
FUNDAMENTALS OF MODERN VLSI DEVICES

YUAN TAUR

TAK H. NING



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

40 West 20th Street, New York, NY 10011-4211, USA

www.cambridge.org

Information on this title: www.cambridge.org/9780521550567

© Cambridge University Press 1998

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1998

9th printing 2006

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Taur, Yuan, 1946–

Fundamentals of modern VLSI devices / Yuan Taur, Tak H. Ning.

p. cm.

ISBN 0-521-55056-4. – ISBN 0-521-55959-6 (pbk.)

1. Metal oxide semiconductors, Complementary. 2. Bipolar
transistors. 3. Integrated circuits – Very large scale integration.

I. Ning, Tak H. 1943– . II. Title.

TK7871.99.M44T38 1998

621.39'5 – dc21

98-16162

CIP

ISBN-13 978-0-521-55056-7 hardback

ISBN-10 0-521-55056-4 hardback

ISBN-13 978-0-521-55959-1 paperback

ISBN-10 0-521-55959-6 paperback

Cambridge University Press has no responsibility for
the persistence or accuracy of URLs for external or
third-party Internet Web sites referred to in this book
and does not guarantee that any content on such
Web sites is, or will remain, accurate or appropriate.

CONTENTS

<i>Physical Constants and Unit Conversions</i>	<i>page xi</i>
<i>List of Symbols</i>	<i>xiii</i>
<i>Preface</i>	<i>xxi</i>
1 INTRODUCTION	1
1.1 Evolution of VLSI Device Technology	1
1.2 Modern VLSI Devices	4
1.2.1 Modern CMOS Transistors	4
1.2.2 Modern Bipolar Transistors	5
1.3 Scope and Brief Description of the Book	5
2 BASIC DEVICE PHYSICS	9
2.1 Electrons and Holes in Silicon	9
2.1.1 Energy Bands in Silicon	9
2.1.2 n-Type and p-Type Silicon	12
2.1.3 Carrier Transport in Silicon	19
2.1.4 Basic Equations for Device Operation	24
2.2 p-n Junctions	29
2.2.1 Built-in Potential and Applied Potential	29
2.2.2 Abrupt Junctions	31
2.2.3 The Diode Equation	38
2.2.4 Current-Voltage Characteristics	42
2.2.5 Time-Dependent and Switching Characteristics	50
2.2.6 Diffusion Capacitance	57
2.3 MOS Capacitors	58
2.3.1 Surface Potential: Accumulation, Depletion, and Inversion	58

2.3.2	Electrostatic Potential and Charge Distribution in Silicon	63
2.3.3	Capacitances in an MOS Structure	68
2.3.4	Polysilicon Work Function and Depletion Effects	74
2.3.5	MOS Under Nonequilibrium and Gated Diodes	78
2.3.6	Charge in Silicon Dioxide and at the Silicon–Oxide Interface	82
2.3.7	Effect of Interface Traps and Oxide Charge on Device Characteristics	86
2.4	High-Field Effects	90
2.4.1	Impact Ionization and Avalanche Breakdown	90
2.4.2	Band-to-Band Tunneling	94
2.4.3	Tunneling into and Through Silicon Dioxide	95
2.4.4	Injection of Hot Carriers from Silicon into Silicon Dioxide	97
2.4.5	High-field Effects in Gated Diodes	99
2.4.6	Dielectric Breakdown	100
	Exercises	106
3	MOSFET DEVICES	112
3.1	Long-Channel MOSFETs	113
3.1.1	Drain-Current Model	114
3.1.2	MOSFET I – V Characteristics	117
3.1.3	Subthreshold Characteristics	125
3.1.4	Substrate Bias and Temperature Dependence of Threshold Voltage	129
3.1.5	MOSFET Channel Mobility	132
3.1.6	MOSFET Capacitances and Inversion-Layer Capacitance Effect	135
3.2	Short-Channel MOSFETs	139
3.2.1	Short-Channel Effect	139
3.2.2	Velocity Saturation	149
3.2.3	Channel Length Modulation	154
3.2.4	Source–Drain Series Resistance	158
3.2.5	MOSFET Breakdown	160
	Exercises	161

