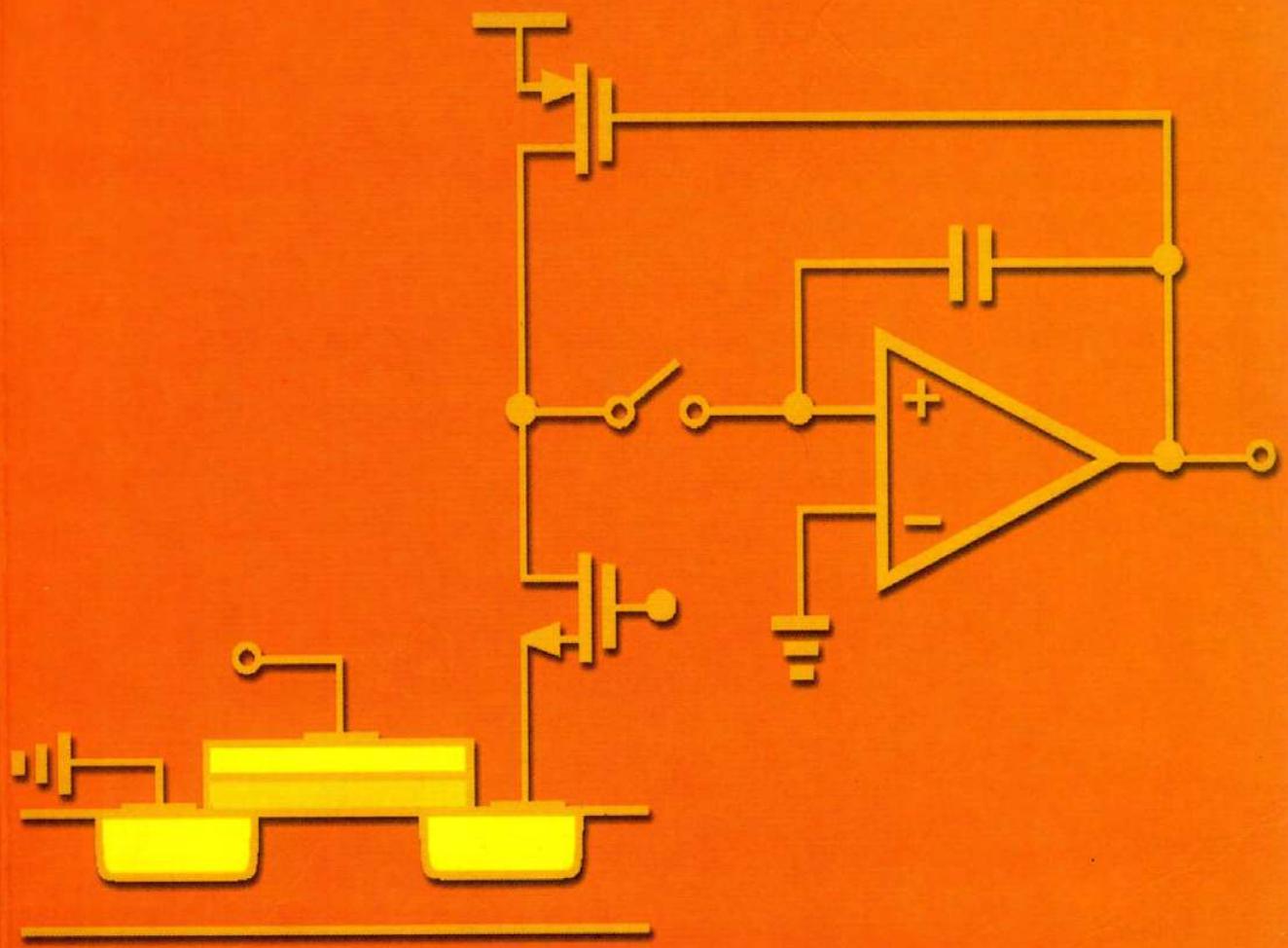


# Design of Analog CMOS Integrated Circuits



Behzad Razavi



McGRAW-HILL INTERNATIONAL EDITION  
Electrical Engineering Series

# **Design of Analog CMOS Integrated Circuits**

# **Design of Analog CMOS Integrated Circuits**

Behzad Razavi

Professor of Electrical Engineering  
University of California, Los Angeles



Boston Burr Ridge, IL Dubuque, IA Madison, WI New York  
San Francisco St. Louis Bangkok Bogotá Caracas Kuala  
Lumpur Lisbon London Madrid Mexico City  
Milan Montreal New Delhi Santiago Seoul Singapore  
Sydney Taipei Toronto

# McGraw-Hill Series in Electrical and Computer Engineering

## **Senior Consulting Editor**

Stephen W. Director, University of Michigan, Ann Arbor

*Circuits and Systems*

*Communications and Signal Processing*

*Computer Engineering*

*Control Theory and Robotics*

*Electromagnetics*

*Electronics and VLSI Circuits*

*Introductory*

*Power*

*Antennas, Microwaves, and Radar*

## **Previous Consulting Editors**

Ronald N. Bracewell, Colin Cherry, James F. Gibbons, Willis W. Harman, Hubert Heffner, Edward W. Herold, John G. Linvill, Simon Ramo, Ronald A. Rohrer, Anthony E. Siegman, Charles Susskind, Frederick E. Terman, John G. Truxal, Ernst Weber, and John R. Whinnery



DESIGN OF ANALOG CMOS INTEGRATED CIRCUITS  
International Edition 2001

Exclusive rights by McGraw-Hill Book Co – Singapore, for manufacture and export. This book cannot be re-exported from the country to which it is sold by McGraw-Hill. The International Edition is not available in North America.

Published by McGraw-Hill, an imprint of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2001, by The McGraw-Hill Companies, Inc. All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written consent of The McGraw-Hill Companies, Inc., including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

10 09 08 07 06  
20 09 08 07 06 05 04 03 02  
FC BJE

ISBN 0-07-118815-0 (softcover edition)  
ISBN 0-07-118839-8 (hardcover edition)

**Library of Congress Cataloging-in-Publication Data**

Razavi, Behzad.

Design of analog CMOS integrated circuits / Behzad Razavi.

p. cm.

ISBN 0-07-238032-2

1. Linear integrated circuit—Design and construction. 2. Metal oxide semiconductors, Complimentary. I. Title.

TK7874.654 R39 2001

621.39'732—dc21

00-044789

[www.mhhe.com](http://www.mhhe.com)

**When ordering this title, use ISBN 0-07-118815-0 (softcover) or ISBN 0-07-118839-8 (hardcover)**

Printed in Singapore

*To the memory of my parents*



## *About the Author*

---

Behzad Razavi received the B.Sc. degree in electrical engineering from Sharif University of Technology in 1985 and the M.Sc. and Ph.D. degrees in electrical engineering from Stanford University in 1988 and 1992, respectively. He was with AT&T Bell Laboratories and subsequently Hewlett-Packard Laboratories until 1996. Since September 1996, he has been an Associate Professor and subsequently a Professor of electrical engineering at University of California, Los Angeles. His current research includes wireless transceivers, frequency synthesizers, phase-locking and clock recovery for high-speed data communications, and data converters.

Professor Razavi served as an Adjunct Professor at Princeton University, Princeton, NJ, from 1992 to 1994, and at Stanford University in 1995. He is a member of the Technical Program Committees of the Symposium on VLSI Circuits and the International Solid-State Circuits Conference (ISSCC), in which he is the chair of the Analog Subcommittee. He has also served as Guest Editor and Associate Editor of the IEEE Journal of Solid-State Circuits, IEEE Transactions on Circuits and Systems, and International Journal of High Speed Electronics.

Professor Razavi received the Beatrice Winner Award for Editorial Excellence at the 1994 ISSCC, the best paper award at the 1994 European Solid-State Circuits Conference, the best panel award at the 1995 and 1997 ISSCC, the TRW Innovative Teaching Award in 1997, and the best paper award at the IEEE Custom Integrated Circuits Conference in 1998. He is the author of *Principles of Data Conversion System Design* (IEEE Press, 1995), and *RF Microelectronics* (Prentice Hall, 1998), and the editor of *Monolithic Phase-Locked Loops and Clock Recovery Circuits* (IEEE Press, 1996).



# *Preface*

---

In the past two decades, CMOS technology has rapidly embraced the field of analog integrated circuits, providing low-cost, high-performance solutions and rising to dominate the market. While silicon bipolar and III-V devices still find niche applications, only CMOS processes have emerged as a viable choice for the integration of today's complex mixed-signal systems. With channel lengths projected to scale down to  $0.03\text{ }\mu\text{m}$ , CMOS technology will continue to serve circuit design for probably another two decades.

Analog circuit design itself has evolved with the technology as well. High-voltage, high-power analog circuits containing a few tens of transistors and processing small, continuous-time signals have gradually been replaced by low-voltage, low-power systems comprising thousands of devices and processing large, mostly discrete-time signals. For example, many analog techniques used only ten years ago have been abandoned because they do not lend themselves to low-voltage operation.

This book deals with the analysis and design of analog CMOS integrated circuits, emphasizing fundamentals as well as new paradigms that students and practicing engineers need to master in today's industry. Since analog design requires both intuition and rigor, each concept is first introduced from an intuitive perspective and subsequently treated by careful analysis. The objective is to develop both a solid foundation and methods of analyzing circuits by inspection so that the reader learns what approximations can be made in which circuits and how much error to expect in each approximation. This approach also enables the reader to apply the concepts to bipolar circuits with little additional effort.

I have taught most of the material in this book both at UCLA and in industry, polishing the order, the format, and the content with every offering. As the reader will see throughout the book, I follow four "golden rules" in writing (and teaching): (1) I explain *why* the reader needs to know the concept that is to be studied; (2) I put myself in the reader's position and predict the questions that he/she may have while reading the material for the first time; (3) With Rule 2 in mind, I pretend to know only as much as the (first-time) reader and try to "grow" with him/her, thereby experiencing the same through process; (4) I begin with the "core" concept in a simple (even imprecise) language and gradually add necessary modifications to arrive at the final (precise) idea. The last rule is particularly important in teaching circuits because it allows the reader to observe the evolution of a topology and hence learn both analysis and synthesis.

The text comprises 18 chapters whose contents and order are carefully chosen to provide a natural flow for both self-study and classroom adoption in quarter or semester systems.

Unlike some other books on analog design, we cover only a *bare minimum* of MOS device physics at the beginning, leaving more advanced properties and fabrication details for later chapters. To an expert, the elementary device physics treatment may appear oversimplified, but my experience suggests that (a) first-time readers simply do not absorb the high-order device effects and fabrication technology before they study circuits because they do not see the relevance; (b) if properly presented, even the simple treatment proves adequate for a substantial coverage of basic circuits; (c) readers learn advanced device phenomena and processing steps much more readily *after* they have been exposed to a significant amount of circuit analysis and design.

Chapter 1 provides the reader with motivation for learning the material in this book.

Chapter 2 describes basic physics and operation of MOS devices.

Chapters 3 through 5 deal with single-stage and differential amplifiers and current mirrors, respectively, developing efficient analytical tools for quantifying the behavior of basic circuits by inspection.

Chapters 6 and 7 introduce two imperfections of circuits, namely, frequency response and noise. Noise is treated at an early stage so that it “sinks in” as the reader accounts for its effects in subsequent circuit developments.

Chapters 8 through 10 describe feedback, operational amplifiers, and stability in feedback systems, respectively. With the useful properties of feedback analyzed, the reader is motivated to design high-performance, stable op amps and understand the trade-offs between speed, precision, and power dissipation.

Chapters 11 through 13 deal with more advanced topics: bandgap references, elementary switched-capacitor circuits, and the effect of nonlinearity and mismatch. These three subjects are included here because they prove essential in most analog and mixed-signal systems today.

Chapters 14 and 15 concentrate on the design of oscillators and phase-locked loops, respectively. In view of the wide usage of these circuits, a detailed study of their behavior and many examples of their operation are provided.

Chapter 16 is concerned with high-order MOS device effects and models, emphasizing the circuit design implications. If preferred, this chapter can directly follow Chapter 2 as well. Chapter 17 describes CMOS fabrication technology with a brief overview of layout design rules.

Chapter 18 presents the layout and packaging of analog and mixed-signal circuits. Many practical issues that directly impact the performance of the circuit are described and various techniques are introduced.

The reader is assumed to have a basic knowledge of electronic circuits and devices, e.g., *pn* junctions, the concept of small-signal operation, equivalent circuits, and simple biasing. For a senior-level elective course, Chapters 1 through 8 can be covered in a quarter and Chapters 1 through 10 in a semester. For a first-year graduate course, Chapters 1 through 11 plus one of Chapters 12 through 15 can be taught in one quarter, and the first 16 chapters in one semester.

The problem sets at the end of each chapter are designed to extend the reader’s understanding of the material and complement it with additional practical considerations. A solutions manual is available for instructors.

# *Acknowledgments*

---

Writing a book begins with a great deal of excitement. However, after two years of relentless writing, drawing, and revising, when the book exceeds 600 pages and it is almost impossible to make the equations and subscripts and superscripts in the last chapter consistent with those in the first, the author begins to feel the streaks of insanity, realizing that the book will never finish without the support of many other people.

This book has benefited from the contributions of many individuals. A number of UCLA students read the first draft and the preview edition sentence by sentence. In particular, Alireza Zolfaghari, Ellie Cijvat, and Hamid Rafati meticulously read the book and found several hundred errors (some quite subtle). Also, Emad Hegazi, Dawei Guo, Alireza Razzaghi, Jafar Savoj, and Jing Tian made helpful suggestions regarding many chapters. I thank all.

Many experts in academia and industry read various parts of the book and provided useful feedback. Among them are Brian Brandt (National Semiconductor), Matt Corey (National Semiconductor), Terri Fiez (Oregon State University), Ian Galton (UC San Diego), Ali Hajimiri (Caltech), Stacy Ho (Analog Devices), Yin Hu (Texas Instruments), Shen-Iuan Liu (National Taiwan University), Joe Lutsky (National Semiconductor), Amit Mehrotra (University of Illinois, Urbana-Champaign), David Robertson (Analog Devices), David Su (T-Span), Tao Sun (National Semiconductor), Robert Taft (National Semiconductor), and Masoud Zargari (T-Span). Jason Woo (UCLA) patiently endured and answered my questions about device physics. I thank all.

Ramesh Harjani (University of Minnesota), John Nyenhius (Purdue University), Norman Tien (Cornell University), and Mahmoud Wagdy (California State University, Long Beach) reviewed the book proposal and made valuable suggestions. I thank all.

My wife, Angelina, has made many contributions to this book, from typing chapters to finding numerous errors and raising questions that made me reexamine my own understanding. I am very grateful to her.

The timely production of the book was made possible by the hard work of the staff at McGraw-Hill, particularly, Catherine Fields, Michelle Flomenhoft, Heather Burbridge, Denise Santor-Mitzit, and Jim Labeots. I thank all.

I learned analog design from two masters: Mehrdad Sharif-Bakhtiar (Sharif University of Technology) and Bruce Wooley (Stanford University) and it is only appropriate that I express my gratitude to them here. What I inherited from them will be inherited by many generations of students.

Behzad Razavi  
July 2000

# *Brief Contents*

---

<b>1</b>	<b>Introduction to Analog Design . . . . .</b>	<b>1</b>
<b>2</b>	<b>Basic MOS Device Physics . . . . .</b>	<b>9</b>
<b>3</b>	<b>Single-Stage Amplifiers . . . . .</b>	<b>47</b>
<b>4</b>	<b>Differential Amplifiers . . . . .</b>	<b>100</b>
<b>5</b>	<b>Passive and Active Current Mirrors . . . . .</b>	<b>135</b>
<b>6</b>	<b>Frequency Response of Amplifiers . . . . .</b>	<b>166</b>
<b>7</b>	<b>Noise . . . . .</b>	<b>201</b>
<b>8</b>	<b>Feedback . . . . .</b>	<b>246</b>
<b>9</b>	<b>Operational Amplifiers . . . . .</b>	<b>291</b>
<b>10</b>	<b>Stability and Frequency Compensation . . . . .</b>	<b>345</b>
<b>11</b>	<b>Bandgap References . . . . .</b>	<b>377</b>
<b>12</b>	<b>Introduction to Switched-Capacitor Circuits . . . . .</b>	<b>405</b>
<b>13</b>	<b>Nonlinearity and Mismatch . . . . .</b>	<b>448</b>
<b>14</b>	<b>Oscillators . . . . .</b>	<b>482</b>
<b>15</b>	<b>Phase-Locked Loops . . . . .</b>	<b>532</b>
<b>16</b>	<b>Short-Channel Effects and Device Models . . . . .</b>	<b>579</b>
<b>17</b>	<b>CMOS Processing Technology . . . . .</b>	<b>604</b>
<b>18</b>	<b>Layout and Packaging . . . . .</b>	<b>631</b>
	<b>Index . . . . .</b>	<b>677</b>



# *Contents*

---

About the Author.....	vii
Preface.....	ix
Acknowledgments.....	xi
<b>1 Introduction to Analog Design.....</b>	<b>1</b>
1.1 Why Analog?.....	1
1.2 Why Integrated?.....	6
1.3 Why CMOS?.....	6
1.4 Why This Book?.....	7
1.5 General Concepts .....	7
1.5.1 Levels of Abstraction .....	7
1.5.2 Robust Analog Design .....	7
<b>2 Basic MOS Device Physics .....</b>	<b>9</b>
2.1 General Considerations .....	10
2.1.1 MOSFET as a Switch.....	10
2.1.2 MOSFET Structure .....	10
2.1.3 MOS Symbols .....	12
2.2 MOS I/V Characteristics .....	13
2.2.1 Threshold Voltage.....	13
2.2.2 Derivation of I/V Characteristics .....	15
2.3 Second-Order Effects .....	23
2.4 MOS Device Models .....	28
2.4.1 MOS Device Layout.....	28
2.4.2 MOS Device Capacitances .....	29
2.4.3 MOS Small-Signal Model .....	33
2.4.4 MOS SPICE models .....	36
2.4.5 NMOS versus PMOS Devices .....	37
2.4.6 Long-Channel versus Short-Channel Devices.....	38

<b>3 Single-Stage Amplifiers .....</b>	<b>4</b>
3.1 Basic Concepts .....	4
3.2 Common-Source Stage .....	48
3.2.1 Common-Source Stage with Resistive Load .....	48
3.2.2 CS Stage with Diode-Connected Load .....	51
3.2.3 CS Stage with Current-Source Load .....	58
3.2.4 CS Stage with Triode Load .....	59
3.2.5 CS Stage with Source Degeneration .....	61
3.3 Source Follower.....	61
3.4 Common-Gate Stage .....	71
3.5 Cascode Stage .....	81
3.5.1 Folded Cascode .....	91
3.6 Choice of Device Models .....	93
<b>4 Differential Amplifiers .....</b>	<b>100</b>
4.1 Single-Ended and Differential Operation.....	100
4.2 Basic Differential Pair .....	100
4.2.1 Qualitative Analysis .....	100
4.2.2 Quantitative Analysis .....	100
4.3 Common-Mode Response.....	111
4.4 Differential Pair with MOS Loads.....	121
4.5 Gilbert Cell.....	121
<b>5 Passive and Active Current Mirrors .....</b>	<b>13</b>
5.1 Basic Current Mirrors .....	13
5.2 Cascode Current Mirrors.....	13
5.3 Active Current Mirrors.....	14
5.3.1 Large-Signal Analysis .....	14
5.3.2 Small-Signal Analysis .....	15
5.3.3 Common-Mode Properties .....	15
<b>6 Frequency Response of Amplifiers .....</b>	<b>16</b>
6.1 General Considerations .....	16
6.1.1 Miller Effect .....	16
6.1.2 Association of Poles with Nodes .....	16
6.2 Common-Source Stage .....	17
6.3 Source Followers.....	17
6.4 Common-Gate Stage .....	18
6.5 Cascode Stage .....	18
6.6 Differential Pair.....	18
Appendix A: Dual of Miller's Theorem.....	19

<b>7 Noise .....</b>	<b>201</b>
7.1 Statistical Characteristics of Noise .....	201
7.1.1 Noise Spectrum .....	203
7.1.2 Amplitude Distribution .....	206
7.1.3 Correlated and Uncorrelated Sources .....	207
7.2 Types of Noise .....	209
7.2.1 Thermal Noise .....	209
7.2.2 Flicker Noise .....	215
7.3 Representation of Noise in Circuits .....	218
7.4 Noise in Single-Stage Amplifiers .....	224
7.4.1 Common-Source Stage .....	225
7.4.2 Common-Gate Stage.....	228
7.4.3 Source Followers.....	231
7.4.4 Cascode Stage .....	232
7.5 Noise in Differential Pairs.....	233
7.6 Noise Bandwidth.....	239
<b>8 Feedback .....</b>	<b>246</b>
8.1 General Considerations .....	246
8.1.1 Properties of Feedback Circuits .....	247
8.1.2 Types of Amplifiers.....	254
8.2 Feedback Topologies .....	258
8.2.1 Voltage-Voltage Feedback .....	258
8.2.2 Current-Voltage Feedback .....	263
8.2.3 Voltage-Current Feedback .....	266
8.2.4 Current-Current Feedback .....	269
8.3 Effect of Loading .....	270
8.3.1 Two-Port Network Models .....	270
8.3.2 Loading in Voltage-Voltage Feedback .....	272
8.3.3 Loading in Current-Voltage Feedback .....	275
8.3.4 Loading in Voltage-Current Feedback .....	278
8.3.5 Loading in Current-Current Feedback .....	281
8.3.6 Summary of Loading Effects .....	283
8.4 Effect of Feedback on Noise.....	284
<b>9 Operational Amplifiers.....</b>	<b>291</b>
9.1 General Considerations .....	291
9.1.1 Performance Parameters .....	291
9.2 One-Stage Op Amps.....	296
9.3 Two-Stage Op Amps .....	307
9.4 Gain Boosting .....	309
9.5 Comparison .....	313
9.6 Common-Mode Feedback.....	314

