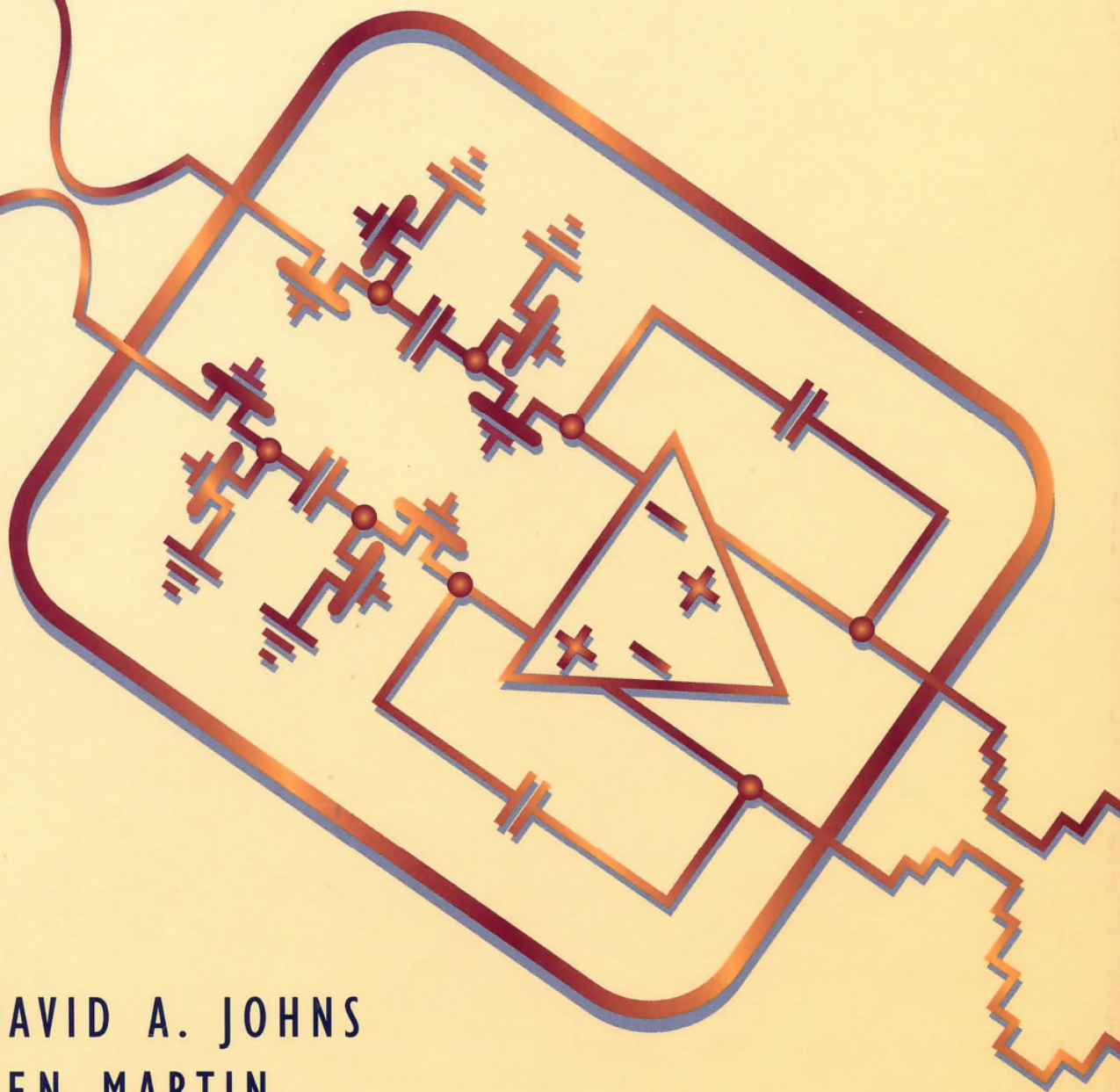


ANALOG INTEGRATED CIRCUIT DESIGN

DAVID A. JOHNS
KEN MARTIN



ANALOG INTEGRATED CIRCUIT DESIGN

David Johns

Ken Martin

University of Toronto



John Wiley & Sons, Inc.

New York • Chichester • Brisbane
Toronto • Singapore • Weinheim

<i>Acquisitions Editor</i>	Charity Robey
<i>Marketing Manager</i>	Jay Kirsch
<i>Production Manager</i>	Lucille Buonocore
<i>Senior Production Editor</i>	Tracey Kuehn
<i>Designer</i>	Kevin Murphy
<i>Manufacturing Manager</i>	Mark Cirillo
<i>Illustration Editor</i>	Sigmund Malinowski

This book was set in Times Roman by Publication Services and printed and bound by R.R. Donnelley/Crawfordsville. The cover was printed by Lehigh Press.

