_n as follows:

$$i_D = k'_n \frac{W}{L} \left[(v_{GS} - V_t)v_{DS} - \frac{1}{2}v_{DS}^2 \right] \quad (\text{Triode region}) \quad (4.5a)$$

$$i_D = \frac{1}{2} k'_n \frac{W}{L} (v_{GS} - V_t)^2 \quad (\text{Saturation region}) \quad (4.6a)$$

In this book we will use the forms with $(\mu_n C_{ox})$ and with k'_n interchangeably.

From Eqs. (4.5) and (4.6) we see that the drain current is proportional to the ratio of the channel width W to the channel length L , known as the **aspect ratio** of the MOSFET. The values of W and L can be selected by the circuit designer to obtain the desired i - v characteristics. For a given fabrication process, however, there is a minimum channel length, L_{\min} . In fact, the minimum channel length that is possible with a given fabrication process is used to characterize the process and is being continually reduced as technology advances. For instance, at the time of this writing (2003) the state-of-the-art in MOS technology is a 0.13- μm process, meaning that for this process the minimum channel length possible is 0.13 μm . There also is a minimum value for the channel width W . For instance, for the 0.13- μm process just mentioned, W_{\min} is 0.16 μm . Finally, we should note that the oxide thickness t_{ox} scales down with L_{\min} . Thus, for a 1.5- μm technology, t_{ox} is 25 nm, but the modern 0.13- μm technology mentioned above has $t_{ox} = 2$ nm.

EXAMPLE 4.1

Consider a process technology for which $L_{\min} = 0.4 \mu\text{m}$, $t_{ox} = 8 \text{ nm}$, $\mu_n = 450 \text{ cm}^2/\text{V}\cdot\text{s}$, and $V_t = 0.7 \text{ V}$.

- (a) Find C_{ox} and k'_n .
- (b) For a MOSFET with $W/L = 8 \mu\text{m}/0.8 \mu\text{m}$, calculate the values of V_{GS} and $V_{DS\min}$ needed to operate the transistor in the saturation region with a dc current $I_D = 100 \mu\text{A}$.
- (c) For the device in (b), find the value of V_{GS} required to cause the device to operate as a 1000- Ω resistor for very small v_{DS} .



Solution

(a)

$$\begin{aligned} C_{ox} &= \frac{\epsilon_{ox}}{t_{ox}} = \frac{3.45 \times 10^{-11}}{8 \times 10^{-9}} = 4.32 \times 10^{-3} \text{ F/m}^2 \\ &= 4.32 \text{ fF}/\mu\text{m}^2 \\ k'_n &= \mu_n C_{ox} = 450 (\text{cm}^2/\text{V}\cdot\text{s}) \times 4.32 (\text{fF}/\mu\text{m}^2) \\ &= 450 \times 10^8 (\mu\text{m}^2/\text{V}\cdot\text{s}) \times 4.32 \times 10^{-15} (\text{F}/\mu\text{m}^2) \\ &= 194 \times 10^{-6} (\text{F}/\text{V}\cdot\text{s}) \\ &= 194 \mu\text{A}/\text{V}^2 \end{aligned}$$

(b) For operation in the saturation region,

$$i_D = \frac{1}{2} k'_n \frac{W}{L} (v_{GS} - V_t)^2$$

Thus,

$$100 = \frac{1}{2} \times 194 \times \frac{8}{0.8} (V_{GS} - 0.7)^2$$

which results in

$$V_{GS} - 0.7 = 0.32 \text{ V}$$

or

$$V_{GS} = 1.02 \text{ V}$$

and

$$V_{DS\min} = V_{GS} - V_t = 0.32 \text{ V}$$

(c) For the MOSFET in the triode region with v_{DS} very small,

$$i_D \equiv k'_n \frac{W}{L} (v_{GS} - V_t) v_{DS}$$

from which the drain-to-source resistance r_{DS} can be found as

$$\begin{aligned} r_{DS} &\equiv \left. \frac{v_{DS}}{i_D} \right|_{\text{small } v_{DS}} \\ &= 1 / \left[k'_n \frac{W}{L} (V_{GS} - V_t) \right] \end{aligned}$$

Thus

$$1000 = \frac{1}{194 \times 10^{-6} \times 10 (V_{GS} - 0.7)}$$

which yields

$$V_{GS} - 0.7 = 0.52 \text{ V}$$

Thus,

$$V_{GS} = 1.22 \text{ V}$$



EXERCISES

- 4.2** For a 0.8- μm process technology for which $t_{ox} = 15 \text{ nm}$ and $\mu_n = 550 \text{ cm}^2/\text{V}\cdot\text{s}$, find C_{ox} , k'_n , and the overdrive voltage $V_{OV} \equiv V_{GS} - V_t$ required to operate a transistor having $W/L = 20$ in saturation with $I_D = 0.2 \text{ mA}$. What is the minimum value of V_{DS} needed?