9.7 Input Range Limitations .....	325
9.8 Slew Rate .....	326
9.9 Power Supply Rejection .....	334
9.10 Noise in Op Amps .....	336
<b>10 Stability and Frequency Compensation .....</b>	<b>345</b>
10.1 General Considerations .....	345
10.2 Multipole Systems .....	349
10.3 Phase Margin .....	351
10.4 Frequency Compensation .....	355
10.5 Compensation of Two-Stage Op Amps .....	361
10.5.1 Slew in Two-Stage Op Amps .....	368
10.6 Other Compensation Techniques .....	369
<b>11 Bandgap References .....</b>	<b>377</b>
11.1 General Considerations .....	377
11.2 Supply-Independent Biasing .....	377
11.3 Temperature-Independent References .....	381
11.3.1 Negative-TC Voltage .....	381
11.3.2 Positive-TC Voltage .....	382
11.3.3 Bandgap Reference .....	384
11.4 PTAT Current Generation .....	390
11.5 Constant- $G_m$ Biasing .....	392
11.6 Speed and Noise Issues .....	393
11.7 Case Study .....	397
<b>12 Introduction to Switched-Capacitor Circuits .....</b>	<b>405</b>
12.1 General Considerations .....	405
12.2 Sampling Switches .....	410
12.2.1 MOSFETS as Switches .....	410
12.2.2 Speed Considerations .....	414
12.2.3 Precision Considerations .....	417
12.2.4 Charge Injection Cancellation .....	421
12.3 Switched-Capacitor Amplifiers .....	423
12.3.1 Unity-Gain Sampler/Buffer .....	424
12.3.2 Noninverting Amplifier .....	432
12.3.3 Precision Multiply-by-Two Circuit .....	438
12.4 Switched-Capacitor Integrator .....	439
12.5 Switched-Capacitor Common-Mode Feedback .....	442
<b>13 Nonlinearity and Mismatch .....</b>	<b>448</b>
13.1 Nonlinearity .....	448
13.1.1 General Considerations .....	448

13.1.2 Nonlinearity of Differential Circuits .....	452
13.1.3 Effect of Negative Feedback on Nonlinearity .....	454
13.1.4 Capacitor Nonlinearity .....	457
13.1.5 Linearization Techniques .....	458
13.2 Mismatch .....	463
13.2.1 Offset Cancellation Techniques .....	471
13.2.2 Reduction of Noise by Offset Cancellation .....	476
13.2.3 Alternative Definition of CMRR .....	478
<b>14 Oscillators.....</b>	<b>482</b>
14.1 General Considerations .....	482
14.2 Ring Oscillators .....	484
14.3 LC Oscillators .....	495
14.3.1 Crossed-Coupled Oscillator .....	499
14.3.2 Colpitts Oscillator .....	502
14.3.3 One-Port Oscillators .....	505
14.4 Voltage-Controlled Oscillators .....	510
14.4.1 Tuning in Ring Oscillators .....	512
14.4.2 Tuning in LC Oscillators .....	521
14.5 Mathematical Model of VCOs .....	525
<b>15 Phase-Locked Loops.....</b>	<b>532</b>
15.1 Simple PLL .....	532
15.1.1 Phase Detector .....	532
15.1.2 Basic PLL Topology .....	533
15.1.3 Dynamics of Simple PLL .....	542
15.2 Charge-Pump PLLs .....	549
15.2.1 Problem of Lock Acquisition .....	549
15.2.2 Phase/Frequency Detector and Charge Pump .....	550
15.2.3 Basic Charge-Pump PLL .....	556
15.3 Nonideal Effects in PLLs .....	562
15.3.1 PFD/CP Nonidealities .....	562
15.3.2 Jitter in PLLs .....	567
15.4 Delay-Locked Loops .....	569
15.5 Applications .....	572
15.5.1 Frequency Multiplication and Synthesis .....	572
15.5.2 Skew Reduction .....	574
15.5.3 Jitter Reduction .....	576
<b>16 Short-Channel Effects and Device Models.....</b>	<b>579</b>
16.1 Scaling Theory .....	579
16.2 Short-Channel Effects .....	583

16.2.1	Threshold Voltage Variation . . . . .	583
16.2.2	Mobility Degradation with Vertical Field . . . . .	585
16.2.3	Velocity Saturation . . . . .	587
16.2.4	Hot Carrier Effects . . . . .	589
16.2.5	Output Impedance Variation with Drain-Source Voltage . . . . .	589
16.3	MOS Device Models . . . . .	591
16.3.1	Level 1 Model . . . . .	592
16.3.2	Level 2 Model . . . . .	593
16.3.3	Level 3 Model . . . . .	595
16.3.4	BSIM Series . . . . .	596
16.3.5	Other Models . . . . .	597
16.3.6	Charge and Capacitance Modeling . . . . .	598
16.3.7	Temperature Dependence . . . . .	599
16.4	Process Corners . . . . .	599
16.5	Analog Design in a Digital World . . . . .	600
<b>17</b>	<b>CMOS Processing Technology . . . . .</b>	<b>604</b>
17.1	General Considerations . . . . .	604
17.2	Wafer Processing . . . . .	605
17.3	Photolithography . . . . .	606
17.4	Oxidation . . . . .	608
17.5	Ion Implantation . . . . .	608
17.6	Deposition and Etching . . . . .	611
17.7	Device Fabrication . . . . .	611
17.7.1	Active Devices . . . . .	611
17.7.2	Passive Devices . . . . .	616
17.7.3	Interconnects . . . . .	624
17.8	Latch-Up . . . . .	627
<b>18</b>	<b>Layout and Packaging . . . . .</b>	<b>631</b>
18.1	General Layout Considerations . . . . .	631
18.1.1	Design Rules . . . . .	632
18.1.2	Antenna Effect . . . . .	634
18.2	Analog Layout Techniques . . . . .	635
18.2.1	Multifinger Transistors . . . . .	635
18.2.2	Symmetry . . . . .	637
18.2.3	Reference Distribution . . . . .	642
18.2.4	Passive Devices . . . . .	644
18.2.5	Interconnects . . . . .	653
18.3	Substrate Coupling . . . . .	660
<b>Index . . . . .</b>		<b>677</b>

# *Introduction to Analog Design*

---

## 1.1 Why Analog?

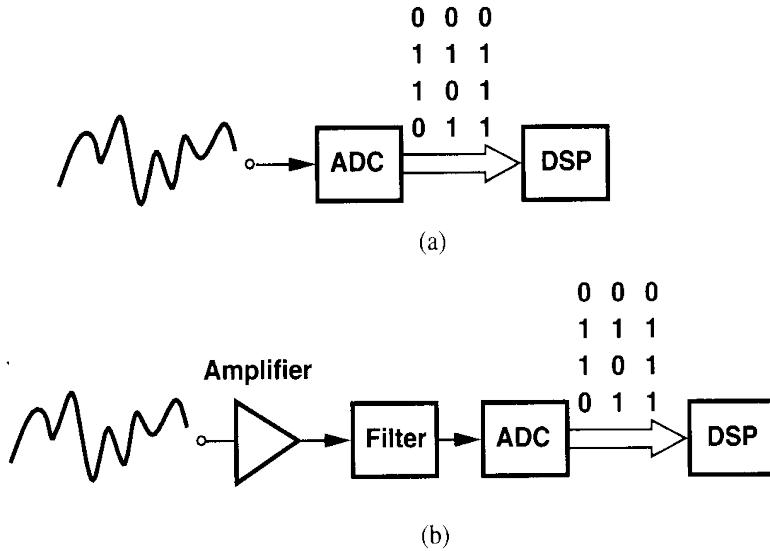
It was in the early 1980s that many experts predicted the demise of analog circuits. Digital signal processing algorithms were becoming increasingly more powerful while advances in integrated-circuit (IC) technology provided compact, efficient implementation of these algorithms in silicon. Many functions that had traditionally been realized in analog form were now easily performed in the digital domain, suggesting that, with enough capability in IC fabrication, all processing of signals would eventually occur digitally. The future looked quite bleak to analog designers and they were seeking other jobs.

But, why are analog designers in such great demand today? After all, digital signal processing and IC technologies have advanced tremendously since the early 1980s, making it possible to realize processors containing millions of transistors and performing billions of operations per second. Why did this progress not confirm the earlier predictions?

While many types of signal processing have indeed moved to the digital domain, analog circuits have proved *fundamentally* necessary in many of today's complex, high-performance systems. Let us consider a few applications where it is very difficult or even impossible to replace analog functions with their digital counterparts regardless of advances in technology.

**Processing of Natural Signals** Naturally occurring signals are analog—at least at a macroscopic level. A high-quality microphone picking up the sound of an orchestra generates a voltage whose amplitude may vary from a few microvolts to hundreds of millivolts. The photocells in a video camera produce a current that is as low as a few electrons per microsecond. A seismographic sensor has an output voltage ranging from a few microvolts for very small vibrations of the earth to hundreds of millivolts for heavy earthquakes. Since all of these signals must eventually undergo extensive processing in the digital domain, we observe that each of these systems consists of an analog-to-digital converter (ADC) and a digital signal processor (DSP) [Fig. 1.1(a)]. The design of ADCs for high speed, high precision, and low power dissipation is one of many difficult challenges in analog design.

In practice, the electrical version of natural signals may be prohibitively small for direct digitization by the ADC. The signals are also often accompanied by unwanted, out-of-band

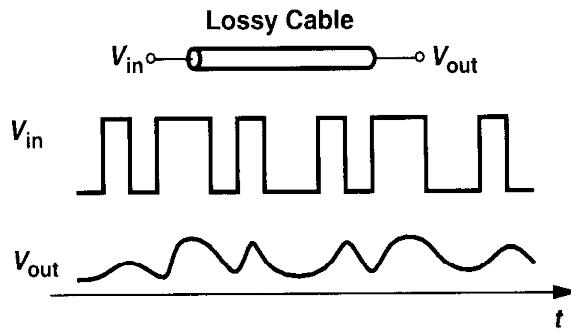


**Figure 1.1** (a) Digitization of a natural signal, (b) addition of amplification and filtering for higher sensitivity.

interferers. The front end of Fig. 1.1(a) may therefore be modified as shown in Fig. 1.1(b), where an amplifier boosts the signal level and an analog filter suppresses the out-of-band components. The design of high-performance amplifiers and filters is also a topic of active research today.

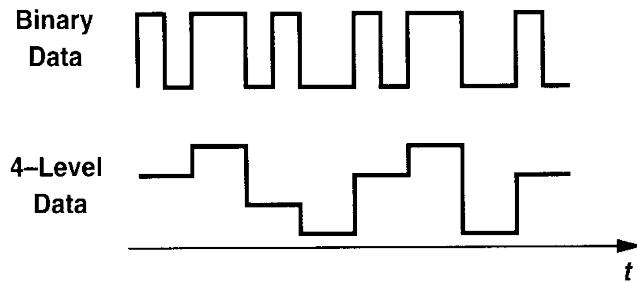
**Digital Communications** Binary data generated by various systems must often be transmitted over long distances. For example, computer networks in large office buildings may transmit the data over cables that are hundreds of meters long.

What happens if a high-speed stream of binary data travels through a long cable? As illustrated in Fig. 1.2, the signal experiences both attenuation and “distortion,” no longer resembling a digital waveform. Thus, a receiver similar to that of Fig. 1.1(b) may be necessary here.



**Figure 1.2** Attenuation and distortion of data through a lossy cable.

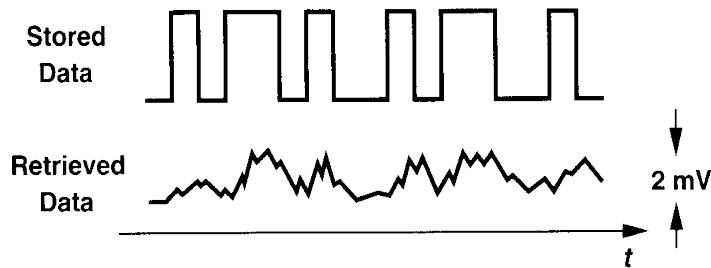
In order to improve the quality of communication, the above system may incorporate “multi-level”—rather than binary—signals. For example, if, as shown in Fig. 1.3, every two consecutive bits in the sequence are grouped and converted to one of four levels, then



**Figure 1.3** Use of multi-level signalling to reduce the required bandwidth.

each level is twice as long as a bit period, demanding only *half* the bandwidth required for transmission of the binary stream. Utilized extensively in today's communication systems, multi-level signals necessitate a digital-to-analog converter (DAC) in the transmitter to produce multiple levels from the grouped binary data and an ADC in the receiver to determine which level has been transmitted. The key point here is that increasing the number of levels relaxes the bandwidth requirements while demanding a higher precision in the DAC and the ADC.

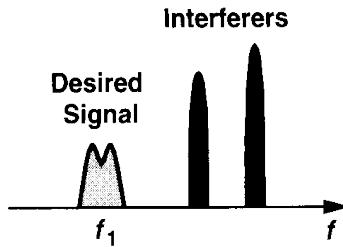
**Disk Drive Electronics** The data stored magnetically on a computer hard disk is in binary form. However, when the data is read by a magnetic head and converted to an electrical signal, the result appears as shown in Fig. 1.4. The amplitude is only a few millivolts, the noise content is quite high, and the bits experience substantial distortion.



**Figure 1.4** Data stored in and retrieved from a hard disk.

Thus, as illustrated in Fig. 1.1, the signal is amplified, filtered, and digitized for further processing. Depending on the overall system architecture, the analog filter in this case may in fact serve to remove a significant portion of the noise and the distortion of the signal. The design of each of these building blocks poses great challenges as the speed of computers and their storage media continues to increase every year. For example, today's disk drives require a speed of 500 Mb/s.

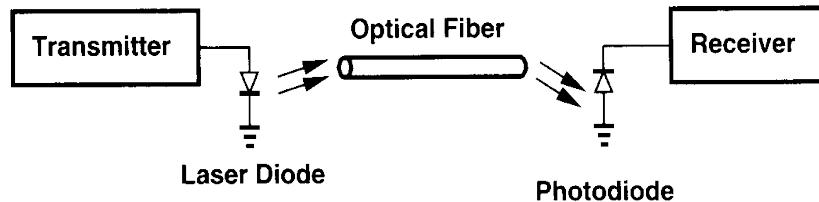
**Wireless Receivers** The signal picked up by the antenna of a radio-frequency (RF) receiver, e.g., a pager or a cellular telephone, exhibits an amplitude of only a few microvolts and a center frequency of 1 GHz or higher. Furthermore, the signal is accompanied by large



**Figure 1.5** Signal and interferers received by the antenna of a wireless receiver.

interferers (Fig. 1.5). The receiver must therefore amplify the low-level signal with minimal noise, operate at a high frequency, and withstand large unwanted components. Note that these requirements are necessary even if the desired signal is not in “analog” form. The trade-offs between noise, frequency of operation, tolerance of interferers, power dissipation, and *cost* constitute the principal challenge in today’s wireless industry.

**Optical Receivers** For transmission of high-speed data over very long distances, cables generally prove inadequate because of their limited bandwidth and considerable attenuation. Thus, as illustrated in Fig. 1.6, the data is converted to light by means of a laser diode and transmitted over an optical fiber, which exhibits an extremely wide band and a very low

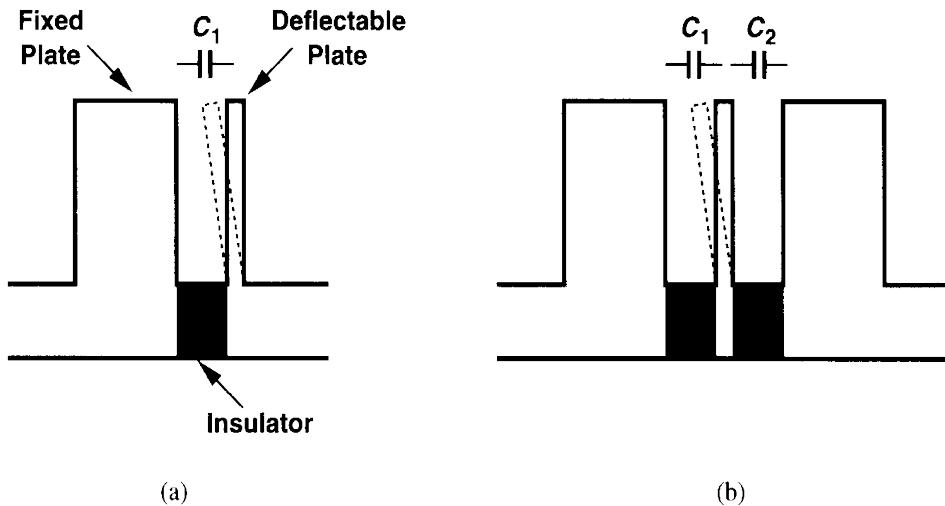


**Figure 1.6** Optical fiber system.

loss. At the receive end, the light is converted to a small electrical current by a photodiode. The receiver must then process a low-level signal at a very high speed, requiring low-noise, broadband circuit design. For example, state-of-the-art optical receivers operate in the range of 10 to 40 Gb/s.

**Sensors** Mechanical, electrical, and optical sensors play a critical role in our lives. For example, video cameras incorporate an array of photodiodes to convert an image to current and ultrasound systems use an acoustic sensor to generate a voltage proportional to the amplitude of the ultrasound waveform. Amplification, filtering, and A/D conversion are essential functions in these applications.

An interesting example of sensors is the accelerometers employed in automobiles to activate air bags. When the vehicle hits an obstacle, the drop in the speed is measured as acceleration and, if exceeding a certain threshold, it triggers the air bag release mechanism. Modern accelerometers are based on a variable capacitor consisting of a fixed plate and a deflectable plate [Fig. 1.7(a)]. The deflection and hence the value of the capacitor are proportional to the acceleration, requiring a circuit that accurately measures the change in capacitance. The design of such interface circuits is quite difficult because for typical



**Figure 1.7** (a) Simple accelerometer, (b) differential accelerometer.

accelerations, the interplate capacitance may change by less than 1%, demanding a high precision in the measurement. In practice, the structure of Fig. 1.7(b) is used to provide two capacitors that change in opposite directions, reducing the task to the measurement of the *difference* between two capacitances rather than the absolute value of one.

**Microprocessors and Memories** Today's microprocessors and memories draw upon a great deal of analog design expertise. Many issues related to the distribution and timing of data and clocks across a large chip or among chips mandate that high-speed signals be viewed as analog waveforms. Furthermore, nonidealities in signal and power interconnects on the chip as well as package parasitics require a solid understanding of analog design. In addition, semiconductor memories employ high-speed "sense amplifiers" extensively, necessitating many analog techniques. For these reasons, it is often said "high-speed digital design is in fact analog design."

The foregoing applications demonstrate the wide and inevitable spread of analog circuits in modern industry. But, why is analog design difficult? We make the following observations.

- (1) Whereas digital circuits entail primarily one trade-off between speed and power dissipation, analog design must deal with a multi-dimensional trade-off consisting of speed, power dissipation, gain, precision, supply voltage, etc.
- (2) With the speed and precision required in processing analog signals, analog circuits are much more sensitive to noise, crosstalk, and other interferers than are digital circuits.
- (3) Second-order effects in devices influence the performance of analog circuits much more heavily than that of digital circuits.
- (4) The design of high-performance analog circuits can rarely be automated, usually requiring that every device be “hand-crafted.” By contrast, many digital circuits are automatically synthesized and laid out.
- (5) Despite tremendous progress, modeling and simulation of many effects in analog circuits continue to pose difficulties, forcing the designers to draw upon experience and intuition when analyzing the results of a simulation.
- (6) An important thrust in today’s semiconductor industry is to design analog circuits in mainstream IC technologies used to fabricate digital products. Developed and characterized for digital applications,

such technologies do not easily lend themselves to analog design, requiring novel circuits and architectures to achieve a high performance.

## 1.2 Why Integrated?

The idea of placing multiple electronic devices on the same substrate was conceived in the late 1950s. In 40 years, the technology has evolved from producing simple chips containing a handful of components to fabricating memories accommodating more than one billion transistors as well as microprocessors comprising more than 10 million devices. As Gordon Moore (one of the founders of Intel) predicted in the early 1970s, the number of transistors per chip has continued to double approximately every one and a half years. At the same time, the minimum dimension of transistors has dropped from about  $25\text{ }\mu\text{m}$  in 1960 to about  $0.18\text{ }\mu\text{m}$  in the year 2000, resulting in a tremendous improvement in the speed of integrated circuits.