4	CMOS DEVICE DESIGN	164
4.1	MOSFET Scaling	164
4.1.1	Constant-Field Scaling	164
4.1.2	Generalized Scaling	167
4.1.3	Nonscaling Effects	170
4.2	Threshold Voltage	173
4.2.1	Threshold-Voltage Requirement	173
4.2.2	Nonuniform Doping	177
4.2.3	Channel Profile Design	184
4.2.4	Quantum Effect on Threshold Voltage	194
4.2.5	Discrete Dopant Effects on Threshold Voltage	200
4.3	MOSFET Channel Length	202
4.3.1	Various Definitions of Channel Length	202
4.3.2	Extraction of the Effective Channel Length	204
4.3.3	Physical Meaning of Effective Channel Length	211
	Exercises	221
5	CMOS PERFORMANCE FACTORS	224
5.1	Basic CMOS Circuit Elements	224
5.1.1	CMOS Inverters	224
5.1.2	CMOS NAND and NOR Gates	232
5.1.3	Inverter and NAND Layouts	237
5.2	Parasitic Elements	240
5.2.1	Source–Drain Resistance	240
5.2.2	Parasitic Capacitances	244
5.2.3	Gate Resistance	247
5.2.4	Interconnect R and C	250
5.3	Sensitivity of CMOS Delay to Device Parameters	257
5.3.1	Propagation Delay and Delay Equation	257
5.3.2	Delay Sensitivity to Channel Width, Length, and Gate Oxide Thickness	264
5.3.3	Sensitivity of Delay to Power-Supply Voltage and Threshold Voltage	269
5.3.4	Sensitivity of Delay to Parasitic Resistance and Capacitance	272

5.3.5	Delay of Two-Way NAND and Body Effect	275
5.4	Performance Factors of Advanced CMOS Devices	280
5.4.1	SOI CMOS	280
5.4.2	Velocity Overshoot Effect	283
5.4.3	SiGe or Strained Si MOSFET	285
5.4.4	Low-Temperature CMOS	287
	Exercises	290
6	BIPOLAR DEVICES	292
6.1	n-p-n Transistors	292
6.1.1	Basic Operation of a Bipolar Transistor	296
6.1.2	Modifying the Simple Diode Theory for Describing Bipolar Transistors	296
6.2	Ideal Current-Voltage Characteristics	302
6.2.1	Collector Current	304
6.2.2	Base Current	306
6.2.3	Current Gains	309
6.2.4	Ideal I_C - V_{CE} Characteristics	311
6.3	Characteristics of a Typical n-p-n Transistor	312
6.3.1	Effect of Emitter and Base Series Resistances	313
6.3.2	Effect of Base-Collector Voltage on Collector Current	315
6.3.3	Collector Current Falloff at High Currents	320
6.3.4	Nonideal Base Current at Low Currents	323
6.4	Bipolar Device Models for Circuit and Time-Dependent Analyses	328
6.4.1	Basic dc Model	328
6.4.2	Basic ac Model	330
6.4.3	Small-Signal Equivalent-Circuit Model	332
6.4.4	Emitter Diffusion Capacitance	335
6.4.5	Charge-Control Analysis	336
6.5	Breakdown Voltages	342
6.5.1	Common-Base Current Gain in the Presence of Base-Collector Junction Avalanche	342
6.5.2	Saturation Currents in a Transistor	345
6.5.3	Relation between BV_{CEO} and BV_{CBO}	346
	Exercises	347

INTRODUCTION

1

Since the invention of the bipolar transistor in 1947, there has been an unprecedented growth of the semiconductor industry, with an enormous impact on the way people work and live. In the last twenty years or so, by far, the strongest growth area of the semiconductor industry has been in silicon very-large-scale-integration (VLSI) technology. The sustained growth in VLSI technology is fueled by the continued shrinking of transistors to ever smaller dimensions. The benefits of miniaturization – higher packing densities, higher circuit speeds, and lower power dissipation – have been key in the evolutionary progress leading to today's computers and communication systems that offer superior performance, dramatically reduced cost per function, and much reduced physical size, in comparison with their predecessors. On the economic side, the integrated-circuit (IC) business has grown worldwide in sales from \$1 billion in 1970 to \$20 billion in 1984 and is projected to reach \$185 billion in 1997. The electronics industry is now among the largest industries in terms of output as well as employment in many nations. The importance of microelectronics in the economic, social, and even political development throughout the world will no doubt continue to ascend. The large worldwide investment in VLSI technology constitutes a formidable driving force that will all but guarantee the continued progress in IC integration density and speed, for as long as physical principles will allow.