Ans. $2.3 \text{ fF}/\mu\text{m}^2; 127 \mu\text{A}/\text{V}^2; 0.40 \text{ V}; 0.40 \text{ V}$

- 4.3** Use the expression for operation in the triode region to show that an *n*-channel MOSFET operated with an overdrive voltage $V_{OV} \equiv V_{GS} - V_t$ and having a small V_{DS} across it behaves approximately as a linear resistance r_{DS} ,

$$r_{DS} = 1/\left[k'_n \frac{W}{L} V_{OV}\right]$$

Calculate the value of r_{DS} obtained for a device having $k'_n = 100 \mu\text{A}/\text{V}^2$ and $W/L = 10$ when operated with an overdrive voltage of 0.5 V.

Ans. $2 \text{ k}\Omega$

4.1.7 The *p*-Channel MOSFET

A *p*-channel enhancement-type MOSFET (PMOS transistor), fabricated on an *n*-type substrate with p^+ regions for the drain and source, has holes as charge carriers. The device operates in the same manner as the *n*-channel device except that v_{GS} and v_{DS} are negative and the threshold voltage V_t is negative. Also, the current i_D enters the source terminal and leaves through the drain terminal.

PMOS technology originally dominated MOS manufacturing. However, because NMOS devices can be made smaller and thus operate faster, and because NMOS historically required lower supply voltages than PMOS, NMOS technology has virtually replaced PMOS. Nevertheless, it is important to be familiar with the PMOS transistor for two reasons: PMOS devices are still available for discrete-circuit design, and more importantly, both PMOS and NMOS transistors are utilized in **complementary MOS** or **CMOS** circuits, which is currently the dominant MOS technology.

4.1.8 Complementary MOS or CMOS

As the name implies, complementary MOS technology employs MOS transistors of both polarities. Although CMOS circuits are somewhat more difficult to fabricate than NMOS, the availability of complementary devices makes possible many powerful circuit-design possibilities. Indeed, at the present time CMOS is the most widely used of all the IC technologies. This statement applies to both analog and digital circuits. CMOS technology has virtually replaced designs based on NMOS transistors alone. Furthermore, at the time of this writing (2003), CMOS technology has taken over many applications that just a few years ago were possible only with bipolar devices. Throughout this book, we will study many CMOS circuit techniques.

Figure 4.9 shows a cross-section of a CMOS chip illustrating how the PMOS and NMOS transistors are fabricated. Observe that while the NMOS transistor is implemented directly in the *p*-type substrate, the PMOS transistor is fabricated in a specially created *n* region, known as an ***n* well**. The two devices are isolated from each other by a thick region of oxide that functions as an insulator. Not shown on the diagram are the connections made to the *p*-type body and to the *n* well. The latter connection serves as the body terminal for the PMOS transistor.

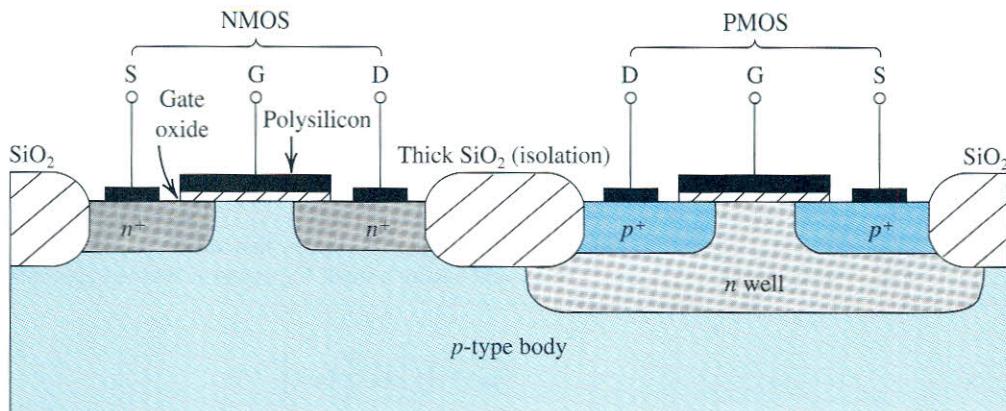


FIGURE 4.9 Cross-section of a CMOS integrated circuit. Note that the PMOS transistor is formed in a separate n -type region, known as an n well. Another arrangement is also possible in which an n -type body is used and the n device is formed in a p well. Not shown are the connections made to the p -type body and to the n well; the latter functions as the body terminal for the p -channel device.