Driven by primarily the memory and microprocessor market, integrated-circuit technologies have also embraced analog design extensively, affording a complexity, speed, and precision that would be impossible to achieve using discrete implementations. Analog and mixed analog/digital integrated circuits containing tens of thousands of devices now routinely appear in consumer products. We can no longer build a discrete prototype to predict the behavior and performance of modern analog circuits.

## 1.3 Why CMOS?

The idea of metal-oxide-silicon field-effect transistors (MOSFETs) was patented by J. E. Lilienfeld in the early 1930s—well before the invention of the bipolar transistor. Owing to fabrication limitations, however, MOS technologies became practical much later, in the early 1960s, with the first several generations producing only *n*-type transistors. It was in the mid-1960s that complementary MOS (CMOS) devices (i.e., both *n*-type and *p*-type transistors) were introduced, initiating a revolution in the semiconductor industry.

CMOS technologies rapidly captured the digital market: CMOS gates dissipated power only during switching and required very few devices, two attributes in sharp contrast to their bipolar or GaAs counterparts. It was also soon discovered that the dimensions of MOS devices could be scaled down more easily than those of other types of transistors. Furthermore, CMOS circuits proved to have a lower fabrication cost.

The next obvious step was to apply CMOS technology to analog design. The low cost of fabrication and the possibility of placing both analog and digital circuits on the same chip so as to improve the overall performance and/or reduce the cost of packaging made CMOS technology attractive. However, MOSFETs were quite slower and noisier than bipolar transistors, finding limited application.

How did CMOS technology come to dominate the analog market as well? The principal force was device scaling because it continued to improve the speed of MOSFETs. The intrinsic speed of MOS transistors has increased by more than three orders of magnitude in the past 30 years, becoming comparable with that of bipolar devices even though the latter

have also been scaled (but not as fast). Multi-gigahertz analog CMOS circuits are now in production.

## 1.4 Why This Book?

The design of analog circuits itself has evolved together with the technology and the performance requirements. As the device dimensions shrink, the supply voltage of integrated circuits drops, and analog and digital circuits are fabricated on one chip, many design issues arise that were unimportant only a decade ago. Such trends demand that the analysis and design of circuits be accompanied by an in-depth understanding of their advantages and disadvantages with respect to new technology-imposed limitations.

Good analog design requires intuition, rigor, and creativity. As analog designers, we must wear our engineer's hat for a quick and intuitive understanding of a large circuit, our mathematician's hat for quantifying subtle, yet important effects in a circuit, and our artist's hat for inventing new circuit topologies.

This book describes modern analog design from both intuitive and rigorous angles. It also fosters the reader's creativity by carefully guiding him/her through the evolution of each circuit and presenting the thought process that occurs during the development of new circuit techniques.

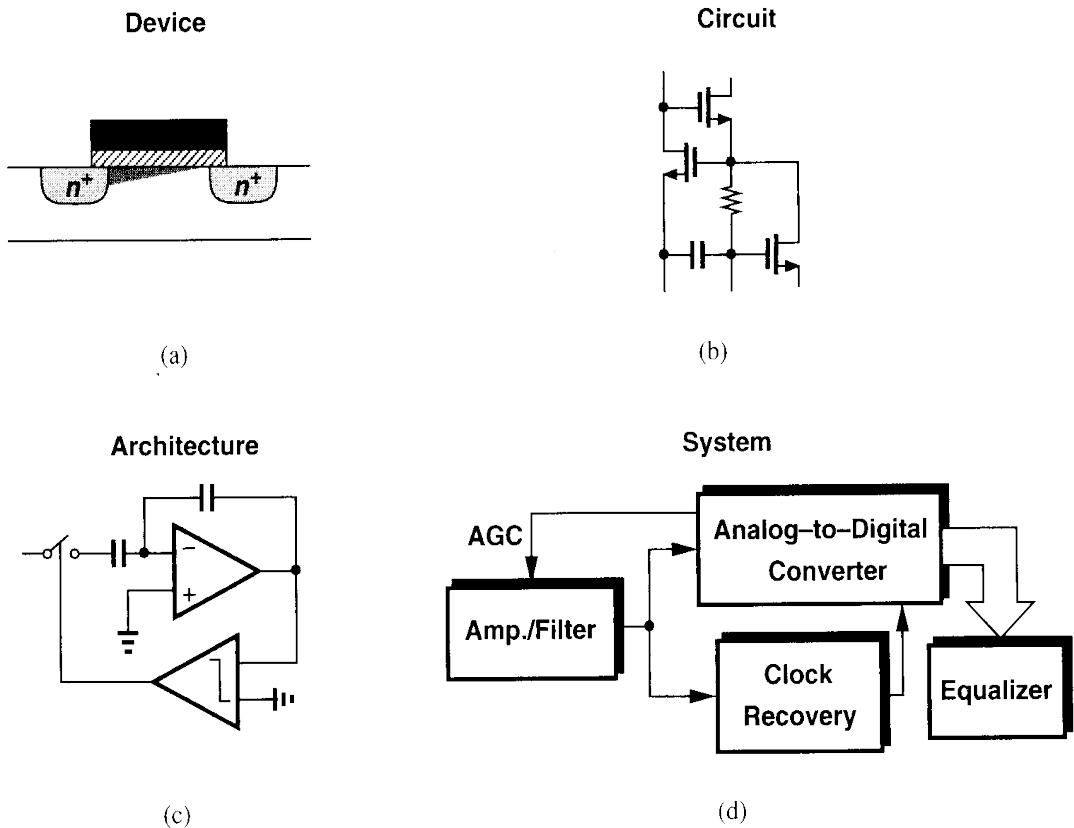
## 1.5 General Concepts

### 1.5.1 Levels of Abstraction

Analysis and design of integrated circuits often require thinking at various levels of abstraction. Depending on the effect or quantity of interest, we may study a complex circuit at device physics level, transistor level, architecture level, or system level. In other words, we may consider the behavior of individual devices in terms of their internal electric fields and charge transport [Fig. 1.8(a)], the interaction of a group of devices according to their electrical characteristics [Fig. 1.8(b)], the function of several building blocks operating as a unit [Fig. 1.8(c)], or the performance of the system in terms of that of its constituent subsystems [Fig. 1.8(d)]. Switching between levels of abstraction becomes necessary in both understanding the details of the operation and optimizing the overall performance. In fact, in today's IC industry, the interaction between all groups, from device physicists to system designers, is essential to achieving a high performance and a low cost. In this book, we begin with device physics and develop increasingly more complex circuit topologies.

### 1.5.2 Robust Analog Design

Many device and circuit parameters vary with the fabrication process, supply voltage, and ambient temperature. We denote these effects by PVT and design circuits such that their performance remains in an acceptable range for a specified range of PVT variations. For example, the supply voltage may vary from 2.7 V to 3.3 V and the temperature from 0° to 70°. Robust analog design in CMOS technology is a challenging task because device parameters vary significantly from wafer to wafer.



**Figure 1.8** Abstraction levels in circuit design: (a) device level, (b) circuit level, (c) architecture level, (d) system level.

### 1.5.3 Notations

The voltages and currents in integrated circuits typically contain a bias component and a signal component. While it is desirable to employ a notation that distinguishes between these quantities, in practice other difficulties arise. For example, if the drain bias current of a transistor is denoted by  $I_D$  and the drain signal current by  $i_D$ , then the Laplace transform of  $i_D$ ,  $I_D(s)$ , may be confused with  $I_D$  unless it is always accompanied by  $s$ . Furthermore, it is confusing to write the low-frequency gain of a circuit as  $v_{out}/v_{in} = -g_m R_D$  and the high-frequency gain as  $V_{out}/V_{in} = -g_m R_D/(1 + R_D C_{LS})$ .

In this book, we denote most voltages and currents by uppercase letters, making it clear from the context which component they represent. For example,  $I_D$ ,  $V_{GS}$ , and  $V_X$  denote bias, signal, or bias+signal quantities. For input and output voltages, we use  $V_{in}$  and  $V_{out}$ , respectively.

## *Basic MOS Device Physics*

---

In studying the design of integrated circuits, one of two extreme approaches can be taken: (1) begin with quantum mechanics and understand solid-state physics, semiconductor device physics, device modeling, and finally the design of circuits; (2) treat each semiconductor device as a black box whose behavior is described in terms of its terminal voltages and currents and design circuits with little attention to the internal operation of the device. Experience shows that neither approach is optimum. In the first case, the reader cannot see the relevance of all of the physics to designing circuits, and in the second, he/she is constantly mystified by the contents of the black box.

In today's IC industry, a solid understanding of semiconductor devices is essential, more so in analog design than in digital design because in the former, transistors are not considered as simple switches and many of their second-order effects directly impact the performance. Furthermore, as each new generation of IC technologies scales the devices, these effects become more significant. Since the designer must often decide which effects can be neglected in a given circuit, insight into device operation proves invaluable.

In this chapter, we study the physics of MOSFETs at an elementary level, covering the bare minimum that is necessary for basic analog design. The ultimate goal is still to develop a circuit model for each device by formulating its operation, but this is accomplished with a good understanding of the underlying principles. After studying many analog circuits in Chapters 3 through 13 and gaining motivation for a deeper understanding of devices, we return to the subject in Chapter 16 and deal with other aspects of MOS operation.

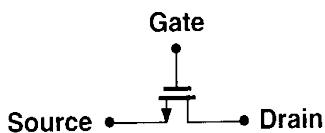
We begin our study with the structure of MOS transistors and derive their I/V characteristics. Next, we describe second-order effects such as body effect, channel-length modulation, and subthreshold conduction. We then identify the parasitic capacitances of MOSFETs, derive a small-signal model, and present a simple SPICE model. We assume that the reader is familiar with such basic concepts as doping, mobility, and *pn* junctions.

## 2.1 General Considerations

### 2.1.1 MOSFET as a Switch

Before delving into the actual operation of the MOSFET, we consider a simplistic model of the device so as to gain a feeling for what the transistor is expected to be and which aspects of its behavior are important.

Shown in Fig. 2.1 is the symbol for an *n*-type MOSFET, revealing three terminals: gate (G), source (S), and drain (D). The latter two are interchangeable because the device is



**Figure 2.1** Simple view of a MOS device.

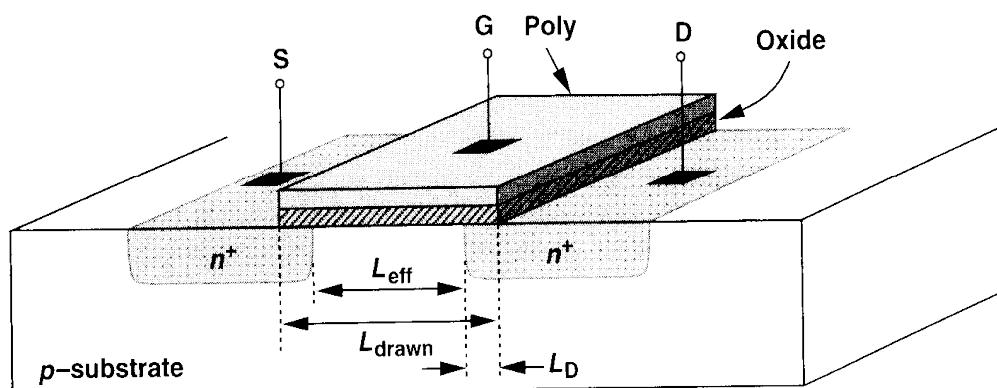
symmetric. When operating as a switch, the transistor “connects” the source and the drain together if the gate voltage,  $V_G$ , is “high” and isolates the source and the drain if  $V_G$  is “low.”

Even with this simplified view, we must answer several questions. For what value of  $V_G$  does the device turn on? In other words, what is the “threshold” voltage? What is the resistance between S and D when the device is on (or off)? How does this resistance depend on the terminal voltages? Can we always model the path between S and D by a simple linear resistor? What limits the speed of the device?

While all of these questions arise at the circuit level, they can be answered only by analyzing the structure and physics of the transistor.

### 2.1.2 MOSFET Structure

Fig. 2.2 shows a simplified structure of an *n*-type MOS (NMOS) device. Fabricated on a *p*-type substrate (also called the “bulk” or the “body”), the device consists of two heavily-doped *n* regions forming the source and drain terminals, a heavily-doped (conductive) piece



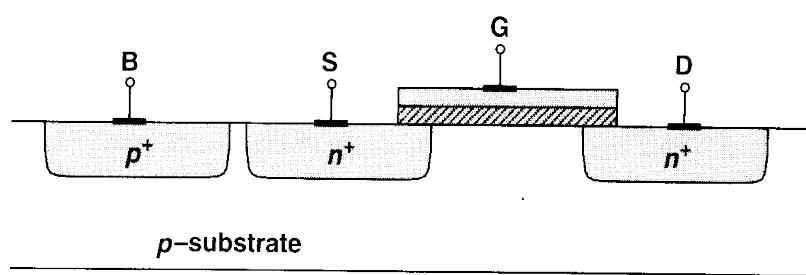
**Figure 2.2** Structure of a MOS device.

of polysilicon<sup>1</sup> (often simply called “poly”) operating as the gate, and a thin layer of silicon dioxide ( $\text{SiO}_2$ ) insulating the gate from the substrate. The useful action of the device occurs in the substrate region under the gate oxide. Note that the structure is symmetric with respect to S and D.

The dimension of the gate along the source-drain path is called the length,  $L$ , and that perpendicular to the length is called the width,  $W$ . Since during fabrication the S/D junctions “side-diffuse,” the actual distance between the source and the drain is slightly less than  $L$ . To avoid confusion, we write,  $L_{\text{eff}} = L_{\text{drawn}} - 2L_D$ , where  $L_{\text{eff}}$  is the “effective” length,  $L_{\text{drawn}}$  is the total length,<sup>2</sup> and  $L_D$  is the amount of side diffusion. As we will see later,  $L_{\text{eff}}$  and the gate oxide thickness,  $t_{\text{ox}}$ , play an important role in the performance of MOS circuits. Consequently, the principal thrust in MOS technology development is to reduce both of these dimensions from one generation to the next without degrading other parameters of the device. Typical values at the time of this writing are  $L_{\text{eff}} \approx 0.15 \mu\text{m}$  and  $t_{\text{ox}} \approx 50 \text{ \AA}$ . In the remainder of this book, we denote the effective length by  $L$ .

If the MOS structure is symmetric, why do we call one  $n$  region the source and the other the drain? This becomes clear if the source is defined as the terminal that provides the charge carriers (electrons in the case of NMOS devices) and the drain as the terminal that collects them. Thus, as the voltages at the three terminals of the device vary, the source and the drain may exchange roles. These concepts are practiced in the problems at the end of the chapter.

We have thus far ignored the substrate on which the device is fabricated. In reality, the substrate potential greatly influences the device characteristics. That is, the MOSFET is a *four-terminal* device. Since in typical MOS operation the S/D junction diodes must be reverse-biased, we assume the substrate of NMOS transistors is connected to the most negative supply in the system. For example, if a circuit operates between zero and 3 volts,  $V_{\text{sub,NMOS}} = 0$ . The actual connection is usually provided through an ohmic  $p^+$  region, as depicted in the side view of the device in Fig. 2.3.

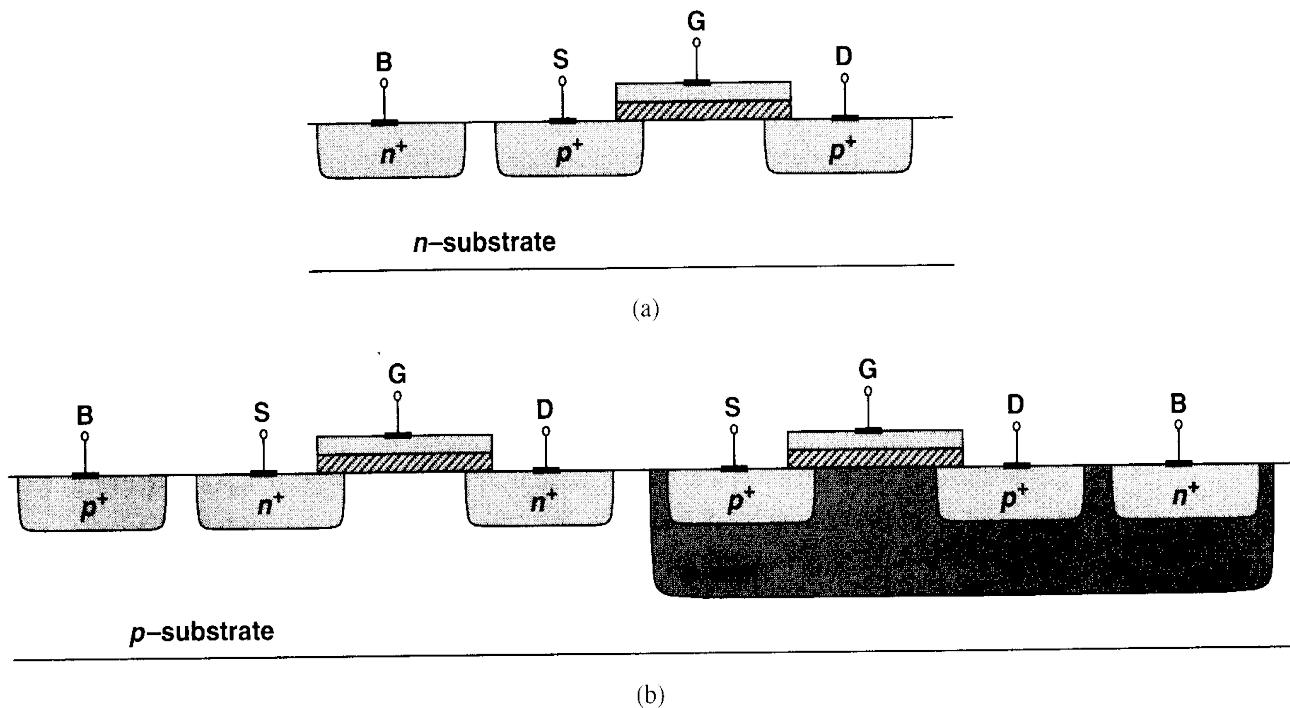


**Figure 2.3** Substrate connection.

In complementary MOS (CMOS) technologies, both NMOS and PMOS transistors are available. From a simplistic view point, the PMOS device is obtained by negating all of

<sup>1</sup>Polysilicon is silicon in amorphous (non-crystal) form. As explained in Chapter 17, when the gate silicon is grown on top of the oxide, it cannot form a crystal.

<sup>2</sup>The subscript “drawn” is used because this is the dimension that we draw in the layout of the transistor (Section 2.4.1).



**Figure 2.4** (a) Simple PMOS device, (b) PMOS inside an *n*-well.

the doping types (including the substrate) [Fig. 2.4(a)], but in practice, NMOS and PMOS devices must be fabricated on the same wafer, i.e., the same substrate. For this reason, one device type can be placed in a “local substrate,” usually called a “well.” In most of today’s CMOS processes, the PMOS device is fabricated in an *n*-well [Fig. 2.4(b)]. Note that the *n*-well must be connected to a potential such that the S/D junction diodes of the PMOS transistor remain reverse-biased under all conditions. In most circuits, the *n*-well is tied to the most positive supply voltage. For the sake of brevity, we sometimes call NMOS and PMOS devices “NFETs” and “PFETs,” respectively.

Fig. 2.4(b) indicates an interesting difference between NMOS and PMOS transistors: while all NFETs share the same substrate, each PFET can have an independent *n*-well. This flexibility of PFETs is exploited in some analog circuits.

### 2.1.3 MOS Symbols

The circuit symbols used to represent NMOS and PMOS transistors are shown in Fig. 2.5. The symbols in Fig. 2.5(a) contain all four terminals, with the substrate denoted by “B” (bulk) rather than “S” to avoid confusion with the source. The source of the PMOS device is positioned on top as a visual aid because it has a higher potential than its gate. Since in most circuits the bulk terminals of NMOS and PMOS devices are tied to ground and  $V_{DD}$  respectively, we usually omit these connections in drawing [Fig. 2.5(b)]. In digital circuits it is customary to use the “switch” symbols depicted in Fig. 2.5(c) for the two types, but we prefer those in Fig. 2.5(b) because the visual distinction between S and D proves helpful in understanding the operation of circuits.

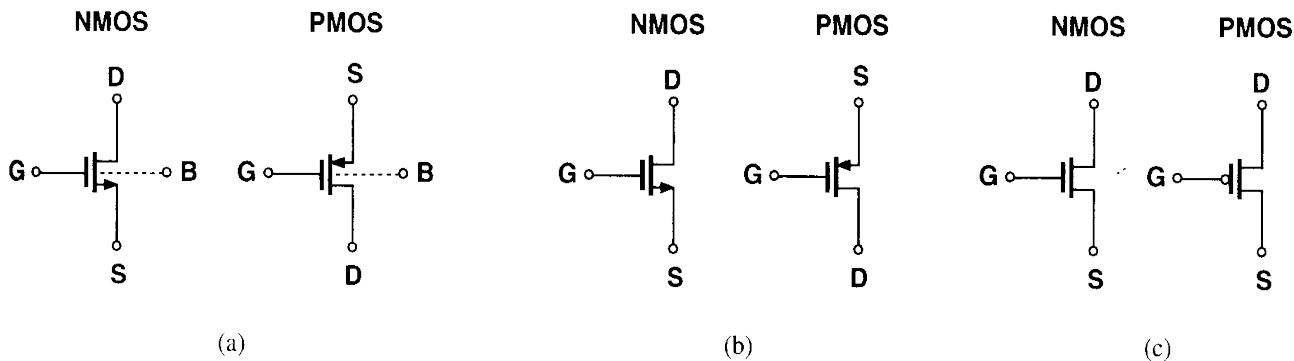


Figure 2.5 MOS symbols.