Recognizing the importance of preserving what has been written, it is a policy of John Wiley & Sons, Inc., to have books of enduring value published in the United States printed on acid-free paper, and we exert our best efforts to that end.

The paper in this book was manufactured by a mill whose forest management programs include sustained yield harvesting of its timberlands. Sustained yield harvesting principles ensure that the number of trees cut each year does not exceed the amount of new growth.

Copyright © 1997, by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Sections 107 and 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging-in-Publication Data:

Johns, David, 1958—
 Analog integrated circuit design / David Johns, Ken Martin.
 p. cm.
 Includes bibliographical references.
 ISBN 0-471-14448-7 (cloth : alk. paper)
 I. Linear integrated circuits—Design and construction.
 I. Martin, Kenneth W. (Kenneth W.) 1952— . II. Title.
 TK7874.J65 1996
 621.3815—dc20

96-34365
 CIP

ISBN 0-471-14448-7

Printed in the United States of America

10 9 8 7 6

biased is smaller. This voltage drop is dependent on the metal used; for aluminum it might be around 0.5 V. More importantly, when the diode is forward biased, there is no minority-charge storage in the lightly doped n^- region. Thus, the small-signal model of a forward-biased Schottky diode has $C_d = 0$ (with reference to Fig. 1.4). The absence of this diffusion capacitance makes the diode much faster. It is particularly faster when turning off, because it is not necessary to remove the minority charge first. Rather, it is only necessary to discharge the depletion capacitance through about 0.2 V.

Schottky diodes have been used extensively in bipolar logic circuits. They are also used in a number of high-speed analog circuits, particularly those realized in gallium arsenide (GaAs) technologies, rather than silicon technologies.

1.2 MOS TRANSISTORS

Presently, the most popular technology for realizing microcircuits makes use of MOS transistors. Unlike most bipolar junction transistor (BJT) technologies, which make dominant use of only one type of transistor (npn transistors in the case of BJT processes³), MOS circuits normally use two complementary types of transistors—n-channel and p-channel. While n-channel devices conduct with a positive gate voltage, p-channel devices conduct with a negative gate voltage. Moreover, electrons are used to conduct current in n-channel transistors, while holes are used in p-channel transistors. Microcircuits containing both n-channel and p-channel transistors are called CMOS circuits, for *complementary MOS*. The acronym MOS stands for *metal-oxide semiconductor*, which historically denoted the gate, insulator, and channel region materials, respectively. However, most present CMOS technologies utilize polysilicon gates rather than metal gates.

Before CMOS technology became widely available, most MOS processes made use of only n-channel transistors (NMOS). However, often two different types of n-channel transistors could be realized. One type, enhancement n-channel transistors, is similar to the n-channel transistors realized in CMOS technologies. Enhancement transistors require a positive gate-to-source voltage to conduct current. The other type, depletion transistors, conduct current with a gate-source voltage of 0 V. Depletion transistors were used to create high-impedance loads in NMOS logic gates.

A typical cross section of an n-channel enhancement-type MOS transistor is shown in Fig. 1.6. With no voltage applied to the gate, the n^+ source and drain regions are separated by the p^- substrate. The separation between the drain and the source is called the channel length, L . In present MOS technologies, the minimum channel length is typically between 0.3 μm and 1.0 μm . It should be noted that there is no physical difference between the drain and the source.⁴ *The source terminal of an*

3. Most BJT technologies can also realize low-speed lateral pnp transistors. Normally these would only be used to realize current sources as they have low gains and poor frequency responses. Recently, bipolar technologies utilizing high-speed vertical pnp transistors, as well as high-speed npn transistors, have become available and are growing in popularity. These technologies are called complementary bipolar technologies.

4. Large MOS transistors used for power applications might not be realized with symmetric drain and source junctions.

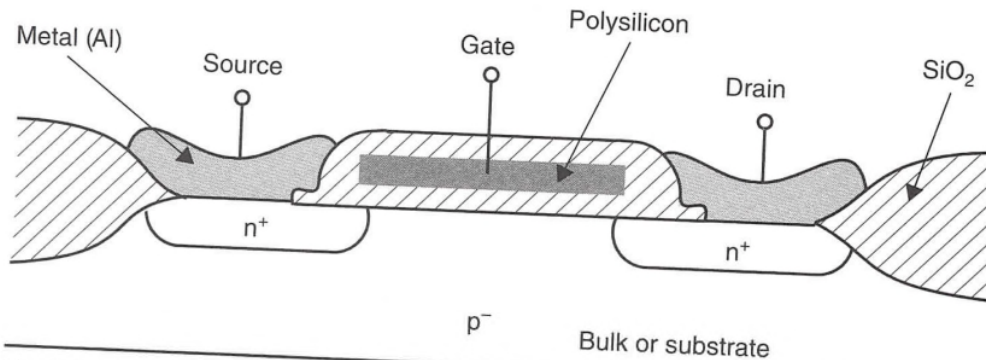


Fig. 1.6 A cross section of a typical n-channel transistor.