1.1 EVOLUTION OF VLSI DEVICE TECHNOLOGY

An excellent account of the evolution of the metal–oxide–semiconductor field-effect transistor (MOSFET), from its initial concept to VLSI applications in the mid-1980s, can be found in the paper by Sah (Sah, 1988). Figure 1.1 gives a chronology of the major milestone events in the development of VLSI technology. The bipolar transistor technology was developed early on and was applied to the first integrated-circuit memory in mainframe computers in the 1960s. Bipolar transistors have been used all along where raw circuit speed is most important, for bipolar circuits remain the fastest at the individual-circuit level. However, the large power dissipation of

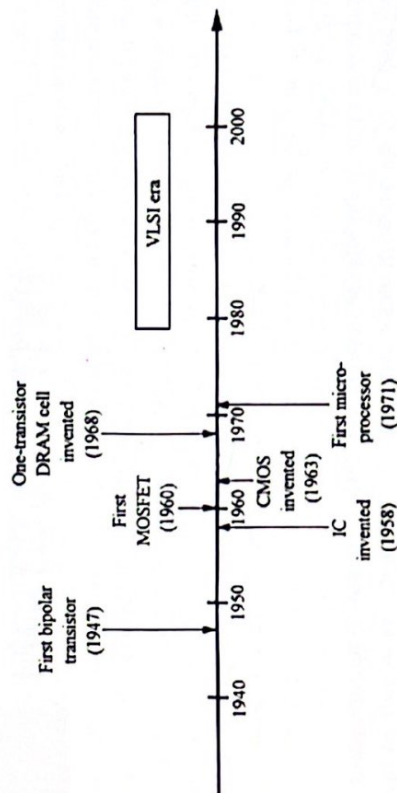


FIGURE 1.1. A brief chronology of the major milestones in the development of VLSI.

bipolar circuits has severely limited their integration level, to about 10^4 circuits per chip. This integration level is quite low by today's VLSI standard.

The idea of modulating the surface conductance of a semiconductor by the application of an electric field was first reported in 1930. However, early attempts to fabricate a surface-field-controlled device were not successful because of the presence of large densities of surface states which effectively shielded the surface potential from the influence of an external field. The first MOSFET on a silicon substrate using SiO_2 as the gate insulator was fabricated in 1960 (Kahng and Atalla, 1960). During the 1960s and 1970s, n-channel and p-channel MOSFETs were widely used, along with bipolar transistors, for implementing circuit functions on a silicon chip. Although the MOSFET devices were slow compared to the bipolar devices, they had a higher layout density and were relatively simple to fabricate; the simplest MOSFET chip could be made using only four masks and a single doping step. However, just like bipolar circuits, single-polarity MOSFET circuits suffered from large standby power dissipation, and hence were limited in the level of integration on a chip.

The major breakthrough in the level of integration came in 1963 with the invention of CMOS (complementary MOS) (Wanlass and Sah, 1963), in which both n-channel and p-channel MOSFETs are constructed simultaneously on the same substrate. A CMOS circuit typically consists of an n-channel MOSFET and a p-channel MOSFET connected in series between the power-supply terminals, so that there is negligible standby power dissipation. Significant power is dissipated only during switching of the circuit (i.e., only when the circuits are active.) By cleverly designing the "switch activities" of the circuits on a chip to minimize active power dissipation, engineers have been able to integrate hundreds of millions of CMOS transistors on a single chip and still have the chip readily air-coolable. Until recently, the integration level of CMOS was not limited by chip-level power dissipation, but by chip fabrication technology. Another advantage of CMOS circuits

comes from the ratioless, full rail-to-rail logic swing, which improves the noise margin and makes a CMOS chip easier to design.

As linear dimensions reached the $0.5\text{-}\mu\text{m}$ level in the early 1990s, the performance advantage of bipolar transistors was outweighed by the significantly greater circuit density of CMOS devices. The system performance benefit of integrated functionality superseded that of raw transistor performance, and practically all the VLSI chips in production today are based on CMOS technology. Bipolar transistors are used only where raw circuit speed makes an important difference. Consequently, bipolar transistors are usually used in small-size bipolar-only chips, or in so-called BiCMOS chips where most of the functions are implemented using CMOS transistors and only a relatively small number are implemented using bipolar transistors.