## 2.2 MOS I/V Characteristics

In this section, we analyze the generation and transport of charge in MOSFETs as a function of the terminal voltages. Our objective is to derive equations for the I/V characteristics such that we can elevate our abstraction from device physics level to circuit level.

### 2.2.1 Threshold Voltage

Consider an NFET connected to external voltages as shown in Fig. 2.6(a). What happens as the gate voltage,  $V_G$ , increases from zero? Since the gate and the substrate form a capacitor,

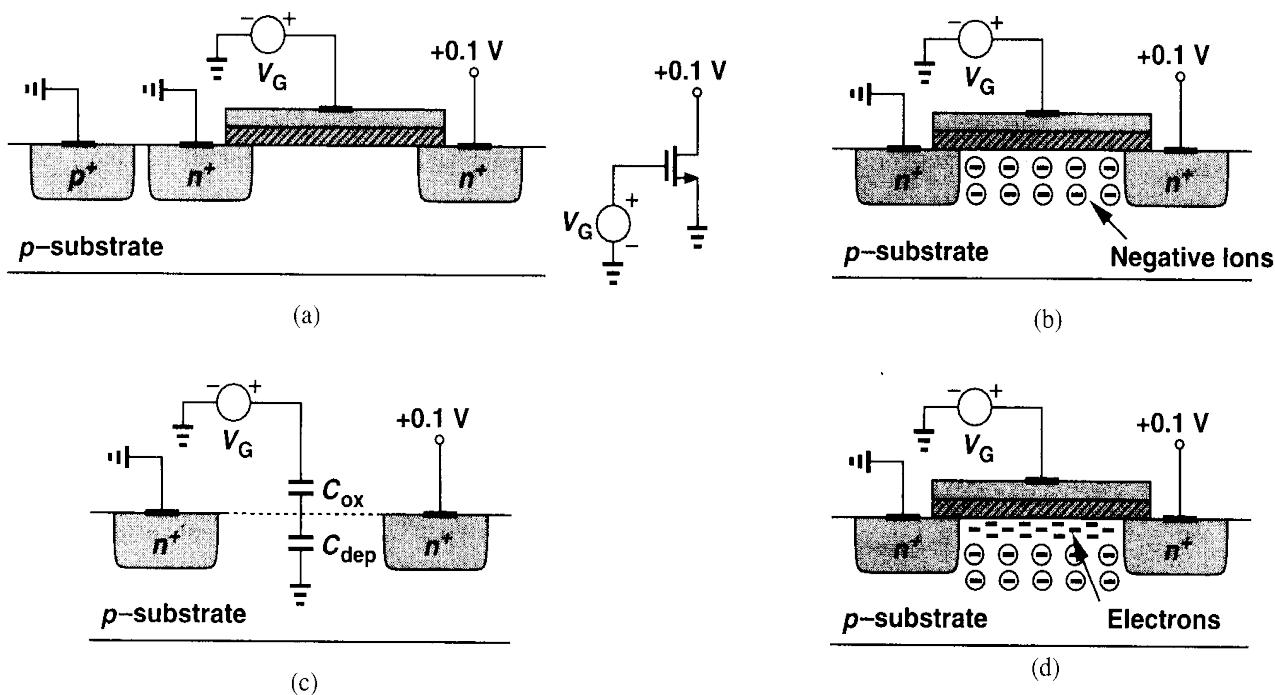


Figure 2.6 (a) A MOSFET driven by a gate voltage, (b) formation of depletion region, (c) onset of inversion, (d) formation of inversion layer.

as  $V_G$  becomes more positive, the holes in the  $p$ -substrate are repelled from the gate area, leaving negative ions behind so as to mirror the charge on the gate. In other words, a depletion region is created [Fig. 2.6(b)]. Under this condition, no current flows because no charge carriers are available.

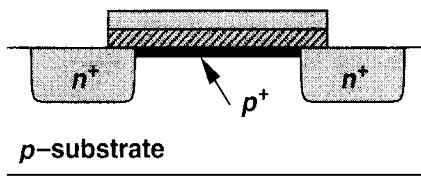
As  $V_G$  increases, so do the width of the depletion region and the potential at the oxide-silicon interface. In a sense, the structure resembles two capacitors in series: the gate oxide capacitor and the depletion region capacitor [Fig. 2.6(c)]. When the interface potential reaches a sufficiently positive value, electrons flow from the source to the interface and eventually to the drain. Thus, a “channel” of charge carriers is formed under the gate oxide between S and D, and the transistor is “turned on.” We also say the interface is “inverted.” The value of  $V_G$  for which this occurs is called the “threshold voltage,”  $V_{TH}$ . If  $V_G$  rises further, the charge in the depletion region remains relatively constant while the channel charge density continues to increase, providing a greater current from S to D.

In reality, the turn-on phenomenon is a gradual function of the gate voltage, making it difficult to define  $V_{TH}$  unambiguously. In semiconductor physics, the  $V_{TH}$  of an NFET is usually defined as the gate voltage for which the interface is “as much  $n$ -type as the substrate is  $p$ -type.” It can be proved [1] that<sup>3</sup>

$$V_{TH} = \Phi_{MS} + 2\Phi_F + \frac{Q_{dep}}{C_{ox}}, \quad (2.1)$$

where  $\Phi_{MS}$  is the difference between the work functions of the polysilicon gate and the silicon substrate,  $\Phi_F = (kT/q) \ln(N_{sub}/n_i)$ ,  $q$  is electron charge,  $N_{sub}$  is the doping concentration of the substrate,  $Q_{dep}$  is the charge in the depletion region, and  $C_{ox}$  is the gate oxide capacitance per unit area. From  $pn$  junction theory,  $Q_{dep} = \sqrt{4q\epsilon_{si}|\Phi_F|N_{sub}}$ , where  $\epsilon_{si}$  denotes the dielectric constant of silicon. Since  $C_{ox}$  appears very frequently in device and circuit calculations, it is helpful to remember that for  $t_{ox} \approx 50 \text{ \AA}$ ,  $C_{ox} \approx 6.9 \text{ fF}/\mu\text{m}^2$ . The value of  $C_{ox}$  can then be scaled proportionally for other oxide thicknesses.

In practice, the “native” threshold value obtained from the above equation may not be suited to circuit design, e.g.,  $V_{TH} = 0$  and the device does not turn off for  $V_G \geq 0$ . For this reason, the threshold voltage is typically adjusted by implantation of dopants into the channel area during device fabrication, in essence altering the doping level of the substrate near the oxide interface. For example, as shown in Fig. 2.7, if a thin sheet of  $p^+$  is created, the gate voltage required to deplete this region increases.



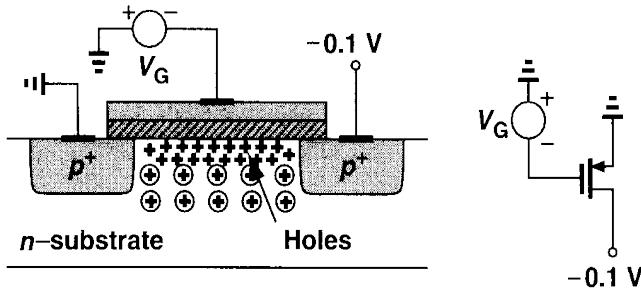
**Figure 2.7** Implantation of  $p^+$  dopants to alter the threshold.

The above definition is not directly applicable to the *measurement* of  $V_{TH}$ . In Fig. 2.6(a), only the drain current can indicate whether the device is “on” or “off,” thus failing to reveal at what  $V_{GS}$  the interface is as much  $n$ -type as the bulk is  $p$ -type. As a result, the calculation

<sup>3</sup>Charge trapping in the oxide is neglected here.

of  $V_{TH}$  from I/V measurements is somewhat ambiguous. We return to this point later but assume in our preliminary analysis that the device turns on *abruptly* for  $V_{GS} \geq V_{TH}$ .

The turn-on phenomenon in a PMOS device is similar to that of NFETs but with all of the polarities reversed. As shown in Fig. 2.8, if the gate-source voltage becomes sufficiently



**Figure 2.8** Formation of inversion layer in a PFET.

*negative*, an inversion layer consisting of holes is formed at the oxide-silicon interface, providing a conduction path between the source and the drain.

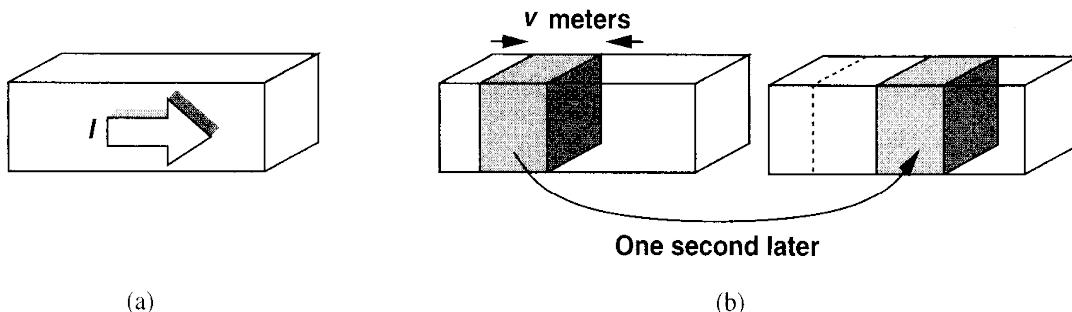
## 2.2.2 Derivation of I/V Characteristics

In order to obtain the relationship between the drain current of a MOSFET and its terminal voltages, we make two observations.

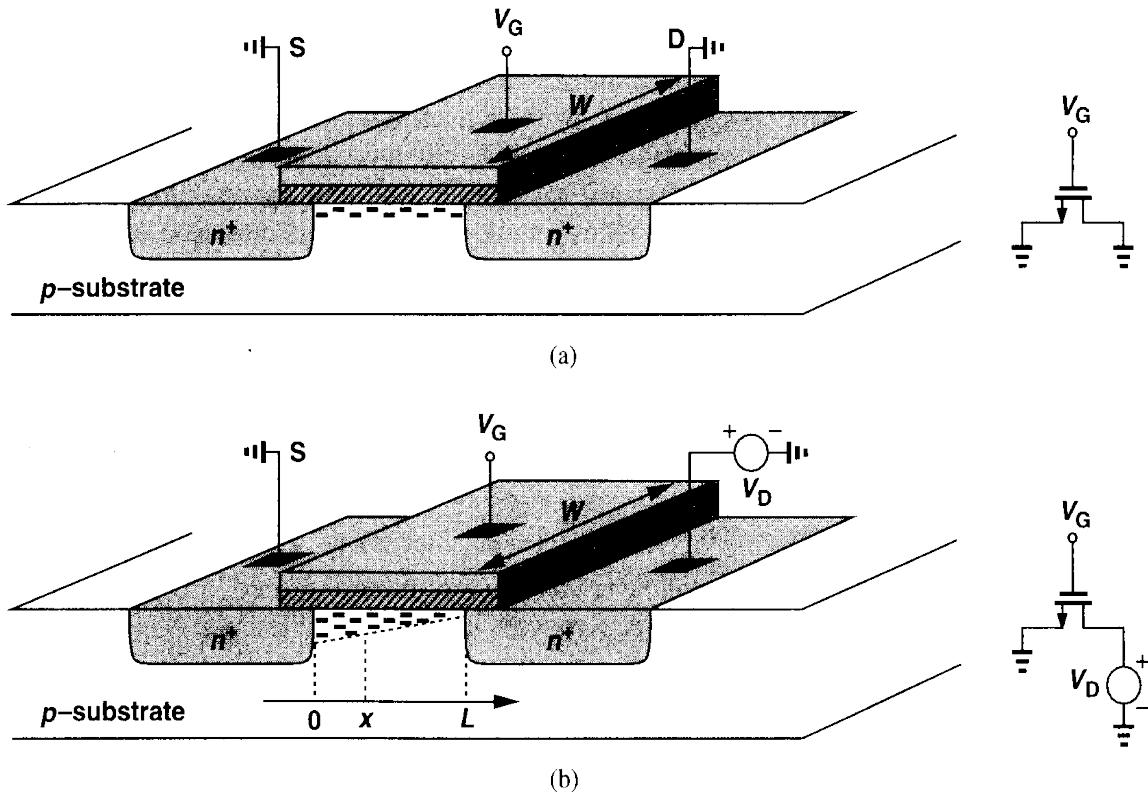
First, consider a semiconductor bar carrying a current  $I$  [Fig. 2.9(a)]. If the charge density along the direction of current is  $Q_d$  coulombs per meter and the velocity of the charge is  $v$  meters per second, then

$$I = Q_d \cdot v. \quad (2.2)$$

To understand why, we measure the total charge that passes through a cross section of the bar in unit time. With a velocity  $v$ , all of the charge enclosed in  $v$  meters of the bar must flow through the cross section in one second [Fig. 2.9(b)]. Since the charge density is  $Q_d$ , the total charge in  $v$  meters equals  $Q_d \cdot v$ . This lemma proves useful in analyzing semiconductor devices.



**Figure 2.9** (a) A semiconductor bar carrying a current  $I$ , (b) snapshots of the carriers one second apart.



**Figure 2.10** Channel charge with (a) equal source and drain voltages, (b) unequal source and drain voltages.

Second, consider an NFET whose source and drain are connected to ground [Fig. 2.10(a)]. What is the charge density in the inversion layer? Since we assume the onset of inversion occurs at  $V_{GS} = V_{TH}$ , the inversion charge density produced by the gate oxide capacitance is proportional to  $V_{GS} - V_{TH}$ . For  $V_{GS} \geq V_{TH}$ , any charge placed on the gate must be mirrored by the charge in the channel, yielding a uniform channel charge density (charge per unit length) equal to

$$Q_d = WC_{ox}(V_{GS} - V_{TH}), \quad (2.3)$$

where  $C_{ox}$  is multiplied by  $W$  to represent the total capacitance per unit length.

Now suppose, as depicted in Fig. 2.10(b), the drain voltage is greater than zero. Since the channel potential varies from zero at the source to  $V_D$  at the drain, the local voltage difference between the gate and the channel varies from  $V_G$  to  $V_G - V_D$ . Thus, the charge density at a point  $x$  along the channel can be written as

$$Q_d(x) = WC_{ox}[V_{GS} - V(x) - V_{TH}], \quad (2.4)$$

where  $V(x)$  is the channel potential at  $x$ .

From (2.2), the current is given by

$$I_D = -WC_{ox}[V_{GS} - V(x) - V_{TH}]v, \quad (2.5)$$

where the negative sign is inserted because the charge carriers are negative and  $v$  denotes the velocity of the electrons in the channel. For semiconductors,  $v = \mu E$ , where  $\mu$  is the mobility of charge carriers and  $E$  is the electric field. Noting that  $E(x) = -dV/dx$  and representing the mobility of electrons by  $\mu_n$ , we have

$$I_D = WC_{ox}[V_{GS} - V(x) - V_{TH}]\mu_n \frac{dV(x)}{dx}, \quad (2.6)$$

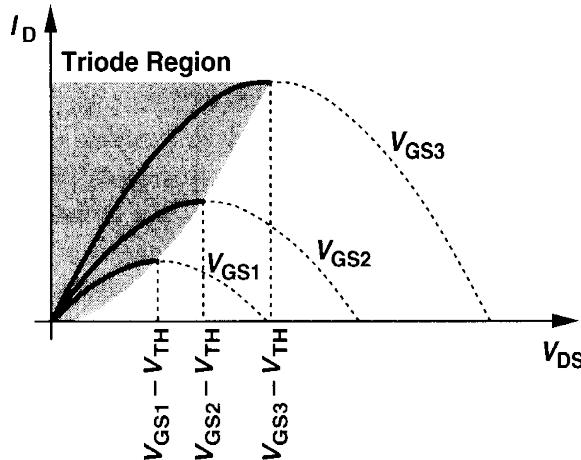
subject to boundary conditions  $V(0) = 0$  and  $V(L) = V_{DS}$ . While  $V(x)$  can be easily found from this equation, the quantity of interest is in fact  $I_D$ . Multiplying both sides by  $dV$  and performing integration, we obtain

$$\int_{x=0}^L I_D dx = \int_{V=0}^{V_{DS}} WC_{ox}\mu_n[V_{GS} - V(x) - V_{TH}]dV. \quad (2.7)$$

Since  $I_D$  is constant along the channel:

$$I_D = \mu_n C_{ox} \frac{W}{L} \left[ (V_{GS} - V_{TH})V_{DS} - \frac{1}{2}V_{DS}^2 \right]. \quad (2.8)$$

Note that  $L$  is the effective channel length.



**Figure 2.11** Drain current versus drain-source voltage in the triode region.

Fig. 2.11 plots the parabolas given by (2.8) for different values of  $V_{GS}$ , indicating that the “current capability” of the device increases with  $V_{GS}$ . Calculating  $\partial I_D / \partial V_{DS}$ , the reader can show that the peak of each parabola occurs at  $V_{DS} = V_{GS} - V_{TH}$  and the peak current is

$$I_{D,max} = \frac{1}{2}\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2. \quad (2.9)$$

We call  $V_{GS} - V_{TH}$  the “overdrive voltage”<sup>4</sup> and  $W/L$  the “aspect ratio.” If  $V_{DS} \leq V_{GS} - V_{TH}$ , we say the device operates in the “triode region.”<sup>5</sup>

<sup>4</sup>Sometimes called the “effective voltage.”

<sup>5</sup>This is also called the “linear region.”

Equations (2.8) and (2.9) serve as the foundation for analog CMOS design, describing the dependence of  $I_D$  upon the constant of the technology,  $\mu_n C_{ox}$ , the device dimensions,  $W$  and  $L$ , and the gate and drain potentials with respect to the source. Note that the integration in (2.7) assumes  $\mu_n$  and  $V_{TH}$  are independent of  $x$  and the gate and drain voltages, an approximation that we will revisit in Chapter 16.

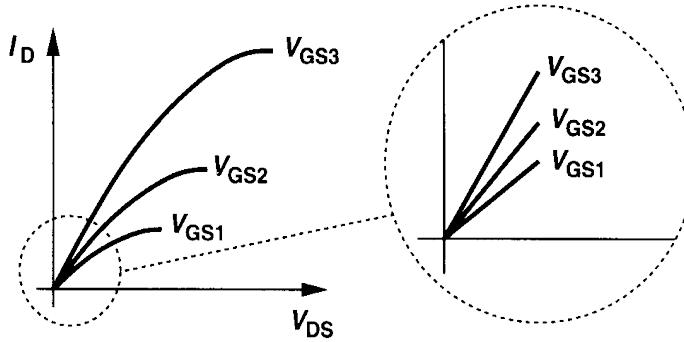
If in (2.8),  $V_{DS} \ll 2(V_{GS} - V_{TH})$ , we have

$$I_D \approx \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH}) V_{DS}, \quad (2.10)$$

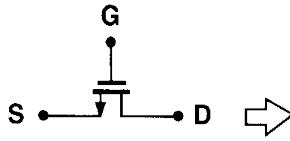
that is, the drain current is a *linear* function of  $V_{DS}$ . This is also evident from the characteristics of Fig. 2.11 for small  $V_{DS}$ : as shown in Fig. 2.12, each parabola can be approximated by a straight line. The linear relationship implies that the path from the source to the drain can be represented by a linear resistor equal to

$$R_{on} = \frac{1}{\mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})}. \quad (2.11)$$

A MOSFET can therefore operate as a resistor whose value is controlled by the overdrive voltage [so long as  $V_{DS} \ll 2(V_{GS} - V_{TH})$ ]. This is conceptually illustrated in Fig. 2.13. Note that in contrast to bipolar transistors, a MOS device may be on even if it carries no



**Figure 2.12** Linear operation in deep triode region.

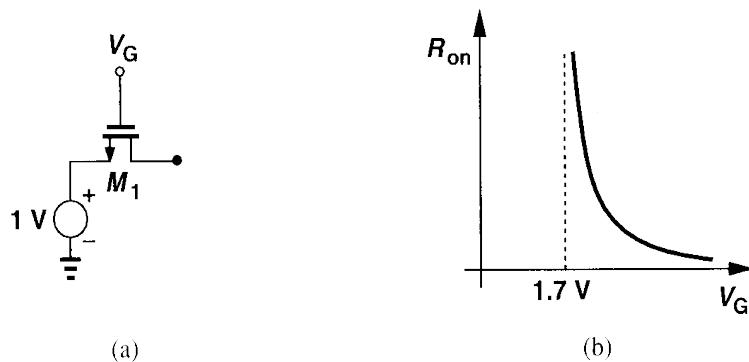


**Figure 2.13** MOSFET as a controlled linear resistor.

current. With the condition  $V_{DS} \ll 2(V_{GS} - V_{TH})$ , we say the device operates in deep triode region.