4.1.9 Operating the MOS Transistor in the Subthreshold Region

The above description of the n -channel MOSFET operation implies that for $v_{GS} < V_t$, no current flows and the device is cut off. This is not entirely true, for it has been found that for values of v_{GS} smaller than but close to V_t , a small drain current flows. In this **subthreshold region** of operation the drain current is exponentially related to v_{GS} , much like the i_C-v_{BE} relationship of a BJT, as will be shown in the next chapter.

Although in most applications the MOS transistor is operated with $v_{GS} > V_t$, there are special, but a growing number of, applications that make use of subthreshold operation. In this book, we will not consider subthreshold operation any further and refer the reader to the references listed in Appendix F.

4.2 CURRENT-VOLTAGE CHARACTERISTICS

Building on the physical foundation established in the previous section for the operation of the enhancement MOS transistor, we present in this section its complete current-voltage characteristics. These characteristics can be measured at dc or at low frequencies and thus are called static characteristics. The dynamic effects that limit the operation of the MOSFET at high frequencies and high switching speeds will be discussed in Section 4.8.

4.2.1 Circuit Symbol

Figure 4.10(a) shows the circuit symbol for the n -channel enhancement-type MOSFET. Observe that the spacing between the two vertical lines that represent the gate and the channel indicates the fact that the gate electrode is insulated from the body of the device. The polarity of the p -type substrate (body) and the n channel is indicated by the arrowhead on the line representing the body (B). This arrowhead also indicates the polarity of the transistor, namely, that it is an n -channel device.

Although the MOSFET is a symmetrical device, it is often useful in circuit design to designate one terminal as the source and the other as the drain (without having to write S and D beside the terminals). This objective is achieved in the modified circuit symbol shown in Fig. 4.10(b). Here an arrowhead is placed on the source terminal, thus distinguishing it from

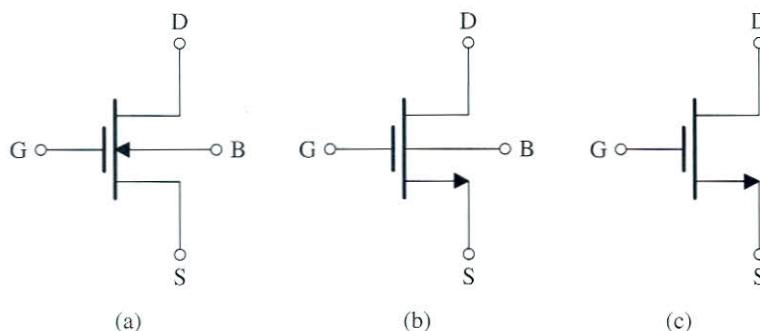


FIGURE 4.10 (a) Circuit symbol for the n -channel enhancement-type MOSFET. (b) Modified circuit symbol with an arrowhead on the source terminal to distinguish it from the drain and to indicate device polarity (i.e., n channel). (c) Simplified circuit symbol to be used when the source is connected to the body or when the effect of the body on device operation is unimportant.

the drain terminal. The arrowhead points in the normal direction of current flow and thus indicates the polarity of the device (i.e., n channel). Observe that in the modified symbol, there is no need to show the arrowhead on the body line. Although the circuit symbol of Fig. 4.10(b) clearly distinguishes the source from the drain, in practice it is the polarity of the voltage impressed across the device that determines source and drain; *the drain is always positive relative to the source in an n-channel FET.*

In applications where the source is connected to the body of the device, a further simplification of the circuit symbol is possible, as indicated in Fig. 4.10(c). This symbol is also used in applications when the effect of the body on circuit operation is not important, as will be seen later.

