Microelectronics

AN INTEGRATED APPROACH



ROGER T. HOWE CHARLES G. SODINI

Prentice Hall Electronics and VLSI Series—Charles G. Sodini, Series Editor

TI 1037

015276247
EN6I
Repl

Library of Congress Cataloging-in-Publication Data

Howe, Roger Thomas.
Microelectronics : an integrated approach / Roger T. Howe, Charles G. Sodini.
p. cm.
Includes bibliographical references and index.
ISBN 0-13-588518-3
1. Microelectronics. 2. Solid state electronics. I. Sodini, Charles G. II. Title.
TK7874.H688 1997
621.3815--dc20 96-17867
CIP

Acquisitions editor: Eric Svendsen
Production service: TKM Productions
Editorial/production supervision: Sharyn Vitrano
Copy editor: TKM Productions
Interior design: Lisa Delgado
Cover design: Joseph Sengotta
Art director: Amy Rosen
Managing editor: Bayani Mendoza DeLeon
Editor-in-chief: Marcia Horton
Director of production and manufacturing: David W. Riccardi
Manufacturing buyer: Julia Meehan
Editorial assistant: Kathryn Cassino



© 1997 by Prentice-Hall, Inc.
Simon & Schuster / A Viacom Company
Upper Saddle River, New Jersey 07458

All rights reserved. No part of this book may
be reproduced, in any form or by any means,
without permission in writing from the publisher.

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Printed in the United States of America

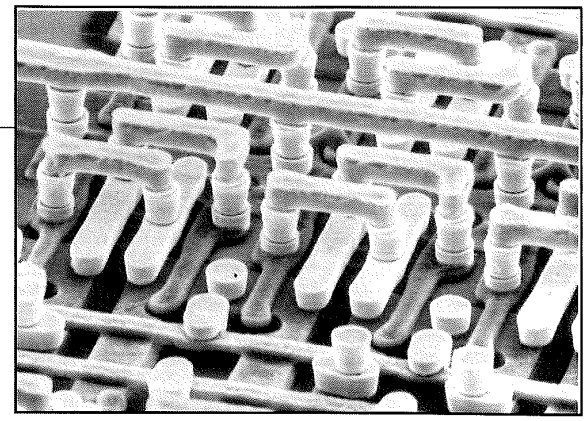
10 9 8 7 6 5

ISBN 0-13-588518-3

Prentice-Hall International (UK) Limited, London
Prentice-Hall of Australia Pty. Limited, Sydney
Prentice-Hall Canada Inc., Toronto
Prentice-Hall Hispanoamericana, S.A., Mexico
Prentice-Hall of India Private Limited, New Delhi
Prentice-Hall of Japan, Inc., Tokyo
Simon & Schuster Asia Pte. Ltd., Singapore
Editora Prentice-Hall do Brasil, Ltda., Rio de Janeiro

chapter 2

Semiconductor Physics and IC Technology



This scanning electron micrograph (SEM) photograph shows the complexity of modern integrated circuit interconnect technology. The interlevel dielectrics have been etched away to show the metallization patterns. (Courtesy of International Business Machines Corporation. Unauthorized use not permitted.)

Microelectronic engineers need a working knowledge of the physical processes underlying the behavior of electronic devices. This chapter provides a short introduction to semiconductor physics, with the goal of developing the basic concepts and terminology needed to model electronic devices described in Chapters 3, 4, 6, and 7. In addition a basic familiarity with IC fabrication is essential to understand the device structures in later chapters. This chapter includes a short overview of the essential features of this powerful technology.

Chapter Objectives

- The bond model for silicon—the electron and hole as mobile charges.
- The control of electron and hole concentrations through doping.
- Drift and diffusion of electrons and holes.
- IC processes and the connection between mask layout and device cross section.
- Sheet resistance and the integrated-circuit resistor.

➤ 2.1 PURE SEMICONDUCTORS

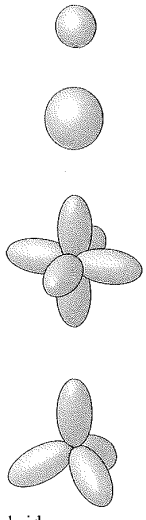
To begin our discussion of semiconductor physics, we consider the dominant semiconductor used in modern microelectronics—crystalline silicon. Silicon is in Group IV of the periodic table, with atomic number 14. Recalling basic atomic chemistry, its 14 electrons occupy the first three energy levels. The first three orbitals, $1s$, $2s$, and $2p$, are filled by 10 electrons, leaving four electrons to occupy the $3s$ and $3p$ orbitals. In order to minimize the overall

$(1s)^2$ (filled)

$(2s)^2$ (filled)

$(2p)^6$ (filled)

$(3sp)^4$



4 tetrahedral hybrid orbitals, half-filled with 4 valence electrons

electronic orbitals in silicon

energy, the 3*s* and 3*p* orbitals hybridize to form four, tetrahedral 3*sp* orbitals. The electronic configuration of silicon is

$$(1s)^2\ (2s)^2\ (2p)^6\ (3sp)^4$$

Each of the four outer hybrid orbitals have one electron and are capable of forming a bond with a neighboring atom. These four electrons that participate in bonding are called **valence electrons**. The inset on the left shows the electronic configuration of silicon.

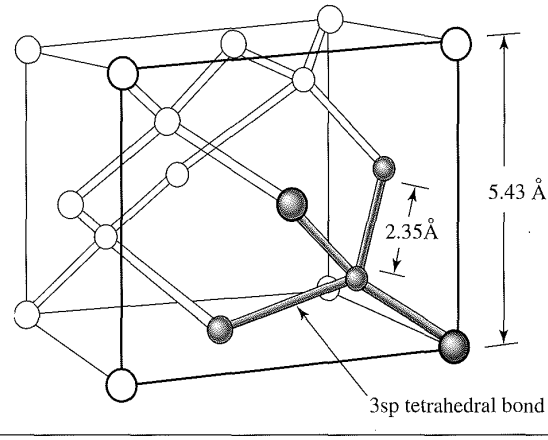
The crystalline form of silicon has the diamond crystal structure. Each silicon ion in the crystal has four nearest neighbors that are aligned with the tetrahedral 3*sp* hybrid orbitals. Each ion contributes an electron to fill the hybrid orbital. Figure 2.1 shows the unit cell of silicon, with the tetrahedral bonds from one atom highlighted. The edge of the unit cell is $a_o = 5.43$ Å.

How many atoms should be assigned to the unit cell? The eight corner atoms are shared among the eight cells touching the corner and therefore, contribute one atom to the unit cell. The six atoms on the faces of the cube are shared with adjacent cells and thus, contribute three atoms. Finally, there are four atoms that lie entirely within the cube, for a total of eight atoms in the unit cell.

The atomic density of crystalline silicon, in atoms per cubic centimeter, is

$$N_{Si} = 8 \cdot n_{cell} = 8 \cdot \frac{1}{a_o^3} = \frac{8}{(5.43 \times 10^{-8}\text{cm})^3} = 5.00 \times 10^{22}\text{cm}^{-3} \qquad (2.1)$$

where n_{cell} is the number of unit cells per cm^3. Since each atom has four tetrahedral bonds and contributes one electron to each bond, the concentration of valence electrons is $4\,N_{Si} = 2 \times 10^{23}\ \text{cm}^{-3}$.



➤ **Figure 2.1** Unit cell of crystalline silicon (After R. S. Muller and T.I. Kamins, *Device Electronics for Integrated Circuits,* 2nd ed., Wiley, 1986.)

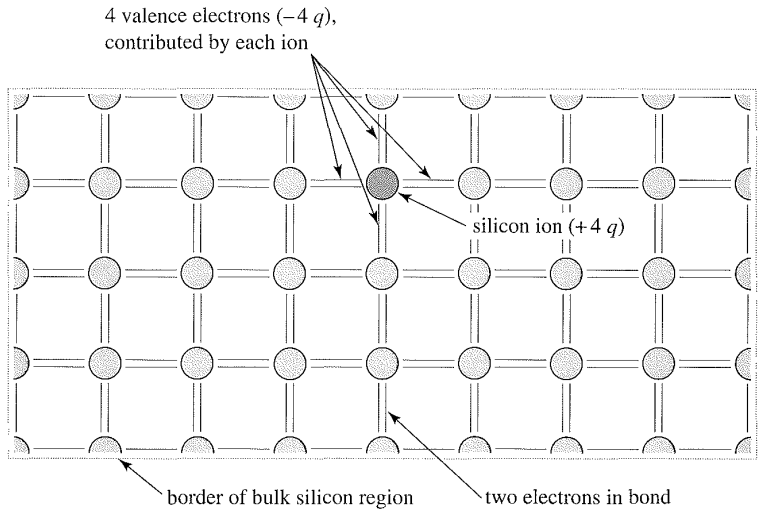
2.1.1 Electrons, Holes, and the Bond Model

We now investigate the electronic properties of the bonded network of silicon atoms in Fig. 2.1, which we refer to as the silicon **lattice**. In order to sketch the silicon lattice conveniently, we use the two-dimensional collapsed view shown in Fig. 2.2. The portion of the lattice shown in Fig. 2.2 is in the interior, or **bulk**, of the silicon crystal away from any surfaces. Each valence electron in a bond is symbolized by a connecting line, which means that two lines are used to represent each tetrahedral bond between silicon atoms. Note that this convention differs from the one used in chemistry, where a single line represents both electrons (in a single bond).

We refer to this abstract picture of the silicon crystal as the **bond model.** The silicon ions (nucleus and 10 inner electrons) each have a charge of $(+4\ q)$, where $q = 1.6 \times 10^{-19}$ C is the magnitude of the charge of an electron. The four valence electrons per ion in the surrounding four bonds contribute a charge of $(-4\ q)$, making the crystal electrically neutral. In Fig. 2.2, all of the bonds are shown to be complete, which is true only for a perfect crystal at a temperature of absolute zero (0 degrees Kelvin).

An **intrinsic semiconductor** is one in which there are no impurities. A perfect crystal of intrinsic silicon at absolute zero has all of its valence electrons tied up in bonds since this is the lowest energy configuration. Since there are no "loose" or "free" electrons to carry current, intrinsic silicon is an insulator at absolute zero.

At temperatures above absolute zero, lattice vibrations occur that can impart enough energy to the valence electrons to break a very small percentage of the bonds. Each broken bond results in a mobile electron that can move around the crystal. This process is known as **thermal generation**. Note that the electron is still confined to the



➢ **Figure 2.2** Collapsed two-dimensional silicon lattice at absolute zero, with all bonds complete. This portion of the crystal is in the interior, or bulk, away from surfaces. It is electrically neutral since the four valence electrons balance the net charge on the silicon ion.

crystal since much more energy is required for it to escape into the surrounding ambient.

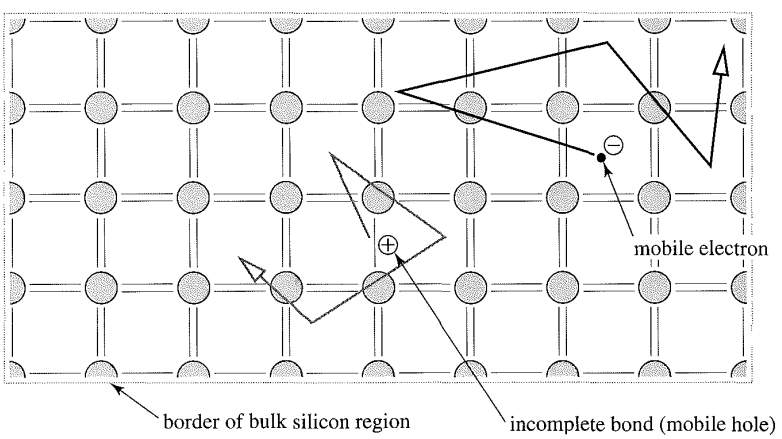
Figure 2.3 shows the silicon lattice after one bond has been broken due to thermal generation. The mobile electron has a negative charge $-q = -1.6 \times 10^{-19}$ C and leaves behind an incomplete bond that takes on a charge of $+q$. At room temperature, the electron moves quite rapidly and undergoes frequent collisions as suggested by the trajectory sketched in Fig. 2.3.

Since electrons have particle-like behavior in a vacuum, it seems plausible that the mobile electron resulting from the broken bond is capable of moving through the crystal lattice. For example, a cathode-ray tube used in a computer monitor works by steering electron trajectories with electric fields. What is surprising is that the positive charge arising from the broken bond is also mobile with similar properties to the electron, except that it has an equal and opposite charge. Figure 2.3 depicts that the positive charge is also free to move around the lattice. This mobile positive charge is called a **hole**.

2.1.2 Limitations of the Bond Model

The bond model is useful in visualizing the creation of electron-hole pairs. However, it has some shortcomings. Electrons and holes in crystals have both particle-like and wave-like behavior that can only be understood from the perspective of quantum mechanics. The bond model necessarily depicts both the electron and the hole as localized in space, which is not correct.

Fortunately, it is not necessary to take a solid-state physics course prior to studying semiconductor devices and circuits. This text considers that electrons and holes can effectively be considered as *classical* particles that obey Newton's Laws and have unit negative and positive charge. Quantum mechanics has a negligible effect on the



➤ **Figure 2.3** Electron and hole resulting from breaking one bond in intrinsic silicon.

accuracy of models for microelectronic devices that have dimensions of 0.1 μm or larger.[1]

2.2 GENERATION, RECOMBINATION, AND THERMAL EQUILIBRIUM

We have seen that mobile electrons and holes can be created in pure, also known as **intrinsic**, silicon by the breaking of bonds due to thermal generation. The thermal generation rate $G_{th}(T)$, with units of $cm^{-3}s^{-1}$, increases exponentially with temperature. Another type of generation process is **optical generation**, with rate G_{op}. Optical generation occurs when a bond absorbs a photon with sufficient energy to break it. The minimum energy for bond-breaking in silicon is around 1.1 eV, corresponding to a wavelength of around 1 μm (in the near-infrared region of the optical spectrum.) In contrast to thermal generation, the optical generation rate is essentially independent of temperature. The total generation rate G is the sum of the thermal and optical rates

$$G = G_{th}(T) + G_{op} \tag{2.2}$$

If mobile electrons and holes are continuously generated, a balancing process is necessary that continuously removes them. **Recombination** is the elimination of an electron and hole from the crystal by completing one of the broken bonds. Recombination processes are classified according to what happens to the energy released when the bond forms.

Thermal recombination—the energy goes to vibrations in the crystal lattice. This is equivalent to saying the silicon crystal is heated.

Optical recombination—the energy goes to photon emission. Optical recombination is of negligible importance in silicon. In other semiconductors such as gallium arsenide, optical recombination is a very significant component of the total recombination rate. In these materials, light emission due to optical recombination is the basic principle behind light-emitting diodes and laser diodes.

For recombination to occur, both an electron and a hole are necessary. The recombination rate will be small if there are few electrons despite an abundance of holes or vice versa. Therefore, it is reasonable that the recombination rate is proportional to the *product* of the electron and hole concentrations

$$R = k(n \cdot p) \tag{2.3}$$

where n is the electron concentration (cm^{-3}), p is the hole concentration (cm^{-3}), and k is the recombination rate constant (cm^3s^{-1}). In contrast, the generation rate G is independent of the electron and hole concentrations since the number of broken bonds is a tiny fraction of the total number of bonds in the crystal.

If the environment of the silicon crystal (optical radiation flux, electric field intensity, temperature, etc.) is *constant,* then the electron and hole concentrations must eventually reach stable values that are set by the balance between the recombi-

steady-state

1. Current research is exploring **nanoelectronic** devices with critical dimensions on the order of 10 nm or less. Quantum effects are crucial to understanding their basic operation.

nation and generation rates. We define this situation as **steady-state**. The total recombination rate and total generation rates must be equal for steady-state conditions

$$R = k(n \cdot p) = G = G_{th}(T) + G_{op} \tag{2.4}$$

Given the particular environmental conditions that determine $G_{th}(T)$ and G_{op}, the product of the electron and hole concentrations is

$$n \cdot p = \frac{G_{th}(T) + G_{op}}{k} \tag{2.5}$$

thermal equilibrium

An idealized, special case of the steady-state condition called **thermal equilibrium** is of great importance in semiconductor physics. Thermal equilibrium is defined as the prolonged absence of external energy sources. The silicon sample is completely isolated from optical or electrical sources (e.g., power supplies and light sources) by placing it in a "black box" at a constant temperature. Some optical generation occurs in addition to thermal generation, due to black-body radiation at the sample temperature. For the case of thermal equilibrium, Eq. (2.5) reduces to

$$n_o \cdot p_o = \frac{G_{th}^o(T) + G_{op}^o(T)}{k_o} \tag{2.6}$$

where all quantities are subscripted or superscripted with "o" to denote thermal equilibrium. The right-hand side of this equation is a function of temperature only (the single variable in thermal equilibrium). Defining this function as $n_i^2(T)$, we can re-express Eq. (2.6) as the ***mass-action law***

mass-action law

$$\boxed{n_o \cdot p_o = n_i^2(T)} \tag{2.7}$$

This important result relates the product of the thermal equilibrium electron and hole concentrations to a constant that depends only on temperature. The unusual name for this equation is borrowed from chemistry, since Eq. (2.7) is analogous to an equilibrium equation for a chemical reaction. For example, the equilibrium between hydrogen ions (concentration $[H^+]$) and hydroxyl ions (concentration $[OH^-]$) reacting to form water is expressed as

$$[H^+] \cdot [OH^-] = K_{eq}(T) \tag{2.8}$$

where $K_{eq}(T)$ is the equilibrium constant.

In pure silicon in thermal equilibrium, the electron and hole concentrations are equal since the electrons and holes are created in pairs when the bonds are broken. Therefore, it follows from substitution in Eq. (2.7)

$$n_o \cdot p_o = n_o^2 = p_o^2 = n_i^2(T) \tag{2.9}$$

and consequently, the electron and hole concentrations in pure silicon are both equal to the **intrinsic concentration** $n_i(T)$ in thermal equilibrium. Before proceeding fur-

ther, it is worthwhile to consider the order of magnitude of the electron and hole concentrations in pure silicon at room temperature

$$\boxed{n_i \cong 10^{10}\ \text{cm}^{-3} \text{at } 300\ \text{K}} \qquad (2.10)$$

intrinsic concentration

Given that the number of bonds is $4 \times N_{Si} = 2 \times 10^{23}\ \text{cm}^{-3}$, at room temperature only an extremely small fraction (1 in 20 trillion) of the bonds are broken. As temperature increases, the intrinsic concentration increases exponentially—approximately doubling for every 10 °C rise over room temperature.

2.3 DOPING

Silicon would be of little use if the hole and electron concentrations were both equal to $10^{10}\ \text{cm}^{-3}$. Fortunately, the hole and electron concentrations can be controlled by adding minute quantities of specific elements to the silicon crystal. There are a variety of ways to incorporate these impurities into the silicon lattice using a process called **doping**. We will first consider the case of **donors**, which are elements in Group V of the periodic table.

2.3.1 Donors

The bond model is useful for visualizing how donor atoms control the concentration of electrons in silicon. Typical donors for silicon are phosphorus (P), arsenic (As), and antimony (Sb), all of which are in Group V of the periodic table. Figure 2.4 shows the bond model of silicon with the addition of an arsenic atom on a substitutional lattice site. Since arsenic has five valence electrons, it is able to form bonds with all four neighboring silicon atoms. The fifth valence electron is left after the bonds are

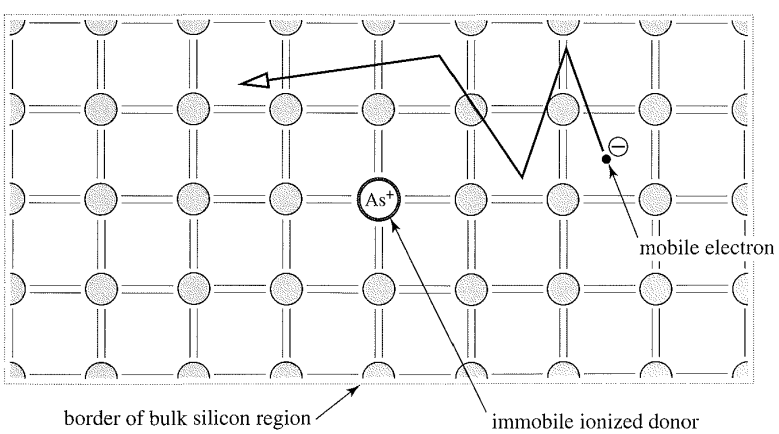


Figure 2.4 Bond model with arsenic incorporated into a substitutional site in a bulk region of the silicon crystal. Due to thermal excitation, arsenic loses (donates) its fifth valence electron as a mobile electron resulting in an immobile positive charge.

formed and is very loosely bound to the arsenic atom. For temperatures $T \gtrsim 100$ K = –173° C, there is enough thermal energy for this fifth electron to leave and move through the lattice, leaving behind an immobile, positively charged arsenic ion.

The charges in the bulk silicon region in Fig. 2.4 are the key to quantifying the electron and hole concentrations for arsenic-doped silicon. The region was electrically neutral before the arsenic was incorporated into the lattice. After the arsenic is added and ionizes, which occurs for $T \gtrsim 100$ K = –173° C, the region is still neutral since the negative electron charge is balanced by the positive charge on the arsenic ion. Adding up all the charge densities (units: C / cm^3) with their proper signs, we require that the net charge density ρ is zero:

$$\rho = -qn_0 + qp_o + qN_d = q(p_o + N_d - n_o) = 0 \tag{2.11}$$

The concentration of dopants is extremely low and their presence in the lattice does not significantly alter the mass-action law in thermal equilibrium. Solving Eq. (2.7) for the hole concentration and substituting into Eq. (2.11), we find that

$$\frac{n_i^2}{n_o} + N_d - n_o = 0 \tag{2.12}$$

By multiplying both sides by the electron concentration n_o and rearranging, the result is a quadratic equation

$$n_o^2 - N_d n_o - n_i^2 = 0 \tag{2.13}$$

Using the quadratic formula, we find that the electron concentration in silicon doped with donors is

$$n_o = \frac{N_d + \sqrt{N_d^2 + 4n_i^2}}{2} = \frac{N_d}{2} + \frac{N_d}{2}\sqrt{1 + \frac{4n_i^2}{N_d^2}} \tag{2.14}$$

where we have taken the positive root since the electron concentration must be positive. The hole concentration p_o can be found from the electron concentration using the mass-action law in Eq. (2.7).

In practice, the donor concentration is much greater than the intrinsic concentration $N_d \gg n_i$. The square root in Eq. (2.14) is approximately one and the electron concentration n_o is

n-type electron concentration

$$\boxed{n_o \cong \frac{N_d}{2} + \frac{N_d}{2} = N_d} \tag{2.15}$$

Silicon doped with $N_d \gg n_i$ is called **extrinsic**. The electron concentration n_o in extrinsic silicon is determined by the donor concentration and is not a function of temperature-dependent thermal generation. The result in Eq. (2.15) can be easily remembered since it corresponds to "one electron per donor," which can be seen from the bond model in Fig. 2.4. The relatively large donor concentration means the thermally generated electrons can be neglected, as quantified by Eq. (2.14).

The thermal equilibrium hole concentration in silicon doped with donors can be found from the mass-action law

$$p_o = \frac{n_i^2}{n_o} \cong \frac{n_i^2}{N_d} \qquad p_o = \frac{10^{20}}{N_d}\ \text{cm}^{-3} \qquad (T = 300\text{K}) \tag{2.16}$$

n-type hole concentration

For example, for a donor concentration of $N_d = 10^{14}$ cm^{-3}, the hole concentration at room temperature is only $p_o = 10^6$ cm^{-3}—a factor of 100 million lower than the electron concentration! By adding donor impurities to the silicon sample, the mass-action law in Eq. (2.7) depresses the hole concentration. Few holes remain in equilibrium since recombination with the large number of electrons removes most of the them. Silicon doped with donors is referred to as **n-type**, in which electrons are the **majority carriers** (by far the most numerous) and holes are the **minority carriers.**

The lower limit of useful donor concentration is set by the requirement that $N_d \gg n_i$ in Eq. (2.14), so $N_{d_{min}} \cong 10^{12}$ cm^{-3}. On the opposite extreme, there is a maximum amount of substitutional impurities that the silicon lattice can absorb without forming a new material phase with a different crystal structure and electronic properties. This maximum varies among the donor elements, but a typical value is around $N_{d_{max}} \cong 10^{19}$ cm^{-3}. This corresponds to about 1 in 5000 lattice sites occupied by a donor since the atomic density of silicon is $N_{Si} = 5 \times 10^{22}$ cm^{-3}.

➤ EXAMPLE 2.1 Electron and Hole Concentrations in n-type Silicon

Find the concentration of holes and electrons for a bulk silicon sample in thermal equilibrium at room temperature that has a donor concentration $N_d = 10^{14}$ cm^{-3}.