n-channel transistor is defined as whichever of the two terminals has a lower voltage. For a p-channel transistor, the source would be the terminal with the higher voltage. When a transistor is turned on, current flows from the drain to the source in an n-channel transistor and from the source to the drain in a p-channel transistor. In both cases, the true carriers travel from the source to drain, but the current directions are different because n-channel carriers (electrons) are negative, whereas p-channel carriers (holes) are positive.

The gate is normally realized using polysilicon, which is heavily doped noncrystalline (or amorphous) silicon. Polysilicon gates are used nowadays (instead of metal) because polysilicon allows the dimensions of the transistor to be realized much more accurately during the patterning of the transistor, which involves what is called a self-aligned process. This higher geometric accuracy results in smaller, faster transistors.

The gate is physically separated from the surface of the silicon by a thin insulator made of silicon dioxide (SiO_2). Thus, the gate is electrically isolated from the channel and affects the channel (and hence, the transistor current) only through electrostatic coupling, similar to capacitive coupling. A typical thickness for the SiO_2 insulator between the gate and the channel is presently from $0.01 \mu\text{m}$ to $0.03 \mu\text{m}$. Since the gate is electrically isolated from the channel, it never conducts dc current. Indeed, the excellent isolation results in leakage currents being almost undetectable. However, because of the inherent capacitances in MOS transistors, transient gate currents do exist when gate voltages are quickly changing.

Normally the p^- substrate (or bulk) is connected to the most negative voltage in a microcircuit. In analog circuits, this might be the negative power supply, but in digital circuits it is normally ground or 0 V. This connection results in all transistors placed in the substrate being surrounded by reverse-biased junctions, which electrically isolate the transistors and thereby prevent conduction through the substrate between transistors (unless, of course, they are connected together through some other means).

Symbols for MOS Transistors

Many symbols have been used to represent MOS transistors. Figure 1.7 shows some of the symbols that have been used to represent n-channel MOS transistors. The symbol in Fig. 1.7(a) is often used; note that there is nothing in the symbol to specify

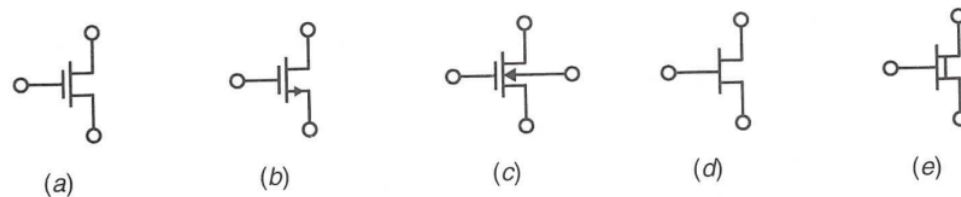


Fig. 1.7 Commonly used symbols for n-channel transistors.

whether the transistor is n-channel or p-channel. A common rule is to assume, when in doubt, that the transistor is an n-channel enhancement transistor. Figure 1.7(b) is the most commonly used symbol for an n-channel enhancement transistor and is used throughout this text. The arrow pointing outward on the source indicates that the transistor is n-channel, similar to the convention used for npn transistors, and indicates the direction of hole current.

MOS transistors are actually four-terminal devices, with the substrate being the fourth terminal. In n-channel devices, the p^- substrate is normally connected to the most negative voltage in the microcircuit, whereas for p-channel devices, the n^- substrate is normally connected to the most positive voltage. In these cases the substrate connection is normally not shown in the symbol. However, for CMOS technologies, at least one of the two types of transistors will be formed in a *well* substrate that need not be connected to one of the power supply nodes. For example, an n-well process would form n-channel transistors in a p^- substrate encompassing the entire microcircuit, while the p-channel transistors would be formed in many n-well substrates. In this case, most of the n-well substrates would be connected to the most positive power supply, while some might be connected to other nodes in the circuit (often the well is connected to the source of a transistor that is not connected to the power supply). In these cases, the symbol shown in Fig. 1.7(c) can be used to show the substrate connection explicitly. It should be noted that this case is not encountered often in digital circuits and is more common in analog circuits. Sometimes, in the interest of simplicity, the isolation of the gate is not explicitly shown, as is the case of the symbol of Fig. 1.7(d). This simple notation is more common for digital circuits in which a large number of transistors are present. Since this symbol is also used for JFET transistors, it will never be used to represent MOS transistors in this text. The last symbol, shown in Fig. 1.7(e), denotes an n-channel depletion transistor. The extra line is used to indicate that a physical channel exists for a 0-V gate-source voltage. Depletion transistors were used in older NMOS technologies but are not typically available in CMOS processes.