4.2.2 The i_D - v_{DS} Characteristics

Figure 4.11(a) shows an n -channel enhancement-type MOSFET with voltages v_{GS} and v_{DS} applied and with the normal directions of current flow indicated. This conceptual circuit can be used to measure the i_D-v_{DS} characteristics, which are a family of curves, each measured at a constant v_{GS} . From the study of physical operation in the previous section, we expect each of the i_D-v_{DS} curves to have the shape shown in Fig. 4.6. This indeed is the case, as is evident from Fig. 4.11(b), which shows a typical set of i_D-v_{DS} characteristics. A thorough understanding of the MOSFET terminal characteristics is essential for the reader who intends to design MOS circuits.

The characteristic curves in Fig. 4.11(b) indicate that there are three distinct regions of operation: the **cutoff region**, the **triode region**, and the **saturation region**. The saturation region is used if the FET is to operate as an amplifier. For operation as a switch, the cutoff and triode regions are utilized. The device is cut off when $v_{GS} < V_t$. To operate the MOSFET in the triode region we must first induce a channel,

$$v_{GS} \geq V_r \quad (\text{Induced channel}) \quad (4.8)$$

and then keep v_{DS} small enough so that the channel remains continuous. This is achieved by ensuring that the gate-to-drain voltage is

$$v_{GD} \geq V, \quad (\text{Continuous channel}) \quad (4.9)$$

This condition can be stated explicitly in terms of v_{DS} by writing $v_{GD} = v_{GS} + v_{SD} = v_{GS} - v_{DS}$; thus,

$$v_{GS} - v_{DS} > V_t$$

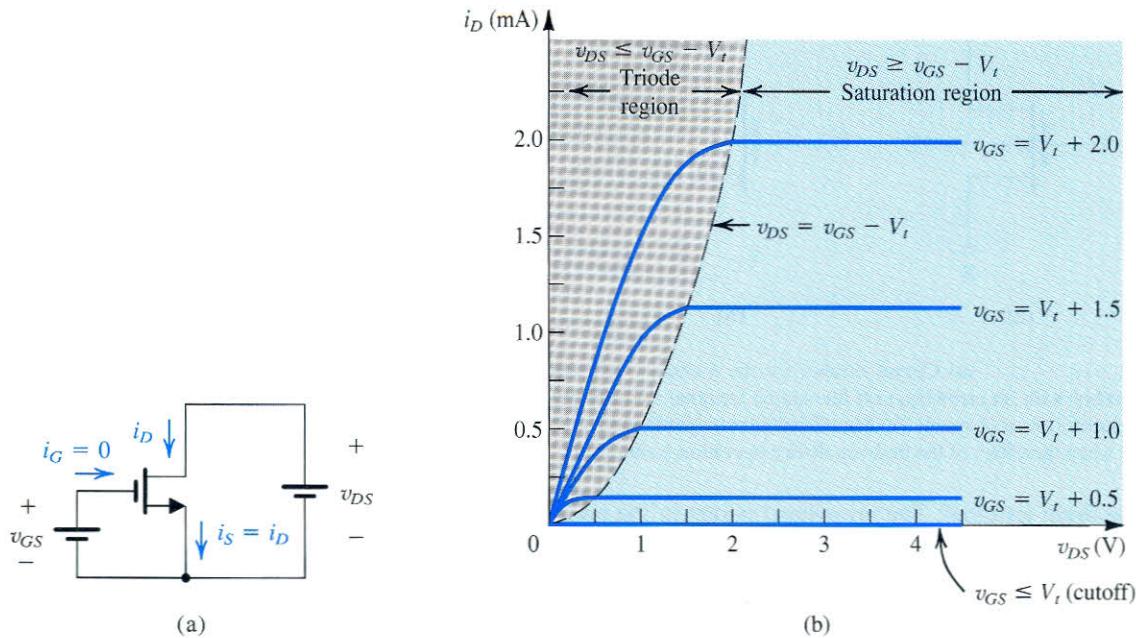


FIGURE 4.11 (a) An *n*-channel enhancement-type MOSFET with v_{GS} and v_{DS} applied and with the normal directions of current flow indicated. (b) The i_D - v_{DS} characteristics for a device with $k'_n (W/L) = 1.0 \text{ mA/V}^2$.

which can be rearranged to yield

$$v_{DS} < v_{GS} - V_t \quad (\text{Continuous channel}) \quad (4.10)$$

Either Eq. (4.9) or Eq. (4.10) can be used to ascertain triode-region operation. In words, the *n*-channel enhancement-type MOSFET operates in the triode region when v_{GS} is greater than V_t and the drain voltage is lower than the gate voltage by at least V_t volts.