SOLUTION

Using Eq. (2.14), the electron concentration is

$$n_o = \frac{10^{14}}{2} + \frac{10^{14}}{2}\sqrt{1 + \frac{4 \cdot 10^{20}}{10^{28}}} = 1.00000001 \times 10^{14} \quad \text{cm}^{-3}$$

The hole concentration p_o can be found from the mass-action law in Eq. (2.7) since the sample is in thermal equilibrium

$$p_o = \frac{n_i^2}{n_o} = \frac{(10^{10})^2}{1.00000001 \times 10^{14}} = 9.9999999 \times 10^5\ \text{cm}^{-3}$$

Verify that the bulk silicon region is electrically neutral by adding the charge densities of the holes, ionized donors, and electrons

$$\rho = q(p_o + N_d - n_o) = q(9.9999999 \times 10^5 + 1 \times 10^{14} - 1.00000001 \times 10^{14}) = 0$$

Note that since N_d exceeds n_i by four orders of magnitude, we could have used Eq. (2.15) to find $n_o \cong N_d = 10^{14}$ cm^{-3}, and then used the mass-action law to find the hole concentration to be $p_o \cong 10^6$ cm^{-3}.

2.3.2 Acceptors

A substitutional impurity from Group III of the periodic table that has only three valence electrons is called an **acceptor**. The usual acceptor atom for silicon is boron (B). Figure 2.5 is the bond model's picture of doping by acceptors.

Since the Group III boron atom only has three valence electrons, it can complete only three of the four bonds with its silicon neighbors. However, an electron in an adjacent bond requires only a small amount thermal energy to jump over and complete the fourth bond. This process of "accepting" an electron accounts for the name of this class of dopants. The resulting hole is free to move around the crystal, whereas the immobile acceptor takes on a negative charge due to the extra electron as shown in Fig. 2.5.

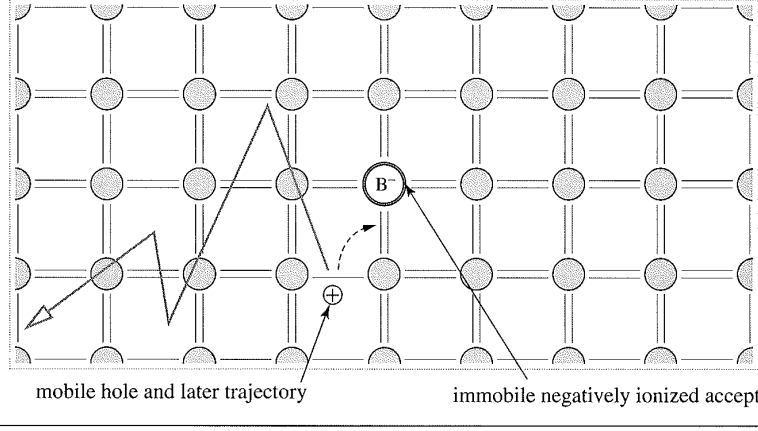
Given that the acceptor concentration is N_a (cm^{-3}), the equilibrium hole and electron concentrations can be found using the same approach used for donors. The bulk silicon region is electrically neutral and the hole, electron, and ionized acceptor charge densities must therefore sum to zero:

$$\rho = -qn_o + qp_o - qN_a = q(p_o - n_o - N_a) = 0 \tag{2.17}$$

The electron concentration can be eliminated by substituting $n_o = n_i^2 / p_o$ from the mass-action law in Eq. (2.7)

$$p_o - \frac{n_i^2}{p_o} - N_a = 0 \tag{2.18}$$

Using the quadratic formula to solve this equation yields an expression for the hole concentration in silicon doped with acceptors



➤ **Figure 2.5** Bond model showing boron on a substitutional lattice site in a bulk region of the silicon crystal. A hole is created when the incomplete fourth bond is filled by acceptance of an electron from a neighboring bond.

$$p_o = \frac{N_a}{2} + \frac{N_a}{2}\sqrt{1 + \frac{4n_i^2}{N_a^2}} \tag{2.19}$$

For the typical case where N_a is orders of magnitude larger than n_i, the intrinsic hole concentration, the square root term is approximately one and the hole concentration is

$$\boxed{p_o \cong N_a} \tag{2.20}$$

p-type hole concentration

By applying the mass-action law, we find that the equilibrium electron concentration at room temperature is given by

$$\boxed{n_o = \frac{n_i^2}{p_o} \cong \frac{n_i^2}{N_a} = \frac{10^{20}}{N_a}\ \text{cm}^{-3} \text{ at } 300\ \text{K}} \tag{2.21}$$

p-type electron concentration

Silicon doped with acceptors is called **p-type**, in which holes are the majority carriers and electrons are the minority carriers. P-type silicon is **complementary** to n-type silicon, in which the relationship between holes and electrons is reversed.

Due to the mass-action law, doping with $N_a = 10^{15}\text{cm}^{-3}$ of boron results in a hole concentration $p_o = 10^{15}\text{cm}^{-3}$ and electron concentration $n_o = 10^5\ \text{cm}^{-3}$. As with donors, the minimum useful acceptor concentration to have extrinsic, p-type silicon is determined by the intrinsic concentration n_i. At room temperature $N_{a_{min}} \approx 10^{12}\ \text{cm}^{-3}$ for $N_a \gg n_i$. The maximum concentration of substitutional boron acceptors in silicon is on the order of $N_{a_{max}} \approx 10^{19}\ \text{cm}^{-3}$ since a boron-silicon compound begins to precipitate at higher concentrations.

2.3.3 Donors and Acceptors: Compensation

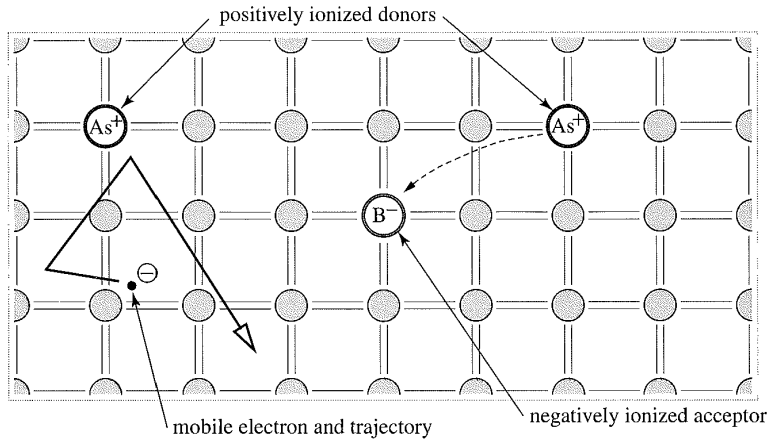
Silicon is often doped with both donors *and* acceptors. To find the equilibrium electron and hole concentrations, first consider the bond model in Fig. 2.6. Using the same reasoning used for the donor-only and acceptor-only cases, it can be seen that charge neutrality applies to the bulk silicon region in Fig. 2.6. The charge densities from electrons, holes, ionized donors, and ionized acceptors must therefore sum to zero

$$\rho = -qn_o + qp_o + qN_d - qN_a = q(p_o - n_o + N_d - N_a) = 0 \tag{2.22}$$

When $N_d > N_a$, we can substitute for the hole concentration in thermal equilibrium $p_o = n_i^2 / n_o$ from the mass-action law and find

$$n_o^2 - (N_d - N_a)\, n_o - n_i^2 = 0 \tag{2.23}$$

This quadratic equation is identical to Eq. (2.13) with the **net** donor concentration $N_d - N_a$ replacing the donor concentration. The solution can be found by substitution into Eq. (2.14)



➤ **Figure 2.6** Bond model showing a bulk section of silicon that is doped with both arsenic (donors) and boron (acceptors) with $N_d > N_a$. Donated electrons ionize the acceptors by completing their half-filled bonds, with the excess $N_d - N_a$ becoming mobile electrons. This process is called **compensation.**

$$n_o = \frac{(N_d - N_a)}{2} + \frac{(N_d - N_a)}{2}\sqrt{1 + \frac{4n_i^2}{(N_d - N_a)^2}} \tag{2.24}$$

where we assume that the donor concentration is greater than the acceptor concentration $N_d > N_a$. In practice, the excess of donors over acceptors $N_d - N_a$ will usually greatly exceed the intrinsic concentration since separately $N_d \gg n_i$ and $N_a \gg n_i$. For example, if the donor concentration exceeds the acceptor concentration by only 5% at a typical doping level of about $10^{15}\,\text{cm}^{-3}$, then

$$N_d - N_a = 1.05 \times 10^{15}\text{cm}^{-3} - 1.0 \times 10^{15}\text{cm}^{-3} = 5 \times 10^{13}\text{cm}^{-3} \gg n_i \tag{2.25}$$

Assuming $N_d - N_a \gg n_i$, we find that

compensated n-type electron concentration

$$\boxed{n_o \cong N_d - N_a \qquad (N_d > N_a)} \tag{2.26}$$

The physical interpretation of Eq. (2.26) is that electrons donated by the donors first ionize all the acceptors, with only the net donor concentration in excess of the acceptor concentration contributing mobile electrons to the crystal. This process is called **compensation**.

Since the sample is in thermal equilibrium, the minority hole concentration can be found by the mass-action law

compensated n-type hole concentration

$$\boxed{p_o \cong \frac{n_i^2}{N_d - N_a} \qquad (N_d > N_a)} \tag{2.27}$$

At first glance, the fact that there are negligible holes in a sample doped with acceptors seems counter-intuitive. However, the mass-action law depresses the hole concentration since the (majority) electron concentration far exceeds n_i.

For the case where the acceptor concentration is greater than the donor concentration, similar reasoning shows that the sample is p-type with

$$p_o \cong N_a - N_d \qquad n_o \cong \frac{n_i^2}{N_a - N_d} \qquad (N_a > N_d) \tag{2.28}$$

compensated p-type carrier concentration

In summary, a silicon crystal that is doped with one type of impurities may be compensated by doping with an excess of impurities of the opposite type. In section 2.5, we will see that compensation is a cornerstone of integrated device fabrication.

➤ EXAMPLE 2.2 Electron and Hole Concentrations for Compensated Silicon

a) Find the thermal equilibrium electron and hole concentrations at room temperature for a bulk silicon region doped with donors at $N_d = 10^{14}\text{cm}^{-3}$ and with acceptors at $N_a = 5 \times 10^{13}\text{cm}^{-3}$.

SOLUTION

The sample is n-type since the donor concentration exceeds the acceptor concentration. Since the net donor concentration $N_d - N_a = 10^{14}\ \text{cm}^{-3} - 5 \times 10^{13}\ \text{cm}^{-3} = 5 \times 10^{13}\ \text{cm}^{-3} \gg n_i$, we can equate the electron concentration to the net donor concentration (Eq. 2.26)

$$n_o \cong N_d - N_a = 5 \times 10^{13}\ \text{cm}^{-3}$$

The mass-action law in Eq. (2.7) determines that the minority hole concentration is

$$p_o \cong \frac{n_i^2}{N_d - N_a} = 2 \times 10^6\ \text{cm}^{-3}$$

To verify that the region is electrically neutral, we add the contributions of the positive holes and ionized donors with the negative electrons and negative ionized acceptors

$$\rho = q(p_o + N_d - n_o - N_a) = q(2 \times 10^6 + 10^{14} - 5 \times 10^{13} - 5 \times 10^{13}) \approx 0$$

Note: The charge density would have been exactly zero if the exact expression for the compensated electron concentration in Eq. (2.24) had been used.

(b) Now consider a bulk n-type silicon sample with $N_d = 10^{14}\ \text{cm}^{-3}$. We would like to add acceptors to make the hole concentration $p_o = 5 \times 10^{13}\ \text{cm}^{-3}$. What acceptor doping concentration N_a is needed? What is the electron concentration?

SOLUTION

The desired hole concentration is much greater than the intrinsic concentration; therefore, the compensated sample is p-type. Using Eq. (2.28), we can solve for the acceptor concentration

$$p_o \cong N_a - N_d \quad \rightarrow \quad N_a \cong p_o + N_d = 5 \times 10^{13}\,\text{cm}^{-3} + 10^{14}\,\text{cm}^{-3} = 1.5 \times 10^{14}\,\text{cm}^{-3}$$

The electron concentration follows from the mass-action law

$$n_o = \frac{n_i^2}{p_o} = \frac{10^{20}}{5 \times 10^{13}} = 2 \times 10^{6}\,\text{cm}^{-3}$$

➤ 2.4 CARRIER TRANSPORT

At room temperature, the holes and electrons in silicon are in agitated, random motion. Their average velocity is called the **thermal velocity** v_{th}, and is about 10^7 cm/s—only a factor of 3000 less than the speed of light! Holes and electrons collide frequently by scattering off ionized impurities, crystal defects, surfaces, and each other. A typical value for the average **collision time** τ_c at room temperature in bulk silicon is about 1×10^{-13} s = 0.1 picoseconds. In between collisions, carriers on average will travel a distance

$$\lambda = v_{th}\,\tau_c = 10\ \text{nm} = 0.01\ \mu\text{m} \tag{2.29}$$

where λ is defined as the **mean free path**. State-of-the-art silicon microelectronic devices have dimensions that are at least 0.1 μm, but are typically much larger. Therefore, holes and electrons undergo many collisions while traversing a device structure. In the next section, this important fact will allow us to find simple equations that describe the average transport of charge carriers due to electric fields or to concentration gradients.

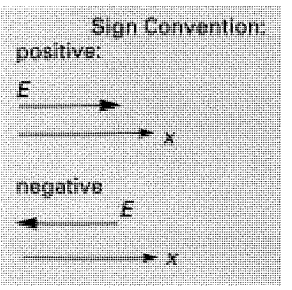
2.4.1 Drift Velocity

In thermal equilibrium, the electrons and holes in a region of the silicon lattice have no net motion. In other words, the average position of all the electrons (or separately, all of the holes) does not change with time. Figure 2.7(a) illustrates this situation for electrons. The random trajectories of three electrons along the x axis are shown in Fig. 2.7(a) for a time interval $\Delta t = 7\,\tau_c$, during which seven collisions occur. To show clearly the electron motion, the x-axis trajectories are spread out and all trajectories are shown originating at $x = x_i$. The average change in position after the interval Δt is

$$\overline{\Delta x} = \frac{(x_{f,1} - x_i) + (x_{f,2} - x_i) + (x_{f,3} - x_i)}{3} \approx 0 \tag{2.30}$$

using the approximate values for the position changes in Fig. 2.7(a). Of course, the average position change for electrons in a region of the silicon lattice will be taken

over a huge number of electrons. Since the trajectories are random, $\overline{\Delta x}$ will be zero after any time interval $\Delta t \gg \tau_c$.



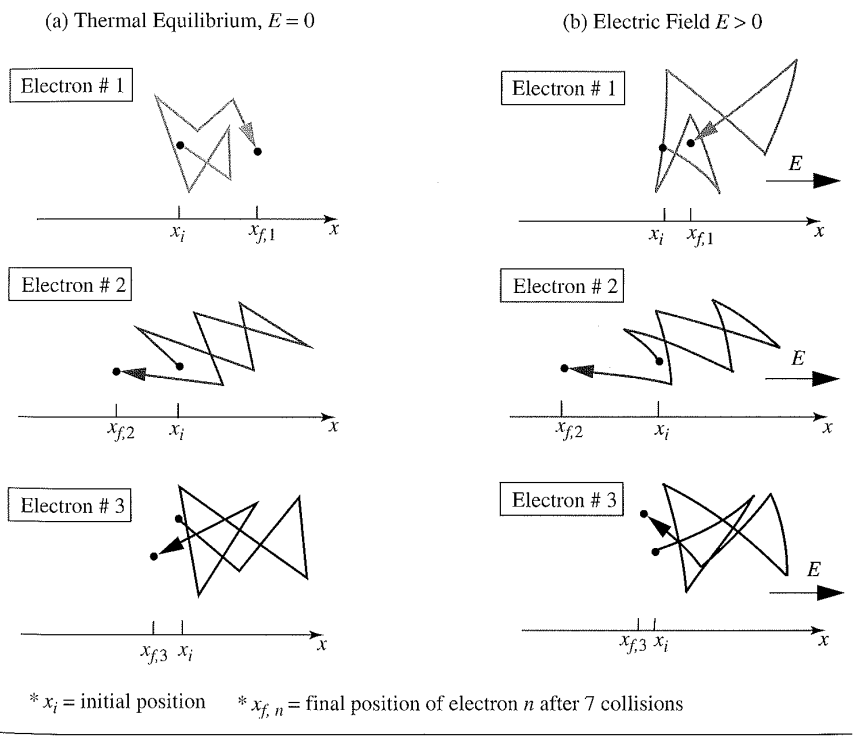
The situation with an applied electric field E is different, as shown in Fig. 2.7(b). It is important to first be sure of the sign for the electric field and other directional variables. By convention, the electric field is positive when it points in the same direction as the reference direction for the x axis. Thus, the electric field E in Fig. 2.7(b) is positive. The electrons are subject to an electrostatic force

$$F_{el} = -qE \tag{2.31}$$

that points in the $-x$ direction. During their "free flight" in between collisions, the electrons with mass m_n experience a constant acceleration $a = F_{el} / m_n = -qE/m_n$ that deflects their trajectories toward the $-x$ direction as shown in Fig. 2.7(b).

Averaging the position changes over the interval Δt, we find

$$\overline{\Delta x} = \frac{(x_{f,1} - x_i) + (x_{f,2} - x_i) + (x_{f,3} - x_i)}{3} < 0 \tag{2.32}$$



➤ **Figure 2.7** (a) Trajectories for three electrons in thermal equilibrium (zero applied electric field) for seven collisions. (b) Trajectories for the same three electrons for an electric field $E > 0$. All of the trajectories are drawn with the same initial position x_i.

using the three sample trajectories in Fig. 2.7(b). Dividing the average position change by the time interval, we define the **drift velocity** of the electrons as

$$v_{dn} = \frac{\overline{\Delta x}}{\overline{\Delta t}} \tag{2.33}$$

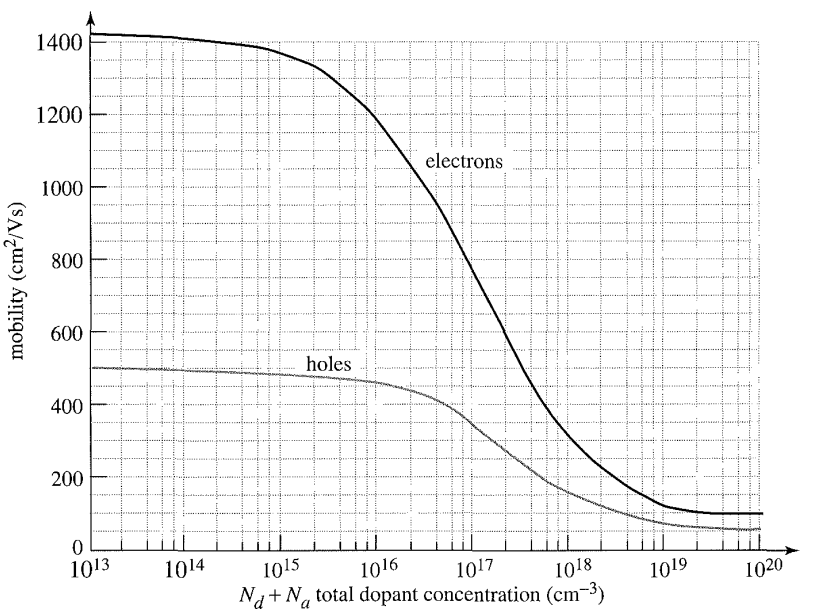
which is negative for $E > 0$. Note that the drift velocity is the average net velocity for a huge number of electrons in a region of the silicon lattice. Experimentally, the drift velocity v_{dn} is found to be proportional to the electric field

electron drift velocity

$$\boxed{v_{dn} = -\mu_n E} \tag{2.34}$$

where μ_n is the **electron mobility** with units of $cm^2/(Vs)$.

The electron mobility is a function of both the temperature and total doping concentration. Mobility decreases with increasing temperature since thermal agitation causes lattice vibrations to scatter electrons. Ionized dopants also scatter electrons and, therefore, reduce the collision time. This means the mobility is a decreasing function of $N_d + N_a$. Figure 2.8 gives experimental measurements of electron mobility for $N_d + N_a = 10^{13}$ to 10^{20} cm^{-3} at room temperature. It is useful to have a rough number for electron mobility in silicon for approximate calculations. Substituting a typical total dopant concentration of 2.5×10^{16} cm^{-3}, Fig. 2.8 gives an electron mobility of $\mu_n \approx 1000$ $cm^2/(Vs)$.



➤ **Figure 2.8** Linear-log plot of electron and hole mobilities at room temperature, as functions of the total doping concentration $N_d + N_a$. After R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed., Wiley, 1986.

Holes also drift in an electric field—the holes drifting *with* the field on average, because the positive charge causes them to accelerate in the same direction as E between collisions. The drift velocity for holes is also proportional to the electric field

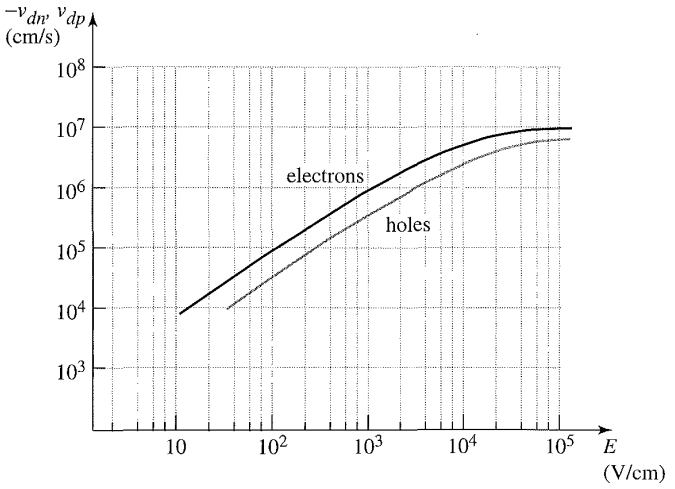
$$v_{dp} = \mu_p E \tag{2.35}$$

hole drift velocity

where μ_p is the **hole mobility**. In silicon, hole mobilities are lower than electron mobilities. They range from 100–500 $cm^2/(Vs)$ at room temperature as shown in Fig. 2.8. For a typical dopant concentration of 2.5×10^{16} cm^{-3}, the "round number" for hole mobility is $\mu_p \approx 400$ $cm^2/(Vs)$.

The linear relationship between drift velocity and electric field breaks down when the electric field is high enough that the drift velocity approaches $v_{sat} = 10^7$cm/s. Figure 2.9 shows $|v_d|$ as a function of electric field for electrons and holes and illustrates that both carriers have a maximum drift velocity of about 10^7 cm/s. This phenomenon is known as **velocity saturation**.

Advances in integrated-circuit technology are enabling ever-smaller electronic devices, a trend that is increasing the internal electric fields. Velocity saturation is an important phenomenon for accurate modeling of VLSI device characteristics. In this book, we will assume that the linear relationships between electric field and drift velocity in Eqs. (2.34) and (2.35) are valid unless velocity saturation is specifically mentioned.



➤ **Figure 2.9** Log-log plot of the drift velocity as a function of electric field for electrons and holes, showing the linear region where the velocity is proportional to the field and velocity saturation at high fields. After R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed., Wiley, 1986.

2.4.2 Drift Current Density

Drift of charged particles results in the transport of charge through the silicon lattice. We quantify the transport of charge by the **current density** J, which is defined as the rate at which charge passes through a reference plane in the $+x$ direction per unit area. The units for current density are C/(s cm^2) or A/cm^2. The sign convention for current density is illustrated in Fig. 2.10. J is positive when positive charge passes in the $+x$ direction. Since the charge carriers can be positive (holes) or negative (electrons), there are two ways that J can be positive.

1. Holes can pass through the reference plane in the $+x$ direction.
2. Electrons can pass through the reference plane in the $-x$ direction.

Note that either case results in a net transport of positive charge through the reference plane in the $+x$ direction making the current density J positive.

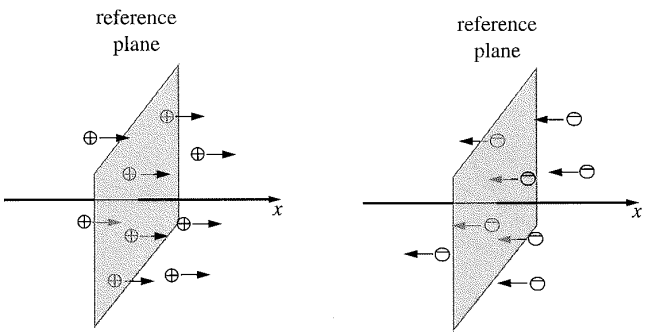
The hole current density due to drift in an electric field is expressed as J_p^{dr}— subscript p designates holes and superscript dr, drift. This can be expressed in terms of the electric field by directly applying the definition of current density, as shown in Fig. 2.11(a).