### Example 2.1

For the arrangement in Fig. 2.14(a), plot the on-resistance of  $M_1$  as a function of  $V_G$ . Assume  $\mu_n C_{ox} = 50 \mu\text{A/V}^2$ ,  $W/L = 10$ , and  $V_{TH} = 0.7 \text{ V}$ . Note that the drain terminal is open.



**Figure 2.14**

### Solution

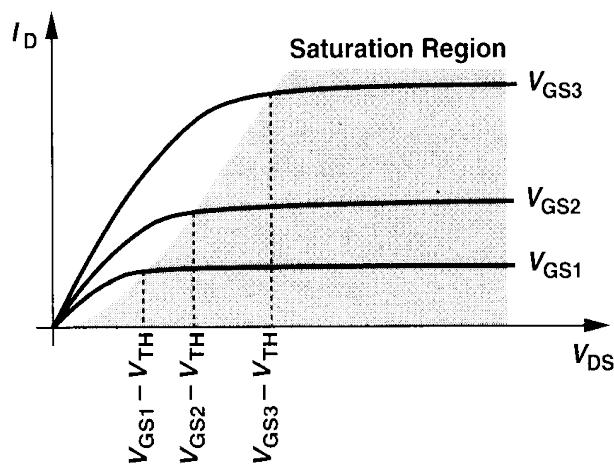
Since the drain terminal is open,  $I_D = 0$  and  $V_{DS} = 0$ . Thus, if the device is on, it operates in the deep triode region. For  $V_G < 1 \text{ V} + V_{TH}$ ,  $M_1$  is off and  $R_{on} = \infty$ . For  $V_G > 1 \text{ V} + V_{TH}$ , we have

$$R_{on} = \frac{1}{50 \mu\text{A}/\text{V}^2 \times 10(V_G - 1 \text{ V} - 0.7 \text{ V})}. \quad (2.12)$$

The result is plotted in Fig. 2.14(b).

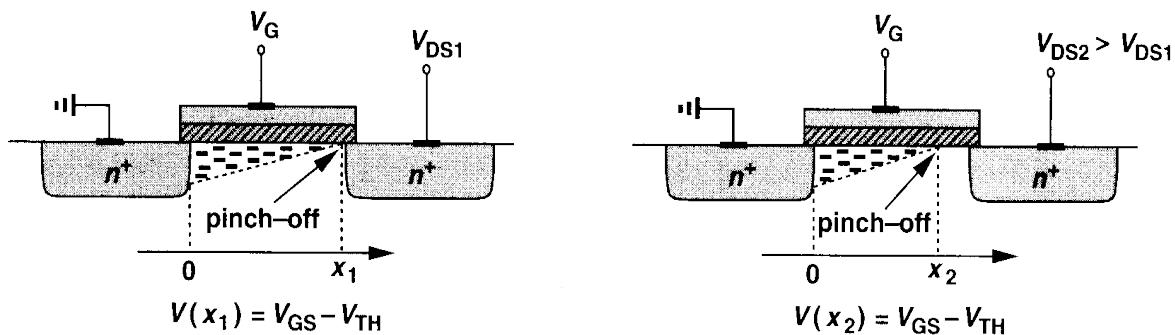
The utility of MOSFETs as controllable resistors and hence switches plays a crucial role in many analog circuits. This is studied in Chapter 12.

What happens if in Fig. 2.11 the drain-source voltage exceeds  $V_{GS} - V_{TH}$ ? In reality, the drain current does *not* follow the parabolic behavior for  $V_{DS} > V_{GS} - V_{TH}$ . In fact, as shown in Fig. 2.15,  $I_D$  becomes relatively constant and we say the device operates in the “saturation” region.<sup>6</sup> To understand this phenomenon, recall from (2.4) that the local



**Figure 2.15** Saturation of drain current.

<sup>6</sup>Note the difference between saturation in bipolar and MOS devices.



**Figure 2.16** Pinch-off behavior.

density of inversion layer charge is proportional to  $V_{GS} - V(x) - V_{TH}$ . Thus, if  $V(x)$  approaches  $V_{GS} - V_{TH}$ , then  $Q_d(x)$  drops to zero. In other words, as depicted in Fig. 2.16, if  $V_{DS}$  is slightly greater than  $V_{GS} - V_{TH}$ , then the inversion layer stops at  $x \leq L$ , and we say the channel is “pinched off.” As  $V_{DS}$  increases further, the point at which  $Q_d$  equals zero gradually moves toward the source. Thus, at some point along the channel, the local potential difference between the gate and the oxide-silicon interface is not sufficient to support an inversion layer.

With the above observations, we re-examine (2.7) for a saturated device. Since  $Q_d$  is the density of *mobile* charge, the integral on the left-hand side of (2.7) must be taken from  $x = 0$  to  $x = L'$ , where  $L'$  is the point at which  $Q_d$  drops to zero, and that on the right from  $V(x) = 0$  to  $V(x) = V_{GS} - V_{TH}$ . As a result:

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L'} (V_{GS} - V_{TH})^2, \quad (2.13)$$

indicating that  $I_D$  is relatively independent of  $V_{DS}$  if  $L'$  remains close to  $L$ .

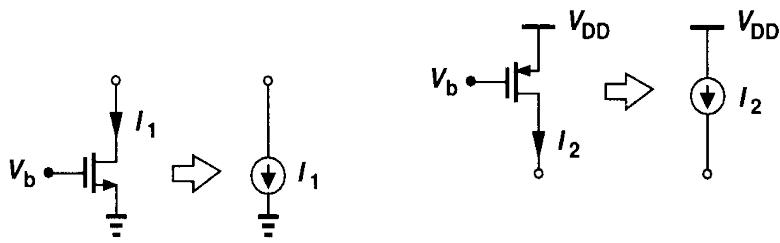
For PMOS devices, Eqs. (2.8) and (2.13) are respectively written as

$$I_D = -\mu_p C_{ox} \frac{W}{L} \left[ (V_{GS} - V_{TH})V_{DS} - \frac{1}{2} V_{DS}^2 \right] \quad (2.14)$$

and

$$I_D = -\frac{1}{2} \mu_p C_{ox} \frac{W}{L'} (V_{GS} - V_{TH})^2. \quad (2.15)$$

The negative sign appears here because we assume  $I_D$  flows from the drain to the source, whereas holes flow in the reverse direction. Since the mobility of holes is about one-half to one-fourth of the mobility of electrons, PMOS devices suffer from lower “current drive” capability.



**Figure 2.17** Saturated MOSFETs operating as current sources.

With the approximation  $L \approx L'$ , a saturated MOSFET can be used as a current source connected between the drain and the source (Fig. 2.17), an important component in analog design. Note that the current sources inject current into ground or draw current from  $V_{DD}$ . In other words, only one terminal of each current source is “floating.”

Since a MOSFET operating in saturation produces a current in response to its gate-source overdrive voltage, we may define a figure of merit that indicates how well a device converts a voltage to a current. More specifically, since in processing signals we deal with the *changes* in voltages and currents, we define the figure of merit as the change in the drain current divided by the change in the gate-source voltage. Called the “transconductance” and denoted by  $g_m$ , this quantity is expressed as:

$$g_m = \left. \frac{\partial I_D}{\partial V_{GS}} \right|_{V_{DS}, \text{const.}} \quad (2.16)$$

$$= \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH}). \quad (2.17)$$

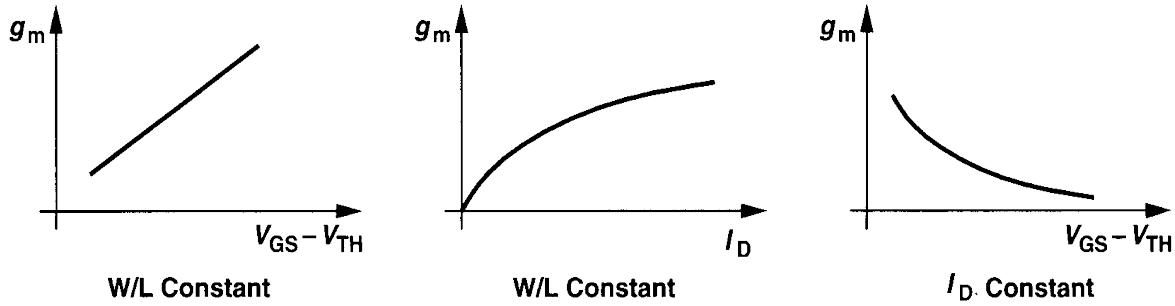
In a sense,  $g_m$  represents the sensitivity of the device: for a high  $g_m$ , a small change in  $V_{GS}$  results in a large change in  $I_D$ . Interestingly,  $g_m$  in the saturation region is equal to the inverse of  $R_{on}$  in deep triode region.

The reader can prove that  $g_m$  can also be expressed as

$$g_m = \sqrt{2\mu_n C_{ox} \frac{W}{L} I_D} \quad (2.18)$$

$$= \frac{2I_D}{V_{GS} - V_{TH}}. \quad (2.19)$$

Plotted in Fig. 2.18, each of the above expressions proves useful in studying the behavior of  $g_m$  as a function of one parameter while other parameters remain constant. For example, (2.17) suggests that  $g_m$  increases with the overdrive if  $W/L$  is constant whereas (2.19) implies that  $g_m$  decreases with the overdrive if  $I_D$  is constant. The concept of transconductance

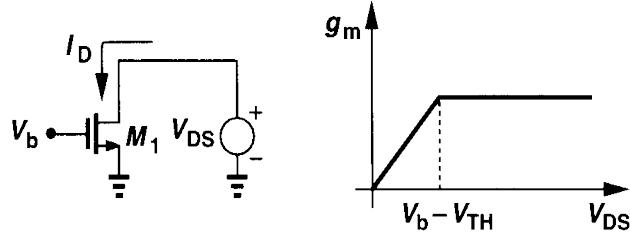


**Figure 2.18** MOS transconductance as a function of overdrive and drain current.

can also be applied to a device operating in the triode region, as illustrated in the following example.

### Example 2.2

For the arrangement shown in Fig. 2.19, plot the transconductance as a function of  $V_{DS}$ .



**Figure 2.19**

### Solution

It is simpler to study  $g_m$  as  $V_{DS}$  decreases from infinity. So long as  $V_{DS} \geq V_b - V_{TH}$ ,  $M_1$  is in saturation,  $I_D$  is relatively constant, and, from (2.18), so is  $g_m$ . For  $V_{DS} < V_b - V_{TH}$ ,  $M_1$  is in the triode region and:

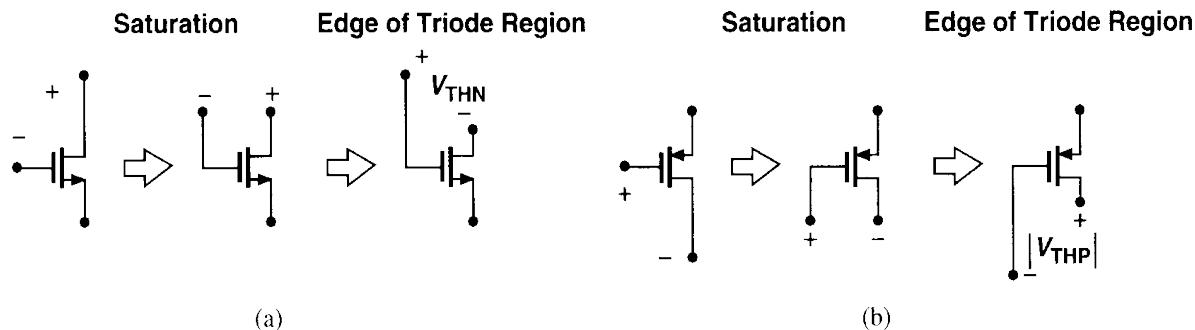
$$g_m = \frac{\partial}{\partial V_{GS}} \left\{ \frac{1}{2} \mu_n C_{ox} \frac{W}{L} [2(V_{GS} - V_{TH})V_{DS} - V_{DS}^2] \right\} \quad (2.20)$$

$$= \mu_n C_{ox} \frac{W}{L} V_{DS}. \quad (2.21)$$

Thus, as plotted in Fig. 2.19, the transconductance drops if the device enters the triode region. For amplification, therefore, we usually employ MOSFETs in saturation.

---

The distinction between saturation and triode regions can be confusing, especially for PMOS devices. Intuitively, we note that the channel is pinched off if the difference between the gate and drain voltages is not sufficient to create an inversion layer. As depicted conceptually in Fig. 2.20, as  $V_G - V_D$  of an NFET drops below  $V_{TH}$ , pinch-off occurs. Similarly,



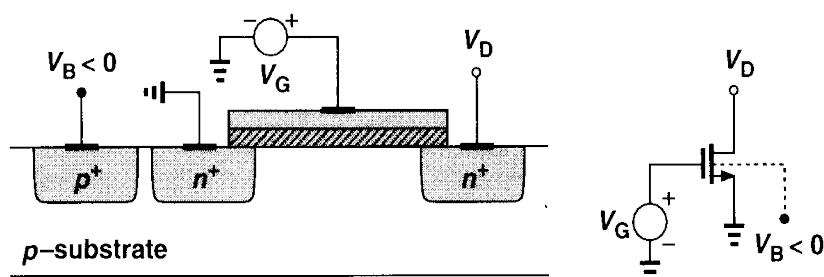
**Figure 2.20** Conceptual visualization of saturation and triode regions.

if  $V_D - V_G$  of a PFET is not large enough ( $< |V_{THP}|$ ), the device is saturated. Note that this view does not require knowledge of the source voltage. This means we must know a priori which terminal operates as the drain.

## 2.3 Second-Order Effects

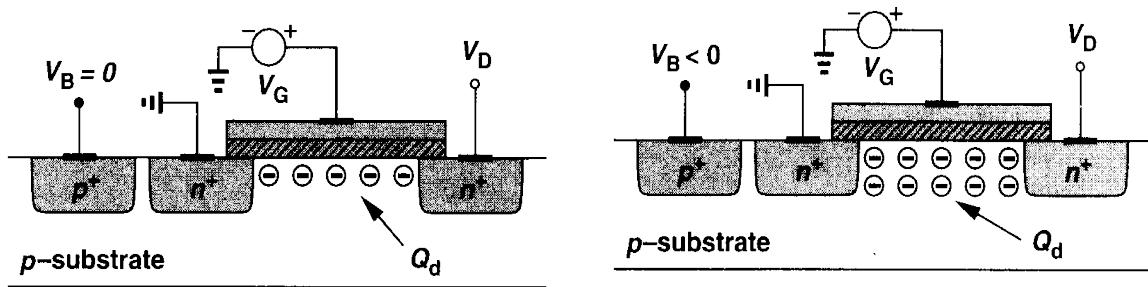
Our analysis of the MOS structure has thus far entailed various simplifying assumptions, some of which are not valid in many analog circuits. In this section, we describe three second-order effects that are essential in our subsequent circuit analyses. Other phenomena that appear in submicron devices are studied in Chapter 16.

**Body Effect** In the analysis of Fig. 2.10, we tacitly assumed that the bulk and the source of the transistor were tied to ground. What happens if the bulk voltage of an NFET drops below the source voltage (Fig. 2.21)? Since the S and D junctions remain reverse-biased, we surmise that the device continues to operate properly but certain characteristics may



**Figure 2.21** NMOS device with negative bulk voltage.

change. To understand the effect, suppose  $V_S = V_D = 0$ , and  $V_G$  is somewhat less than  $V_{TH}$  so that a depletion region is formed under the gate but no inversion layer exists. As  $V_B$  becomes more negative, more holes are attracted to the substrate connection, leaving a larger negative charge behind, i.e., as depicted in Fig. 2.22, the depletion region becomes wider. Now recall from Eq. (2.1) that the threshold voltage is a function of the total charge in the depletion region because the gate charge must mirror  $Q_d$  before an inversion layer is



**Figure 2.22** Variation of depletion region charge with bulk voltage.

formed. Thus, as  $V_B$  drops and  $Q_d$  increases,  $V_{TH}$  also increases. This is called the “body effect” or the “backgate effect.”

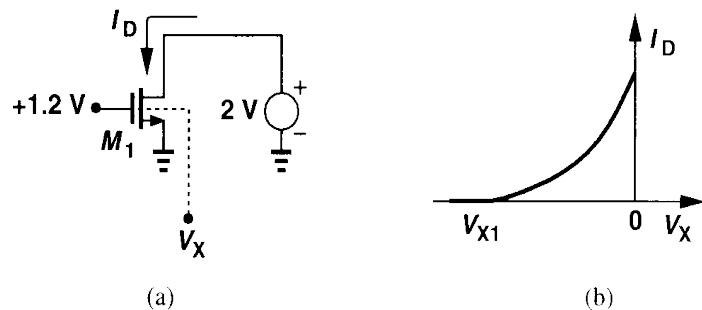
It can be proved that with body effect:

$$V_{TH} = V_{TH0} + \gamma \left( \sqrt{|2\Phi_F + V_{SB}|} - \sqrt{|2\Phi_F|} \right), \quad (2.22)$$

where  $V_{TH0}$  is given by (2.1),  $\gamma = \sqrt{2q\epsilon_{si}N_{sub}/C_{ox}}$  denotes the body effect coefficient, and  $V_{SB}$  is the source-bulk potential difference [1]. The value of  $\gamma$  typically lies in the range of 0.3 to 0.4 V<sup>1/2</sup>.

### Example 2.3

In Fig. 2.23(a), plot the drain current if  $V_X$  varies from  $-\infty$  to 0. Assume  $V_{TH0} = 0.6$  V,  $\gamma = 0.4$  V<sup>1/2</sup>, and  $2\Phi_F = 0.7$  V.



**Figure 2.23**

### Solution

If  $V_X$  is sufficiently negative, the threshold voltage of  $M_1$  exceeds 1.2 V and the device is off. That is,

$$1.2 \text{ V} = 0.6 + 0.4 \left( \sqrt{0.7 - V_{X1}} - \sqrt{0.7} \right), \quad (2.23)$$

and hence  $V_{X1} = -4.76$  V. For  $V_{X1} < V_X < 0$ ,  $I_D$  increases according to

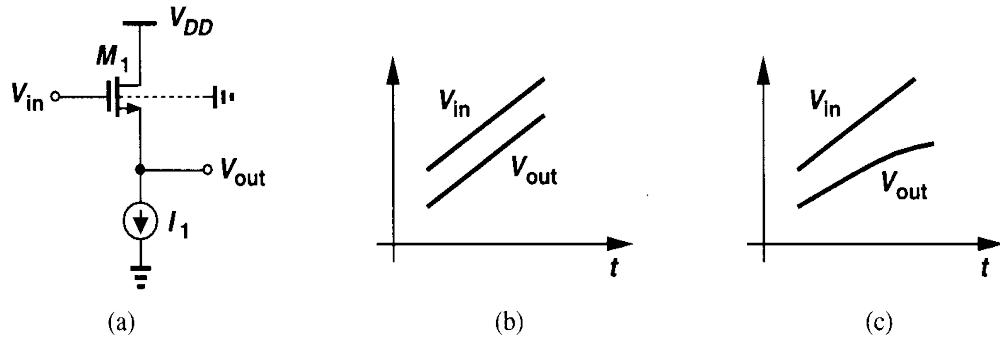
$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} \left[ V_{GS} - V_{TH0} - \gamma \left( \sqrt{2\Phi_F - V_X} - \sqrt{2\Phi_F} \right) \right]^2. \quad (2.24)$$

Fig. 2.23(b) shows the resulting behavior.

For body effect to manifest itself, the bulk potential,  $V_{sub}$ , need not change: if the source voltage varies with respect to  $V_{sub}$ , the same phenomenon occurs. For example, consider the circuit in Fig. 2.24(a), first ignoring body effect. We note that as  $V_{in}$  varies,  $V_{out}$  closely follows the input because the drain current remains equal to  $I_1$ . In fact, we can write

$$I_1 = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{in} - V_{out} - V_{TH})^2, \quad (2.25)$$

concluding that  $V_{in} - V_{out}$  is constant if  $I_1$  is constant [Fig. 2.24(b)].



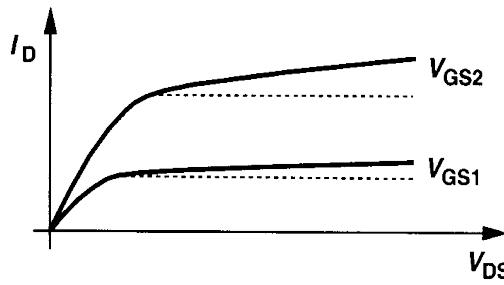
**Figure 2.24** (a) A circuit in which the source-bulk voltage varies with input level, (b) input and output voltages with no body effect, (c) input and output voltages with body effect.

Now suppose the substrate is tied to ground and body effect is significant. Then, as  $V_{in}$  and hence  $V_{out}$  become more positive, the potential difference between the source and the bulk increases, raising the value of  $V_{TH}$ . Eq. (2.25) therefore implies that  $V_{in} - V_{out}$  must increase so as to maintain  $I_D$  constant [Fig. 2.24(c)].

Body effect is usually undesirable. The change in the threshold voltage, e.g., as in Fig. 2.24(a), often complicates the design of analog (and even digital) circuits. Device technologists balance  $N_{sub}$  and  $C_{ox}$  to obtain a reasonable value for  $\gamma$ .