Figure 1.8 shows some commonly used symbols for p-channel transistors. In this text, the symbol of Fig. 1.8(a) will be most often used. The symbol in Fig. 1.8(c) is

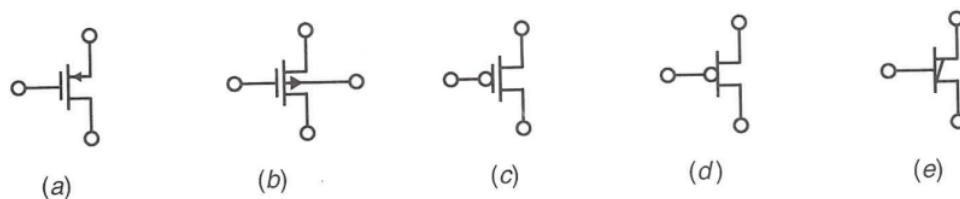


Fig. 1.8 Commonly used symbols for p-channel transistors.

sometimes used in digital circuits, where the circle indicates that a low voltage on the gate turns the transistor on, as opposed to a high voltage for an n -channel transistor (Fig. 1.7(a)). The symbols of Fig. 1.8(d) or Fig. 1.8(e) might be used in larger circuits where many transistors are present, to simplify the drawing somewhat. They will not be used in this text.

Basic Operation

The basic operation of MOS transistors will be described with respect to an n -channel transistor. First, consider the simplified cross sections shown in Fig. 1.9, where the source, drain, and substrate are all connected to ground. In this case, the MOS transistor operates similarly to a capacitor. The gate acts as one plate of the capacitor, and the surface of the silicon, just under the thin insulating SiO_2 , acts as the other plate.

If the gate voltage is very negative, as shown in Fig. 1.9(a), positive charge will be attracted to the channel region. Since the substrate was originally doped p^- , this negative gate voltage has the effect of simply increasing the channel doping to p^+ ,

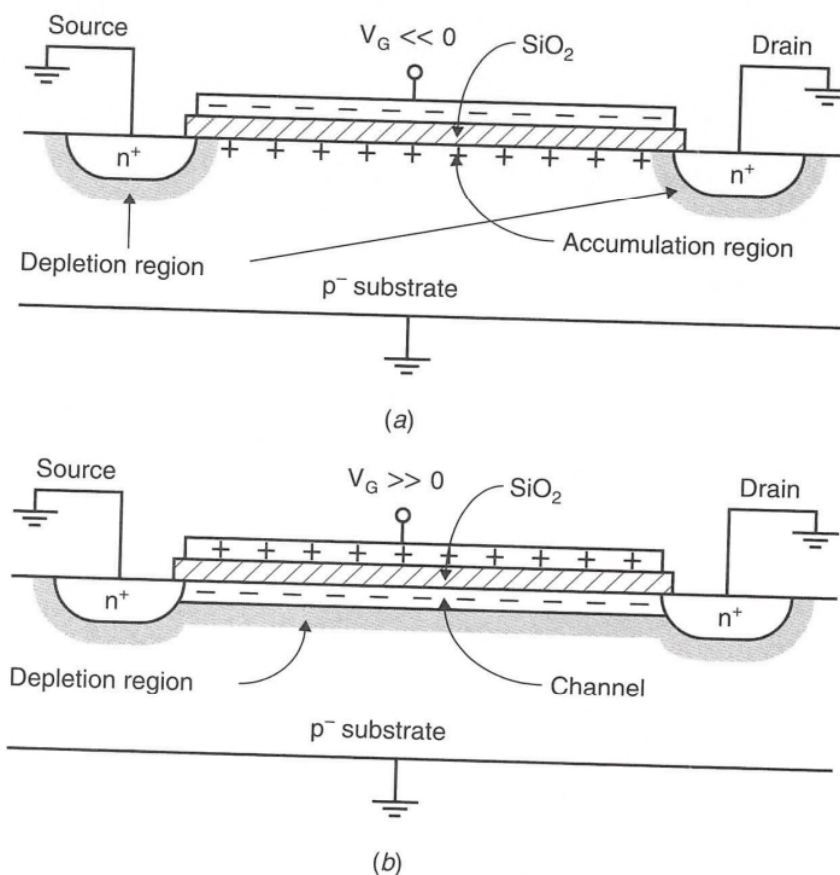


Fig. 1.9 An n -channel MOS transistor. (a) $V_G \ll 0$, resulting in an accumulated channel (no current flow); (b) $V_G \gg 0$, and the channel is present (current flow possible from drain to source).