Over a time interval Δt, holes in the volume $\Delta V = A(v_{dp}\,\Delta t)$ will drift across the reference plane in the $+x$ direction. The number of holes ΔN_p that cross the plane in the interval Δt is given by the product of the concentration p of holes and the volume ΔV:

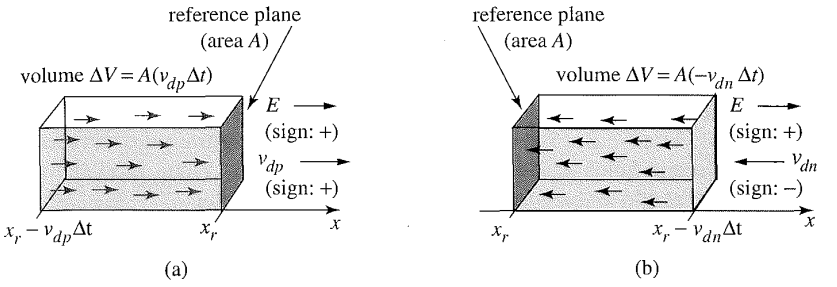
$$\Delta N_p = p\Delta V \tag{2.36}$$

The hole charge drifting across the reference plane ΔQ_p is the product of ΔN_p and the charge per hole q. According to the definition of current density

$$J_p^{dr} = \frac{\Delta Q_p}{A\Delta t} = \frac{qp\Delta V}{A\Delta t} = \frac{qpAv_{dp}\Delta t}{A\Delta t} = qpv_{dp} \tag{2.37}$$



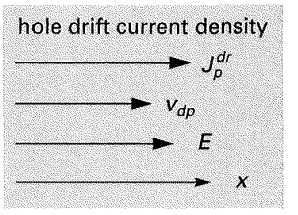
➤ **Figure 2.10** Sign convention for current density J. Both cases represent positive values of J since there is a net transport of positive charge in the $+x$ direction.



➤ **Figure 2.11** (a) The definition of current density applied to drifting holes, and (b) to drifting electrons. The reference plane is located at $x = x_r$ for this example. For an electric field pointed in the $+x$ direction, holes drift through the reference plane from left to right, whereas electrons drift from left to right. The current density $J^{dr} > 0$ for both cases.

Substituting for the drift velocity from Eq. (2.35), we find that

$$\boxed{J_p^{dr} = qp\mu_p E \text{ for } (v_{dp} \ll v_{sat})} \tag{2.38}$$



For electrons, the drift velocity is opposite in direction to the electric field as shown in Fig. 2.11(b). On average, electrons in a volume $\Delta V = A(-v_{dn}\,\Delta t)$ drift across the reference plane in the time interval Δt. The number of electrons crossing the plane is therefore $\Delta N_n = n\Delta V$, and the electron charge is equal to

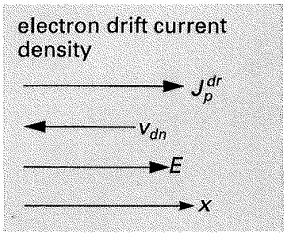
$$\Delta Q_n = (-q)\,\Delta N_n = -qnA\,(-v_{dn}\Delta t) \tag{2.39}$$

The electrons drift across the reference plane in the $-x$ direction as shown in Fig. 2.11(b). Therefore, the drift current density is given by

$$J_n^{dr} = -\left[\frac{\Delta Q_n}{A\Delta t}\right] \tag{2.40}$$

Substituting the electron charge from Eq. (2.39) and the drift velocity from Eq. (2.34), the electron drift current density is given by

$$J_n^{dr} = -\left[\frac{(qnAv_{dn}\Delta t)}{A\Delta t}\right] = -qnv_{dn} = -qn\,(-\mu_n E) = qn\mu_n E \tag{2.41}$$



Therefore, the electron drift current points in the same direction as the electric field. Restating Eq. (2.41)

$$\boxed{J_n^{dr} = qn\mu_n E, \text{ for } |v_{dn}| \ll v_{sat}}, \tag{2.42}$$

➤ EXAMPLE 2.3 Drift Current Density

Figure Ex2.3A shows a slab of p-type silicon ($N_a = 10^{16}$ cm^{-3}) under the influence of a uniform electric field E. Plot the drift current density as a function of E over the range -5×10^5 V/cm $< E < 5 \times 10^5$ V/cm. Note: Velocity saturation occurs for holes drifting in electric fields with magnitudes in excess of 10^5 V/cm. Find the voltage across the slab at which the current reaches its maximum value.

SOLUTION

For small electric fields, we can use Eq. (2.38) to find the relationship between E and the drift current density

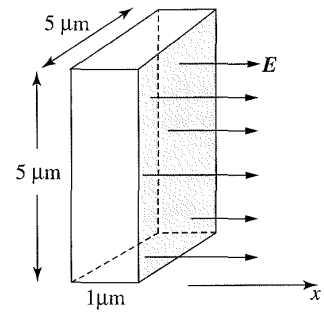
$$J_p^{dr} = qp\mu_p E = qN_a \mu_p E \qquad v_{dp} \ll v_{sat}$$

where we have substituted $p = N_a$ for the hole concentration. Even though the holes are drifting with the electric field and the slab is not in thermal equilibrium, their concentration remains approximately equal to the acceptor concentration. As we will see later in this chapter, metal coatings on the left and right faces of the silicon slab can supply and remove the drifting carriers.

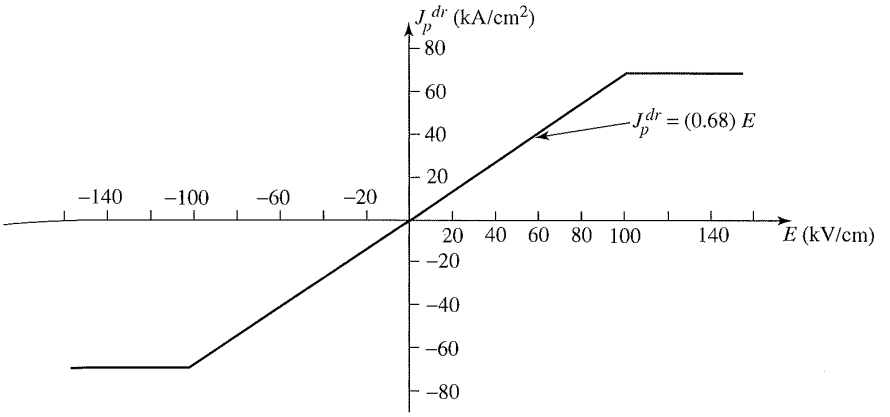
From Fig. 2.8, the hole mobility is $\mu_p = 425$ cm^2/(Vs) and the low-field drift current density is $J_p^{dr} = (0.68)\,E$ (units: J: A/cm^2, E: V/cm).

According to Fig. 2.9, this linear relationship is valid until the hole drift velocity saturates, which occurs at electric fields greater than $E_{crit} \approx 10^5$ V/cm = 100 kV/cm. Since we don't have a model for velocity saturation, we limit the drift current density at $J_{p,max}^{dr} = (0.68) \cdot E_{crit} = 68$ kA/cm^2 as shown in Fig. Ex2.3B

Substituting for the thickness of the slab, the voltage needed to saturate the drifting holes is $E_{crit} \times 1\ \mu$m = (10^5V/cm) (10^{-4} cm) = 10 V. The maximum current is given by the product of $J_{p,max}^{dr}$ and the area of the slab: $I_{max} = (68$ kA/cm$^2)$ $(5 \times 5 \times 10^{-8}$ cm$^2) = 17$ mA.



➤ **Figure Ex2.3A** P-type silicon slab with uniform applied electric field E.



➢ **Figure Ex2.3B** Drift current density through slab as a function of the applied electric field.

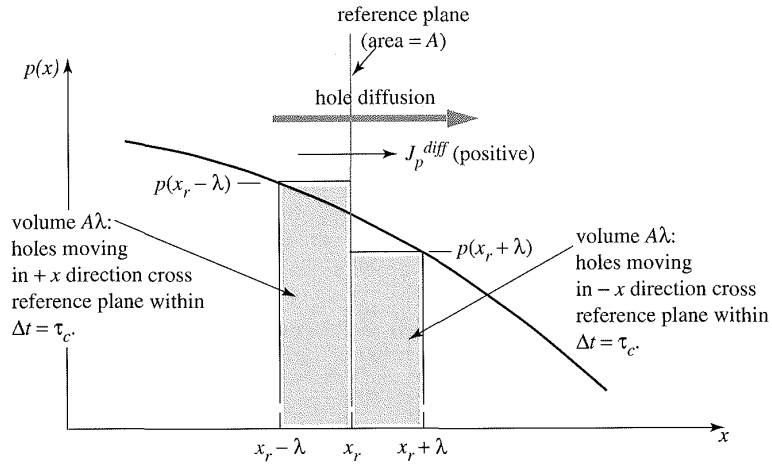
2.4.3 Diffusion Current Density

Diffusion is the process where gradients in concentration cause a net transport of particles. For example, suppose an ammonia bottle is opened in one corner of a room in which the air is at room temperature. Even if the air in the room is perfectly still, the random motions (with frequent collisions) of the air and ammonia molecules enable the transport of ammonia throughout the room. The net transport will be from regions of high concentration in the corner with the open bottle, to regions of low concentration in the opposite corner. If the bottle is capped, the concentration of ammonia will equalize throughout the room after a period of time. Diffusion reduces and eventually eliminates all gradients in ammonia concentration throughout the room.

Since electrons and holes also undergo random motion with frequent collisions, gradients in the electron or hole concentrations result in diffusion transport. In order to model diffusion, we consider holes having a concentration $p(x)$, which decreases with x as shown in Fig. 2.12. On the average, holes crossing the reference plane at $x = x_r$ started within one mean free path λ of x_r. The diffusion current density J_p^{diff} of holes across the reference plane has its origin in the difference in the number of holes in the two regions of volume $\Delta V = A\lambda$ to the left and right of the reference plane. Only half of the holes starting in either region will cross the reference plane in the time interval τ_c, while the other half are moving away from the reference plane. Therefore, the hole diffusion current density at $x = x_r$ is given by

$$J_p^{diff}(x=x_r) = q\left[\frac{\frac{1}{2}p\,(x = x_r - \lambda)\,A\lambda - \frac{1}{2}p\,(x = x_r + \lambda)\,A\lambda}{A\tau_c}\right] \tag{2.43}$$

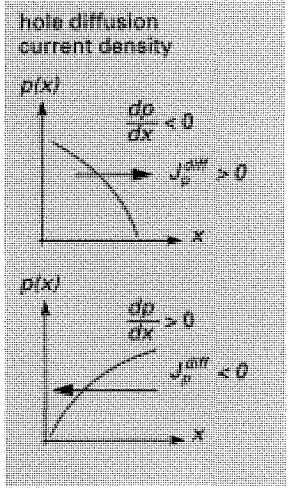
where $A\lambda$ is the volume. Note that holes crossing from the right to the left, opposite to the $+x$ direction, enter into Eq. (2.43) with a negative sign.



➤ **Figure 2.12** One-dimensional picture for quantifying the hole diffusion current density as a function of the gradient in hole concentration.

For Eq. (2.43) to be a good approximation, the mean free path (about 0.01 μm at room temperature) must be much smaller than the distance over which $p(x)$ is varying. This requirement is usually satisfied for microelectronic device structures. Note that we have highly exaggerated the gradient in the hole concentration in Fig. 2.12 to make the difference in the hole population across the reference plane clearly visible. Using first-order Taylor expansions for $p(x = x_r + \lambda)$ and $p(x = x_r - \lambda)$ and simplifying, we find that

$$J_p^{diff}(x = x_r) = \frac{q\lambda}{\tau_c}\left(\frac{\left(p(x_r) - \left.\frac{dp}{dx}\right|_{x_r}\lambda\right) - \left(p(x_r) + \left.\frac{dp}{dx}\right|_{x_r}\lambda\right)}{2}\right) \tag{2.44}$$



Canceling the hole concentration and collecting terms, Eq. (2.44) reduces to

$$J_p^{diff}(x = x_r) = -q\left(\frac{\lambda^2}{\tau_c}\right)\left.\frac{dp}{dx}\right|_{x = x_r} = -qD_p\left.\frac{dp}{dx}\right|_{x = x_r} \tag{2.45}$$

In Eq. (2.45), the term λ^2/τ_c is identified as the **hole diffusion coefficient** D_p, with units of cm^2/s. The diffusion current density for holes is

$$\boxed{J_p^{diff} = -qD_p\frac{dp}{dx}} \tag{2.46}$$

The proportionality of the diffusion current density to the concentration gradient is known as **Fick's Law** and is a fundamental result for diffusion processes. The negative sign in Eq. (2.46) reflects the fact that particles diffuse from regions of high con-

centration toward regions of low concentration. The concentration gradient dp/dx in Fig. 2.12 is negative, so the holes diffuse in the $+x$ direction and the diffusion current density is positive.

For the case of the electron concentration gradient shown in Fig. 2.13, we find the net transport of charge across the reference plane in the $+x$ direction as

$$J_n^{diff}(x=x_r) = (-q)\left[\frac{\frac{1}{2}n(x=x_r-\lambda)A\lambda - \frac{1}{2}n(x=x_r+\lambda)A\lambda}{A\tau_c}\right] \tag{2.47}$$

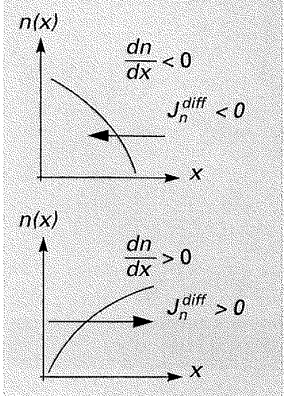
This is identical in form to Eq. (2.43) for holes, except we have accounted for the negative charge on the electrons crossing the reference plane. Again, use first-order Taylor expansions for the concentrations at $x_r - \lambda$ and $x_r + \lambda$

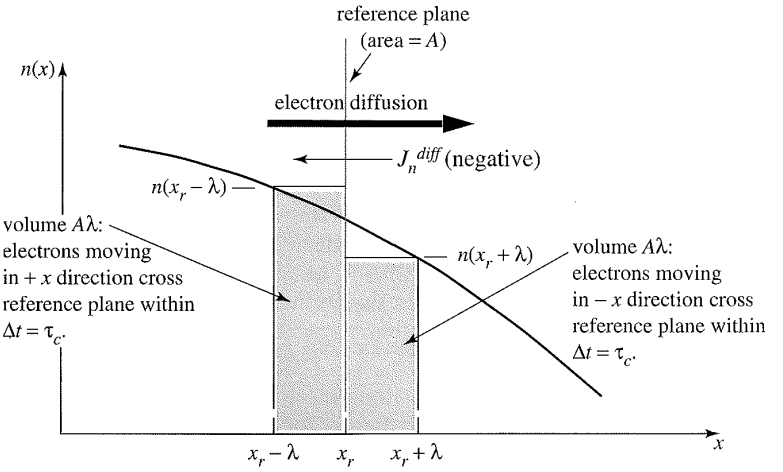
$$J_n^{diff}(x=x_r) = \frac{(-q)\lambda}{\tau_c}\left(\frac{\left(n(x_r) - \left.\frac{dn}{dx}\right|_{x_r}\lambda\right) - \left(n(x_r) + \left.\frac{dn}{dx}\right|_{x_r}\lambda\right)}{2}\right) \tag{2.48}$$

Collecting terms, the electron diffusion current density can be expressed by

$$\boxed{J_n^{diff}(x=x_r) = q\left(\frac{\lambda^2}{\tau_c}\right)\left.\frac{dn}{dx}\right|_{x=x_r} = qD_n\left.\frac{dn}{dx}\right|_{x=x_r}} \tag{2.49}$$

electron diffusion current density





➤ **Figure 2.13** One-dimensional picture for quantifying the electron diffusion current density as a function of the gradient in electron concentration. Note that the electrons diffuse left-to-right (in the $+x$ direction), but that J_n^{diff} points in the $-x$ direction and is negative.

where the **electron diffusion coefficient** is identified as D_n.The negative gradient in electron concentration ($dn/dx < 0$) shown in Fig. 2.12 means the electron diffusion current density is negative (i.e., it points in the $-x$ direction). The physical reason behind this result is that the electrons diffuse from regions of high concentration to regions of low concentration in the $+x$ direction in Fig. 2.13. However, each electron carries a charge $-q$, which results in J_n^{diff} pointing in the $-x$ direction.

The diffusion coefficients for holes and electrons are functions of doping concentration and temperature since these are determined by the mean free path and collision time. The mobility μ and diffusion coefficient D were derived from physical models based on large numbers of carriers undergoing random thermal motion with frequent collisions. In a solid-state physics course, it is shown that the two constants are proportional and their ratio is equal to

$$\frac{D}{\mu} = \frac{kT}{q} \tag{2.50}$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's Constant. This important result is called the **Einstein Relation**—named after Albert Einstein, who was active in solid-state physics in addition to his other interests. Eq. (2.50) holds separately for holes and for electrons

Einstein Relation

$$\boxed{\frac{D_p}{\mu_p} = \frac{kT}{q} \quad \text{and} \quad \frac{D_n}{\mu_n} = \frac{kT}{q}} \tag{2.51}$$

One implication of the Einstein Relation is that only the mobility need be measured. The diffusion coefficient can be found directly from Eq. (2.51). Using room-temperature values for the electron and hole mobilities in silicon (1000 and 400 cm^2/(Vs)), typical round numbers for the electron and hole diffusion coefficients are

$$D_n = 25 \text{ cm}^2/\text{s} \text{ and } D_p = 10 \text{ cm}^2/\text{s} \tag{2.52}$$

The **thermal voltage** is defined as the quantity kT/q:

thermal voltage

$$V_{th} = \frac{kT}{q} \tag{2.53}$$

The thermal voltage has units of Volts and will appear in many device physics equations. It ranges from 25 mV in a cool room in which the temperature is 17 °C or 62 °F to 26 mV in a somewhat warm room at 28.5 °C or 83 °F. In this book, we will choose either $V_{th} = 25$ mV or 26 mV—whichever results in "round numbers."

2.4.4 Total Current Density

As we have seen in the previous two sections, electrons and holes drift in an electric field and diffuse due to concentration gradients. The total current density J is the sum of all four components

$$\begin{aligned} J &= J_n + J_p = J_n^{dr} + J_n^{diff} + J_p^{dr} + J_p^{diff} \\ J &= qn\mu_n E + qD_n \frac{dn}{dx} + qp\mu_p E - qD_p \frac{dp}{dx} \end{aligned} \tag{2.54}$$

The total electron and total hole current densities J_n and J_p are important results, therefore they are given separately

$$J_n = J_n^{dr} + J_n^{diff} = qn\mu_n E + qD_n \frac{dn}{dx} \quad (2.55)$$

$$J_p = J_p^{dr} + J_p^{diff} = qp\mu_p E - qD_p \frac{dp}{dx} \quad (2.56)$$

electron and hole total current densities

In most cases, we will be able to simplify Eqs. (2.55) and (2.56) for a particular device structure and focus on the dominant components of the current densities.

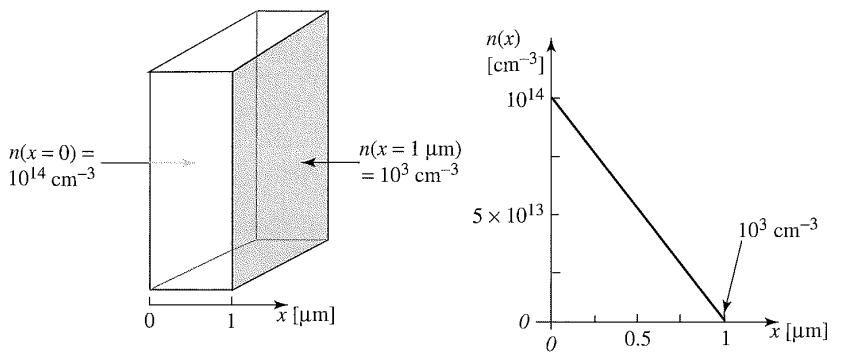
➤ EXAMPLE 2.4 Diffusion Current Density

Figure Ex2.4 shows a 1 μm-thick slab of p-type silicon having $N_a = 10^{17}$ cm^{-3}. By continuously injecting minority electrons on the left side at $x = 0$ and extracting them from the right side at $x = 1$ μm, we have set up a linear gradient in the electron concentration $n(x)$. What is the electron diffusion current density in the slab?

SOLUTION

On the linear plot of $n(x)$, we note that the electron concentration at $x = 1$ μm, 10^3 cm^{-3}, is indistinguishable from zero. For a total doping concentration of 10^{17}cm^{-3}, Fig. 2.8 indicates that the *electron* mobility is $\mu_n = 750$ cm^2/(Vs). Substituting into the Einstein Relation from Eq. (2.51), we find that the electron diffusion coefficient is

$$D_n = V_{th} \cdot \mu_n = (25 \times 10^{-3}\text{V}) \cdot (750\text{cm}^2/\text{Vs}) = 18.75\,\text{cm}^2/\text{s}$$



➤ **Figure Ex2.4** Electron concentration through a slab of p-type silicon.

We apply Eq. (2.49) to calculate the electron the diffusion current density in the silicon slab

$$J_n^{diff} = qD_n \frac{dn}{dx} = (1.6 \times 10^{-19})(18.75)\left(\frac{10^3 - 10^{14}}{10^{-4}}\right) = -3 \text{ A/cm}^2$$

Note that the current density is negative since the electrons are diffusing from left to right and carry a negative charge. The magnitude of the diffusion current density in this example is also small in comparison with typical majority carrier drift current densities. A voltage of less than 1 mV between the left and right faces of the slab is all that is needed to cause a drift current of majority holes of 3 A/cm^2.

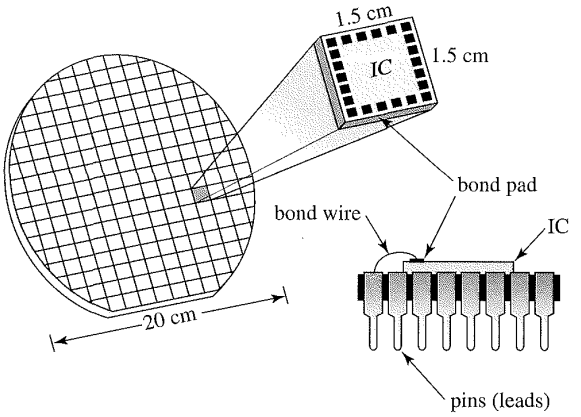
➤ 2.5 SILICON INTEGRATED CIRCUIT TECHNOLOGY

The focus of this text is silicon **integrated** devices and circuits, in which an electronic functional block, subsystem, or entire system is implemented on a single integrated circuit (IC). A basic appreciation of the manufacturing processes used to make ICs will be helpful in later chapters since it will give you an appreciation for the dimensional scales and physical structures of integrated devices. As you will see, the origin of the important "parasitic" circuit elements in ICs is obvious from the device structures.

This section provides a basic introduction to **IC fabrication technology**. This fascinating discipline draws from materials science and engineering, chemical engineering, and applied physics, as well as electrical engineering. By many measures, silicon IC technology is the crowning technological achievement of the late 20th century. Few areas of life have been unaffected by the penetration of silicon ICs such as microprocessors, microcontrollers, and modems. Since the fundamental physical limits of these products are at least two decades away, the early 21st century is certain to see further advances in the core fabrication processes that will have great impact on computers and other electronic systems. In addition, "spin-offs" from IC technology are catalyzing new fields such as micro electromechanical systems (MEMS). Silicon MEMS have display, sensing, or actuating functions. They may incorporate 10^6 micromechanical elements on the silicon chip along with a complex electronic interface. In summary, although the first silicon IC dates to around 1960, the end of the IC revolution is by no means in sight.

Figure 2.14 shows the fundamental unit of IC manufacturing—a silicon **wafer**. The typical wafer diameter is 20 cm in the mid-1990's. The wafer is a nearly perfect crystal that has been polished to be extremely flat. Wafers of this diameter cost about US$50.00 in large quantities. A group of around 24 wafers, called a **lot**, undergoes the same sequence of fabrication steps together. In a complex IC such as a microprocessor or dynamic random-access memory (DRAM), the process has around 200 steps and takes about three weeks to complete for a typical manufacturing schedule.