Advances in lithography and etching technologies have enabled the industry to scale down transistors in physical dimensions, and to pack more transistors in the same chip area. Such progress, combined with a steady growth in chip size, resulted in an exponential growth in the number of transistors and memory bits per chip. The recent trends and future projections in these areas are illustrated in Fig. 1.2. Dynamic random-access memories (DRAMs) have characteristically contained the highest component count of any IC chips. This has been so because of the small size of the one-transistor memory cell (Dennard, 1968) and because of the large and often insatiable demand for more memory in computing systems. It is interesting to note that the entire content of this book can be stored in one 64-Mb DRAM chip, which is in volume production in 1997 and has an area equivalent to a square of about $1.2 \times 1.2\text{ cm}^2$.

One remarkable feature of silicon devices that fuels the rapid growth of the information technology industry is that their speed increases and their cost decreases

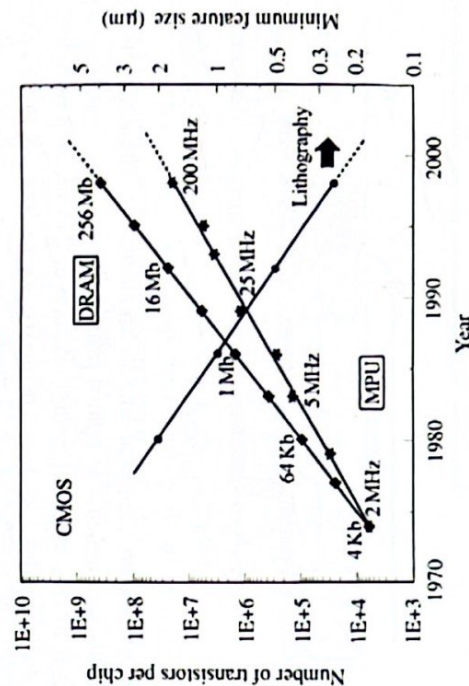


FIGURE 1.2. Trends in lithographic feature size, and number of transistors per chip for DRAM and microprocessor chips.

as their size is reduced. The transistors manufactured today are 20 times faster and occupy less than 1% of the area of those built 20 years ago. This is illustrated in the trend of microprocessor units (MPUs) in Fig. 1.2. The increase in the clock frequency of microprocessors is the result of a combination of improvements in microprocessor architecture and improvements in transistor speed.

1.2 MODERN VLSI DEVICES

It is clear from Fig. 1.2 that modern transistors of practical interest have feature sizes of 0.5 μm and smaller. Although the basic operation principles of large and small transistors are the same, the relative importance of the various device parameters and performance factors for the small-dimension modern transistors is quite different from that for the transistors of the early 1980s or earlier. It is our intention to focus our discussion in this book on the fundamentals of silicon devices of sub-0.5- μm generations.

1.2.1 MODERN CMOS TRANSISTORS

A schematic cross section of modern CMOS transistors, consisting of an n-channel MOSFET and a p-channel MOSFET integrated on the same chip, is shown in Fig. 1.3. A generic process flow for fabricating the CMOS transistors is outlined in Appendix 1. The key physical features of the modern CMOS technology, as illustrated in Fig. 1.3, include: p-type polysilicon gate for the p-channel MOSFET and n-type polysilicon gate for the n-channel MOSFET, refractory metal silicide on the polysilicon gate as well as on the source and drain diffusion regions, and shallow-trench oxide isolation.

In the electrical design of the modern CMOS transistor, the power-supply voltage is reduced with the physical dimensions in some coordinated manner. A great deal of design detail goes into determining the channel length, or separation between the source and drain, accurately, maximizing the on current of the transistor while

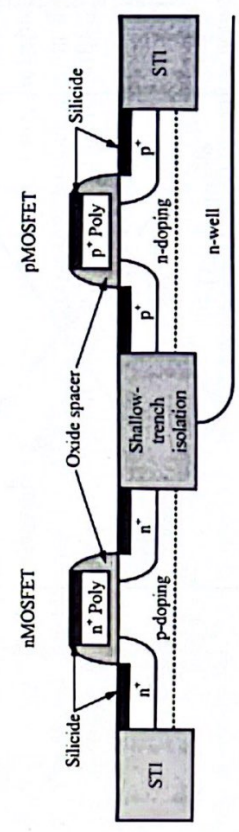


FIGURE 1.3. Schematic device cross section for an advanced CMOS technology.