resulting in what is called an *accumulated channel*. The n^+ source and drain regions are separated from the p^+ -channel region by depletion regions, resulting in the equivalent circuit of two back-to-back diodes. Thus, only leakage current will flow even if one of the source or drain voltages becomes large (unless the drain voltage becomes so large as to cause the transistor to break down).

In the case of a positive voltage being applied to the gate, the opposite situation occurs, as shown in Fig. 1.9(b). For small positive gate voltages, the positive carriers in the channel under the gate are initially repulsed and the channel changes from a p^- doping level to a depletion region. As a more positive gate voltage is applied, the gate attracts negative charge from the source and drain regions, and the channel becomes an n region with mobile electrons connecting the drain and source regions.⁵ In short, a sufficiently large positive gate-source voltage changes the channel beneath the gate to an n region, and the channel is said to be *inverted*.

The gate-source voltage, for which the concentration of electrons under the gate is equal to the concentration of holes in the p^- substrate far from the gate, is commonly referred to as the *transistor threshold voltage* and denoted V_{tn} (for n -channel transistors). For gate-source voltages larger than V_{tn} , there is an n -type channel present, and conduction between the drain and the source can occur. For gate-source voltages less than V_{tn} , it is normally assumed that the transistor is off and no current flows between the drain and the source. However, it should be noted that this assumption of zero drain-source current for a transistor that is off is only an approximation. In fact, for gate voltages around V_{tn} , there is no abrupt current change, and for gate-source voltages slightly less than V_{tn} , small amounts of *subthreshold current* can flow, as discussed in Section 1.3.

When the gate-source voltage, V_{GS} , is larger than V_{tn} , the channel is present. As V_{GS} is increased, the density of electrons in the channel increases. Indeed, the carrier density, and therefore the charge density, is proportional to $V_{GS} - V_{tn}$, which is often called the *effective gate-source voltage* and denoted V_{eff} . Specifically, define

$$V_{eff} \equiv V_{GS} - V_{tn} \quad (1.54)$$

The charge density of electrons is then given by

$$Q_n = C_{ox}(V_{GS} - V_{tn}) = C_{ox}V_{eff} \quad (1.55)$$

Here, C_{ox} is the gate capacitance per unit area and is given by

$$C_{ox} = \frac{K_{ox}\epsilon_0}{t_{ox}} \quad (1.56)$$

where K_{ox} is the relative permittivity of SiO_2 (approximately 3.9) and t_{ox} is the thickness of the thin oxide under the gate. A point to note here is that (1.55) is only accurate when both the drain and the source voltages are zero.

5. The drain and source regions are sometimes called diffusion regions or junctions for historical reasons. This use of the word *junction* is not synonymous with our previous use, in which it designated a pn interface of a diode.

To obtain the total gate capacitance, (1.56) should be multiplied by the effective gate area, WL , where W is the gate width and L is the effective gate length. These dimensions are shown in Fig. 1.10. Thus the total gate capacitance, C_{gs} , is given by

$$C_{gs} = WLC_{ox} \quad (1.57)$$

and the total charge of the channel, Q_{T-n} , is given by

$$Q_{T-n} = WLC_{ox}(V_{GS} - V_{tn}) = WLC_{ox}V_{eff} \quad (1.58)$$

The gate capacitance, C_{gs} , is one of the major load capacitances that circuits must be capable of driving. Gate capacitances are also important when one is calculating *charge injection*, which occurs when a MOS transistor is being turned off because the channel charge, Q_{T-n} , must flow from under the gate out through the terminals to other places in the circuit.