Figure 2.14 shows that when the process is complete, the wafer is tiled with an array of ICs interspersed with a few diagnostic test chips for monitoring the fabrication process. A typical complex IC has an area of about 1.5 cm × 1.5 cm, or about 2 cm^2, on which are several *million* electronic devices. For a 20 cm-diameter wafer, there will be around 150 chips of this area. The cost of the completed wafer is roughly US$2,000, which is a function of the fabrication process. How many of the chips work



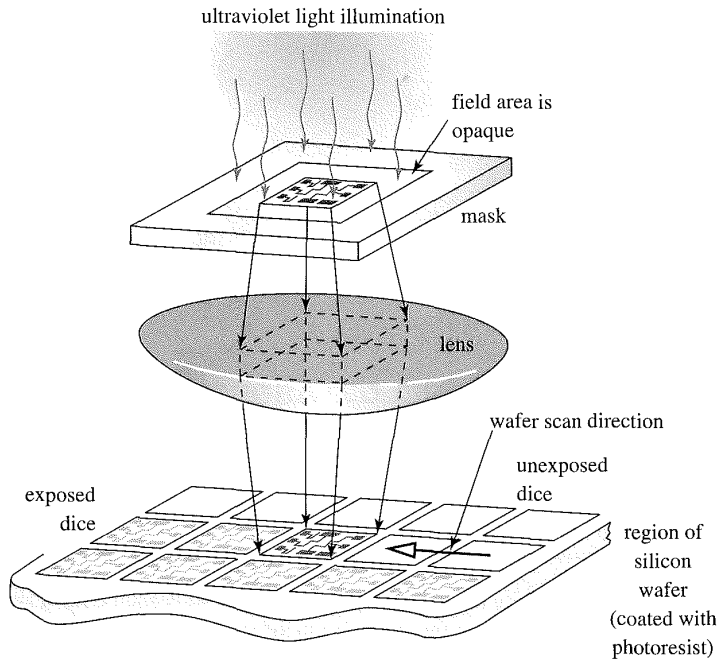
➢ **Figure 2.14** Silicon wafer and inset showing a single IC chip (die), with bond pads, and a side view of a dual in-line (DIP) package, shown with the lid removed and only a single bond wire. Wafer diameter and chip area are typical for mid-1990's large-volume, complex ICs.

(the **yield**) will affect the price at which the manufacturer sells each chip. The yield is a function of the quality control in the manufacturing plant, as well as the design margins built into the process and circuit. For example, a high-performance IC may have dimensions that approach the technological limits with the result that the yield is low. However, the high performance may justify a very high price and make the process profitable despite the low yield. As a particular IC design matures in production, the yield can be predicted to improve to around 90% or higher, which allows the selling price to drop steadily. The IC industry is extremely competitive. Companies aggressively track the **yield curve** of most products by dropping prices as yields increase. This maximizes volume and makes it more difficult for other companies to enter into direct competition on the product.

General-use ICs are packaged in plastic dual in-line (DIP) packages as shown in Fig. 2.14. In recent years, there has been great progress in economical packaging of ICs into high-density **multichip modules** for use in applications such as lap-top computers and cellular phones. Instead of the DIP shown in Fig. 2.14, chips are often encapsulated in low-profile surface mount packages or mounted directly on the circuit board using a "flip-chip" technique. Electrical connections between the board and inverted chip are made through electroplated solder bumps, rather than through bond wires, allowing for many more interconnects per chip.

2.5.1 Photolithography and Pattern Transfer

The essential process in IC fabrication is the imaging, alignment, and transfer of complex patterns onto the silicon wafer. A modern fabrication process uses around 15–20 separate patterning steps to define the transistors, diodes, and several levels of electrical interconnections (i.e., thin-film wires) that make up the IC. A **wafer stepper** is used to transfer patterns from an optical plate called a **mask**, to a photosensitive film coating the top surface of the wafer called **photoresist**. The process shown schematically in Fig. 2.15 depicts the exposure of a single die to the image of the mask.



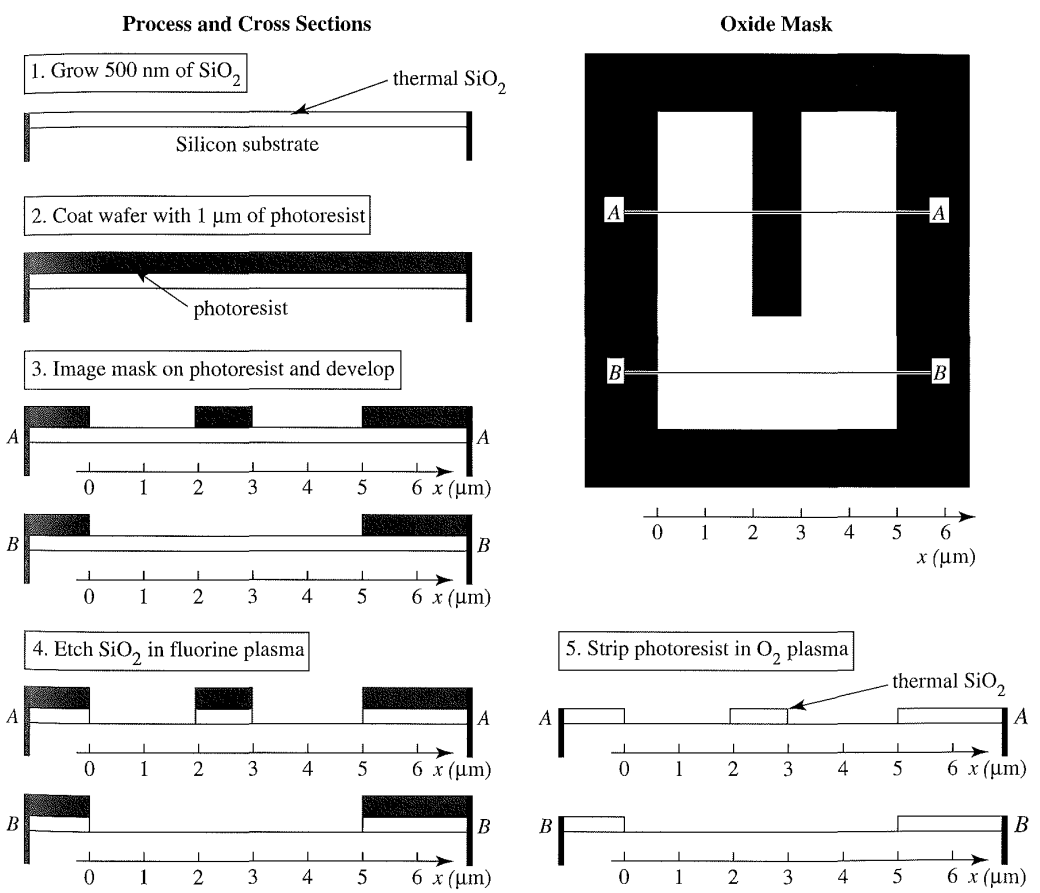
➤ **Figure 2.15** Schematic diagram showing imaging of the mask pattern on one die in a wafer stepper. The mask pattern is shown in gray on exposed dice. Unexposed dice are shown in outline only. In reality, the image in the photoresist is not visible until after development.

After completion of the exposure, the wafer stepper indexes to the next die, automatically aligns to special features previously etched on it, and repeats the exposure. Since the image of the mask pattern must be distortion-free and the error allowed in the alignment is less than 0.1 μm, it is not surprising that a high-volume wafer stepper costs well over US$2 million in the mid-1990s.

Figure 2.16 shows the step-by-step process of how the transfer from the mask to the wafer is accomplished through the photolithography and pattern transfer process. Figure 2.16 also introduces the customary way of visualizing three-dimensional IC structures. The **layout** consists of the series of masks used to create the IC structure. Lines across the layout denote the locations of **cross sections** through the silicon wafer. Typically, only the top few μm of the wafer are needed to show the device structure. In order to draw the cross sections, we must know the sequence of masking and fabrication steps. This sequence is called the **fabrication process**, or **process** for short. In analogy with food preparation, the fabrication process is sometimes called the **recipe**.

The single-mask photolithography and pattern transfer process in Fig. 2.16 involves five steps.

1. A 500 nm-thick film of silicon dioxide, SiO_2, is grown on the silicon wafer. In order to grow the thermal SiO_2 film, the wafer is placed in a furnace at an elevated temperature. Silicon surfaces react with oxygen to form a very high quality insulating film.
2. The wafer is uniformly coated with a 1 μm-thick film of photoresist by applying a mixture of photoresist and solvent, spinning the wafer at several thousand rpm for 30 seconds, and then baking to drive off the solvent.
3. Areas of this film that are exposed to ultraviolet light (UV) during the exposure step dissolve in an alkaline solution—a process known as **development**. Note carefully the differences in cross sections *A-A* and *B-B* after development of the mask image in the photoresist.



➤ **Figure 2.16** Cross sections for transferring the oxide mask pattern to a thermal oxide film. A shorthand description of this process is "grow 500 nm of thermal SiO_2 and pattern using the oxide mask."

4. After baking the photoresist to drive off residual solvents, the wafer is immersed in a vacuum chamber containing a fluorine **plasma**. A typical example of a plasma is the glow discharge in a fluorescent light. The plasma reacts with SiO_2 and etches it much faster than it etches the photoresist film. Once the silicon dioxide is etched away from the exposed areas, the underlying silicon is also etched, but very slowly. This enables the etch to be terminated before significant silicon erosion occurs.
5. The final step is to remove the photoresist in an oxygen plasma. This plasma aggressively attacks organic materials, such as photoresist, but does not etch silicon or silicon dioxide. Comparing the cross sections for steps 3 and 5 in Fig. 2.16, we see that the mask pattern has been successfully transferred to the silicon dioxide film.

As we mentioned, an IC process involves many photolithography and pattern transfer steps on deposited films. In order to avoid repeating steps 1–5 each time we use this procedure, we will summarize Fig. 2.16 by the statement: "Grow 500 nm of SiO_2 and pattern using the oxide mask." The photolithography and pattern transfer sequence can be thought of as a subroutine of the IC fabrication process.

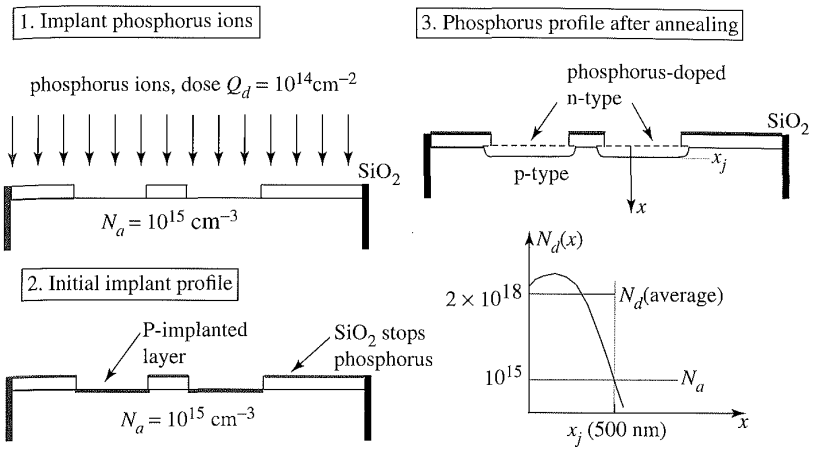
2.5.2 Doping by Ion Implantation

Acceptors and donors can be incorporated into selected regions of the top surface of a silicon wafer by means of **ion implantation**. Ions of boron, arsenic, or phosphorus are extracted from a plasma, accelerated to energies from 20 keV to 3 MeV (eV = electron volt), and formed into a tightly aligned ion beam. The ion beam is scanned over the surface of the silicon wafer. The beam is scanned until the desired **dose** of ions is implanted into the wafer. The ion dose has units of ions per unit area (cm^{-2}). After entering the silicon, the bombarding ions are slowed and then stopped by collisions with the silicon lattice.

Ion implantation creates a damaged region near the surface, since the collisions knock silicon atoms out of their positions in the lattice. The damaged region loses its crystal structure entirely and becomes **amorphous**. By heating the silicon wafer above 900° C, the damaged region recrystallizes, a process which is known as **annealing**. Remarkably, most of the dopant ions in the recrystallized region end up on substitutional lattice sites. During annealing, the implanted dopant ions also diffuse further into the silicon wafer.

Figure 2.17 illustrates how ion implantation can be used to dope specific regions of the wafer. We start with the patterned oxide film from Fig. 2.16, grown on a p-type wafer with a background donor concentration $N_a = 10^{15}$ cm^{-3}. A dose of $Q_d = 10^{14}$ cm^{-2} phosphorous ions is implanted into the wafer. Their energy is such that the phosphorus ions do not penetrate the SiO_2 film as shown in Fig. 2.17. After a thermal cycle, the phosphorus concentration $N_d(x)$ is found to exceed the background donor concentration up to a depth $x_j = 500$ nm, which is called the **junction depth**. Over the past 20 years, extensive research has investigated and modeled the details of dopant concentrations after implantation and annealing. Further discussion of this subject will be left to advanced courses on IC fabrication technology.

In this book, we will simply use the average donor concentration in the layer N_d as defined in the ion profile in Fig. 2.17. Since the dose Q_d of donors (units: cm^{-2}) is located in a layer of depth x_j after annealing, the average concentration is



➤ **Figure 2.17** Phosphorus implantation into the patterned oxide film shown in Fig. 2.16. The average concentration of phosphorus is defined in the plot of $N_d(x)$ after annealing.

$$N_d = \frac{Q_d}{x_j} \tag{2.57}$$

For the case shown in Fig. 2.17, $N_d = (10^{14}\ \text{cm}^{-2})/(5 \times 10^{-5}\ \text{cm}) = 2 \times 10^{18}\ \text{cm}^{-3}$.

2.5.3 Deposited Conducting and Insulating Films

In order to fabricate integrated circuits, we need conducting films that can be patterned into the microscopic wiring called **interconnects**. The interconnects enable electrical signals to be propagated around the IC. In addition, insulating films are needed to enable crossovers between two or more levels of interconnects. In the mid-1990s, three to five levels of interconnect are used for efficient packing and for minimum interconnect lengths in high-density digital ICs. As we will see in Chapter 5, short interconnects are important for achieving high switching speed.

Our first conductor is **polycrystalline silicon** or **polysilicon**. This material is an aggregate of crystallites or grains of single crystal silicon. Thin films of polysilicon are deposited on the surface of a silicon wafer at temperatures around 600 °C in a process known as **chemical vapor deposition** (CVD). Polysilicon is usually deposited with a maximum doping level of phosphorus. As we will see in the next section, polysilicon is a mediocre conductor; however, it has the virtues of being thermally stable and free of contaminants.

Aluminum is the most commonly used conducting film for interconnects. It is deposited by **sputtering**, a process in which material is deposited on the wafer from a nearby target that is bombarded by ions from a plasma discharge. One drawback of aluminum interconnects is that they cannot tolerate temperatures higher than about

450 °C due to degradation of the Al-Si junctions. Other metals used increasingly in ICs are tungsten and copper, both of which can be deposited by CVD processes.

There are several processes for depositing insulating films using CVD. The most common inter-metal insulator is silicon dioxide, SiO_2. Phosphorus and boron are often added during growth in order to allow the film to soften and flow slightly at temperatures around 900° C. This process, called **reflow**, is important in smoothing (or planarizing) the surface of the wafer after several depositions and masking steps have created too rough a surface. Silicon nitride, Si_3N_4, is another CVD insulating film that is able to mask a high-temperature thermal oxidation since it is impermeable to oxygen. In plasma-enhanced CVD processes, SiO_2 and Si_3N_4 can be grown at low temperatures (around 300° C) for use as a final protective layer over the IC.

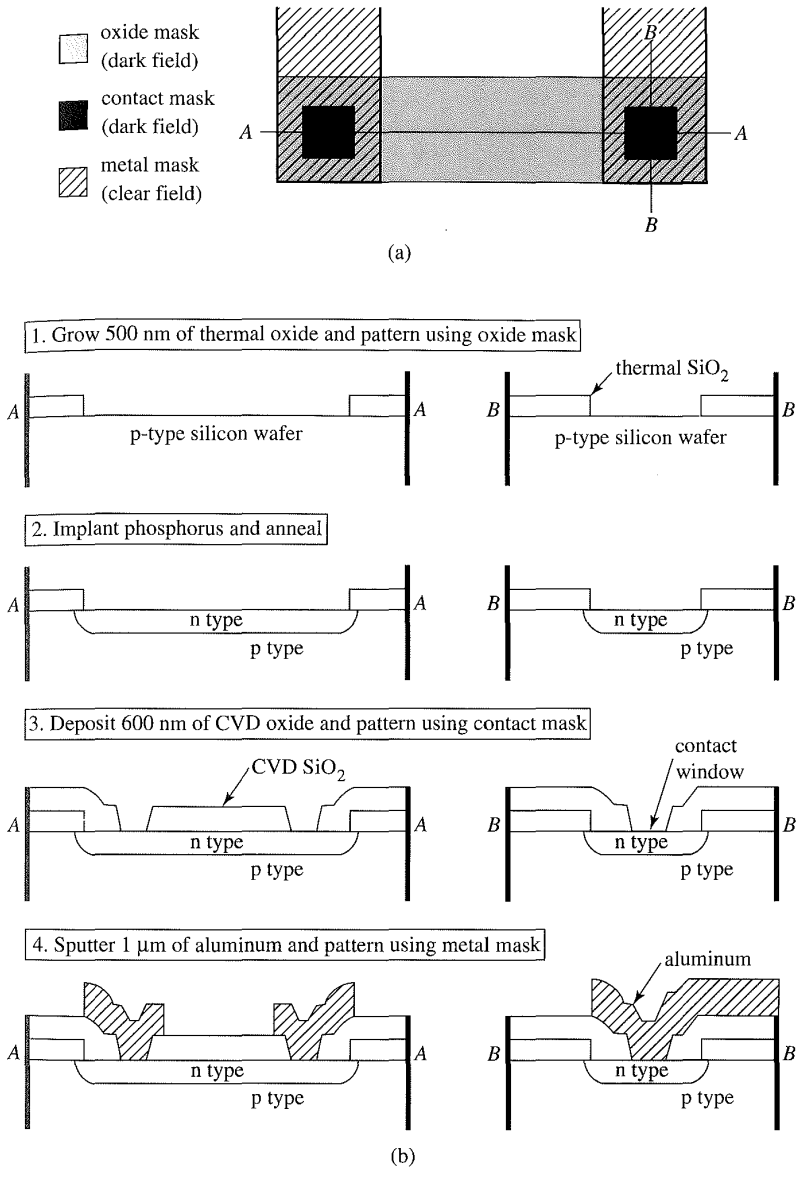
2.5.4 Layouts and Process Flows

You are now familiar with the lithography and etching processes and the basic materials that are the foundation of silicon IC technology. In this section, we will first describe a three-mask process for fabricating an ion-implanted resistor and then describe a more complex process to make an MOS field-effect transistor. You need not understand anything at this point about how these structures function as electronic devices. In following these sample processes, focus on the connection between the layout and the evolving device cross sections.

Figure 2.18(a) depicts the layout of the resistor, as it would be viewed in a computer-aided design (CAD) program. At first glance, you would expect that these patterns would correspond exactly to the actual masks used for photolithography. In fact, a mask that is nearly opaque (referred to as **dark field** mask) will obscure the other masks. Even with the use of color and a variety of "fill" and "border" designs, a nearly opaque mask will make it difficult to see features on the other masks in a layout. Fortunately, there is a simple solution: we display the negative, or complement, of the dark field mask in depicting the layout. **Clear field** masks have few opaque regions and are displayed directly. Of course, we must keep a careful record of which are clear field and which are dark field, so that the correct masks can be made. Many wafer lots have been ruined due to miscommunication on this minor, but critically important point.

In the resistor layout shown in Fig. 2.18(a), the oxide mask is identified as a dark field mask. Since the negative of the oxide mask is shown in the layout, the rectangle is clear on the actual mask. You should note carefully from cross sections *A-A* and *B-B* after step 1 that the oxide has been etched from the region of the wafer exposed by the clear rectangle. If this mask had been clear field, the oxide would have been removed everywhere *except* the rectangle, which would have been dark on the physical mask.

We now describe the process sequence in Fig. 2.18(b). The first step is to grow 500 nm of thermal oxide, which is done by heating the wafer in a furnace with oxygen flowing through it. This film is patterned using the oxide mask and the photolithography and etching process shown in Fig. 2.16. The wafer is then implanted with a dose of phosphorus ions that is sufficiently heavy to compensate the background acceptor concentration in the p-type silicon wafer. The thermal SiO_2 is thick enough to stop the phosphorus ions from penetrating into the underlying silicon substrate and is, therefore, effective at masking the implant.



➤ **Figure 2.18** (a) Layout and (b) process flow with cross sections *A-A* and *B-B* for fabrication of an ion-implanted resistor.

After annealing the wafer at high temperature, the phosphorus diffuses further into the substrate and creates a rectangular n-type region as shown in the cross sections after Step 2 in Fig. 2.18. Note that the phosphorus diffuses laterally under the oxide during the anneal, with the result that the dimensions of the n-type region are somewhat larger than the rectangle on the oxide mask. A chemical-vapor deposition (CVD) process is then used to deposit 600 nm of CVD SiO_2. Contact windows are patterned in this film using a second photolithography and etching step. Note that the contact mask is also a dark field mask. A 1 μm-thick film of aluminum is sputtered and patterned using the metal mask, which is a clear field mask. In order for the aluminum to form good electrical contacts to the n-type silicon region, the wafer must be heated to around 425 °C. It is helpful to observe the differences in the *A-A* and the *B-B* cross sections in the completed structure and to connect these to the layout.

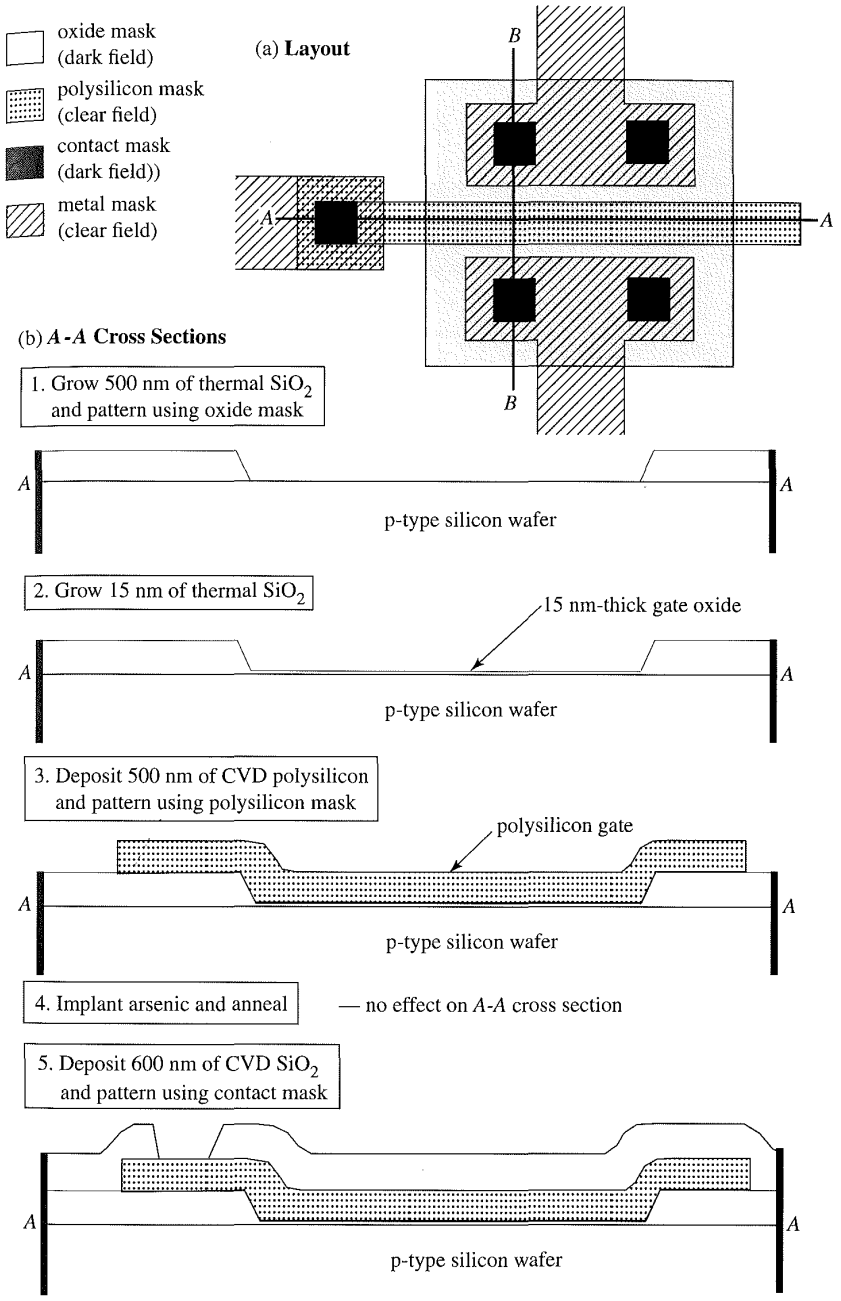
With only a slightly more complicated process, we can fabricate a "starter" version of an MOS field-effect transistor—the core device in digital ICs. Figure 2.19 gives the four mask patterns and the sequence of cross sectional views of the structure. After patterning a window in the initial 500 nm thick SiO_2 film in the first masking step, a 15 nm-thick SiO_2 "gate oxide" layer is grown in the bare silicon regions by a high-temperature oxidation process. Polysilicon is deposited by CVD with n-type doping and is patterned in the second masking step. The polysilicon film serves as the mask for a high-dose arsenic implant that penetrates the gate oxide. After annealing, n-type regions are formed on each side of the polysilicon "gate" that are seen only in the *B-B* cross section perpendicular to the gate. There is some lateral spreading of the arsenic implant during the anneal as seen in the *B-B* cross section. The next step involves the deposition of a CVD SiO_2 film, etching contact windows to the gate and to the two n-type regions adjacent to it. Finally, the MOSFET is completed by deposition and patterning of the aluminum interconnect layer.