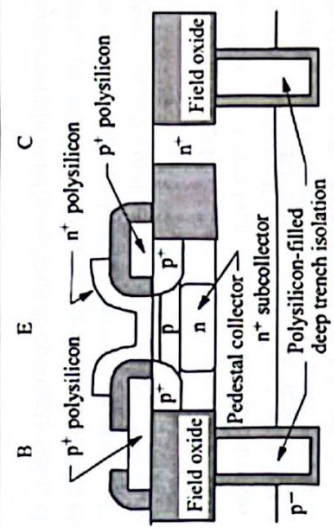


FIGURE 1.4. Schematic cross section of a modern n-p-n bipolar transistor.

maintaining an adequately low off current, minimizing variation of the transistor characteristics with process tolerances, and minimizing the parasitic resistances and parasitic capacitances.

1.2.2 MODERN BIPOLAR TRANSISTORS

The schematic cross section of a modern silicon bipolar transistor is shown in Fig. 1.4. The process outline for fabricating this transistor is shown in Appendix 2. The salient features of this transistor include: shallow-trench field oxide and deep-trench isolation, polysilicon emitter, polysilicon base contact which is self-aligned to the emitter contact, and a pedestal collector which is doped to the desired level only directly underneath the emitter.

Unlike CMOS, the power-supply voltage for a bipolar transistor is usually kept constant as the transistor physical dimensions are reduced. Without the ability to reduce the operating voltage, electrical breakdown is a severe concern in the design of modern bipolar transistors. In designing a modern bipolar transistor, a lot of effort is spent tailoring the doping profile of the various device regions in order to maintain adequate breakdown-voltage margins while maximizing the device performance. At the same time, unlike the bipolar transistors before the early 1980s when the device performance was mostly limited by the device physical dimensions practical at the time, a modern bipolar transistor often has its performance limited by its current-density capability and not by its physical dimensions. Attempts to improve the current-density capability of a transistor usually lead to reduced breakdown voltages.

1.3 SCOPE AND BRIEF DESCRIPTION OF THE BOOK

In writing this book, it is our goal to address the factors governing the performance of modern VLSI devices in depth. This is carried out by first discussing the

The metal-oxide-semiconductor field-effect transistor (MOSFET) is the building block of VLSI circuits in microprocessors and dynamic memories. Because the current in a MOSFET is transported predominantly by carriers of one polarity only (e.g., electrons in an n-channel device), the MOSFET is usually referred to as a unipolar or majority-carrier device. Throughout this chapter, n-channel MOSFETs are used as an example to illustrate device operation and derive drain-current equations. The results can easily be extended to p-channel MOSFETs by exchanging the dopant types and reversing the voltage polarities.

The basic structure of a MOSFET is shown in Fig. 3.1. It is a four-terminal device with the terminals designated as *gate* (subscript *g*), *source* (subscript *s*), *drain* (subscript *d*), and *substrate* or *body* (subscript *b*). An n-channel MOSFET, or nMOSFET, consists of a p-type silicon substrate into which two n^+ regions, the source and the drain, are formed (e.g., by ion implantation). The gate electrode is usually made of metal or heavily doped polysilicon and is separated from the substrate by a thin silicon dioxide film, the *gate oxide*. The gate oxide is usually formed by thermal oxidation of silicon. In VLSI circuits, a MOSFET is surrounded by a thick oxide called the *field oxide* to isolate it from the adjacent devices. The surface region under the gate oxide between the source and drain is called the *channel* region and is critical for current conduction in a MOSFET. The basic operation of a MOSFET device can be easily understood from the MOS capacitor discussed in Section 2.3. When there is no voltage applied to the gate or when the gate voltage is zero, the p-type silicon surface is either in accumulation or in depletion and there is no current flow between the source and drain. The MOSFET device acts like two back-to-back p-n junction diodes with only low-level leakage currents present. When a sufficiently large positive voltage is applied to the gate, the silicon surface is inverted to n-type, which forms a conducting channel between the n^+ source and drain. If there is a voltage difference between them, an electron current will flow from the source to the drain. A MOSFET device therefore operates like a switch ideally suited for digital circuits. *Since the gate electrode is electrically insulated from the substrate, there is effectively no dc gate current, and the channel is capacitively*