**Channel-Length Modulation** In the analysis of channel pinch-off in Section 2.2, we noted that the actual length of the inverted channel gradually decreases as the potential difference between the gate and the drain increases. In other words, in (2.13),  $L'$  is in fact a function of  $V_{DS}$ . This effect is called “channel-length modulation.” Writing  $L' = L - \Delta L$ , i.e.,  $1/L' \approx (1 + \Delta L/L)/L$ , and assuming a first-order relationship between  $\Delta L/L$  and  $V_{DS}$  such as  $\Delta L/L = \lambda V_{DS}$ , we have, in saturation,

$$I_D \approx \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS}), \quad (2.26)$$



**Figure 2.25** Finite saturation region slope resulting from channel-length modulation.

where  $\lambda$  is the channel-length modulation coefficient. Illustrated in Fig. 2.25, this phenomenon results in a nonzero slope in the  $I_D/V_{DS}$  characteristic and hence a nonideal current source between D and S in saturation. The parameter  $\lambda$  represents the *relative* variation in length for a given increment in  $V_{DS}$ . Thus, for longer channels,  $\lambda$  is smaller.

With channel-length modulation, some of the expressions derived for  $g_m$  must be modified. Equations (2.17) and (2.18) are respectively rewritten as

$$g_m = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})(1 + \lambda V_{DS}). \quad (2.27)$$

$$= \sqrt{\frac{2\mu_n C_{ox} (W/L) I_D}{1 + \lambda V_{DS}}}, \quad (2.28)$$

while Eq. (2.19) remains unchanged.

#### Example 2.4

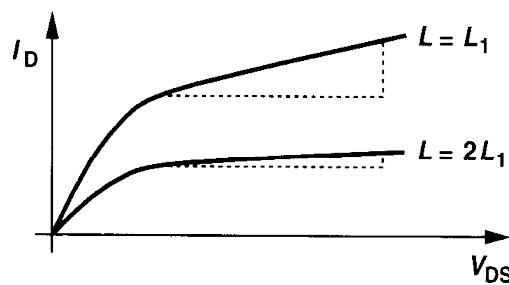
Keeping all other parameters constant, plot  $I_D/V_{DS}$  characteristic of a MOSFET for  $L = L_1$  and  $L = 2L_1$ .

#### Solution

Writing

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS}) \quad (2.29)$$

and  $\lambda \propto 1/L$ , we note that if the length is doubled, the slope of  $I_D$  vs.  $V_{DS}$  is divided by four because  $\partial I_D / \partial V_{DS} \propto \lambda/L \propto 1/L^2$  (Fig. 2.26). For a given gate-source overdrive, a larger  $L$  gives a more



**Figure 2.26** Effect of doubling channel length.

ideal current source while degrading the current capability of the device. Thus,  $W$  may need to be increased proportionally.

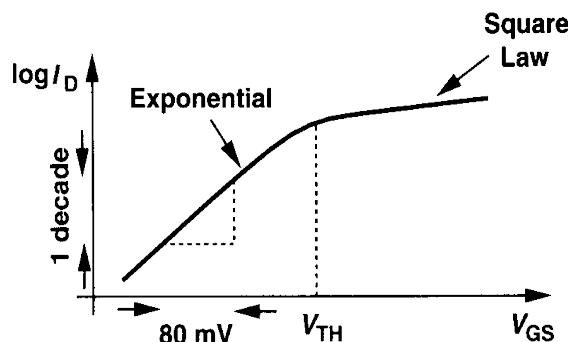
The linear approximation  $\Delta L/L \propto V_{DS}$  becomes less accurate in short-channel transistors, resulting in a *variable* slope in the saturated  $I_D/V_{DS}$  characteristics. We return to this issue in Chapter 16.

The dependence of  $I_D$  upon  $V_{DS}$  in saturation may suggest that the bias current of a MOSFET can be defined by the proper choice of the drain-source voltage, allowing freedom in the choice of  $V_{GS} - V_{TH}$ . However, since the dependence on  $V_{DS}$  is much weaker, the drain-source voltage is not used to set the current. The effect of  $V_{DS}$  on  $I_D$  is usually considered an *error* and it is studied in Chapter 5.

**Subthreshold Conduction** In our analysis of the MOSFET, we have assumed that the device turns off abruptly as  $V_{GS}$  drops below  $V_{TH}$ . In reality, for  $V_{GS} \approx V_{TH}$ , a “weak” inversion layer still exists and some current flows from D to S. Even for  $V_{GS} < V_{TH}$ ,  $I_D$  is finite, but it exhibits an *exponential* dependence on  $V_{GS}$  [2, 3]. Called “subthreshold conduction,” this effect can be formulated for  $V_{DS}$  greater than roughly 200 mV as

$$I_D = I_0 \exp \frac{V_{GS}}{\zeta V_T}, \quad (2.30)$$

where  $\zeta > 1$  is a nonideality factor and  $V_T = kT/q$ . We also say the device operates in “weak inversion.” Except for  $\zeta$ , (2.30) is similar to the exponential  $I_C/V_{BE}$  relationship in a bipolar transistor. The key point here is that as  $V_{GS}$  falls below  $V_{TH}$ , the drain current drops at a finite rate. With typical values of  $\zeta$ , at room temperature  $V_{GS}$  must decrease by approximately 80 mV for  $I_D$  to decrease by one decade (Fig. 2.27). For example, if a



**Figure 2.27** MOS subthreshold characteristics.

threshold of 0.3 V is chosen in a process to allow low-voltage operation, then when  $V_{GS}$  is reduced to zero, the drain current decreases by only a factor of  $10^{3.75}$ . Especially problematic in large circuits such as memories, subthreshold conduction can result in significant power dissipation (or loss of analog information).

It is appropriate at this point to return to the definition of the threshold voltage. One definition is to plot the inverse on-resistance of the device  $R_{on}^{-1} = \mu C_{ox}(W/L)(V_{GS} - V_{TH})$  as a function of  $V_{GS}$  and extrapolate the result to zero, for which  $V_{GS} = V_{TH}$ . In rough calculations, we often view  $V_{TH}$  as the gate-source voltage yielding  $I_D/W = 1\mu A/\mu m$  in saturation. For example, if a device with  $W = 100 \mu m$  operates with  $I_D = 100 \mu A$ , it is in the vicinity of the subthreshold region. This view is nonetheless vague, especially as device length scales down in every technology generation.

We now re-examine Eq. (2.18) for the transconductance of a MOS device operating in the subthreshold region. Is it possible to achieve an arbitrarily high transconductance by increasing  $W$  while maintaining  $I_D$  constant? Is it possible to obtain a *higher* transconductance than that of a bipolar transistor ( $I_C / V_T$ ) biased at the same current? Equation (2.18) was derived from the square-law characteristics  $I_D = (1/2)\mu_n C_{ox}(W/L)(V_{GS} - V_{TH})^2$ . However, if  $W$  increases while  $I_D$  remains constant, then  $V_{GS} \rightarrow V_{TH}$  and the device enters the subthreshold region. As a result, the transconductance is calculated from (2.30) to be  $g_m = I_D/(\zeta V_T)$ , revealing that MOSFETs are inferior to bipolar transistors in this respect.

The exponential dependence of  $I_D$  upon  $V_{GS}$  in subthreshold operation may suggest the use of MOS devices in this regime so as to achieve a higher gain. However, since such conditions are met by only a large device width or low drain current, the speed of subthreshold circuits is severely limited.

**Voltage Limitations** MOSFETs experience various breakdown effects if their terminal voltage differences exceed certain limits. At high gate-source voltages, the gate oxide breaks down irreversibly, damaging the transistor. In short-channel devices, an excessively large drain-source voltage widens the depletion region around the drain so much that it touches that around the source, creating a very large drain current. (This effect is called “punchthrough.”) Other limitations relate to “hot electron effects” and are described in Chapter 16.

## 2.4 MOS Device Models

### 2.4.1 MOS Device Layout

For the developments in subsequent sections, it is beneficial to have some understanding of the layout of a MOSFET. We describe only a simple view here, deferring the fabrication details and structural subtleties to Chapters 17 and 18.

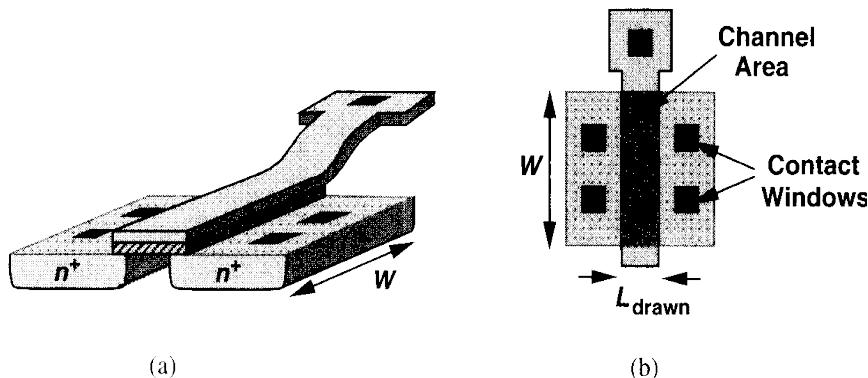
The layout of a MOSFET is determined by both the electrical properties required of the device in the circuit and the “design rules” imposed by the technology. For example,  $W/L$  is chosen to set the transconductance or other circuit parameters, while the minimum  $L$  is dictated by the process. In addition to the gate, the source and drain areas must be defined properly as well.

Shown in Fig. 2.28 are the “bird eye’s view” and the top view of a MOSFET. The gate polysilicon and the source and drain terminals are typically tied to metal (aluminum) wires that serve as interconnects with low resistance and capacitance. To accomplish this, one or more “contact windows” must be opened in each region, filled with metal, and connected to the upper metal wires. Note that the gate poly extends beyond the channel area by some amount to ensure reliable definition of the “edge” of the transistor.

The source and drain junctions play an important role in the performance. To minimize the capacitance of S and D, the total area of each junction must be minimized. We see from Fig. 2.28 that one dimension of the junctions is equal to  $W$ . The other dimension must be large enough to accommodate the contact windows and is specified by the technology design rules.<sup>7</sup>

---

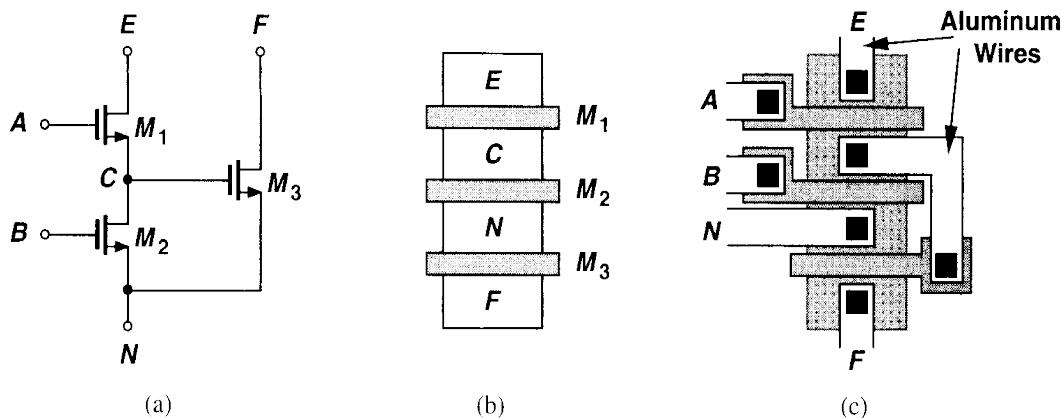
<sup>7</sup>This dimension is typically three to four times the minimum allowable channel length.



**Figure 2.28** Bird's eye and vertical views of a MOS device.

### Example 2.5

Draw the layout of the circuit shown in Fig. 2.29(a).



**Figure 2.29**

### Solution

Noting that  $M_1$  and  $M_2$  share the same S/D junctions at node  $C$  and  $M_2$  and  $M_3$  also do so at node  $N$ , we surmise that the three transistors can be laid out as shown in Fig. 2.29(b). Connecting the remaining terminals, we obtain the layout in Fig. 2.29(c). Note that the gate polysilicon of  $M_3$  cannot be directly tied to the source material of  $M_1$ , thus requiring a metal interconnect.

## 2.4.2 MOS Device Capacitances

The basic quadratic I/V relationships derived in the previous section along with corrections for body effect and channel-length modulation provide a reasonable model for understanding the “dc” behavior of CMOS circuits. In many analog circuits, however, the capacitances associated with the devices must also be taken into account so as to predict the “ac” behavior as well.

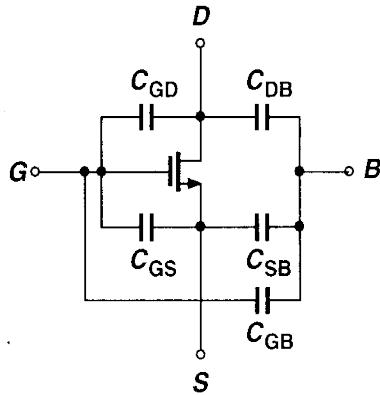


Figure 2.30 MOS capacitances.

We expect that a capacitance exists between every two of the four terminals of a MOSFET (Fig. 2.30).<sup>8</sup> Moreover, the value of each of these capacitances may depend on the bias conditions of the transistor. Considering the physical structure in Fig. 2.31(a), we identify the following. (1) Oxide capacitance between the gate and the channel,  $C_1 = WL C_{ox}$ ; (2) Depletion capacitance between the channel and the substrate,  $C_2 = WL \sqrt{q\epsilon_{si} N_{sub}/(4\Phi_B)}$ ; (3) Capacitance due to the overlap of the gate poly with the source and drain areas,  $C_3$  and  $C_4$ . Owing to fringing electric field lines,  $C_3$  and  $C_4$  cannot be simply written as  $WL_D C_{ox}$ , and are usually obtained by more elaborate calculations. The overlap capacitance per unit width is denoted by  $C_{ov}$ ; (4) Junction capacitance between the source/drain areas and the substrate. As shown in Fig. 2.31(b), this capacitance is usually decomposed into two components: bottom-plate capacitance associated with the bottom of the junction,  $C_j$ , and sidewall capacitance due to the perimeter of the junction,  $C_{jsw}$ . The distinction is necessary because different transistor geometries yield different area and perimeter values for the S/D junctions. We typically specify  $C_j$  and  $C_{jsw}$  as capacitance per unit area and unit length, respectively. Note that each junction capacitance can be expressed as  $C_j = C_{j0}/[1 + V_R/\Phi_B]^m$ , where  $V_R$  is the reverse voltage across the junction,  $\Phi_B$  is the junction built-in potential, and  $m$  is a power typically in the range of 0.3 and 0.4.

<sup>8</sup>The capacitance between S and D is negligible.

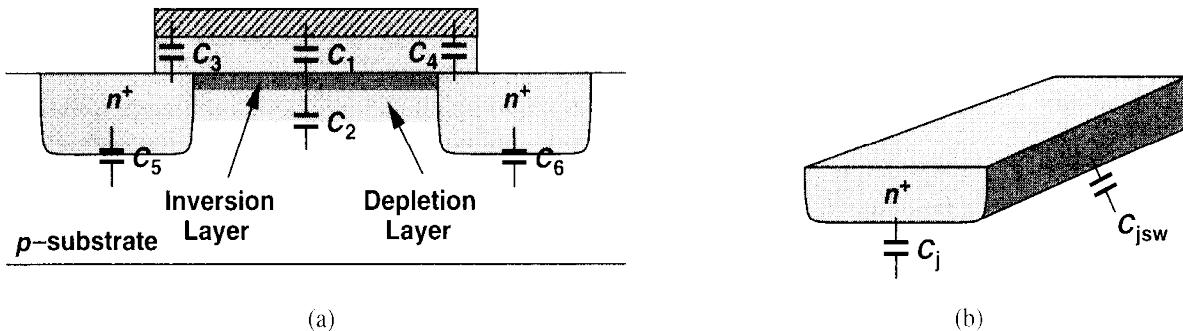
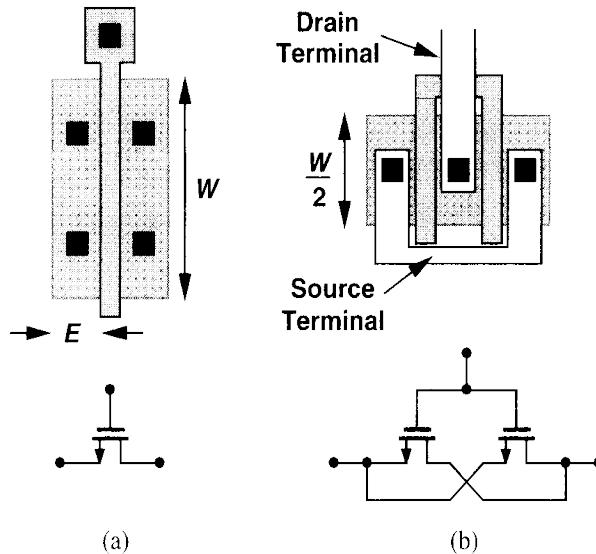


Figure 2.31 (a) MOS device capacitances, (b) decomposition of S/D junction capacitance into bottom-plate and sidewall components.

**Example 2.6**

Calculate the source and drain junction capacitances of the two structures shown in Fig. 2.32.

**Figure 2.32****Solution**

For the transistor in Fig. 2.32(a), we have

$$C_{DB} = C_{SB} = WEC_j + 2(W + E)C_{jsw}, \quad (2.31)$$

whereas for that in Fig. 2.32(b),

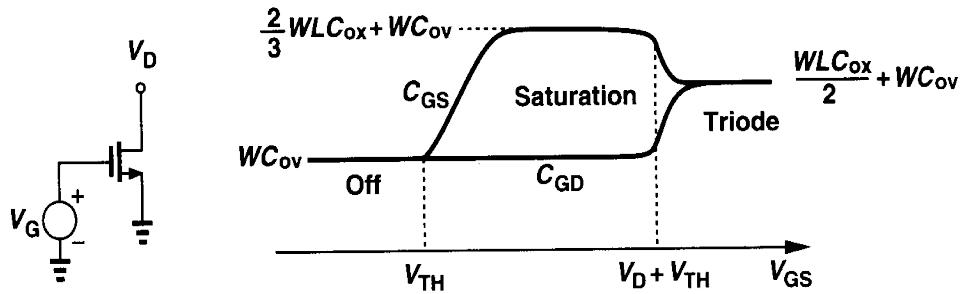
$$C_{DB} = \frac{W}{2}EC_j + 2\left(\frac{W}{2} + E\right)C_{jsw} \quad (2.32)$$

$$C_{SB} = 2\left[\frac{W}{2}EC_j + 2\left(\frac{W}{2} + E\right)C_{jsw}\right] \quad (2.33)$$

$$= WEC_j + 2(W + 2E)C_{jsw}. \quad (2.34)$$

Called a “folded” structure, the geometry in Fig. 2.32(b) exhibits substantially less drain junction capacitance than that in Fig. 2.32(a) while providing the same  $W/L$ .

In the above calculations, we have assumed that the total source or drain perimeter,  $2(W + E)$ , is multiplied by  $C_{jsw}$ . In reality, the capacitance of the sidewall facing the channel may be less than that of the other three sidewalls because of the channel-stop implant (Chapter 17). Nonetheless, we typically assume all four sides have the same unit capacitance. The error resulting from this assumption is negligible because each node in a circuit is connected to a number of other device capacitances as well.



**Figure 2.33** Variation of gate-source and gate-drain capacitances versus  $V_{GS}$ .

We now derive the capacitances between terminals of a MOSFET in different regions of operation. If the device is off,  $C_{GD} = C_{GS} = C_{ov}W$ , and the gate-bulk capacitance consists of the series combination of the gate oxide capacitance and the depletion region capacitance, i.e.,  $C_{GB} = (WLC_{ox})C_d/(WLC_{ox} + C_d)$ , where  $L$  is the effective length and  $C_d = WL\sqrt{q\epsilon_{si}N_{sub}/(4\Phi_F)}$ . The value of  $C_{SB}$  and  $C_{DB}$  is a function of the source and drain voltages with respect to the substrate.

If the device is in deep triode region, i.e., if S and D have approximately equal voltages, then the gate-channel capacitance,  $WLC_{ox}$ , is divided equally between the gate and source terminals and the gate and drain terminals. This is because a change  $\Delta V$  in the gate voltage draws equal amounts of charge from S and D. Thus,  $C_{GD} = C_{GS} = WLC_{ox}/2 + WC_{ov}$ .

If in saturation, a MOSFET exhibits a gate-drain capacitance of roughly  $WC_{ov}$ . The potential difference between the gate and the channel varies from  $V_{GS}$  at the source to  $V_{GS} - V_{TH}$  at the pinch-off point, resulting in a nonuniform vertical electric field in the gate oxide along the channel. It can be proved that the equivalent capacitance of this structure excluding the gate-source overlap capacitance equals  $2WLC_{ox}/3$  [1]. Thus,  $C_{GS} = 2WL_{eff}C_{ox}/3 + WC_{ov}$ . The behavior of  $C_{GD}$  and  $C_{GS}$  in different regions of operation is plotted in Fig. 2.33. Note that the above equations do not provide a smooth transition from one region of operation to another, creating convergence difficulties in simulation programs. This issue is revisited in Chapter 16.