An IC technology is often specified by λ, its minimum feature size. In an MOS technology, λ is the etched polysilicon gate length as shown in the final cross section in Fig. 2.19(b). The trend has been to ever-smaller dimensions, with state-of-the-art production processes at $\lambda < 0.5$ μm in the mid-1990s.

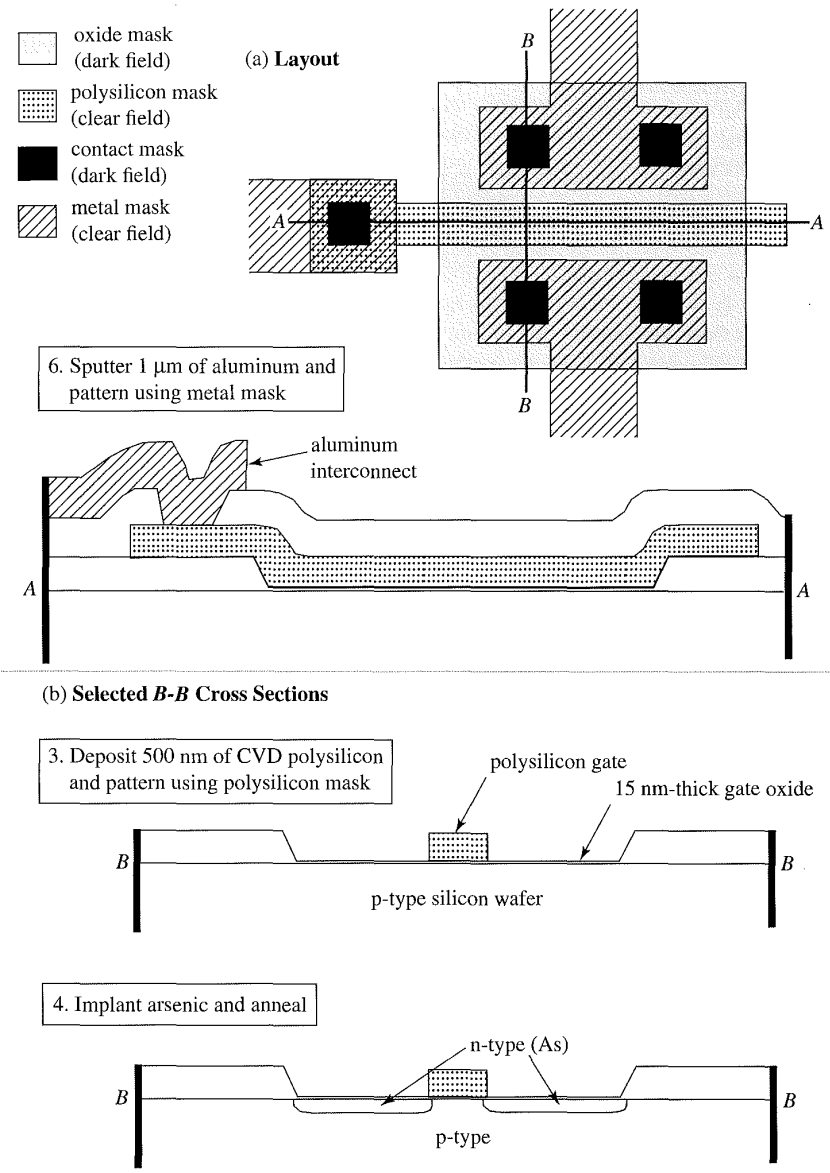
2.5.5 Geometric Design Rules

In Fig. 2.19(a), the contact nests inside both the oxide and metal patterns. The overlaps between patterns on different masks are specified in the **geometric design rules** for the process. These rules allow for systematic changes in the dimensions of the device structures, due to lithography, diffusion, or etching. They also incorporate knowledge of the random variations in these factors and the misalignment tolerances of the mask making and lithography equipment. There are "proximity" rules that specify the separation between the edges of patterns on different masks. Finally, there are rules that define the minimum line and space widths for each mask.

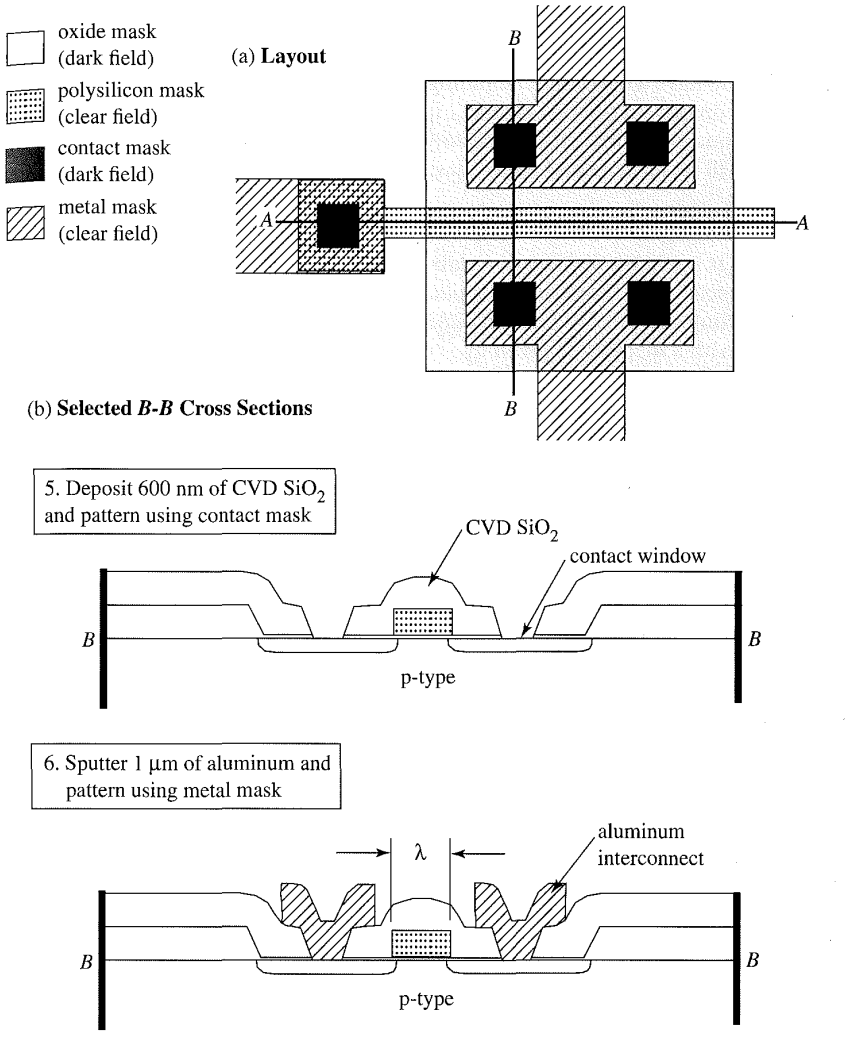
The yield of a particular product in a given process technology will be heavily influenced by the margins built into the design rules. With an IC containing 10^6 MOSFETs, each with at least three contacts, carefully chosen design rules for the contact mask nesting are essential for achieving a balance between high yield and a reasonable die size. Systematic experimentation and modeling of statistical variations are required to optimize the design rules for a particular technology and product family.



➤ **Figure 2.19** (a) Layout and (b) *A-A* cross sections for a simple MOSFET



➤ **Figure 2.19 (cont.)** (a) Layout and final *A-A* section and (b) selected *B-B* cross sections for a MOSFET.



➤ **Figure 2.19 (cont.)** (a) Layout and (b) selected *B-B* cross sections for the simple MOSFET.

➤ 2.6 IC RESISTORS

Now that you have a basic understanding of IC technology, we are prepared to model our first IC device—a silicon resistor. First we must derive Ohm's Law from the fundamental concept of drift current density. We will then re-express the resistance in a form that will be more useful for IC design.

2.6.1 Ohm's Law

Figure 2.20 is a schematic view of a slab of n-type silicon with two sides having metallized contacts. A DC voltage V is applied across the contacts. What is the magnitude of the resulting current I? Our first observation is that there is an electric field E in the silicon with a magnitude given by

$$E = \frac{V}{L} \tag{2.58}$$

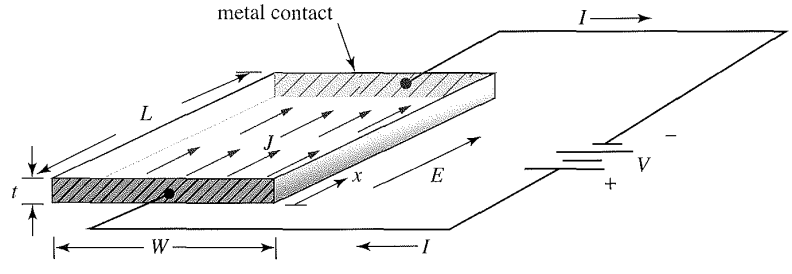
where L is the length of the slab in Fig. 2.20.

The electric field points in the direction of the x coordinate since it points from the contact with the higher potential (the nearer one in Fig. 2.20) to the contact with the lower potential. This electric field will cause electrons and holes in the silicon slab to drift, with the resulting drift current density being given by

$$J = J_n^{dr} + J_p^{dr} = qn\mu_n E + qp\mu_p E = (qn\mu_n + qp\mu_p)\left(\frac{V}{L}\right) \tag{2.59}$$

in which we have assumed that the electric field is low enough that velocity saturation is not an issue. There is no concentration gradient along the slab, since the mobile carriers are supplied and removed by the metal contacts. As a result, diffusion is ignored in Eq. (2.59). Since the current I [units: A] is the product of the current density J [units: A/cm^2] and the cross sectional area of the slab $A = W\,t$, we can find an equation relating current and voltage

$$I = J \cdot A = (qn\mu_n + qp\mu_p)\left(\frac{V}{L}\right)(Wt) = \left[(qn\mu_n + qp\mu_p)\left(\frac{Wt}{L}\right)\right] \cdot V \tag{2.60}$$



➤ **Figure 2.20** n-type silicon slab with length L, width W, and thickness t. A voltage V is applied across two metal contacts at opposite ends.

Ohm's Law states that $I = V/R$ and therefore, we can identify the resistance R of the slab from Eq. (2.60)

$$R = \frac{1}{(qn\mu_n + qp\mu_p)\left(\frac{Wt}{L}\right)} = \left(\frac{1}{(qn\mu_n + qp\mu_p)}\right)\left(\frac{L}{Wt}\right) \tag{2.61}$$

The first term on the right-hand side of Eq. (2.61) is defined as the **resistivity**, ρ [units: Ω cm]:

$$\rho = \frac{1}{(qn\mu_n + qp\mu_p)} \tag{2.62}$$

Substituting for the resistivity, Eq. (2.61) reduces to a simple expression

$$R = \rho\left(\frac{L}{Wt}\right) \tag{2.63}$$

In order to evaluate the resistivity, we must substitute the electron and hole concentrations n and p. The slab in Fig. 2.20 was assumed to be n-type with a donor concentration of N_d. Although the slab is not in thermal equilibrium, the carrier concentrations are not significantly perturbed from their equilibrium values by their constant drift across the slab. Substituting $n = N_d$ and $p = n_i^2/N_d$ for the n-type slab in Eq. (2.62), the resistivity ρ_n of n-type silicon is given by

n-type resistivity

$$\boxed{\rho_n = \left(qN_d\mu_n + \frac{qn_i^2\mu_p}{N_d}\right)^{-1} \cong \frac{1}{qN_d\mu_n}} \tag{2.64}$$

In other words, we can neglect the contribution of minority carriers in the drift current density since their concentration is orders of magnitude smaller than the majority carriers. For a p-type silicon region, holes are the majority carriers and the resistivity ρ_p is

p-type resistivity

$$\boxed{\rho_p = \left(q\frac{n_i^2}{N_a}\mu_n + qN_a\mu_p\right)^{-1} \cong \frac{1}{qN_a\mu_p}} \tag{2.65}$$

For doping concentrations of 10^{13} cm^{-3} to 10^{19} cm^{-3}, the resistivity of silicon ranges from about 500 Ω–cm to 5 mΩ–cm.

2.6.2 Sheet Resistance

Integrated-circuit resistors are fabricated from regions of one doping type in a substrate of opposite type. Figure 2.21 shows the layout and cross section of an n-type, ion-implanted IC resistor, that is fabricated using the process illustrated in Fig. 2.18. Metal interconnects make electrical contact at each end of the implanted region through contact windows etched through the deposited oxide film. The rectangular central portion of the implanted region has length L, width W, and thickness t. The

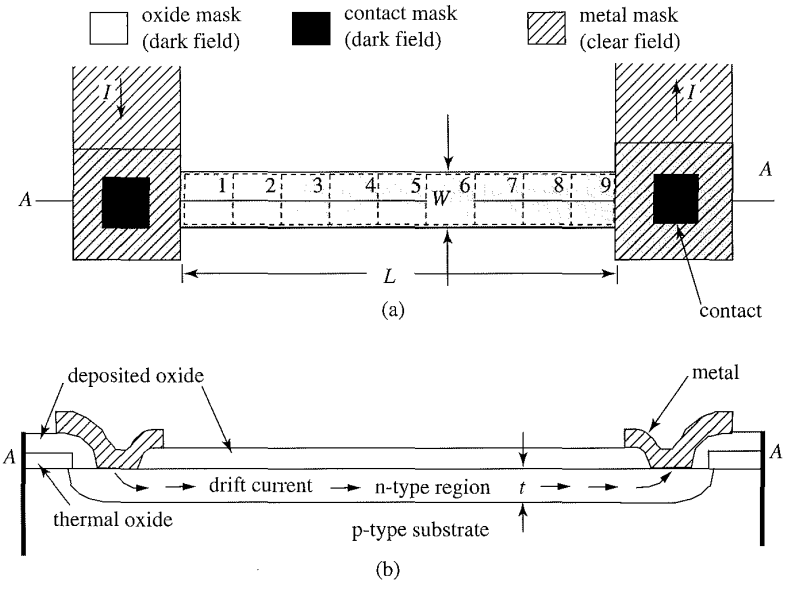
design rules require that the ends of the n-type region are widened to allow room for the contact windows.

If a voltage is applied across the contacts to the resistor, the resulting electric field will cause a lateral electron drift current density as shown in Fig. 2.21(b). As we will see in Chapter 3, if we apply positive voltages to both contacts with respect to the substrate, there is negligible current across the n-p junction between the n-type region and the p-type substrate. Hence, the n-type region is electrically isolated from the substrate. This simple structure hints at some of the difficulties in integrated circuit design, in that a p-n junction is the unwanted by-product of making an ion-implanted resistor. We will revisit the IC resistor in Example 3.6 in Chapter 3 and develop a more accurate model for it, after we have studied the reverse biased p-n junction.

From the previous section, the resistance of the rectangular region is given by

$$R = \rho_n\left(\frac{L}{Wt}\right) = \left(\frac{1}{qN_d\mu_n}\right)\left(\frac{L}{Wt}\right) \tag{2.66}$$

In IC design, it is very convenient to re-express Eq. (2.66) in a form that separates those parameters determined by the fabrication process (mobility μ_n, donor concentration N_d, and thickness t of the implanted layer) from those that are determined by



➤ **Figure 2.21** (a) Layout (b) cross section of an n-type resistor having $L/W = 9$ that is fabricated using the process sequence described in Fig. 2.18. Arrows show direction of electron drift current density for the case of a higher voltage on the left-hand contact.

the layout (length L and width W of the resistor). Collecting terms in Eq. (2.66) accordingly

sheet resistance

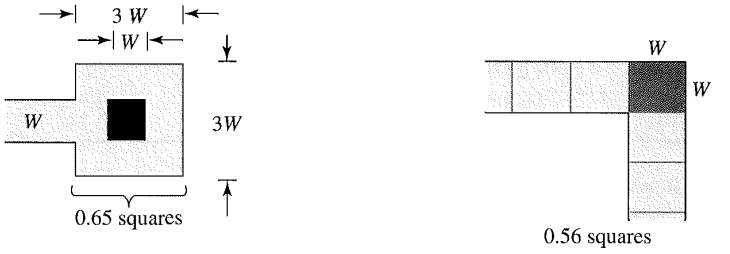
$$R = \left(\frac{1}{qN_d\mu_n t}\right)\left(\frac{L}{W}\right) = R_\square\left(\frac{L}{W}\right) = R_\square N_\square \tag{2.67}$$

in which we have defined the **sheet resistance** $R_\square = (q\,\mu_n N_d t)^{-1}$. The ratio of length to width is also defined as $N_\square = (L/W)$, the number of "squares." The units for sheet resistance are by convention Ω/□, pronounced "ohms per square." Of course, the number of "squares" is dimensionless. The layout in Fig. 2.21(a) illustrates the "square" concept by dividing the central portion into 9 segments of length W. In practical cases, the length L is often an integer multiple of the width W and $N_\square$ will be an integer.

For average doping levels of 10^{15} cm^{-3} to 10^{19} cm^{-3} and a typical layer thickness of $t = 1$ μm, the sheet resistance ranges from 50 kΩ /□ to 10 Ω /□. For other IC materials, heavily phosphorus-doped (n^+) gate polysilicon films are typically 500 nm thick and have $R_\square$ approximately 20 Ω /□. Aluminum interconnections are 1 μm in thickness and have a sheet resistance of around 30 mΩ /□—far lower than for the polysilicon or implanted silicon layers.

Why is sheet resistance such a useful parameter to the IC designer? The answer is that the parameters of the fabrication process such as the resistor thickness, doping concentration, and mobility are all "givens" for an IC designer. The layout shown in Fig. 2.21(a) is all that is under the IC designer's control in a particular process technology.

Thus far, we have neglected the contributions of the contact regions at the ends of the resistor to the total resistance. The details of how the current spreads out from the metal-silicon contact determine the resistance for a particular contact layout. We will neglect any additional resistance due to the metal-silicon junction not being an ideal ohmic contact. For the "dogbone" style contact shown in Fig. 2.22(a), each contact region adds 0.65 squares. It is convenient to express the additional resistances in units of squares, so that they can be counted into the total $N_\square$. In Fig. 2.22(b), the effect of the 2-D distribution of current density at a right angle bend is that the "corner square" contributes only 0.56 squares to the total $N_\square$. Accurate accounting for corners is important, since large-value resistors are typically folded in order to fit compactly into the layout. Example 2.5 illustrates the concept of sheet resistance for resistor layout.



➤ **Figure 2.22** Effective fractional square for (a) a "dogbone" contact area and (b) a corner. After R. C. Jaeger, *Introduction to Microelectronic Fabrication*, Addison-Wesley, 1988.

➤ EXAMPLE 2.5 IC Resistor Layout

(a) Given that a dose $Q_d = 10^{12}$ cm^{-2} of arsenic is implanted into p-type silicon, with a junction depth of $x_j = 400$ nm, estimate the sheet resistance of this n-type layer.

SOLUTION

The average arsenic concentration in the implanted layer is the thickness of the layer (in cm) divided into the dose (per cm^2)

$$N_d = \frac{Q_d}{x_j} = \frac{1 \times 10^{12}}{4 \times 10^{-5}} = 2.5 \times 10^{16}\,\text{cm}^{-3}$$

From Fig. 2.8, the electron mobility for the average concentration is $\mu_n = 1000$ cm^2/Vs. Substituting into the definition of sheet resistance in Eq. (2.67), we find that

$$R_\square = \frac{1}{q\mu_n N_d x_j} = \frac{1}{(1.6 \times 10^{-19})\ (1000)\ (2.5 \times 10^{16})\ (4 \times 10^{-5})} \cong 6\ \text{k}\Omega/\square$$

In letting $x_j = t$ in finding $R_\square$, we assume there is no penetration of the n-type region by the charged layer that forms at the p-n junction. This approximation is reasonable so long as the background acceptor concentration in the substrate, N_a, is much less than the donor concentration N_d in the implanted layer. We will quantify this approximation in Chapter 3. We have also made this same assumption in letting the electron concentration $n = N_d$, rather than $N_d - N_a$. For a typical substrate doping, these approximations lead to small errors.

(b) For this n-type layer, find the length L for a straight resistor (no corners) of value $R = 56$ kΩ. The minimum width $W = 2.5$ μm and the two contact areas have the layout shown in Fig. 2.22(a)

SOLUTION

First, the total number of squares needed is

$$N_\square = R/R_\square = 56\text{k}\Omega\ /\ (6\ \text{k}\Omega/\square) = 9.3\ \square$$

Subtracting off the contributions of the contacts, the rectangular portion of the resistor has

$$(L/W) = N_\square - 2\ (0.65) = 9.3 - 1.3 = 8$$

Therefore, the length $L = 8 \times 2.5\ \mu\text{m} = 20\ \mu\text{m}$.

2.6.3 Statistical Variations in IC Resistors

In order to use the IC resistor as a circuit element, its variation from lot to lot and over temperature must be understood. We will introduce a simple way to quantify the effect of manufacturing variations on the resistance, but will leave the subject of tem-

perature effects to a more advanced course. This section is applicable to other device structures since the sources of variation are fundamental to IC manufacturing.

From Eq. (2.66), the resistance is the product of the sheet resistance and the number of squares, each of which contains parameters or dimensions that vary from chip to chip:

$$R = \left(\frac{1}{qN_d\mu_n t}\right)\left(\frac{L}{W}\right) \tag{2.68}$$

Since a background in probability and statistics is *not* assumed, the variation of a particular quantity, such as the doping concentration N_d, will be expressed simply as

$$N_d = \overline{N_d}(1 \pm \varepsilon_{N_d}) , \tag{2.69}$$

where $\overline{N_d}$ is the average doping and ε_{N_d} is defined as the **normalized uncertainty**. It is convenient to use the normalized uncertainty rather than the absolute uncertainty in the particular quantity. We can determine the range of variation by multiplying the average by the normalized uncertainty. For example, a doping concentration defined by an average of $\overline{N_d} = 10^{17}\ \text{cm}^{-3}$ and a normalized uncertainty of $\varepsilon_{N_d} = 0.025$ means that

$$N_d = (10^{17}\ \text{cm}^{-3})\,(1 \pm 0.025) = 10^{17}\ \text{cm}^{-3} \pm 2.5 \times 10^{15}\ \text{cm}^{-3} \tag{2.70}$$

Returning to Eq. (2.68), we substitute for each term that is subject to manufacturing variations:

$$R = \left(\frac{1}{q\overline{N_d}(1+\varepsilon_{N_d})\,\overline{\mu_n}(1+\varepsilon_{\mu_n})\,\bar{t}\,(1+\varepsilon_t)}\right)\left(\frac{\bar{L}(1+\varepsilon_L)}{\bar{W}(1+\varepsilon_W)}\right) \tag{2.71}$$

It is reasonable that the average resistance should be

$$\bar{R} = \left(\frac{1}{q\overline{N_d}\,\overline{\mu_n}\,\bar{t}}\right)\left(\frac{\bar{L}}{\bar{W}}\right) \tag{2.72}$$

However, how do we predict the normalized uncertainty ε_R given the normalized uncertainties in the parameters and dimensions? Applying an important result from statistics, the normalized uncertainty in the resistance is the square root of the sum of the squares of the normalized uncertainties of each variable

$$\varepsilon_R = \sqrt{\varepsilon_{N_d}^2 + \varepsilon_{\mu_n}^2 + \varepsilon_t^2 + \varepsilon_L^2 + \varepsilon_W^2}\,. \tag{2.73}$$

This result is valid when the variation of one variable is independent of those of the other variables and when all variables follow the Gaussian, or normal, probability distribution. In practice, Eq. (2.73) is a good first-cut estimate for the case where the normalized uncertainties are relatively small (< 0.1). This "statistical" approach is convenient and we will use it to estimate the effects of process and layout variations—even though we are not going to verify the underlying assumptions.