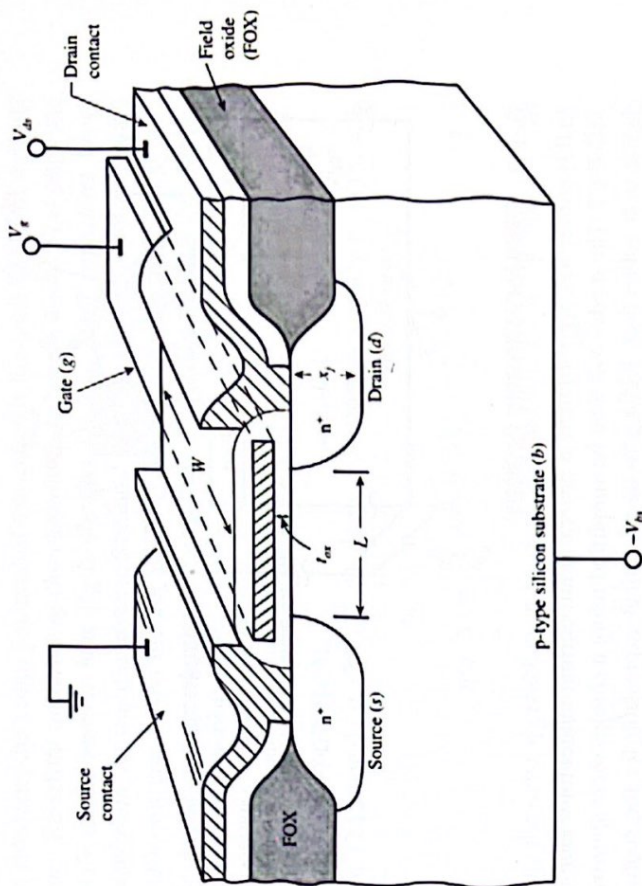


FIGURE 3.1. Three-dimensional view of basic MOSFET device structure. (After Arora, 1993.)

coupled to the gate via the electric field in the oxide (hence the name field-effect transistor).

3.1 LONG-CHANNEL MOSFETs

This section describes the basic characteristics of a long-channel MOSFET, which will serve as the foundation for understanding the more important but more complex short-channel MOSFETs in Section 3.2. First, a general MOSFET current model based on the *gradual channel approximation* (GCA) is formulated in Section 3.1.1. The GCA is valid for most regions of MOSFET operation except beyond the pinch-off or saturation point. A *charge-sheet approximation* is then introduced in Section 3.1.2 to obtain analytical expressions for the source-drain current in the linear and saturation regions. Current characteristics in the subthreshold region are discussed in Section 3.1.3. Section 3.1.4 addresses the threshold-voltage dependence on substrate bias and temperature. Section 3.1.5 presents an empirical model for electron and hole mobilities in a MOSFET channel. Lastly, intrinsic MOSFET capacitances and inversion-layer capacitance effects (neglected in the charge sheet approximation) are covered in Section 3.1.6.

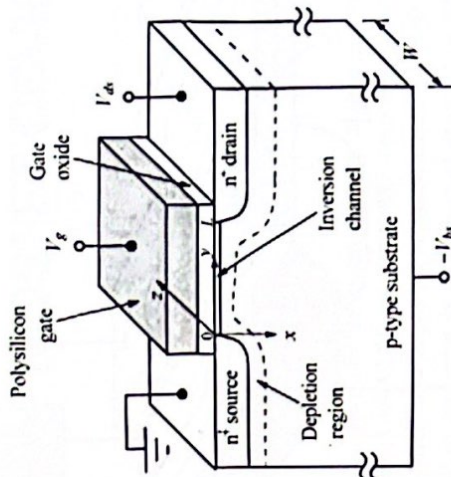


FIGURE 3.2. A schematic MOSFET cross section, showing the axes of coordinates and the bias voltages at the four terminals for the drain-current model.

3.1.1 DRAIN-CURRENT MODEL

In this subsection, we formulate a general drain-current model for a long-channel MOSFET. The model will then be simplified using a *charge-sheet approximation* in the next subsection, leading to an analytical expression for the source-drain current. Figure 3.2 shows the schematic cross section of an n-channel MOSFET in which the source is the n^+ region on the left, and the drain is the n^+ region on the right. A thin oxide film separates the gate from the channel region between the source and drain. We choose an x - y coordinate system consistent with Section 2.3 on MOS capacitors, namely, the x -axis is perpendicular to the gate electrode and is pointing into the p -type substrate with $x = 0$ at the silicon surface. The y -axis is parallel to the channel or the current flow direction, with $y = 0$ at the source and $y = L$ at the drain. L is called the *channel length* and is a key parameter in a MOSFET device. The MOSFET is assumed to be uniform along the z -axis over a distance called the *channel width*, W , determined by the boundaries of the thick field oxide.