The gate-bulk capacitance is usually neglected in the triode and saturation regions because the inversion layer acts as a “shield” between the gate and the bulk. In other words, if the gate voltage varies, the charge is supplied by the source and the drain rather than the bulk.

### Example 2.7

Sketch the capacitances of  $M_1$  in Fig. 2.34 as  $V_X$  varies from zero to 3 V. Assume  $V_{TH} = 0.6$  V and  $\lambda = \gamma = 0$ .

#### Solution

To avoid confusion, we label the three terminals as shown in Fig. 2.34. For  $V_X \approx 0$ ,  $M_1$  is in the triode region,  $C_{EN} \approx C_{EF} = (1/2)WLC_{ox} + WC_{ov}$ , and  $C_{FB}$  is maximum. The value of  $C_{NB}$  is independent of  $V_X$ . As  $V_X$  exceeds 1 V, the role of the source and drain is exchanged [Fig. 2.35(a)],

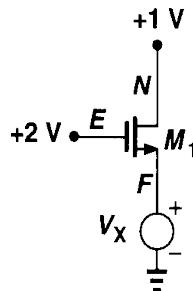


Figure 2.34

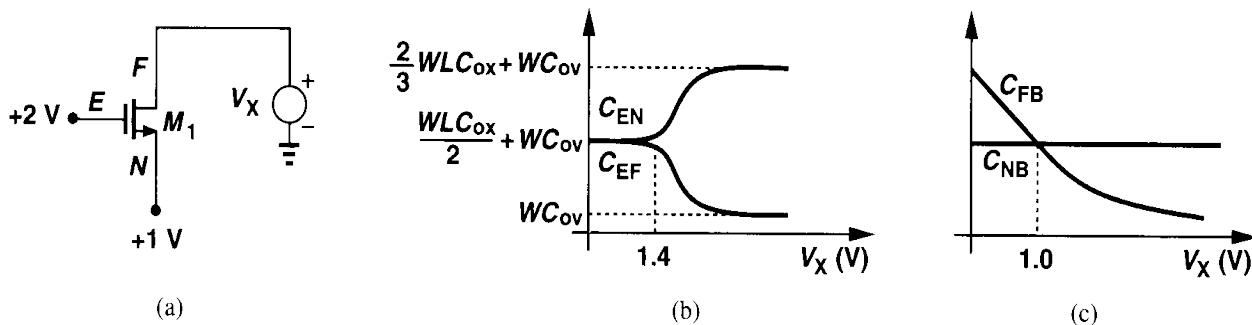


Figure 2.35

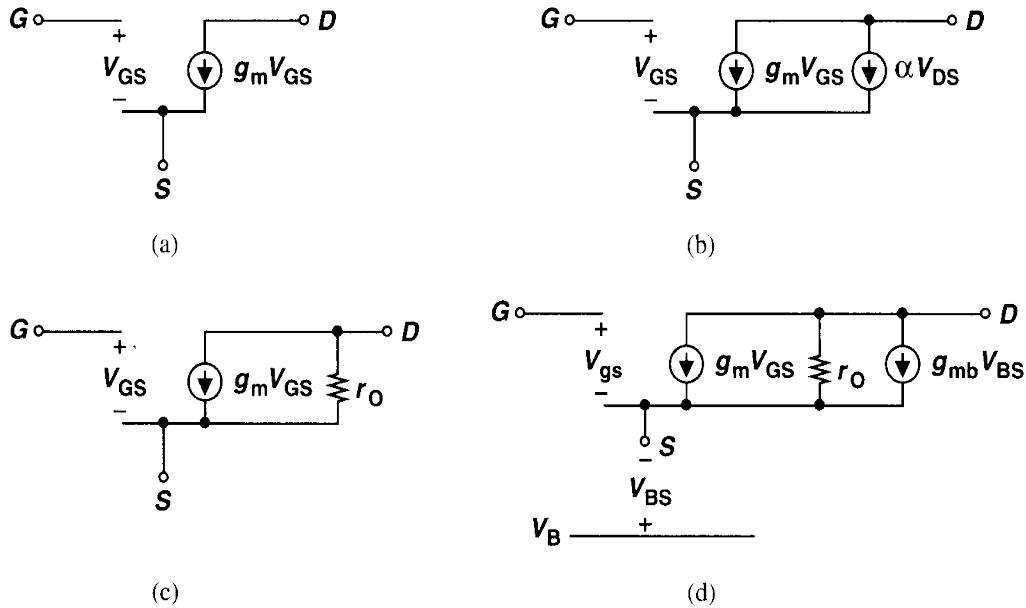
eventually bringing  $M_1$  out of the triode region for  $V_X \geq 2\text{ V} - 0.6\text{ V}$ . The variation of the capacitances is plotted in Figs. 2.35(b) and (c).

### 2.4.3 MOS Small-Signal Model

The quadratic characteristics described by (2.8) and (2.9) along with the voltage-dependent capacitances derived above form the large-signal model of MOSFETs. Such a model proves essential in analyzing circuits in which the signal significantly disturbs the bias points, particularly if nonlinear effects are of concern. By contrast, if the perturbation in bias conditions is small, a small-signal model, i.e., an approximation of the large-signal model around the operating point, can be employed to simplify the calculations. Since in many analog circuits, MOSFETs are biased in the saturation region, we derive the corresponding small-signal model here. For transistors operating as switches, a linear resistor given by (2.11) together with device capacitances serves as a rough small-signal equivalent.

We derive the small-signal model by producing a small increment in a bias point and calculating the resulting increment in other bias parameters. Since the drain current is a function of the gate-source voltage, we incorporate a voltage-dependent current source equal to  $g_m V_{GS}$  [Fig. 2.36(a)]. Note that the low-frequency impedance between G and S is very high. This is the small-signal model of an ideal MOSFET.

Owing to channel-length modulation, the drain current also varies with the drain-source voltage. This effect can also be modeled by a voltage-dependent current source [Fig. 2.36(b)], but a current source whose value linearly depends on the voltage across it is equivalent to



**Figure 2.36** (a) Basic MOS small-signal model, (b) channel-length modulation represented by a dependent current source, (c) channel-length modulation represented by a resistor, (d) body effect represented by a dependent current source.

a linear resistor [Fig. 2.36(c)]. Tied between D and S, the resistor is given by

$$r_O = \frac{\partial V_{DS}}{\partial I_D} \quad (2.35)$$

$$= \frac{1}{\partial I_D / \partial V_{DS}}. \quad (2.36)$$

$$= \frac{1}{\frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2 \cdot \lambda} \quad (2.37)$$

$$\approx \frac{1}{\lambda I_D}. \quad (2.38)$$

As seen throughout this book, the output resistance,  $r_O$ , impacts the performance of many analog circuits. For example,  $r_O$  limits the maximum voltage gain of most amplifiers.

Now recall that the bulk potential influences the threshold voltage and hence the gate-source overdrive. As demonstrated in Example 2.3, with all other terminals held at a constant voltage, the drain current is a function of the bulk voltage. That is, the bulk behaves as a second gate. Modeling this dependence by a current source connected between D and S [Fig. 2.36(d)], we write the value as  $g_{mb} V_{BS}$ , where  $g_{mb} = \partial I_D / \partial V_{BS}$ . In the saturation region,  $g_{mb}$  can be expressed as:

$$g_{mb} = \frac{\partial I_D}{\partial V_{BS}} \quad (2.39)$$

$$= \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH}) \left( -\frac{\partial V_{TH}}{\partial V_{BS}} \right). \quad (2.40)$$

We also have

$$\frac{\partial V_{TH}}{\partial V_{BS}} = -\frac{\partial V_{TH}}{\partial V_{SB}} \quad (2.41)$$

$$= -\frac{\gamma}{2}(2\Phi_F + V_{SB})^{-1/2}. \quad (2.42)$$

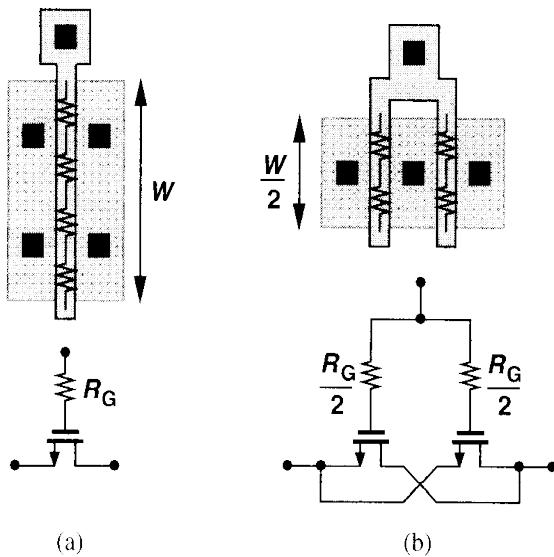
Thus,

$$g_{mb} = g_m \frac{\gamma}{2\sqrt{2\Phi_F + V_{SB}}} \quad (2.43)$$

$$= \eta g_m, \quad (2.44)$$

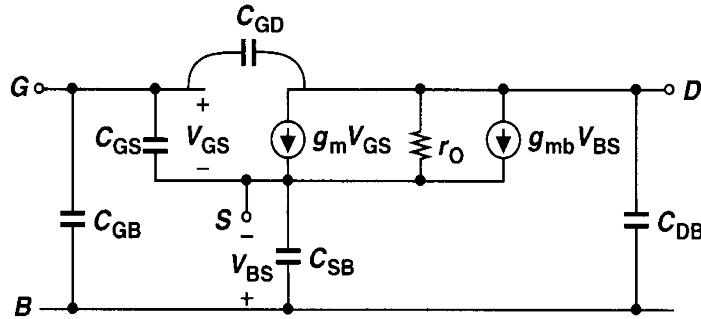
where  $\eta = g_{mb}/g_m$ . As expected,  $g_{mb}$  is proportional to  $\gamma$ . Equation (2.43) also suggests that incremental body effect becomes less pronounced as  $V_{SB}$  increases. Note that  $g_m V_{GS}$  and  $g_{mb} V_{BS}$  have the same polarity, i.e., raising the gate voltage has the same effect as raising the bulk potential.

The model in Fig. 2.36(d) is adequate for most low-frequency small-signal analyses. In reality, each terminal of a MOSFET exhibits a finite ohmic resistance resulting from the resistivity of the material (and the contacts), but proper layout can minimize such resistances. For example, consider the two structures of Fig. 2.32, repeated in Fig. 2.37 along with the gate distributed resistance. We note that folding reduces the gate resistance by a factor of four.



**Figure 2.37** Reduction of gate resistance by folding.

Shown in Fig. 2.38, the complete small-signal model includes the device capacitances as well. The value of each capacitance is calculated according to the equations derived in Section 2.4.2. The reader may wonder how a complex circuit is analyzed intuitively if each transistor must be replaced by the model of Fig. 2.38. The first step is to determine

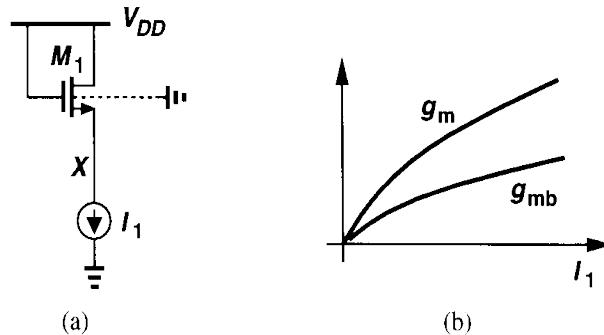


**Figure 2.38** Complete MOS small-signal model.

the *simplest* device model that can represent the role of each transistor with reasonable accuracy. We provide some guidelines for this task at the end of Chapter 3.

### Example 2.8

Sketch  $g_m$  and  $g_{mb}$  of  $M_1$  in Fig. 2.39 as a function of the bias current  $I_1$ .



**Figure 2.39**

### Solution

Since  $g_m = \sqrt{2\mu_n C_{ox}(W/L)I_D}$ , we have  $g_m \propto \sqrt{I_1}$ . The dependence of  $g_{mb}$  upon  $I_1$  is less straightforward. As  $I_1$  increases,  $V_X$  decreases and so does  $V_{SB}$ .

Unless otherwise stated, in this book we assume the bulk of all NFETs is tied to the most negative supply (usually the ground) and that of PFETs to the most positive supply (usually  $V_{DD}$ ).

#### 2.4.4 MOS SPICE models

In order to represent the behavior of transistors in circuit simulations, SPICE requires an accurate model for each device. Over the last two decades, MOS modeling has made tremendous progress, reaching quite sophisticated levels so as to represent high-order effects in short-channel devices.

**Table 2.1** Level 1 SPICE Models for NMOS and PMOS Devices.

NMOS Model			
LEVEL = 1	VTO = 0.7	GAMMA = 0.45	PHI = 0.9
NSUB = 9e+14	LD = 0.08e-6	UO = 350	LAMBDA = 0.1
TOX = 9e-9	PB = 0.9	CJ = 0.56e-3	CJSW = 0.35e-11
MJ = 0.45	MJSW = 0.2	CGDO = 0.4e-9	JS = 1.0e-8
PMOS Model			
LEVEL = 1	VTO = -0.8	GAMMA = 0.4	PHI = 0.8
NSUB = 5e+14	LD = 0.09e-6	UO = 100	LAMBDA = 0.2
TOX = 9e-9	PB = 0.9	CJ = 0.94e-3	CJSW = 0.32e-11
MJ = 0.5	MJSW = 0.3	CGDO = 0.3e-9	JS = 0.5e-8

In this section, we describe the simplest MOS SPICE model, known as “Level 1,” and provide typical values for each parameter in the model corresponding to a 0.5- $\mu\text{m}$  technology. Chapter 16 describes more accurate SPICE models. Table 2.1 shows the model parameters for NMOS and PMOS devices. The parameters are defined as below:

VTO: threshold voltage with zero  $V_{SB}$  (unit: V)

GAMMA: body effect coefficient (unit:  $\text{V}^{1/2}$ )

PHI:  $2\Phi_F$  (unit: V)

TOX: gate oxide thickness (unit: m)

NSUB: substrate doping (unit:  $\text{cm}^{-3}$ )

LD: source/drain side diffusion (unit: m)

UO: channel mobility (unit:  $\text{cm}^2/\text{V}\cdot\text{s}$ )

LAMBDA: channel-length modulation coefficient (unit:  $\text{V}^{-1}$ )

CJ: source/drain bottom-plate junction capacitance per unit area (unit:  $\text{F}/\text{m}^2$ )

CJSW: source/drain sidewall junction capacitance per unit length (unit:  $\text{F}/\text{m}$ )

PB: source/drain junction built-in potential (unit: V)

MJ: exponent in CJ equation (unitless)

MJSW: exponent in CJSW equation (unitless)

CGDO: gate-drain overlap capacitance per unit width (unit:  $\text{F}/\text{m}$ )

CGSO: gate-source overlap capacitance per unit width (unit:  $\text{F}/\text{m}$ )

JS: source/drain leakage current per unit area (unit:  $\text{A}/\text{m}^2$ )

#### 2.4.5 NMOS versus PMOS Devices

In most CMOS technologies, PMOS devices are quite inferior to NMOS transistors. For example, due to the lower mobility of holes,  $\mu_p C_{ox} \approx 0.25 \mu_n C_{ox}$  in modern processes, yielding low current drive and transconductance. Moreover, for given dimensions and bias currents, NMOS transistors exhibit a higher output resistance, providing more ideal current sources and higher gain in amplifiers. For these reasons, it is preferred to incorporate NFETs rather than PFETs wherever possible.

### 2.4.6 Long-Channel versus Short-Channel Devices

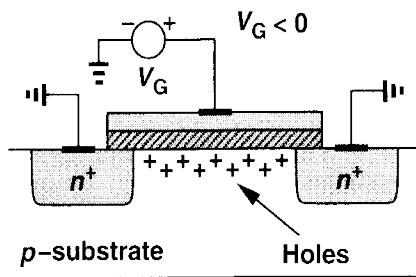
In this chapter, we have employed a very simple view of MOSFETs so as to understand the basic principles of their operation. Most of our treatment is valid for “long-channel” devices, e.g., transistors having a minimum length of about  $4 \mu\text{m}$ . Many of the relationships derived here must be reexamined and revised for short-channel MOSFETs. Furthermore, the SPICE models necessary for simulation of today’s devices need to be much more sophisticated than the Level 1 model. For example, the intrinsic gain,  $g_m r_O$ , calculated from the device parameters in Table 2.1 is quite higher than actual values. These issues are studied in Chapter 16.

The reader may wonder why we begin with a simplistic view of devices if such a view does not lead to a high accuracy in predicting the performance of circuits. The key point is that the simple model provides a great deal of intuition that is necessary in analog design. As we will see throughout this book, we often encounter a trade-off between intuition and rigor, and our approach is to establish the intuition first and gradually complete our understanding so as to achieve rigor as well.

## Appendix A: Behavior of MOS Device as a Capacitor

In this chapter, we have limited our treatment of MOS devices to a basic level. However, the behavior of a MOSFET as a capacitor merits some attention. Recall that if the source, drain, and bulk of an NFET are grounded and the gate voltage rises, an inversion layer begins to form for  $V_{GS} \approx V_{TH}$ . We also noted that for  $0 < V_{GS} < V_{TH}$ , the device operates in the subthreshold region.

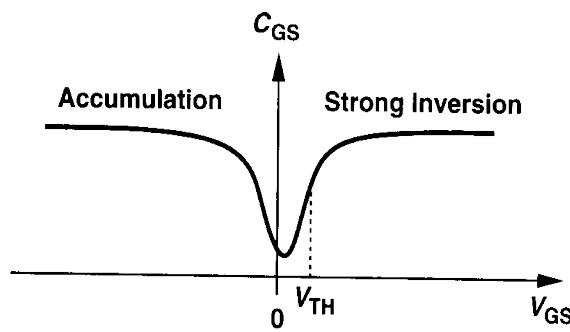
Now consider the NFET of Fig. 2.40. The transistor can be considered a two-terminal



**Figure 2.40** NMOS operating in accumulation mode.

device and hence its capacitance can be examined for different gate voltages. Let us begin with a very *negative* gate-source voltage. The negative potential on the gate attracts the holes in the substrate to the oxide interface. We say the MOSFET operates in the “accumulation” region. The two-terminal device can be viewed as a capacitor having a unit-area capacitance of  $C_{ox}$  because the two “plates” of the capacitor are separated by  $t_{ox}$ .

As  $V_{GS}$  rises, the density of holes at the interface falls, a depletion region begins to form under the oxide, and the device enters weak inversion. In this mode, the capacitance consists of the series combination of  $C_{ox}$  and  $C_{dep}$ . Finally, as  $V_{GS}$  exceeds  $V_{TH}$ , the oxide-silicon interface sustains a channel and the unit-area capacitance returns to  $C_{ox}$ . Figure 2.41 plots the behavior.

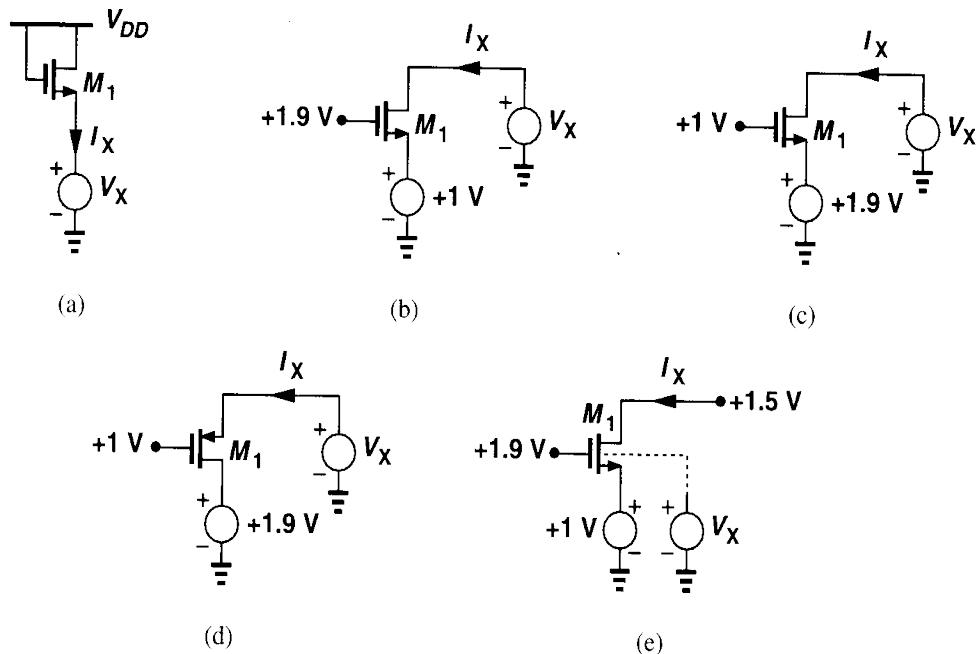


**Figure 2.41** Capacitance-voltage characteristic of an NMOS device.

## Problems

Unless otherwise stated, in the following problems, use the device data shown in Table 2.1 and assume  $V_{DD} = 3$  V where necessary.