Next, if the drain voltage is increased above 0 V, a drain-source potential difference exists. This difference results in current flowing from the drain to the source.⁶ The relationship between V_{DS} and the drain-source current, I_D , is the same as for a resistor, assuming V_{DS} is small. This relationship is given [Sze, 1981] by

$$I_D = \mu_n Q_n \frac{W}{L} V_{DS} \quad (1.59)$$

where $\mu_n \cong 0.06 \text{ m}^2/\text{Vs}$ is the mobility of electrons near the silicon surface, and Q_n is the charge concentration of the channel per unit area (looking from the top down). Note that as the channel length increases, the drain-source current decreases, whereas this current increases as either the charge density or the transistor width increases. Using (1.58) and (1.59) results in

$$I_D = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{tn}) V_{DS} = \mu_n C_{ox} \frac{W}{L} V_{eff} V_{DS} \quad (1.60)$$

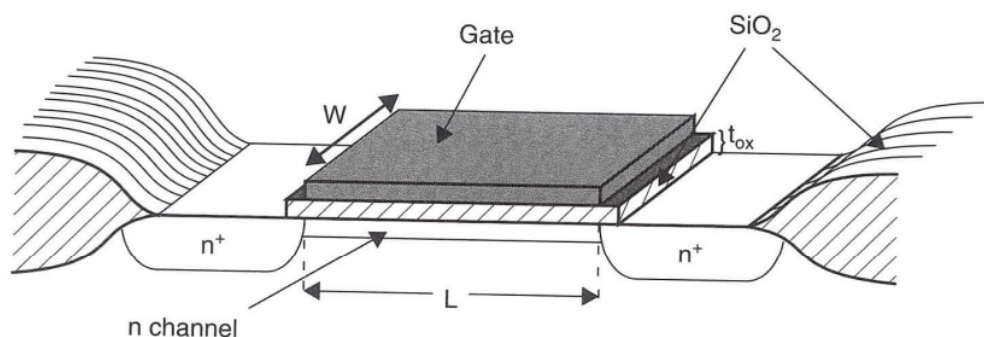


Fig. 1.10 The important dimensions of a MOS transistor.

6. The current is actually conducted by negative carriers (electrons) flowing from the source to the drain. Negative carriers flowing from source to drain results in a positive current from drain to source, I_{DS} .

where it should be emphasized that this relationship is only valid for drain-source voltages near zero (i.e., V_{DS} much smaller than V_{eff}).

As the drain-source voltage increases, the channel charge concentration decreases at the drain end. This decrease is due to the smaller gate-to-channel voltage difference across the thin gate oxide as one moves closer to the drain. In other words, since the drain voltage is assumed to be at a higher voltage than the source, there is an increasing voltage gradient from the source to the drain, resulting in a smaller gate-to-channel voltage near the drain. Since the charge density at a distance x from the source end of the channel is proportional to $V_G - V_{ch}(x) - V_{tn}$, as $V_G - V_{ch}(x)$ decreases, the charge density also decreases.⁷ This effect is illustrated in Fig. 1.11.

Note that at the drain end of the channel, we have

$$V_G - V_{ch}(L) = V_{GD} \quad (1.61)$$

For small V_{DS} , we saw from (1.60) that I_D was linearly related to V_{DS} . However, as V_{DS} increases, and the charge density decreases near the drain, the relationship becomes nonlinear. In fact, the linear relationship for I_D versus V_{DS} flattens for larger V_{DS} , as shown in Fig. 1.12.

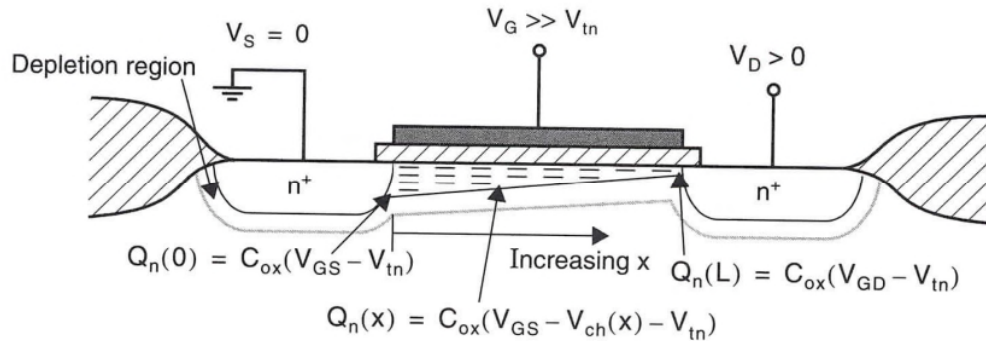


Fig. 1.11 The channel charge density for $V_{DS} > 0$.