Conventionally, the source voltage is defined as the ground potential. The drain voltage is V_{ds} , the gate voltage is V_g , and the p -type substrate is biased at $-V_{bs}$. Initially, we assume $V_{bs} = 0$, i.e., the substrate contact is grounded to the source potential. Later on, we will discuss the effect of substrate bias on MOSFET characteristics. The p -type substrate is assumed to be uniformly doped with an acceptor concentration N_a .

As defined in Section 2.3, $\psi(x, y)$ is the band bending, or intrinsic potential, at (x, y) with respect to the bulk intrinsic potential. We further assume that $V(y)$ is the electron quasi-Fermi potential at a point y along the channel with respect to the Fermi potential of the n^+ source. Since there is no bias between the source

and the substrate, the Fermi potential of the source is the same as that of the bulk substrate. Therefore, in terms of the *channel-to-substrate diode*, $V(y)$ plays the same role as the reverse bias V_R in Section 2.3.5 on MOS capacitors under nonequilibrium. As was discussed in Section 2.2.3, the quasi-Fermi potential stays essentially constant across the depletion region, i.e., $V(y)$ does not change with x in the direction perpendicular to the surface. At the drain end of the channel, $V(y = L) = V_{ds}$.

3.1.1.1 INVERSION CHARGE DENSITY AS A FUNCTION OF QUASI-FERMI POTENTIAL

From Eq. (2.150) and Eq. (2.187), the electron concentration at any point (x, y) is given by

$$n(x, y) = \frac{n_i^2}{N_a} e^{q(\psi - V)/kT}. \quad (3.1)$$

Following the same approach as in Section 2.3.2, one obtains an expression for the electric field similar to that of Eq. (2.153):

$$\mathcal{E}^2(x, y) = \left(\frac{d\psi}{dx} \right)^2 = \frac{2kTN_a}{\epsilon_{si}} \left[\left(e^{-q\psi/kT} + \frac{q\psi}{kT} - 1 \right) + \frac{n_i^2}{N_a^2} \left(e^{-qV/kT} (e^{q\psi/kT} - 1) - \frac{q\psi}{kT} \right) \right]. \quad (3.2)$$

The condition for surface inversion, Eq. (2.190), becomes

$$\psi(0, y) = V(y) + 2\psi_B, \quad (3.3)$$

which is a function of y . From Eq. (2.191), the maximum depletion layer width is

$$W_{dm}(y) = \sqrt{\frac{2\epsilon_{si}[V(y) + 2\psi_B]}{qN_a}}, \quad (3.4)$$

which is also a function of y .

3.1.1.2 GRADUAL-CHANNEL APPROXIMATION

One of the key assumptions in any 1-D MOSFET model is the *gradual channel approximation (GCA)*, which assumes that the variation of the electric field in the y -direction (along the channel) is much less than the corresponding variation in the x -direction (perpendicular to the channel) (Pao and Sah, 1966). This allows us to reduce Poisson's equation to the 1-D form (x -component only) as in Eq. (2.147). The GCA is valid for most of the channel regions except beyond the *pinch-off* point,

which will be discussed later. One further assumes that both the hole current and the generation and recombination current are negligible, so that the current continuity equation can be applied to the electron current in the y -direction. In other words, the total drain-to-source current I_{ds} is the same at any point along the channel. From Eq. (2.45), the electron current density at a point (x, y) is

$$J_n(x, y) = -q\mu_n n(x, y) \frac{dV(y)}{dy}, \quad (3.5)$$

where $n(x, y)$ is the electron density, and μ_n is the electron mobility in the channel. The carrier mobility in the channel is generally much lower than the mobility in the bulk, due to additional surface scattering mechanisms, as will be addressed in Section 3.1.5. *With $V(y)$ defined as the quasi-Fermi potential, i.e., playing the role of ϕ_n in Eq. (2.45), Eq. (3.5) includes both the drift and diffusion currents.* The total current at a point y along the channel is obtained by multiplying Eq. (3.5) with the channel width W and integrating over the depth of the inversion layer. The integration is carried out from $x = 0$ to x_i , the bottom of the inversion layer where $\psi = \psi_B$:

$$I_{ds}(y) = qW \int_0^{x_i} \mu_n n(x, y) \frac{dV}{dy} dx. \quad (3.6)$$

There is a sign change, as we define $I_{ds} > 0$ to be the drain-to-source current in the $-y$ direction. Since V is a function of y only, dV/dy can be taken outside the integral. We also assume that μ_n can be taken outside the integral by defining an *effective mobility*, μ_{eff} , at some average gate and drain fields. What remains in the integral is the electron concentration, $n(x, y)$. Its integration over the inversion layer gives the inversion charge per unit gate area, Q_i :

$$Q_i(y) = -q \int_0^{x_i} n(x, y) dx. \quad (3.7)$$

Equation (3.6) then becomes

$$I_{ds}(y) = -\mu_{eff} W \frac{dV}{dy} Q_i(y) = -\mu_{eff} W \frac{dV}{dy} Q_i(V). \quad (3.8)$$

In the last step, Q_i is expressed as a function of V ; V is interchangeable with y , since V is a function of y only. Multiplying both sides of Eq. (3.8) by dy and integrating from 0 to L (source to drain) yield

$$\int_0^L I_{ds} dy = \mu_{eff} W \int_0^{V_{ds}} [-Q_i(V)] dV. \quad (3.9)$$

Current continuity requires that I_{ds} be a constant, independent of y . Therefore, the drain-to-source current is

$$I_{ds} = \mu_{eff} \frac{W}{L} \int_0^{V_{ds}} [-Q_i(V)] dV. \quad (3.10)$$

3.1.1.3 PAO AND SAH'S DOUBLE INTEGRAL

An alternative form of $Q_i(V)$ can be derived if $n(x, y)$ is expressed as a function of (ψ, V) using Eq. (3.1), i.e.,

$$n(x, y) = n(\psi, V) = \frac{n_i^2}{N_a} e^{q(\psi - V)/kT}, \quad (3.11)$$

and substituted into Eq. (3.7):

$$\begin{aligned} Q_i(V) &= -q \int_{\psi_B}^{\psi_s} n(\psi, V) \frac{dx}{d\psi} d\psi \\ &= -q \int_{\psi_B}^{\psi_s} \frac{(n_i^2/N_a) e^{q(\psi - V)/kT}}{\mathcal{E}(\psi, V)} d\psi. \end{aligned} \quad (3.12)$$

Here, ψ_s is the surface potential at $x = 0$ and $\mathcal{E}(\psi, V) = -d\psi/dx$ is given by the square root of Eq. (3.2). Substituting Eq. (3.12) into Eq. (3.10) yields

$$I_{ds} = q\mu_{eff} \frac{W}{L} \int_0^{V_{ds}} \left(\int_{\psi_B}^{\psi_s} \frac{(n_i^2/N_a) e^{q(\psi - V)/kT}}{\mathcal{E}(\psi, V)} d\psi \right) dV. \quad (3.13)$$

This is referred to as *Paol and Sah's double integral* (Paol and Sah, 1966). The boundary value ψ_s is determined by two coupled equations: Eq. (2.180) and $Q_s = -\epsilon_{si} \mathcal{E}_s(\psi_s)$ or Gauss's law, where $\mathcal{E}_s(\psi_s)$ is obtained by letting $\psi = \psi_s$ in Eq. (3.2). In inversion, only two of the terms in Eq. (3.2) are significant and need to be kept. The merged equation is then

$$\begin{aligned} V_g &= V_{fb} + \psi_s - \frac{Q_s}{C_{ox}} \\ &= V_{fb} + \psi_s + \frac{\sqrt{2\epsilon_{si} k T N_a}}{C_{ox}} \left[\frac{q\psi_s}{kT} + \frac{n_i^2}{N_a^2} e^{q(\psi_s - V)/kT} \right]^{1/2}, \end{aligned} \quad (3.14)$$

which is an implicit equation for $\psi_s(V)$. Equations (3.14) and (3.13) can only be solved numerically.

3.1.2 MOSFET I - V CHARACTERISTICS

In this subsection, we derive the basic expressions for long-channel current in the *linear* and *saturation* regions.