We now turn to the origin of the variations in the terms in Eq. (2.71). The ion implanter will largely determine the run-to-run control on the doping concentration. Mobility variations are due to random local variations in crystal defects. Note that there is also coupling between the mobility and the doping concentration, as shown in Fig. 2.8. This coupling will be ignored in our first-cut estimate of uncertainty in Eq. (2.73). The thickness t of the resistor is subject to variations in the time and local wafer temperature during the furnace annealing process.

The lateral dimensions of the resistor are affected by random linewidth variations from the pattern transfer processes, as shown in Figure 2.23. Defining the uncertainty on one edge as $\delta/2$, the width and length of the resistor are

$$W = \bar{W} \pm \frac{\delta}{2} \pm \frac{\delta}{2} = \bar{W} \pm \delta \text{ and } L = \bar{L} \pm \frac{\delta}{2} \pm \frac{\delta}{2} = \bar{L} \pm \delta, \tag{2.74}$$

where we have added the uncertainties from each edge. The width and length can be rewritten in the form

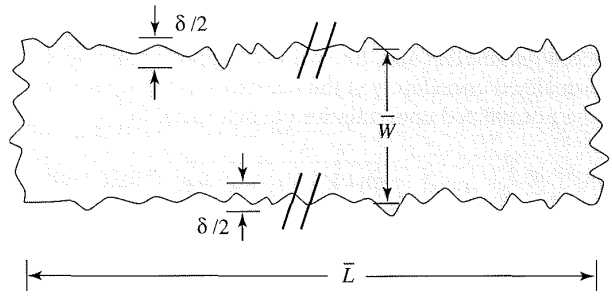
$$W = \bar{W}\left(1 \pm \frac{\delta}{\bar{W}}\right) = \bar{W}(1 \pm \varepsilon_W) \text{ and} \tag{2.75}$$

$$L = \bar{L}\left(1 \pm \frac{\delta}{\bar{L}}\right) = \bar{L}(1 \pm \varepsilon_L)\,. \tag{2.76}$$

From Eqs. (2.75) and (2.76), the normalized uncertainties in the width and length are inversely proportional to their average values:

$$\varepsilon_W = \delta/\bar{W} \text{ and } \varepsilon_L = \delta/\bar{L}. \tag{2.77}$$

Therefore, the designer has control over the normalized uncertainties in width and length and can make trade-offs between the resistor area (the product of width and length) and its normalized uncertainty. It is common for the length to greatly



➤ **Figure 2.23** Highly magnified top view of the resistor showing the uncertainty $\delta/2$ in the edges due to random variations in the photolithography and etching processes. Note that the sketch only shows the ends of the relatively long and narrow resistor.

exceed the width, in which case the normalized uncertainty in the length becomes negligibly small, according to Eq. (2.77)

What are the magnitudes of the process uncertainties that enter into Eq. (2.73)? To answer this question, we must first specify the collection of samples used to evaluate the average and variation. The variation is smallest for structures that are nearby on the same chip, since they are subjected to nearly identical processes. If an entire wafer is considered, then the spread in parameters and dimensions increases. The uncertainties increase further when a wafer lot is evaluated and are largest when measurements are made on samples from many lots. Finally, the IC process equipment and the organization and discipline of the workforce are major factors in setting the baseline levels of process uncertainties.

The statistics on process variations are highly proprietary for each manufacturer, needless to say. However, some rough numbers are useful for gaining an appreciation for the limits of IC technology. Nominally identical, adjacent ion-implanted resistors can have a normalized uncertainty as low as 10^{-3}. With careful attention to the layout, resistor ratios can have a similar level of precision. A further benefit of adjacent resistors is that they are at nearly the same temperature. As a result, their resistances track with changing ambient temperature. When the set of resistors is expanded to include samples from a wafer, a lot, or a series of lots, the normalized uncertainty increases by a factor of 10 to 100—not including the significant variation over temperature. As a result, the precision is too poor to allow a designer to use an IC resistor for circuits where its absolute value must be tightly controlled. The absence of precision integrated resistors has been a major motivator of new approaches to circuit design.

➤ EXAMPLE 2.6 IC Resistor Process Variations

An ion-implanted resistor with the cross section shown in Fig. 2.21 is designed to have a resistance $R = 20\ \text{k}\Omega$. The process parameters and their normalized uncertainties over a series of lots are: $N_d = (4 \times 10^{17}\ \text{cm}^{-3})(1 \pm 0.04)$, $\mu_n = (450\ \text{cm}^2/\text{Vs})(1 \pm 0.02)$, $t = 550\ \text{nm}(1 \pm 0.035)$.

(a) Find the average sheet resistance and its normalized uncertainty.

SOLUTION

Substituting the average values for the parameters into the definition of sheet resistance in Eq. (2.67), we find that

$$\bar{R}_{\square} = \frac{1}{(1.6 \times 10^{-19}\ \text{C})\,(4 \times 10^{17}\ \text{cm}^{-3})\,(450\ \text{cm}^2/\text{Vs})\,(0.55 \times 10^{-4}\ \text{cm})} = 630\ \Omega/\square.$$

The normalized uncertainty in the sheet resistance is given by

$$\varepsilon_{R_{\square}} = \sqrt{\varepsilon_{N_d}^2 + \varepsilon_{\mu_n}^2 + \varepsilon_t^2} = \sqrt{(0.04)^2 + (0.02)^2 + (0.035)^2} = 0.06$$

(b) Over several wafer lots, the linewidth edge variation is $\delta = 0.15\ \mu\text{m}$. For nominal resistor widths of $W = 2\ \mu\text{m}$ and $W = 4\ \mu\text{m}$, find the normalized uncertainty for the resistor.

SOLUTION

The number of squares required is:

$$N_{\square} = \frac{R}{R_{\square}} = \frac{20\ \text{k}\Omega}{630\ \Omega/\text{sq.}} = 31.75$$

Due to the large number of squares, the uncertainty in the length can be neglected. According to Eq. (2.77), the normalized uncertainty in the resistor width is

$$\varepsilon_W = \frac{0.15\ \mu\text{m}}{W} = 0.075\ (W = 2\ \mu\text{m}) \text{ and } \varepsilon_W = 0.038\ (W = 4\ \mu\text{m})$$

Therefore, the normalized uncertainties of the resistor for the two widths are

$$\varepsilon_R = \sqrt{\varepsilon_{R_{\square}}^2 + \varepsilon_W^2} = \sqrt{(0.06)^2 + (0.075)^2} = 0.096\ (W = 2\ \mu\text{m}) \text{ and}$$

$$\varepsilon_R = \sqrt{\varepsilon_{R_{\square}}^2 + \varepsilon_W^2} = \sqrt{(0.06)^2 + (0.038)^2} = 0.0625\ (W = 4\ \mu\text{m})$$

The nominally 20 kΩ IC resistor has an absolute uncertainty of nearly 2 kΩ for the narrow width layout and about 1.25 kΩ for the wider layout. The penalty for the wider layout is an increase in area by a factor of 4 since the length must also be doubled to keep the number of squares at 31.75.

For large resistor values, it is desirable to fold the resistor in order to make a more compact IC structure. We conclude the chapter with a design example which explores the geometry involved in designing for a square layout.



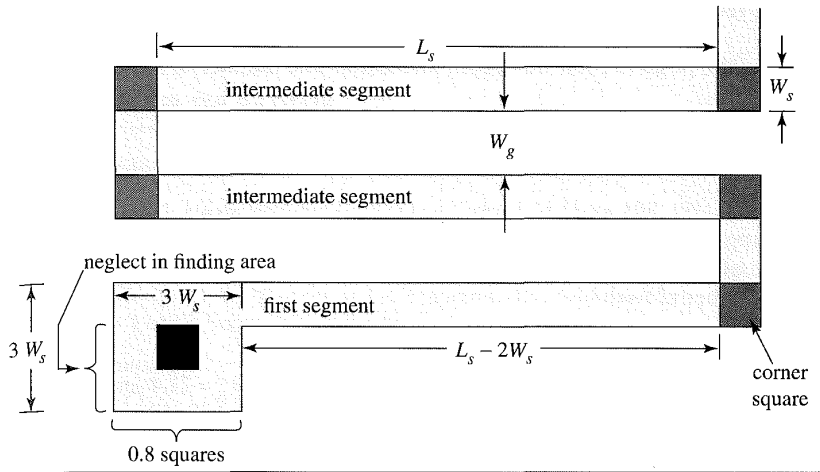
DESIGN EXAMPLE 2.7 Folded IC Resistor Layout

Given an n-type layer with a sheet resistance $R_{\square} = 1\ \text{k}\Omega/\square$, we want to lay out a folded resistor with $R = 75\ \text{k}\Omega$. The geometric design rules for the n-type region specify that the minimum width is $W_s = 2\ \mu\text{m}$ and that the minimum gap between n-type regions is $W_g = 3\ \mu\text{m}$, as defined in the sample layout in Fig. Ex2.7. The contact regions each contribute $0.8\square$. In order to make the resistor easy to pack into the IC layout, we want to fold the resistor.

What is the segment length L_s (defined in the sample layout) that results in a nearly square resistor layout? How much area does the resistor require? You may neglect the portions of the contact areas outside the folded segments in finding the resistor area.

SOLUTION

Assuming that the resistor consists of N_s segments, the total number of squares $N_{\square}$ is



➤ **Figure Ex2.7** Layout of a folded IC resistor with offset contacts.

$$N_\square = 2\,(0.8) + 2\left(\frac{L_s - 2W_s}{W_s}\right) + (N_s - 2)\left(\frac{L_s}{W_s}\right) + (N_s - 1)\left(2\,(0.56) + \frac{W_g}{W_s}\right)$$

in which the first two terms are for the first and last segments (with the contacts), the third term is for the intermediate segments of length L_s, and the final term is for the folds (with the two corner squares contributing 0.56 squares each). Using the minimum width $W_s = 2$ μm and the minimum gap $W_g = 3$ μm, we find

$$N_\square = 1.6 + L_s - 4 + \left(\frac{(N_s - 2)}{2}\right)L_s + (N_s - 1)\,(2.62) = \left(\frac{L_s}{2} + 2.62\right)N_s - 5.02$$

In order to achieve a square layout, Fig. Ex2.7 indicates that

$$L_s + 2W_s = (N_s - 1)\,(W_s + W_g) + W_s = (N_s - 1)\,(5\ \mu\text{m}) + 2\ \mu\text{m}$$

in which the left hand side is the total length of one segment and the right hand side is the length of the side with the folds. Solving for the number of segments, we find that

$$N_s = \frac{L_s + 2W_s + W_g}{W_s + W_g} = \frac{L_s + 7}{5}$$

The total number of squares needed is $N_\square = R/R_\square = 75\ \text{k}\Omega\ /\ (1\ \text{k}\Omega/\square) = 75$. Substituting for N_s in the expression for the total number of squares, we find a quadratic equation for the segment length

$$75 = \left[\frac{L_s}{2} + 2.62\right]\left(\frac{L_s + 7}{5}\right) - 5.02\,,\ \text{which simplifies to}$$

$$L_s^2 + (12.24)\,L_s - 763.32 = 0$$

Taking the positive root, we find $L_s = 22.2$ μm. The number of segments needed is

$$N_s = \frac{L_s + 7}{5} = \frac{22.2 + 7}{5} = 5.84$$

One approach is to use an integer number of folds so we round up to $N_s = 6$. The length of each segment must be modified in order to maintain $N_\square = 75$ using the new number of folds

$$N_\square = 75 = \left(\frac{L_s}{2} + 2.62\right)N_s - 5.02 = \left(\frac{L_s}{2} + 2.62\right)6 - 5.02 \rightarrow L_s = 21.4\ \mu\text{m}$$

The final layout of the folded resistor with $L_s = 21.4$ μm is not a perfect square, since the segments are 25.4 μm long (counting the folds) and the edge with the 6 folds is 27 μm long. An alternative to forcing N_s to be an integer would be to adjust the lengths of the first or the last segments, in order to make the overall layout closer to a square. In practice, the integer number of folds is close enough to a square layout. The area of the resistor is

$$A = (L_s + 2W_s)\,[(N_s - 1)\,(W_s + W_g) + W_s] = (21.4 + 4)\,(5 \cdot 5 + 2) = 686\ \mu\text{m}^2$$

➤ SUMMARY

This chapter has provided a foundation in two subjects that are essential for integrated device and circuit modeling: the physics of holes and electrons in silicon and the technology of IC fabrication. Key concepts that you should understand are:

- The bond model and its description of an electron and a hole in silicon.
- The concept of thermal equilibrium and the mass-action law.
- Hole and electron concentrations for silicon with donors, with acceptors, or with both donors and acceptors.
- Drift velocity and drift current density.
- Diffusive transport and diffusion current density.
- The total electron and hole current densities.
- Lithography and pattern transfer.
- Masked ion implantation of impurities.
- Cross sections through device structures from the layout and the sequence of fabrication steps (process).
- Resistivity and sheet resistance.
- Ion-implanted IC resistors and their layout, including equivalent squares for contact regions and corners.

◆ Normalized uncertainty and its calculation for a function of several variables, given their uncertainties.

FURTHER READING

Semiconductor Physics

1. C. G. Fonstad, *Microelectronic Devices and Circuits*. McGraw-Hill, 1994, Chapters 1 and 2. About the same level.
2. R. F. Pierret, *Semiconductor Fundamentals*, 2nd Ed. Modular Series on Solid State Devices, Vol. I, G. W. Neudeck and R. F. Pierret, eds., Addison-Wesley, 1988, Chapters 1 and 3. Somewhat more advanced.

IC Technology

1. W. Maly, *Atlas of VLSI Technologies*, Addison-Wesley, 1986. Several step-by-step process sequences provide a good intuitive understanding of the connection between layout, process, and device cross sections.
2. R. C. Jaeger, *Introduction to Microelectronic Fabrication*, Modular Series on Solid State Devices, Vol. V, G. W. Neudeck and R. F. Pierret, eds., Addison-Wesley, 1988, Chapters 1 and 2. Early chapters provide a good overview of IC fabrication.

PROBLEMS

Note: Assume that the sample temperature is 27 °C , unless otherwise noted in the problem statement.

Exercises

E2.1 A silicon wafer is doped with donors at a concentration of $N_d = 10^{15}$ cm^{-3}.

(a) What is the electron concentration n_o (cm^{-3}) at room temperature?

(b) What is the hole concentration p_o (cm^{-3}) at room temperature?

E2.2 A silicon wafer is doped with acceptors at a concentration of $N_a = 10^{14}$ cm^{-3}.

(a) What is the electron concentration n_o (cm^{-3}) at room temperature?

(b) What is the hole concentration p_o (cm^{-3}) at room temperature?

E2.3 Given that a silicon wafer is doped with acceptors with $N_a = 10^{14}$ cm^{-3}. We add a donor concentration of $N_d = 7.5 \times 10^{13}$ cm^3 to a particular region.

(a) What type is the region (n or p)?

(b) What is the electron concentration n_o (cm^{-3}) in the region?

(c) What is the hole concentration p_o (cm^{-3}) in the region?

E2.4 Given that a silicon wafer is doped with donors with $N_d = 5 \times 10^{17}$ cm^{-3}. We add an acceptor concentration of $N_a = 5.5 \times 10^{17}$ cm^3 to a particular region.

(a) Which carrier is the majority carrier?

(b) What is the hole concentration p_o (cm^{-3}) in the region?

(c) What is the electron concentration n_o (cm^{-3}) in the region?

E2.5 For a wafer doped with acceptors at $N_a = 10^{14}$ cm^{-3}, what concentration of donors must be added in order to reduce the hole concentration by 90%?

E2.6 In a subatomic particle detector, extremely low doped silicon substrates are needed. What donor concentration is required in order to have an electron concentration $n_o = 8\, n_i$? You will need to account for thermal generation for this problem.

E2.7 In modern VLSI processes, the dimensions of device structures are approaching 1 μm^3. For a donor concentration $N_d = 5 \times 10^{14}\ cm^{-3}$, find the number of electrons and the number of holes in this volume. *Comment*: your answer will show one reason why there is a tendency for doping concentrations to increase as device dimensions are scaled into the submicron region.

E2.8 In a silicon region with donor concentration $N_a = 10^{16}\ cm^{-3}$, we apply an electric field in the $+x$ direction with a magnitude of 10^3 V/cm.

(a) What is the electron drift velocity (magnitude and sign)?

(b) What is the electron drift current density (magnitude and sign)?

(c) How long is required for an electron to drift 1 μm, on average?

(d) How many collisions occur while it is drifting? You can assume that the collision time for the drifting electron is $\tau_c = 0.1$ ps.

E2.9 Repeat E2.8 with a donor concentration of $N_d = 10^{15}\ cm^{-3}$.

E2.10 In a silicon region with acceptor concentration $N_a = 10^{18}\ cm^{-3}$, we apply an electric field in the $+x$ direction with a magnitude of 2×10^3 V/cm.

(a) What is the hole drift velocity (magnitude and sign)?

(b) What is the hole drift current density (magnitude and sign)?

(c) How long is required for a hole to drift 1 μm, on average?

(d) How many collisions occur while it is drifting? You can assume that the collision time for the drifting hole is $\tau_c = 0.05$ ps.

E2.11 A silicon region is doped with acceptors at $N_a = 10^{15}\ cm^{-3}$ and with donors at $N_d = 8 \times 10^{14}\ cm^{-3}$.

(a) What type is the silicon (n or p)?

(b) What electric field must be applied for the magnitude of the drift velocity of the majority carriers to be 4×10^6 cm/s?

(c) What is the drift current density of the majority carriers for the case of the electric field in part (b)?

E2.12 By optical means, we set up a minority hole concentration gradient across a 2 μm-wide silicon region that is given by:

$$p(x) = (10^{18}\ cm^{-4})\, x$$

where x is the coordinate in the direction of the concentration gradient. The donor concentration in the region is $N_d = 10^{16}\ cm^{-3}$.

(a) Find the hole diffusion current density in the region.

(b) How long does a hole take to diffuse across the 2 μm-wide region?

E2.13 The minority hole concentration gradient across a 2.5 μm-wide silicon region is given by:

$$p(x) = (5 \times 10^{18}\ cm^{-4})\ (2.5\ \mu m - x)$$

where x is the coordinate in the direction of the concentration gradient. The donor concentration in the region is $N_d = 5 \times 10^{17}$ cm^{-3}.

(a) Find the hole diffusion current density in the region.

(b) How long does a hole take to diffuse across the 2.5 μm-wide region?

E2.14 In a silicon sample, we establish a minority electron concentration gradient over the region $x \geq 0$ that is given by

$$n(x) = (10^{15}\text{cm}^{-3})\, e^{-x/(2\ \mu\text{m})}$$

The acceptor concentration is $N_a = 10^{17}$ cm^{-3}.

(a) Find the magnitude and sign of the electron diffusion current density at $x = 0$.

(b) Plot the electron diffusion current density over the interval $0 < x < 10$ μm.

E2.15 The gradient in minority hole concentration in a silicon sample over the region $x \geq 0$ is given by

$$p(x) = (10^{13}\text{cm}^{-3})\ (1 - e^{-x/(1\ \mu\text{m})})$$

The donor concentration is $N_d = 5 \times 10^{15}$ cm^{-3}.

(a) What is the hole concentration at $x = 0$?

(b) What is the hole diffusion current density at $x = 0$? Provide an explanation for the apparent discrepancy between parts (a) and (b).

(c) Plot the hole diffusion current density over the interval $0 \leq x \leq 5$ μm.

E2.16 By continuous improvements in fabrication technology, three generations of a microprocessor IC have areas that are given in the table below. The cost of processing a 20 cm-diameter wafer increases with each generation, due to the need for more expensive process equipment.

Technology	Chip Area	Wafer Cost
1st Generation	2×2 cm^2	$1750
2nd Generation	1.5×1.5 cm^2	$2000
3rd Generation	1×1 cm^2	$2250

The yield of good chips for each generation is 75%.

(a) Find the number of good chips on a 20-cm wafer for each generation. Do not consider the complication of dice that intersect the edge of the wafer—an estimate of the chip count per wafer is adequate.

(b) What is the manufacturing cost per good microprocessor chip for each generation?

E2.17 For an implant dose $Q_a = 10^{15}$ cm^{-2} of acceptors and a post-anneal thickness of $t = 0.5$ μm, what is the acceptor concentration in the layer?

E2.18 Given a silicon wafer with an acceptor concentration of $N_a = 10^{16}$ cm^{-3}. We would like to use ion implantation to dope a surface region so that the average acceptor concentration is $N_a = 10^{17}$ cm^{-3}. If the post-anneal thickness of the region is $t = 0.1$ μm, what dose is needed?

E2.19 Starting with a silicon wafer with a donor concentration of $N_d = 10^{17}$ cm^{-3}, we would like to use ion implantation to convert a surface region to p-type, with an average acceptor concentration $N_a = 3 \times 10^{17}$ cm^{-3}. What implant dose is needed, if the post-anneal thickness of the implanted region is 0.25 μm.

E2.20 The five-step process for a two-layer structure made from silicon dioxide and polysilicon is as follows:

1. Grow 0.5 μm of silicon dioxide (SiO_2)
2. Pattern using the oxide mask.
3. Grow 500 Å of SiO_2.
4. Deposit 1 μm of polysilicon.
5. Pattern using the polysilicon mask.

The CAD layout for this structure is shown in Fig. E2.20.

(a) Sketch the cross section *A-A* from the process sequence, assuming that the oxide mask is dark field and the polysilicon mask is clear field.

(b) Repeat part (a), with the oxide mask clear field and the polysilicon mask dark field.

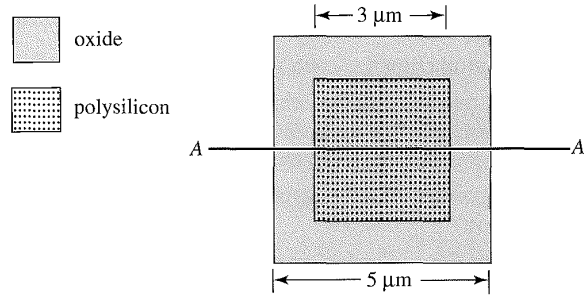
(c) Repeat part (a), with the oxide mask dark field and the polysilicon mask dark field.

(d) Repeat part (a), with the oxide mask clear field and the polysilicon mask clear field.

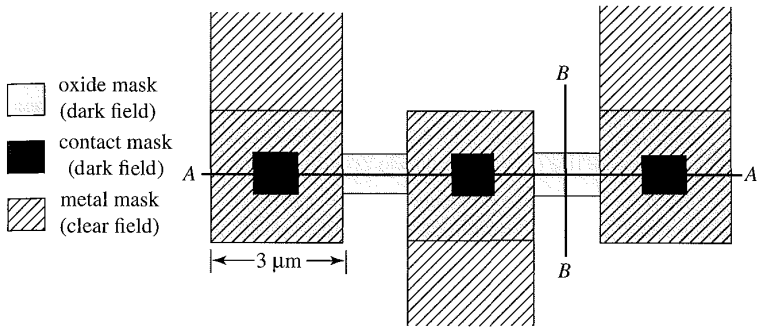
E2.21 The layout in Fig. E2.21 is for the ion-implanted resistor process outlined in Fig. 2.18.

(a) Sketch the cross section *A-A* from the layout in Fig. E2.21.

(b) Sketch the cross section *B-B* from the layout in Fig. E2.21.



➤ **Figure E2.20**



➤ **Figure E2.21**

E2.22 The layout in Fig. E.2.22 is for the simple MOS transistor process described in Fig. 2.19 (b).

(a) Sketch the cross section along *A*-*A*.

(b) Sketch the cross section along *B*-*B*.

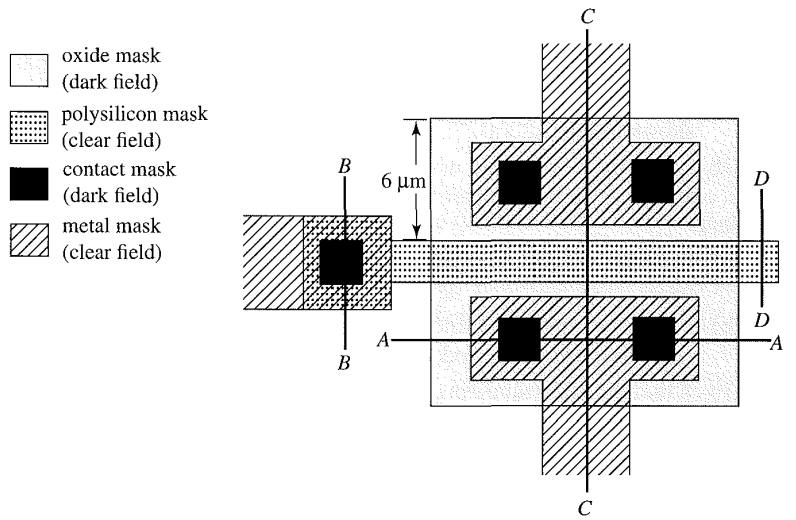
(c) Sketch the cross section along *C*-*C*.