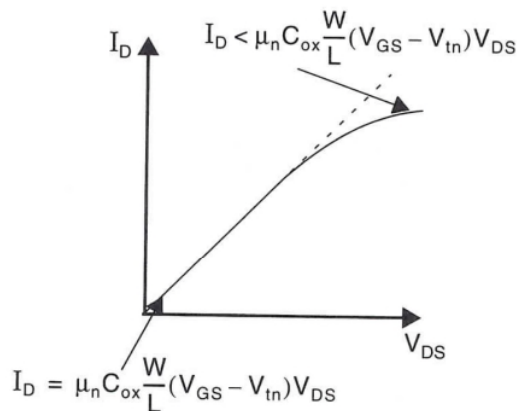


Fig. 1.12 For V_{DS} not close to zero, the I_D versus V_{DS} relationship is no longer linear.

7. $V_G - V_{ch}(x)$ is the gate-to-channel voltage drop at distance x from the source end, with V_G being the same everywhere in the gate, since the gate material is highly conductive.

As the drain voltage is increased, at some point the gate-to-channel voltage at the drain end will decrease to the threshold value V_{tn} —the minimum gate-to-channel voltage needed for n carriers in the channel to exist. Thus, at the drain end, the channel becomes *pinched off*, as shown in Fig. 1.13. This pinch-off occurs at $V_{GD} = V_{tn}$, since the channel voltage at the drain end is simply equal to V_D . Thus, pinch-off occurs for

$$V_{DG} > -V_{tn} \quad (1.62)$$

Denoting V_{DS-sat} as the drain-source voltage when the channel becomes pinched off, we can substitute $V_{DG} = V_{DS} - V_{GS}$ into (1.62) and find an equivalent pinch-off expression

$$V_{DS} > V_{DS-sat} \quad (1.63)$$

where V_{DS-sat} is given⁸ by

$$V_{DS-sat} = V_{GS} - V_{tn} = V_{eff} \quad (1.64)$$

The electron carriers travelling through the pinched-off drain region are velocity saturated, similar to a gas under pressure travelling through a very small tube. If the drain-gate voltage rises above this critical pinch-off voltage of $-V_{tn}$, the charge concentration in the channel remains constant (to a first-order approximation) and the drain current no longer increases with increasing V_{DS} . The result is the current-voltage relationship shown in Fig. 1.14 for a given gate-source voltage. In the region of operation where $V_{DS} > V_{DS-sat}$, the drain current is independent of V_{DS} and is called the *active region*.⁹ The region where I_D changes with V_{DS} is called the *triode region*. When MOS transistors are used in analog amplifiers, they almost always are biased in the active region. When they are used in digital logic gates, they often operate in both regions.

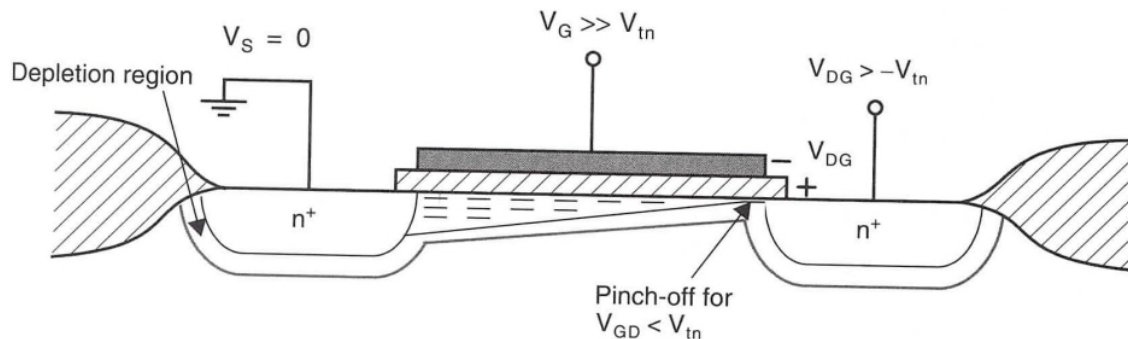


Fig. 1.13 When V_{DS} is increased so that $V_{GD} < V_{tn}$, the channel becomes pinched off at the drain end.

8. Because of the body effect, the threshold voltage at the drain end of the transistor is increased, resulting in the true value of V_{DS-sat} being slightly lower than V_{eff} .

9. Historically, the active region was called the saturation region, but this led to confusion because in the case of bipolar transistors, the saturation region occurs for small V_{CE} , whereas for MOS transistors it occurs for large V_{DS} . The renaming of the saturation region to the active region is becoming widely accepted.

Basic Current Mirrors and Single-Stage Amplifiers

In this chapter, fundamental building blocks are described. These blocks include a variety of current mirrors, single-stage amplifiers with active loads, and differential pairs. A good knowledge of these building blocks is critical to understanding many subjects in the rest of this book and for analog IC design in general. Although some bipolar circuits are described, CMOS mirrors and gain stages are emphasized because modern designs mostly make use of CMOS technology. Fortunately, most of the small-signal analyses presented can be applied to bipolar circuits with little change. In addition, rather than using resistive loads and ac coupling, the gain stages covered are shown with current-mirror active loads since such loads are almost always used in integrated circuits. Finally, single-transistor amplifiers are described here for completeness, although it is assumed that most readers are already somewhat familiar with them.