In the triode region, the i_D - v_{DS} characteristics can be described by the relationship of Eq. (4.5), which we repeat here,

$$i_D = k'_n \frac{W}{L} \left[(v_{GS} - V_t)v_{DS} - \frac{1}{2}v_{DS}^2 \right] \quad (4.11)$$

where $k'_n = \mu_n C_{ox}$ is the process transconductance parameter; its value is determined by the fabrication technology. If v_{DS} is sufficiently small so that we can neglect the v_{DS}^2 term in Eq. (4.11), we obtain for the i_D - v_{DS} characteristics near the origin the relationship

$$i_D \approx k'_n \frac{W}{L} (v_{GS} - V_t) v_{DS} \quad (4.12)$$

This linear relationship represents the operation of the MOS transistor as a linear resistance r_{DS} whose value is controlled by v_{GS} . Specifically, for v_{GS} set to a value V_{GS} , r_{DS} is given by

$$r_{DS} \equiv \frac{v_{DS}}{i_D} \Big|_{\substack{v_{DS} \text{ small} \\ v_{GS} = V_{GS}}} = \left[k'_n \frac{W}{L} (V_{GS} - V_t) \right]^{-1} \quad (4.13)$$

We discussed this region of operation in the previous section (refer to Fig. 4.4). It is also useful to express r_{DS} in terms of the **gate-to-source overdrive voltage**,

$$V_{OV} \equiv V_{GS} - V_t \quad (4.14)$$

as

$$r_{DS} = 1 / \left[k'_n \left(\frac{W}{L} \right) V_{OV} \right] \quad (4.15)$$

Finally, we urge the reader to show that the approximation involved in writing Eq. (4.12) is based on the assumption that $v_{DS} \ll 2V_{OV}$.

To operate the MOSFET in the saturation region, a channel must be induced,

$$v_{GS} \geq V_t \quad (\text{Induced channel}) \quad (4.16)$$

and pinched off at the drain end by raising v_{DS} to a value that results in the gate-to-drain voltage falling below V_t ,

$$v_{GD} \leq V_t \quad (\text{Pinched-off channel}) \quad (4.17)$$

This condition can be expressed explicitly in terms of v_{DS} as

$$v_{DS} \geq v_{GS} - V_t \quad (\text{Pinched-off channel}) \quad (4.18)$$

In words, the *n*-channel enhancement-type MOSFET operates in the saturation region when v_{GS} is greater than V_t and the drain voltage does not fall below the gate voltage by more than V_t volts.

The boundary between the triode region and the saturation region is characterized by

$$v_{DS} = v_{GS} - V_t \quad (\text{Boundary}) \quad (4.19)$$

Substituting this value of v_{DS} into Eq. (4.11) gives the saturation value of the current i_D as

$$i_D = \frac{1}{2} k'_n \frac{W}{L} (v_{GS} - V_t)^2 \quad (4.20)$$

Thus in saturation the MOSFET provides a drain current whose value is independent of the drain voltage v_{DS} and is determined by the gate voltage v_{GS} according to the square-law relationship in Eq. (4.20), a sketch of which is shown in Fig. 4.12. Since the drain current is independent of the drain voltage, the saturated MOSFET behaves as an ideal current source whose value is controlled by v_{GS} according to the nonlinear relationship in Eq. (4.20). Figure 4.13 shows a circuit representation of this view of MOSFET operation in the saturation region. Note that this is a **large-signal equivalent-circuit model**.

Referring back to the i_D-v_{DS} characteristics in Fig. 4.11(b), we note that the boundary between the triode and the saturation regions is shown as a broken-line curve. Since this curve is characterized by $v_{DS} = v_{GS} - V_t$, its equation can be found by substituting for $v_{GS} - V_t$ by v_{DS} in either the triode-region equation (Eq. 4.11) or the saturation-region equation (Eq. 4.20). The result is

$$i_D = \frac{1}{2} k'_n \frac{W}{L} v_{DS}^2 \quad (4.21)$$

It should be noted that the characteristics depicted in Figs. 4.4, 4.11, and 4.12 are for a MOSFET with $k'_n(W/L) = 1.0 \text{ mA/V}^2$ and $V_t = 1 \text{ V}$.

Finally, the chart in Fig. 4.14 shows the relative levels of the terminal voltages of the enhancement-type NMOS transistor for operation in the triode region and in the saturation region.