(d) Sketch the cross section along *D*-*D*.

E2.23 The parameters for an ion-implanted region are: arsenic dose $Q_d = 10^{12}$ cm^{-2}, thickness after anneal $t = 1.1$ μm. The background doping of the p-type substrate is 10^{14} cm^{-2}.

(a) What is the average arsenic concentration in the ion-implanted layer?

(b) What is the sheet resistance in units of Ω/□.



➤ **Figure E2.22**

E2.24 After discovering that boron was implanted at a dose of $Q_a = 10^{11}$ cm^{-2} instead of phosphorus, you try to salvage the lot by compensating the layer back to n-type by means of implanting a heavier dose of phosphorus.

(a) Assuming that the layer thicknesses for the phosphorus and boron are the same at $t = 1$ μm after annealing, what phosphorus dose is necessary to achieve the desired electron concentration?

(b) What is the sheet resistance for the compensated layer? Compare your answer to what the sheet resistance would have been if the original mistake had not been made.

(c) From your answer in part (b), what phosphorus dose is needed to achieve the original sheet resistance?

E2.25 Given a 1 μm-thick implanted layer and a maximum average dopant concentration of 1×10^{20} cm^{-3} in this layer.

(a) What implant dose is needed to reach the maximum concentration? Neglect the background doping in the substrate.

(b) For an n-type layer, what is the minimum sheet resistance? Note that the mobility degrades with increasing doping concentration, as shown in Fig. 2.8.

(c) For a p-type layer, what is the minimum sheet resistance? Note that the mobility degrades with increasing doping concentration, as shown in Fig. 2.8.

E2.26 Given a sheet resistance of 100 Ω/□ in a 1 μm-thick, n-type implanted layer.

(a) What is the average donor concentration in the layer? Neglect the background doping in the substrate.

(b) Sketch the layout for a 2.5 kΩ resistor, using the "dogbone" style contacts shown in Fig. 2.22(a).

E2.27 Given a sheet resistance of 250 Ω/□ in a 1 μm-thick, p-type implanted layer.

(a) What is the average acceptor concentration in the layer? Neglect the background doping in the substrate.

(b) Sketch the layout for a 2 kΩ resistor, using the "dogbone" style contacts shown in Fig. 2.22(a).

E2.28 Given that the normalized uncertainties are: $\varepsilon_{N_d} = 0.025$ in the donor concentration and $\varepsilon_t = 0.02$ in the layer thickness. The uncertainty in the mobility can be neglected.

(a) For an average donor concentration of 10^{16} cm^{-3}, find the variation in N_d.

(b) For an average layer thickness of 500 nm, find the average sheet resistance and its normalized and absolute uncertainties.

E2.29 Given that the uncertainty in the lateral dimension of an ion-implanted resistor is $\delta = 500$ Å and that the sheet resistance is 100 Ω/□.

(a) Neglecting the uncertainty in the sheet resistance, what are the normalized uncertainty and the variation in a 5 kΩ resistor for a resistor width $W = 2$ μm.

(b) How wide must the resistor be made in order to have the normalized uncertainty less than 0.01 (1%)?

E2.30 For a doped polysilicon film, the sheet resistance is 25 Ω/□ for a 5000 Å-thick film. If the thickness of the film is controlled to ±2.5% over the wafer, what is the variation in the resistance of a 20 kΩ resistor?

PROBLEMS

P2.1 A sample of silicon is doped with phosphorus with a concentration of 10^{15} cm^{-3}.

(a) What is the average volume occupied by an electron in this sample, in μm^3?

(b) What is the average volume occupied by a holes in this sample, in μm^3?

P2.2 A region of a silicon wafer is doped with three impurities:

arsenic: 10^{16} cm^{-3}, boron: 1.15×10^{16} cm^{-3}, phosphorus: 2.5×10^{15} cm^{-3}.

Find the electron and hole concentrations n_o and p_o.

P2.3 A sample of silicon is doped with $N_d = 1.1 \times 10^{13}$ cm^{-3} and $N_a = 1 \times 10^{13}$ cm^{-3}.

(a) Find the electron and hole concentrations n_o and p_o at room temperature.

(b) Show that the charge density $\rho = 0$ (exactly).

P2.4 The intrinsic concentration in silicon n_i has a strong temperature dependence, as shown by these measurements over the temperature range for automotive electronic applications:

T (° C)	n_i (cm^{-3})
−55	2.24×10^6
0	1.42×10^9
70	2.82×10^{11}
125	5.06×10^{12}

A course in semiconductor physics shows that the functional dependence of the intrinsic concentration is

$$n_i = A \cdot T^{3/2} e^{-6608.7/T},$$

where T is in Kelvin.

(a) Find the constant A by fitting the function to the measurements above.

(b) How rapidly the intrinsic concentration varies near room temperature is important for device modeling. Find the temperature increase (in degrees Centigrade) needed to double n_i and the temperature decrease needed to halve n_i.

P2.5 A sample of silicon is doped with $N_d = 10^{13}$ cm^{-3}. As the temperature increases, the contribution of thermal generation to the electron and hole concentrations eventually becomes significant. Using the result from P2.4(a) for the intrinsic concentration as a function of temperature,

(a) Plot $n_o(T)$ over the temperature range −55° C < T < 125° C on a log-linear plot.

(b) What is the maximum temperature for which $p_o/n_o < 10^{-3}$?

P2.6 A sample of silicon is doped with $N_a = 2 \times 10^{14}$ cm^{-3} and $N_d = 9 \times 10^{13}$ cm^{-3}. Using the result from P2.4(a) for the intrinsic concentration as a function of temperature,

(a) Plot $p_o(T)$ over the temperature range -55° C $< T < 125^\circ$ C on a log-linear plot.

(b) What is the maximum temperature for which $n_o/p_o < 5 \times 10^{-4}$?

P2.7 Figure P2.7 is a top view of a silicon region which has metal electrical contacts on two sides, across which a voltage $V_A = 2$ V is applied. The silicon layer is doped with arsenic at a concentration of 10^{13} cm^{-3} and is 2 μm thick and you can assume room temperature. You should use the appropriate mobility for the given dopant concentration.

(a) Find the value of the electric field E_x in the silicon.

(b) Find the value of the drift velocity of electrons and of holes in the silicon. How long does is take an electron, on average, to drift across the silicon region.

(c) Find the drift current densities for electrons and for holes. What is the percentage error in the total current I if the minority carrier drift current density is neglected?

(d) What is the value of the resistance R in ohms?

P2.8 If the arsenic concentration is increased to 10^{18}cm^{-3}, repeat P2.7.

P2.9 Plot the maximum current I_{max} across the silicon region in Fig. P2.7 as a function of doping concentration from $N_d = 10^{13}$ cm^{-3} to $N_d = 10^{18}$ cm^{-3}, using a log-log scale. You can assume that the saturation velocity of electrons is $v_{sat} = 10^7$ cm/s and should include the dependence of mobility on doping concentration. Plot the applied voltage $V_{A,max}$ at which I_{max} is reached as a function of doping concentration.

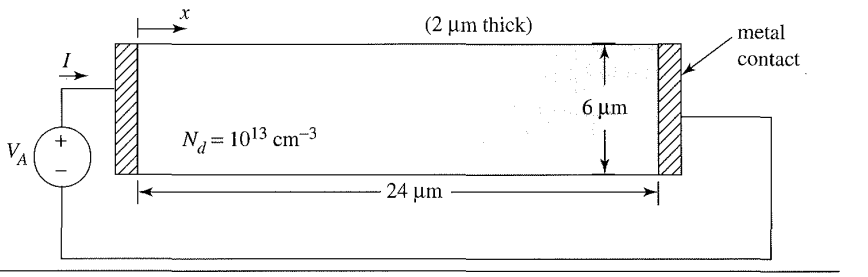
P2.10 Figure P2.10 is a top view of a 1 μm-thick silicon region with variable acceptor concentration, across which a voltage V_A is applied through metal electrical contacts. Assume that the hole saturation velocity is $v_{sat} = 10^7$ cm/s.

(a) For an applied voltage $V_A = 1$ V, plot the electric field $E(x)$. *Hint*: you will find it helpful to consider the silicon region as three series resistors.

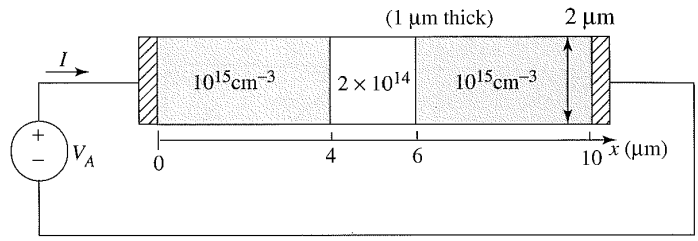
(b) What is the maximum current I_{max} through this structure?

(c) What voltage V_{max} must be applied to reach this maximum current?

(d) Plot the electric field $E(x)$ corresponding to V_{max}.



➤ **Figure P2.7**



➤ **Figure P2.10**

P2.11 Figure P2.11 is a top view of a 1 μm-thick silicon region with variable donor concentration. A voltage V_A is applied through metal electrical contacts.

(a) For an applied voltage $V_A = 1$ V, what is the current I? You may find it helpful to view the silicon region as two resistors in parallel.

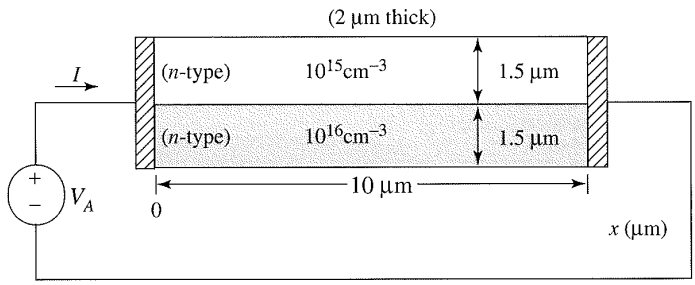
(b) Plot the current as a function of voltage over the range $-20 \text{ V} < V_A < 20$ V. You can assume that the saturation velocity of electrons $v_{sat} = 10^7$ cm/s and should include the variation of mobility with donor concentration.

P2.12 The top view of a 2 μm-thick silicon region with a varying width is shown in Fig. P2.12. The donor concentration is $N_d = 10^{17}$ cm^{-3} throughout. You can assume that 25% of the applied voltage is dropped across the "neck," as long as the saturation velocity is not reached in any part of the region.

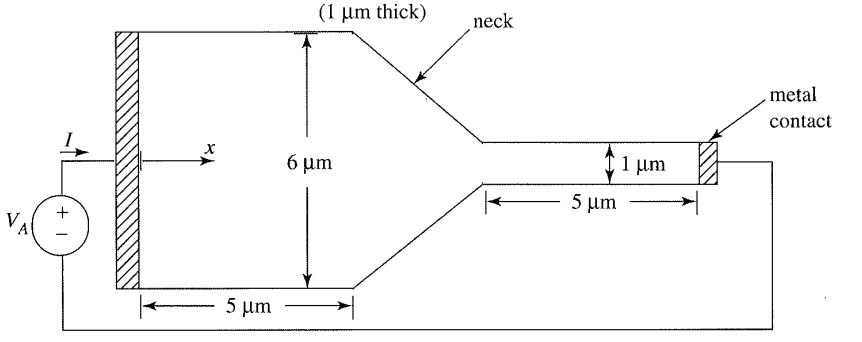
(a) Find the ratio of the current density in the wide region to that in the narrow region, for the case where $V_A = 2$ V.

(b) *Sketch* the electric field $E(x)$ along the center line of the structure, for the case when $V_A = 2$ V. The numerical value in the regions of constant width should be accurate. It will be helpful to view the silicon region as three resistors in series, with the voltage drop across the "neck" region given as 25% of V_A

(c) What is the maximum current I_{max} through this structure? What is the applied voltage when I_{max} is reached?



➤ **Figure P2.11**



➤ **Figure P2.12**

P2.13 In the slab of silicon region in Fig. P2.13, the minority carrier concentration is maintained at its thermal equilibrium value at the bottom surface ($x = 0.3$ μm). The silicon is doped with acceptors at $N_a = 10^{16}$ cm^{-3}.

(a) If the minority carrier concentration at the top surface is 1/10 the majority carrier concentration and we assume a linear gradient from top to bottom, what is the diffusion current density in the slab?

(b) How long does it take for a minority carrier to diffuse across the slab, on average?

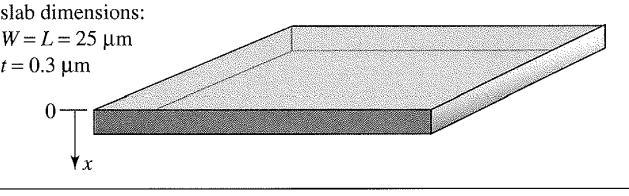
P2.14 When a gradient in minority carrier concentration is set up in a semiconductor region, an identical gradient in the majority carrier concentration is quickly established, in order to keep the silicon region electrically neutral. This situation is shown in the carrier concentration distributions in Fig. P2.14.

(a) Find the minority electron diffusion current density.

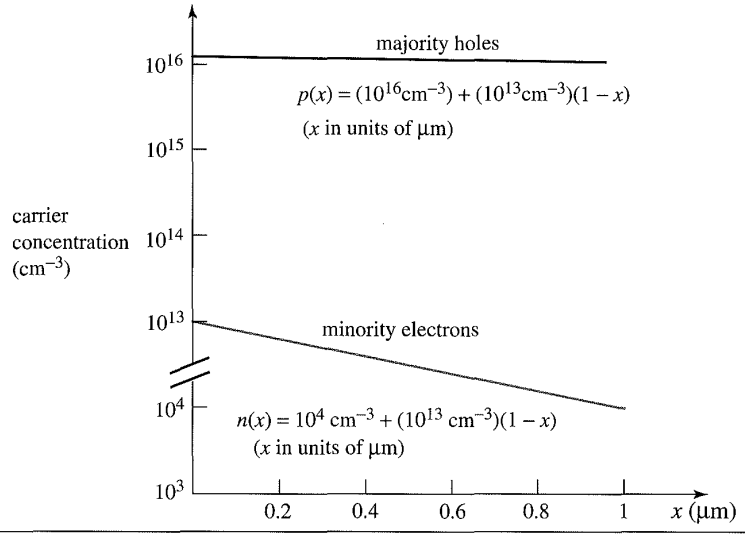
(b) Find the majority hole diffusion current density.

(c) Since the silicon region is open-circuited, the total current density must be zero. Find the required drift current density.

(d) Assuming that the contribution of the minority carriers to the drift current is negligible (which it is), what electric field is required in the sample to generate the drift current density needed from part (c).



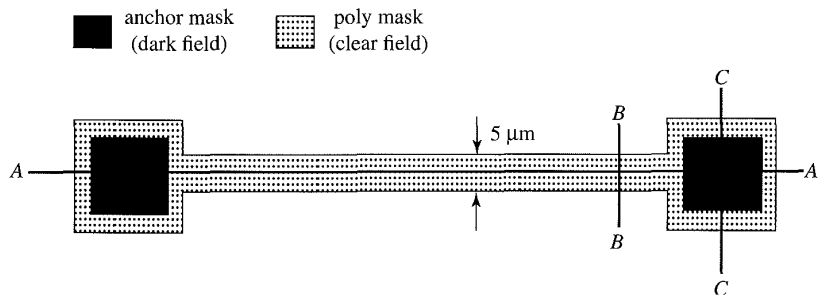
➤ **Figure P2.13**



➤ **Figure P2.14**

P2.15 The process sequence and layout for making a simple polysilicon micromechanical bridge are listed below and in Fig. P2.15. Using IC-based processes, we can make structures for sensing non-electrical inputs, such as acceleration.

1. Deposit 2 μm of CVD SiO_2.
2. Pattern with the anchor mask.
3. Implant phosphorus and anneal.
3. Deposit 2 μm of phosphorus-doped polysilicon and anneal.
4. Pattern with the poly mask.
5. Etch for 4 minutes in hydrofluoric acid (SiO_2 etch rate: 1 μm per minute); rinse wafer in water and dry.



➤ **Figure P2.15**

(a) Draw the cross sections *A-A* after step 4 and after completion of the process.

(b) Draw the cross section *B-B* after step 5 is half completed (2 minutes of HF etching.)

(c) Draw the cross section *C-C* after completion of the process.

P2.16 This problem considers the fabrication of a two-level micromechanical structure, which is accomplished by extending the process in P2.15.

Steps 1-4: same as those in P2.15, with the masks renamed as *anchor1* and *poly1*

5. Deposit 1 μm of SiO_2.
6. Pattern the oxide with the anchor2 mask.
7. Deposit 1 μm of phosphorus-doped polysilicon and anneal.
8. Pattern the second polysilicon layer with the poly2 mask.
9. Etch for 12 minutes in hydrofluoric acid (SiO_2 etch rate: 1 μm per minute); rinse wafer in water and dry.

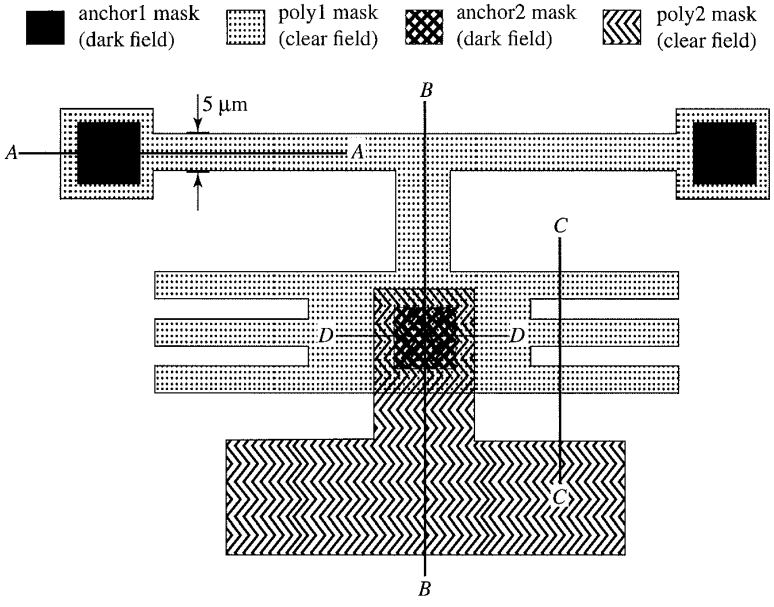
(a) Draw the cross section *A-A* after step 8.

(b) Draw the cross section *B-B* after completion of the process.

(c) Draw the cross section *C-C* after completion of the process.

(d) Draw the cross section *D-D* after completion of the process.

P2.17 Texturing the silicon surface is a way to increase the surface area of a capacitor; the following process and layout are one way to achieve this goal.



➤ **Figure P2.16**

1. Etch 5 μm deep into the silicon substrate using the trench mask*.
2. After stripping the photoresist and cleaning the wafer, grow 500 Å of thermal SiO_2.
3. Deposit 7500 Å of phosphorus-doped polysilicon.
4. Pattern using the polysilicon mask.

(a) Draw the cross section *A-A* after step 2.

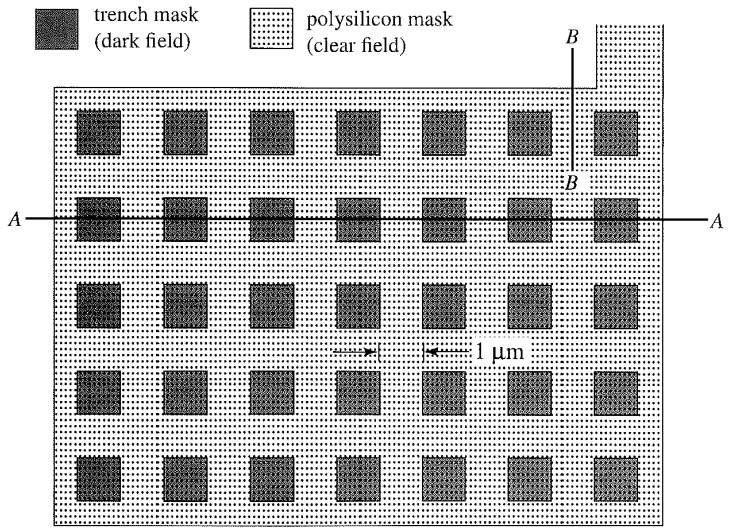
(b) Draw the cross section *A-A* after step 3.

(c) Draw the cross section *B-B* after completion of the process.

P2.18 A simple double metallization process is as follows; the mask layout is shown in Fig. P2.18.

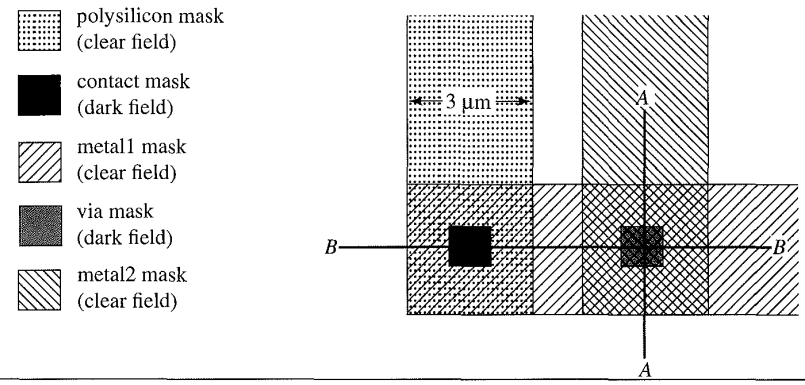
1. Grow 0.5 μm of thermal SiO_2.
2. Deposit 0.5 μm of phosphorus doped polysilicon and pattern using polysilicon mask.
3. Deposit 0.5 μm of CVD SiO_2 and pattern using contact mask.
4. Deposit 1 μm of sputtered Al and pattern using metal1 mask.
5. Deposit 0.5 μm of CVD SiO_2 and pattern using via mask.
6. Deposit 1 μm of sputtered Al and pattern using metal2 mask.

(a) Accurately sketch the cross section *A-A*, assuming that the deposited CVD films conformally coat the surface.



➤ **Figure P2.17**

*A reactive-ion etch is used for the trench etch into the silicon substrate, which is optimized to yield vertical walls in silicon with negligible mask undercutting.



➤ **Figure P2.18**

(b) Accurately sketch the cross section *B-B*, assuming that the deposited CVD films conformally coat the surface.

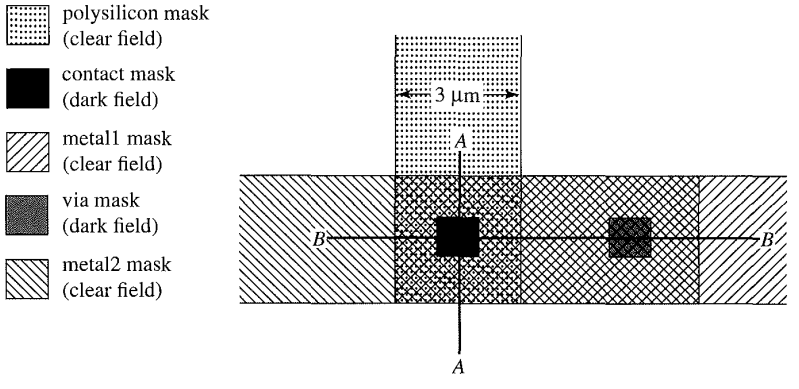
P2.19 Using the double metal process from P2.18, consider the layout in Fig. P2.19.

(a) Accurately sketch the cross section *A-A*, assuming that the deposited CVD films conformally coat the surface.

(b) Accurately sketch the cross section *B-B*, assuming that the deposited CVD films conformally coat the surface. Note that the rough topography will make lithography difficult. The geometric design rules can prevent this situation by prohibiting the stacking of metal1 and metal2.

P2.20 The mask layout in Fig. 2.20 uses the simple MOSFET process in Fig. 2.19(b).

(a) Sketch the cross section along *A-A*.