- 2.1. For  $W/L = 50/0.5$ , plot the drain current of an NFET and a PFET as a function of  $|V_{GS}|$  as  $|V_{GS}|$  varies from 0 to 3 V. Assume  $|V_{DS}| = 3$  V.
- 2.2. For  $W/L = 50/0.5$ , and  $|I_D| = 0.5$  mA, calculate the transconductance and output impedance of both NMOS and PMOS devices. Also, find the “intrinsic gain,” defined as  $g_{mrO}$ .
- 2.3. Derive expressions for  $g_{mrO}$  in terms of  $I_D$  and  $W/L$ . Plot  $g_{mrO}$  as a function of  $I_D$  with  $L$  as a parameter. Note that  $\lambda \propto 1/L$ .
- 2.4. Plot  $I_D$  versus  $V_{GS}$  for an MOS transistor (a) with  $V_{DS}$  as a parameter, (b) with  $V_{BS}$  as a parameter. Identify the break points in the characteristics.
- 2.5. Sketch  $I_X$  and the transconductance of the transistor as a function of  $V_X$  for each circuit in Fig. 2.42 as  $V_X$  varies from 0 to  $V_{DD}$ . For part (a), assume  $V_X$  varies from 0 to 1.5 V.



**Figure 2.42**

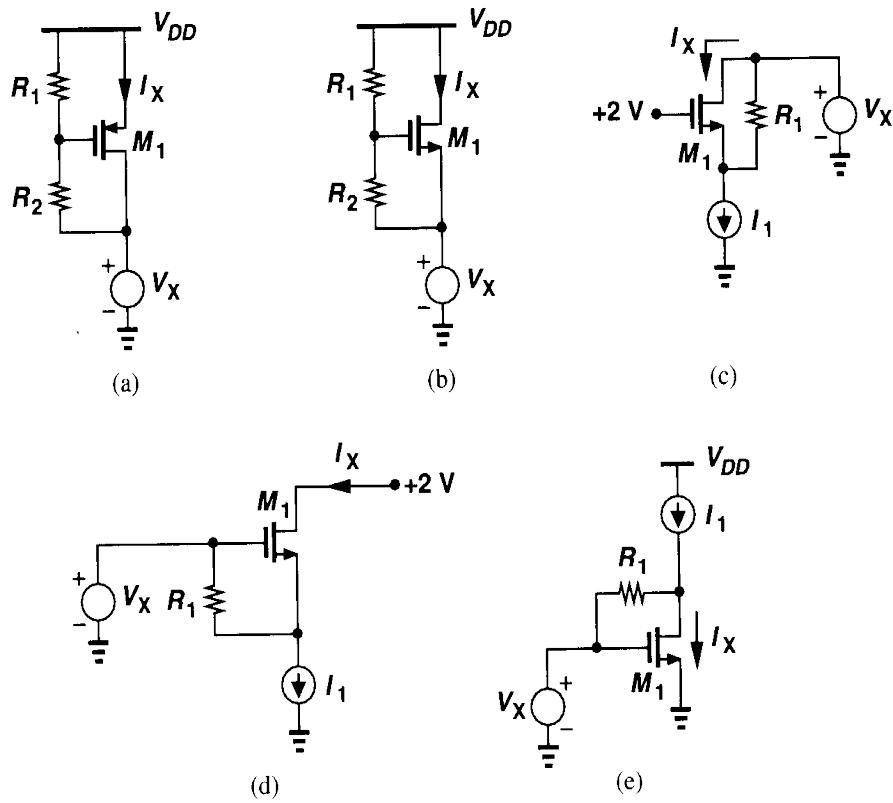


Figure 2.43

- 2.6. Sketch  $I_X$  and the transconductance of the transistor as a function of  $V_X$  for each circuit in Fig. 2.43 as  $V_X$  varies from 0 to  $V_{DD}$ .
- 2.7. Sketch  $V_{out}$  as a function of  $V_{in}$  for each circuit in Fig. 2.44 as  $V_{in}$  varies from 0 to  $V_{DD}$ .
- 2.8. Sketch  $V_{out}$  as a function of  $V_{in}$  for each circuit in Fig. 2.45 as  $V_{in}$  varies from 0 to  $V_{DD}$ .
- 2.9. Sketch  $V_X$  and  $I_X$  as a function of time for each circuit in Fig. 2.46. The initial voltage of  $C_1$  is equal to 3 V.
- 2.10. Sketch  $V_X$  and  $I_X$  as a function of time for each circuit in Fig. 2.47. The initial voltages of  $C_1$  and  $C_2$  are equal to 1 V and 3 V, respectively.
- 2.11. Sketch  $V_X$  as a function of time for each circuit in Fig. 2.48. The initial voltage of each capacitor is shown.
- 2.12. Sketch  $V_X$  as a function of time for each circuit in Fig. 2.49. The initial voltage of each capacitor is shown.
- 2.13. The transit frequency,  $f_T$ , of a MOSFET is defined as the frequency at which the small-signal current gain of the device drops to unity while the source and drain terminals are held at ac ground.
  - (a) Prove that

$$f_T = \frac{g_m}{2\pi(C_{GD} + C_{GS})}. \quad (2.45)$$

Note that  $f_T$  does not include the effect of the S/D junction capacitance.

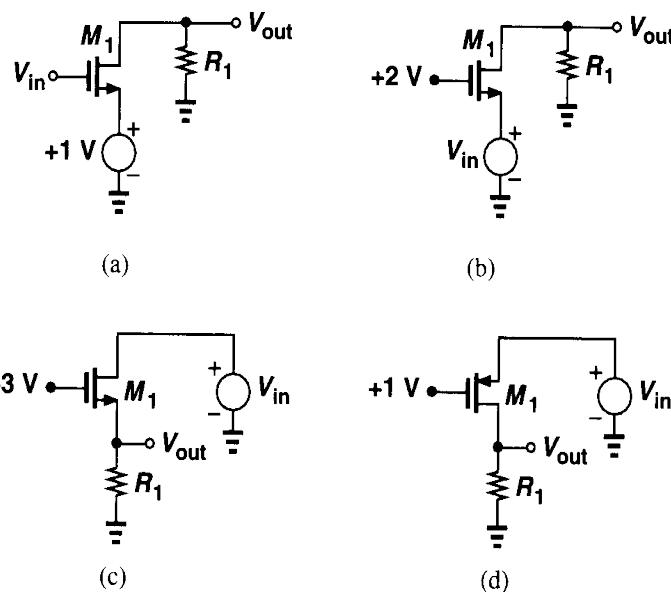


Figure 2.44

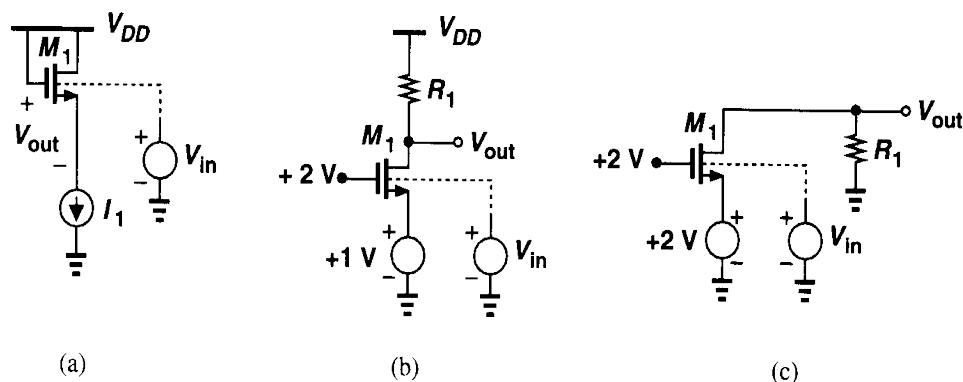


Figure 2.45

- (b) Suppose the gate resistance,  $R_G$ , is significant and the device is modeled as a distributed set of  $n$  transistors each with a gate resistance equal to  $R_G/n$ . Prove that the  $f_T$  of the device is independent of  $R_G$  and still equal to the value given above.
- (c) For a given bias current, the minimum allowable drain-source voltage for operation in saturation can be reduced only by increasing the width and hence the capacitances of the transistor. Using square-law characteristics, prove that

$$f_T = \frac{\mu_n}{2\pi} \frac{V_{GS} - V_{TH}}{L^2}. \quad (2.46)$$

This relation indicates how the speed is limited as a device is designed to operate with lower supply voltages.

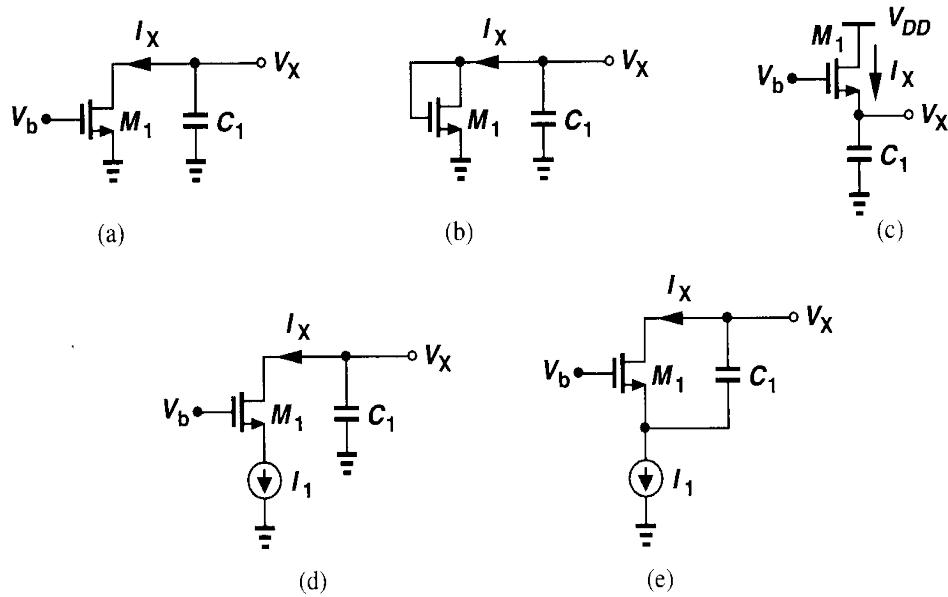


Figure 2.46

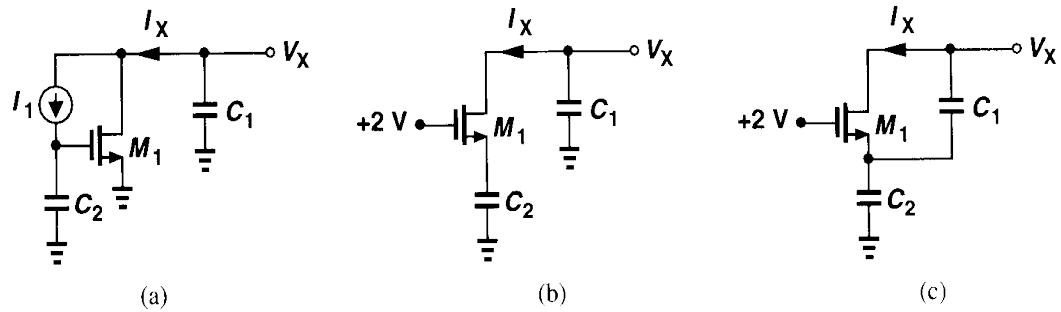


Figure 2.47

- 2.14. Calculate the  $f_T$  of a MOS device in the subthreshold region and compare the result with those obtained in Problem 2.13.
- 2.15. For a saturated NMOS device having  $W = 50 \mu\text{m}$  and  $L = 0.5 \mu\text{m}$ , calculate all of the capacitances. Assume the minimum (lateral) dimension of the S/D areas is  $1.5 \mu\text{m}$  and the device is folded as shown in Fig. 2.32(b). What is the  $f_T$  if the drain current is 1 mA?
- 2.16. Consider the structure shown in Fig. 2.50. Determine  $I_D$  as a function of  $V_{GS}$  and  $V_{DS}$  and prove that the structure can be viewed as a single transistor having an aspect ratio  $W/(2L)$ . Assume  $\lambda = \gamma = 0$ .
- 2.17. For an NMOS device operating in saturation, plot  $W/L$  versus  $V_{GS} - V_{TH}$  if (a)  $I_D$  is constant, (b)  $g_m$  is constant.

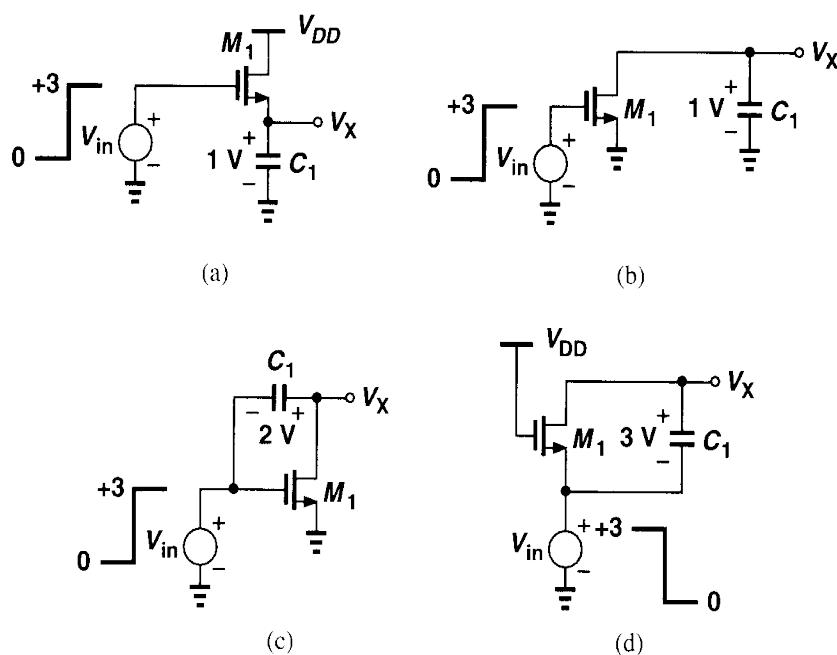


Figure 2.48

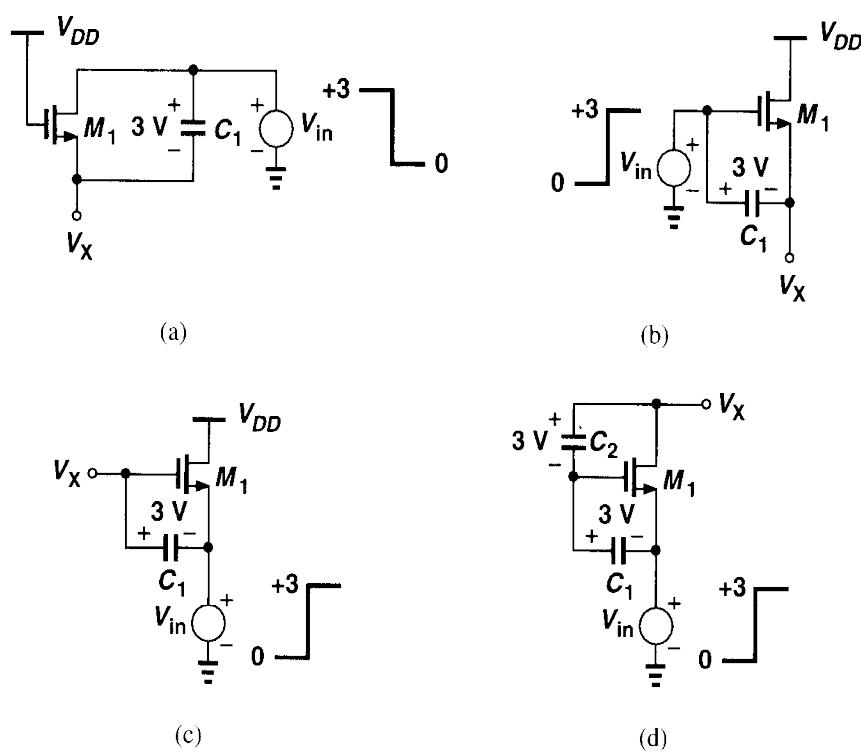


Figure 2.49

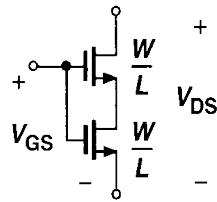


Figure 2.50

- 2.18.** Explain why the structures shown in Fig. 2.51 cannot operate as current sources even though the transistors are in saturation.

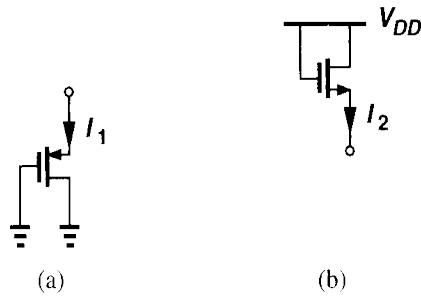


Figure 2.51

- 2.19.** Considering the body effect as “backgate effect,” explain intuitively why  $\gamma$  is directly proportional to  $\sqrt{N_{sub}}$  and inversely proportional to  $C_{ox}$ .
- 2.20.** A “ring” MOS structure is shown in Fig. 2.52. Explain how the device operates and estimate its equivalent aspect ratio. Compare the drain junction capacitance of this structure with that of the devices shown in Fig. 2.32.

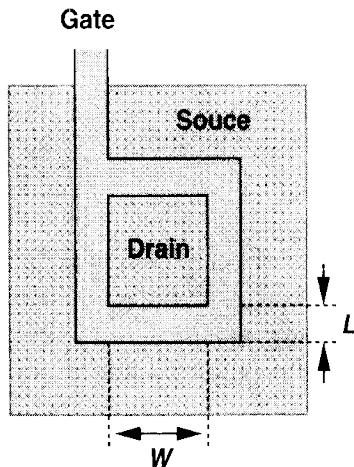


Figure 2.52

- 2.21.** Suppose we have received an NMOS transistor in a package with four unmarked pins. Describe the minimum number of dc measurement steps using an ohmmeter necessary to determine the gate, source/drain, and bulk terminals of the device.
- 2.22.** Repeat Problem 2.21 if the type of the device (NFET or PFET) is not known.

- 2.23. For an NMOS transistor, the threshold voltage is known but  $\mu_n C_{ox}$  and  $W/L$  are not. Assume  $\lambda = \gamma = 0$ . If we cannot measure  $C_{ox}$  independently, is it possible to devise a sequence of dc measurement tests to determine  $\mu_n C_{ox}$  and  $W/L$ ? What if we have two transistors and we know one has twice the aspect ratio of the other?
- 2.24. Sketch  $I_X$  versus  $V_X$  for each of the composite structures shown in Fig. 2.53 with  $V_G$  as a parameter. Also, sketch the equivalent transconductance. Assume  $\lambda = \gamma = 0$ .

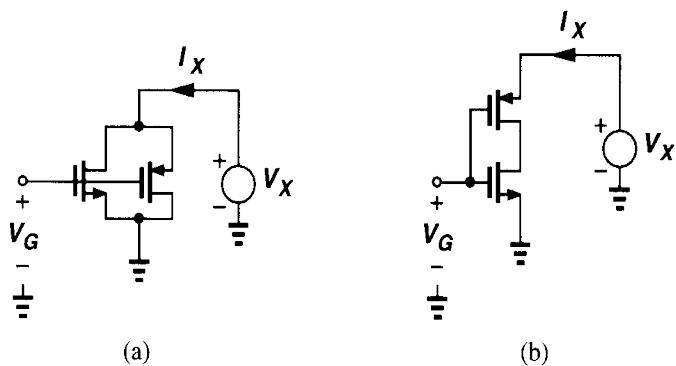


Figure 2.53

- 2.25. An NMOS current source with  $I_D = 0.5$  mA must operate with drain-source voltages as low as 0.4 V. If the minimum required output impedance is  $20$  k $\Omega$ , determine the width and length of the device. Calculate the gate-source, gate-drain, and drain-substrate capacitance if the device is folded as in Fig. 2.32 and  $E = 3$   $\mu$ m.
- 2.26. Consider the circuit shown in Fig. 2.54, where the initial voltage at node X is equal to  $V_{DD}$ . Assuming  $\lambda = \gamma = 0$  and neglecting other capacitances, plot  $V_X$  and  $V_Y$  versus time if (a)  $V_{in}$  is a positive step with amplitude  $V_0 > V_{TH}$ , (b)  $V_{in}$  is a negative step with amplitude  $V_0 = V_{TH}$ .

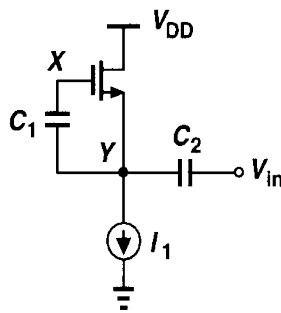


Figure 2.54

- 2.27. An NMOS device operating in the subthreshold region has a  $\zeta$  of 1.5. What variation in  $V_{GS}$  results in a ten-fold change in  $I_D$ ? If  $I_D = 10$   $\mu$ A, what is  $g_m$ ?
- 2.28. Consider an NMOS device with  $V_G = 1.5$  V and  $V_S = 0$ . Explain what happens if we continually decrease  $V_D$  below zero or increase  $V_{sub}$  above zero.

## References

1. R.S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, Second Ed., New York: Wiley, 1986.
2. Y. Tsividis, *Operation and Modeling of the MOS Transistor*, Second Ed., Boston: McGraw-Hill, 1999.
3. Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, New York: Cambridge University Press, 1998.