3.1 SIMPLE CMOS CURRENT MIRROR

A simple CMOS current mirror is shown in Fig. 3.1, in which it is assumed that both transistors are in the active region. If the finite output impedances of the transistors are ignored, and it is assumed that both transistors are the same size, then Q_1 and Q_2 will have the same current since they both have the same gate-source voltage, V_{gs} . However, when finite output impedance is considered, whichever transistor has a larger drain-source voltage will also have a larger current. In addition, the finite output impedance of the transistors will cause the small-signal output impedance of the current mirror, that is, the small-signal impedance looking into the drain of Q_2 , to be less than infinite. To find the output impedance of the current mirror, r_{out} , the small-signal circuit is analyzed after placing a signal source, v_x , at the output node. Then by definition, r_{out} is given by the ratio v_x/i_x , where i_x is the current flowing out of the source and into the drain of Q_2 .

Before finding r_{out} , consider the small-signal model for Q_1 alone, as shown in Fig. 3.2(a). Note that Q_1 is diode connected (i.e., its drain and gate are connected) and that I_{in} does not exist in the small-signal model; I_{in} was replaced with an open

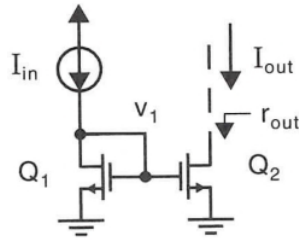
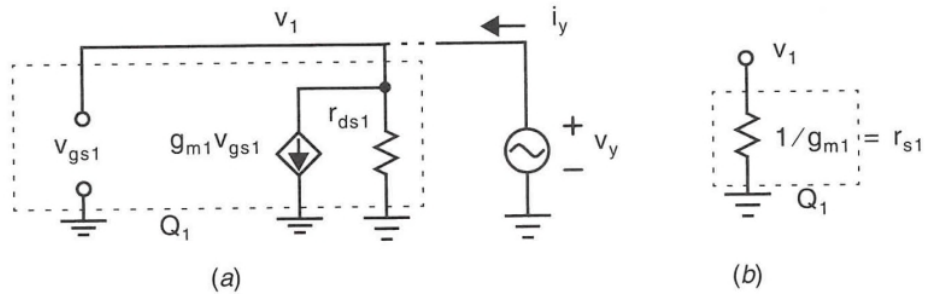


Fig. 3.1 A simple CMOS current mirror.

Fig. 3.2 (a) The small-signal model for Q_1 and (b) the equivalent small-signal model for Q_1 .

circuit because it is an *independent* current source. Also note that a low-frequency small-signal model is used for Q_1 (i.e., all the capacitors are ignored in the model). This small-signal model can be further reduced by finding the *Thévenin-equivalent* circuit. The Thévenin-equivalent output voltage is 0 since the circuit is stable and contains no input signal. This circuit's Thévenin-equivalent output impedance is found by applying a test signal voltage, v_y , at v_1 and measuring the signal current, i_y , as shown. Here, the current i_y is given by

$$i_y = \frac{v_y}{r_{ds1}} + g_{m1}v_{gs1} = \frac{v_y}{r_{ds1}} + g_{m1}v_y \quad (3.1)$$

and recalling that the output impedance is given by v_y/i_y , the output impedance equals $1/g_{m1} \parallel r_{ds1}$. Because typically $r_{ds1} \gg 1/g_{m1}$, we approximate the output impedance to be simply $1/g_{m1}$ (which is also defined to be r_{s1}), which results in the equivalent model shown in Fig. 3.2(b). This same result holds in the bipolar case and is also equivalent to the small-signal model for a diode—hence the name *diode-connected* transistor.

Using the model just described leads to the simplified small-signal model for the overall current mirror, as shown in Fig. 3.3(a), where v_{gs2} has been connected to ground via a resistance of $1/g_{m1}$. Since no current flows through the $1/g_{m1}$ resistor, v_{gs2} equals 0 no matter what voltage level v_x is applied to the current-mirror output. This should come as no surprise, since MOS transistors operate unilaterally at low frequencies. Thus, since $g_{m2}v_{gs2} = 0$, the circuit is simplified to the equivalent small-signal model shown in Fig. 3.3(b). The small-signal output impedance, r_{out} , is simply equal to r_{ds2} .