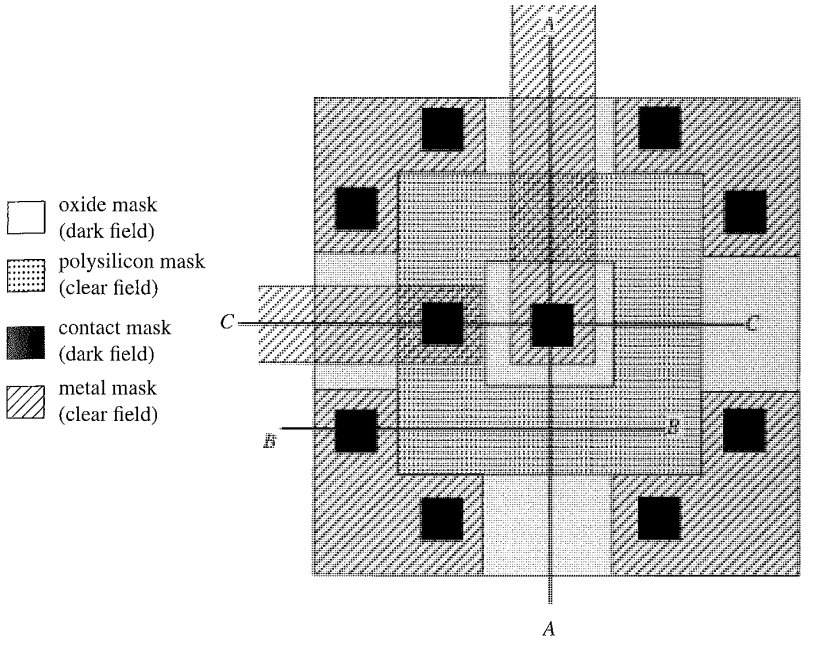
➤ **Figure P2.19**

(b) Sketch the cross section along *B-B*.

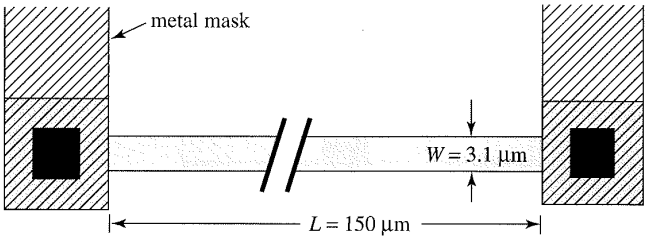
(c) Sketch the cross section along *C-C*.

P2.21 The resistor whose layout is shown in Fig. P2.21 is a test structure on an integrated circuit.

The measured resistance is $R = 22.3\ \text{k}\Omega$. What is the sheet resistance of the ion-implanted layer in ohms per square? You can assume that the contact regions each contribute 0.65 squares.



➤ **Figure P2.20**



➤ **Figure P2.21**

P2.22 The two ion-implanted resistors in Fig. P2.22 have a different style of contact region compared to that shown in Fig. 2.22. The longer one has a measured resistance of $R_1 = 395\ \Omega$, with the shorter one measuring $R_2 = 270\ \Omega$.

From these measurements and the dimensions, what is the sheet resistance of this layer and the effective number of squares for this type of contact region?

P2.23 Given a p-type ion-implanted resistor with an average acceptor concentration of $N_{a1} = 10^{16}\ \text{cm}^{-3}$ over the depth $0 < x < 0.5\ \mu\text{m}$ and $N_{a2} = 7.5 \times 10^{15}\ \text{cm}^{-3}$ over the depth $0.5\ \mu\text{m} < x < 1.25\ \mu\text{m}$. Find the sheet resistance in $\Omega/\square$.

P2.24 Given an n-type ion-implanted layer with thickness $t = 1\ \mu\text{m}$ and average doping concentration $N_d = 10^{17}\ \text{cm}^{-3}$.

(a) What is the sheet resistance?

(b) In order to meet the target sheet resistance of 750 $\Omega/\square$, we can perform a second implant anneal to add additional dopants to a layer of thickness 0.25 μm. Neglecting the change in the overall layer thickness after annealing, what dose and dopant type (n or p) are needed?

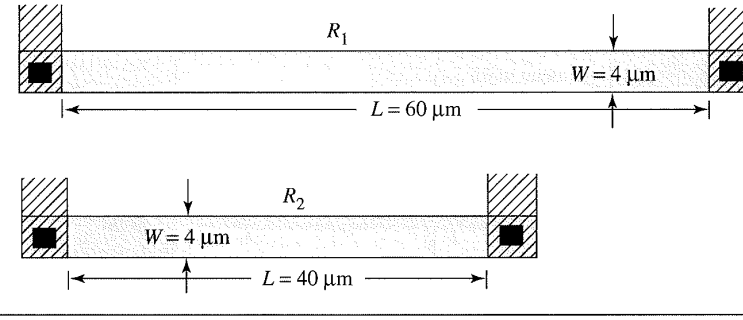
P2.25 The normalized uncertainty in the average acceptor concentration in an implanted layer is $\varepsilon_{N_a} = 0.025$. The uncertainties in the mobility and in the layer thickness ($\bar{t} = 1.4\ \mu\text{m}$) are negligible for this problem. Finally, the dimensions of the resistor pattern have an uncertainty of $\delta = 400$ Å.

(a) For a 120 square resistor with $\overline{N_a} = 10^{17}\ \text{cm}^{-3}$ and $\overline{W} = 2.5\ \mu\text{m}$, what is the average resistance and its uncertainty?

(b) If we increase the doping by a factor of two, what is the width necessary in order to keep the average resistance the same? What is the new uncertainty?

P2.26 It is an attractive idea, at first glance, to compensate a silicon region so that $N_d = N_a$ and $n_o = p_o = n_i$. Assume that the wafer is doped with acceptors at precisely $N_a = 10^{16}\ \text{cm}^{-3}$ and that we can introduce donors uniformly into a 1 μm-thick surface region with a normalized uncertainty of 0.005.

(a) If we wish to guarantee that the surface region stays p-type, what donor concentration should we add? Note that the normalized uncertainty is 0.005.



➤ **Figure P2.22**

(b) What is the range of sheet resistances for the surface region, given the doping in part (a)?

DESIGN PROBLEMS

D2.1 There are often geometric constraints in implementing resistors on an integrated circuit. Figure D2.1 defines the outline of the area in which the 160 kΩ resistor must fit, given the following geometric design rules. The locations for the "dogbone" style contacts are set by the other circuit elements are cannot be moved. The oxide mask pattern can abut the edges of the outline in Fig. D2.1.

Geometric design rules:

minimum width of oxide mask pattern = minimum resistor width= $W = 2\,\mu m$

minimum spacing between oxide patterns = $W_g = 2.5\,\mu m$

Resistor layer characteristics:

doping concentration $N_d = \overline{N_d}(1 \pm \varepsilon_{N_d}) = 10^{17}\ cm^{-3}\ (1 \pm 0.025)$

thickness $t = \bar{t}\,(1 \pm \varepsilon_t) = 0.6\,\mu m\ (1 \pm 0.02)$

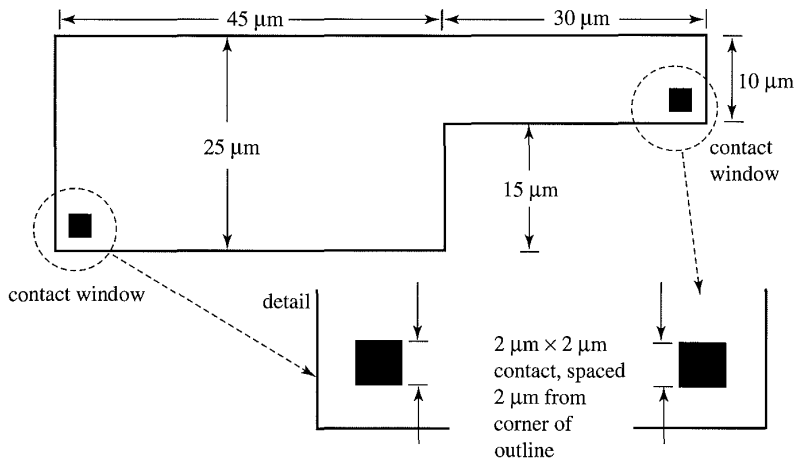
variation in resistor lateral dimension: $\delta = 650$ Å.

(a) Lay out the 160 kΩ resistor to fit inside the outline in Fig. D2.1.

(b) What is the uncertainty in the resistance?

(c) Modify your layout in (a) to minimize the uncertainty, while keeping within the boundaries of the outline. By what factor can you reduce the uncertainty?

D2.2 In some cases, the IC designer aims for the maximum resistance possible in a particular area. In this problem, you will lay out a polysilicon resistor in an area of 100 μm × 120 μm. The contacts should use the "dogbone" style in Fig. 2.22(a) and are



➤ **Figure D2.1**

to be located as shown on Fig. D2.2. The polysilicon resistor can abut the edges of the outline. The geometric design rules and parameters for the polysilicon film (lot-to-lot uncertainties) are:

Geometric design rules:

minimum polysilicon line width $W = 1\ \mu m$

minimum spacing between polysilicon lines = $W_g = 1\ \mu m$

Polysilicon film parameters:

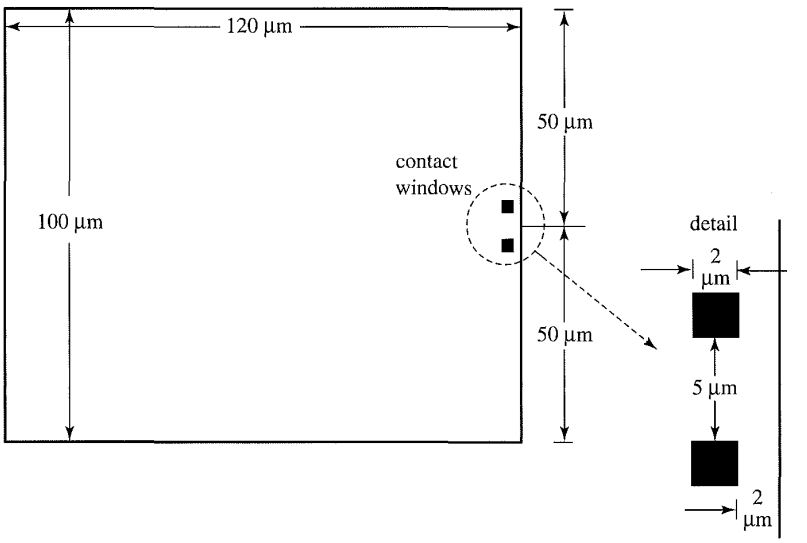
resistivity $\rho = \bar{\rho}(1 \pm \varepsilon_\rho) = (1 m\Omega - cm)\ (1 \pm 0.05)$

thickness $t = \bar{t}(1 \pm \varepsilon_t) = 0.5\ \mu m\ (1 \pm 0.05)$

variation in polysilicon line: $\delta = 500$ Å.

(a) Design the polysilicon resistor layout for the maximum number of squares in this area, given the design rules and fixed location of the contacts.

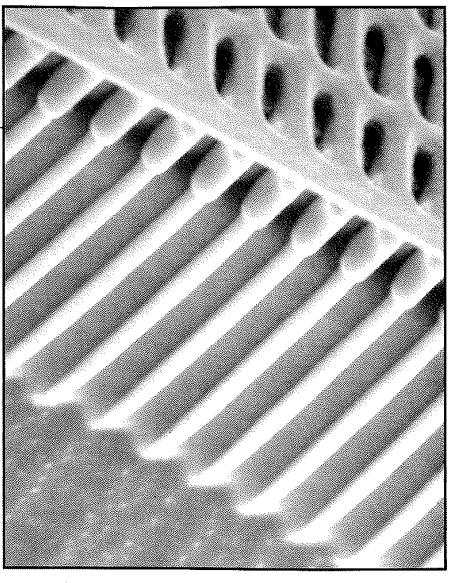
(b) What is the maximum resistance and its variation?



➤ **Figure D2.2**

chapter 3

PN Junction and MOS Electrostatics



SEM of trenches used to form extremely dense MOS capacitors used in state-of-the-art dynamic random access memories. (Courtesy of International Business Machines Corporation. Unauthorized use not permitted.)

With a basic understanding of semiconductor physics and IC technology from Chapter 2, we are ready to tackle two important device structures: the pn junction and the metal-oxide-semiconductor (MOS) capacitor. A good command of electrostatics is needed to find the electrostatic potential, electric field, and charge density in these structures. The electrostatics of the MOS capacitor is the foundation for understanding the MOS field effect transistor in Chapter 4. The electrostatics of the pn junction plays a similar role for the pn diode in Chapter 6 and the bipolar transistor in Chapter 7.

Chapter Objectives

- The interrelationships between charge density, electric field, and potential.
- The boundary conditions for electric field and electrostatic potential.
- The connection between potential and carrier concentration in thermal equilibrium.
- The depletion approximation and its application in simplifying the analysis of the pn junction and MOS capacitor.
- The qualitative spatial variation of charge density, electric field, and potential through the pn junction and MOS capacitor.
- Quantitative models for the depletion widths, electric fields, and potential for both structures.
- The definition of capacitance and its application to the reverse-biased diode and the MOS capacitor.

➤ 3.1 APPLIED ELECTROSTATICS

In introductory physics courses, an overview of the classical theory of electromagnetism is presented. These courses focus on the underlying assumptions, the three dimensional mathematical structure, and the inherent beauty of this remarkably powerful theory. Problem solving is usually limited to highly symmetrical geometries where the various surface integrals can be evaluated exactly. As a result, students often master the mathematics of electromagnetism, but fail to gain an intuitive grasp of the subject. When presented with a simple problem in one dimension, students are often unsure which way the electric field points, let alone grasp the details of the potential variation!

In order to understand semiconductor devices, we must develop a practical, *working knowledge* of electrostatics. Semiconductor device structures have complicated boundaries that make it difficult, if not impossible, to find closed-form analytical solutions in two or three dimensions. Therefore, we will use the one-dimensional electrostatic equations to develop insight and a basic understanding of device operation. For detailed modeling, device physicists solve the three-dimensional equations numerically using simulation programs.

3.1.1 Gauss's Law

We will first collect the basic electrostatic equations that relate charge density, electric field, and electrostatic potential. Gauss's Law connects the electric field E [units: volts per centimeter, V/cm] with the charge density ρ [units: Coulombs per cubic centimeter, C/cm^3]. Its one dimensional, differential form is

Gauss's Law in differential form

$$\boxed{\frac{dE}{dx} = \frac{\rho}{\varepsilon}} \tag{3.1}$$

where ε is the electric permittivity [units: Farads per centimeter, F/cm]. If the electric field is known, we can differentiate it (i.e., find its slope) and use Eq. (3.1) to plot the charge density. In many cases, we will know the charge density ρ and need to find the electric field. Integrating Eq. (3.1) along an interval $x_a \leq x' \leq x$ we find the integral form of Gauss's Law:

Gauss's Law in integral form

$$\boxed{\int_{x_a}^{x} d[\varepsilon E(x')] = \varepsilon E(x) - \varepsilon_a E(x_a) = \int_{x_a}^{x} \rho(x')dx' = Q(x)} \tag{3.2}$$

where $Q(x)$ is the charge in the interval [units: C/cm^2]. The possibility of a change in permittivity due to a material interface in the interval has been accounted for by keeping the permittivity together with the field in the integral. This result is useful since we can often conclude from the distribution of charge in semiconductor device structures that $E(x_a) = 0$ for a particular boundary x_a.

3.1.2 Potential and Poisson's Equation

The electrostatic potential $\phi(x)$ is defined with respect to its value at a point x_o as the integral of the negative of the electric field E from x_o to x:

definition of electrostatic potential

$$\phi(x) - \phi(x_o) = \int_{x_o}^{x} -E(x)dx \tag{3.3}$$

Eq. (3.3) allows us to calculate the potential *difference* between two points given the electric field. To find a numerical value for $\phi(x)$, it is necessary to select an arbitrary reference point x_o where by definition $\phi(x_o) = 0$. We will consider the question of selecting references for $\phi(x)$ in semiconductors later in this chapter.

Differentiation of Eq. (3.3) allows us to find the electric field given the potential

$$E(x) = -\frac{d\phi(x)}{dx} \tag{3.4}$$

The potential can be related to the electric field, and vice versa, through Eqs. (3.3) and (3.4). Since the derivative of the electric field is proportional to the charge density through the differential form of Gauss's Law (Eq. (3.1)), differentiation of Eq. (3.4) and substitution leads to a second-order differential equation for the potential:

Poisson's Equation

$$\frac{d^2\phi(x)}{dx^2} = -\frac{dE(x)}{dx} = -\frac{\rho(x)}{\varepsilon} \tag{3.5}$$

The result, called Poisson's Equation, can be useful since it directly links the potential with the charge distribution.

3.1.3 Intuition and Sign Conventions

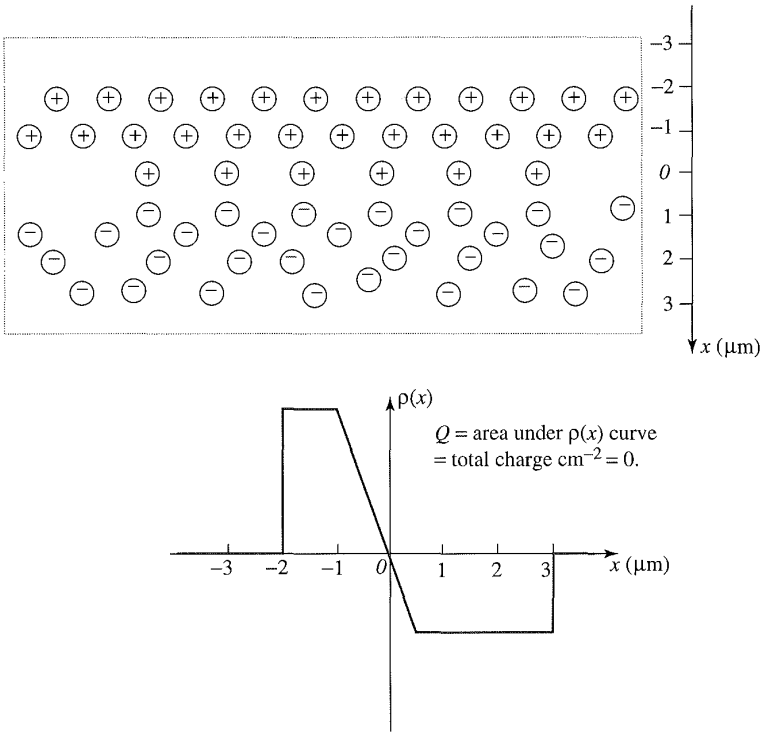
Even though Eqs. (3.1), (3.4), and (3.5) are ordinary differential equations with constant coefficients, the mathematical details of solving them can get in the way of finding the answer. For instance, a sign error will cause the electric field to point in the wrong direction. In this section we will introduce some techniques for sketching the answers *before* solving the equations.

As an example to motivate these techniques, Figure 3.1 sketches the charge density in a material with permittivity ε. Positive charges are located on the top (negative x) and negative charges are on the bottom (positive x). The *total* charge in the region is zero—which is always true for electronic device structures, as we will discover later in this chapter. From Eq. (3.2), the electric field $E(-2\ \mu m)$ at $x = -2\ \mu m$ and the electric field $E(3\ \mu m)$ at $x = 3\ \mu m$ are equal

$$E(3\mu m) - E(-2\mu m) = \int_{-2\mu m}^{3\mu m} \frac{\rho(x)}{\varepsilon} dx = 0 \tag{3.6}$$

In device physics, we are interested in the internal electric field caused by the charge distribution inside the device. For our purposes, any external electric field at the boundaries of the charged region $E(-2\ \mu m) = E(3\ \mu m)$ will be ignored and, for convenience, set to zero.

In this text, the x axis will be drawn horizontally with the positive direction to the right. The physical orientation of the x axis will be specified on the device cross section as in the top of Fig. 3.1. The charge distribution can be integrated graphically to find the electric field, using the integral form of Gauss's Law

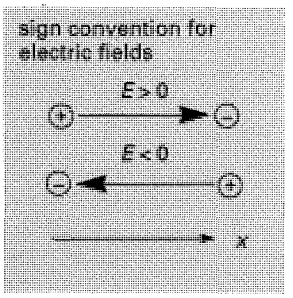


➤ **Figure 3.1** Charge distribution and corresponding charge density $\rho(x)$. The area under the curve is the total charge Q (per cm^2) and is zero for this charge distribution—and for *all* charge distributions in silicon devices in this text.

$$E(x) = E(-2\mu m) + \int_{-2\mu m}^{x} \frac{\rho(x')}{\varepsilon} dx' = \int_{-2\mu m}^{x} \frac{\rho(x')}{\varepsilon} dx' \tag{3.7}$$

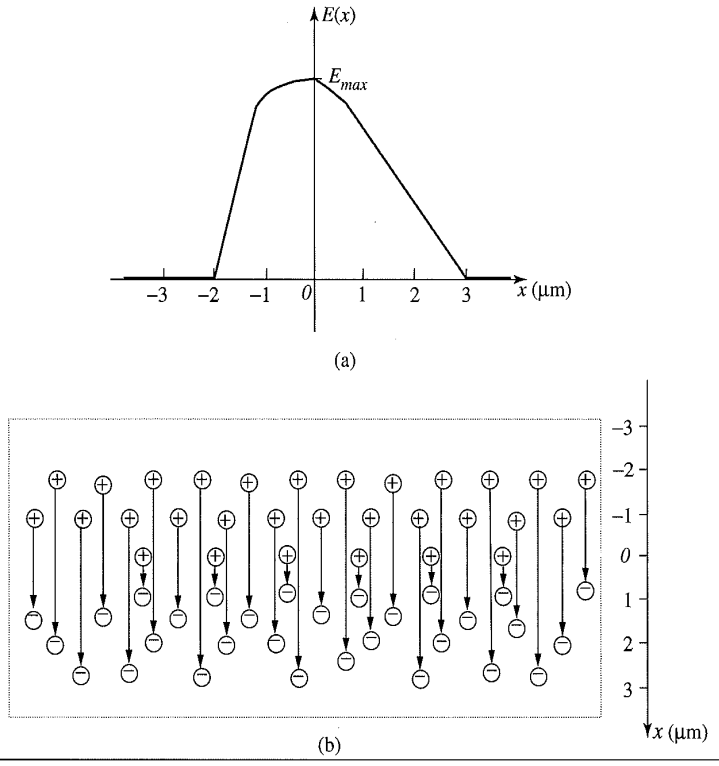
since $E(-2\ \mu m) = 0$. Figure 3.2(a) shows the electric field.

The electric field varies linearly with x for regions with constant charge density and varies quadratically in the interval $-1\ \mu m < x < 0.5\ \mu m$, where the charge density varies linearly. The electric field in Fig. 3.2(a) reaches its maximum value which is equal to



$$E_{max} = \int_{-2\mu m}^{0} \frac{\rho(x')}{\varepsilon} dx' \tag{3.8}$$

The electric field is positive over the entire charged region, meaning that it points in the same direction as the x axis. Note that $E(x)$ is positive in the interval $0 < x < 3\ \mu m$,

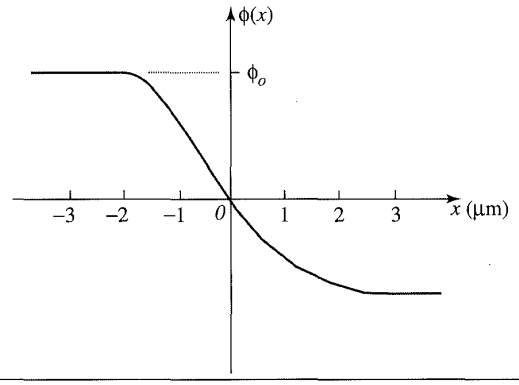


➤ **Figure 3.2** (a) Electric field and (b) qualitative picture of electric field lines connecting the positive and negative charges for the charge distribution in Fig. 3.1.

where the charge density $\rho(x)$ is negative. In order to clarify the connection between charge density and electric field, Figure 3.2(b) shows electric field lines connecting the positive and negative charges. From this figure, we can see that the electric field's magnitude increases (i.e., there are more field lines) as we move from $x = -2$ μm to the origin, after which it decreases as the field lines terminate on the negative charges.

From the electric field $E(x)$, the electrostatic potential $\phi(x)$ can be found by integration from Eq. (3.3). Since we have not yet considered a potential reference in silicon, we will set the potential for $x < -2$ μm to an arbitrary potential ϕ_o in Fig. 3.3. Since the electric field is zero outside the charged region, the potential in this region is constant. To check the sketch of $\phi(x)$, the potential at $x = -2$ μm must be higher than at $x = 3$ μm since the electric field points from regions of high potential (positive charge) to regions of low potential (negative charge)

In summary, the basic features of the electric field and potential can be found from the charge density without excessive use of formal mathematics. In order to understand electronic devices, it is very important to *know* which way the electric field points or whether an answer for the potential "makes sense." There will be



➤ **Figure 3.3** Electrostatic potential for the charge distribution in Fig. 3.1. The reference potential is ϕ_o at $x < -2$ μm.

many opportunities in this chapter, as well as Chapters 4 and 6, to practice the skills needed to master applied electrostatics.

3.1.4 Boundary Conditions for Potential and Electric Field

Electronic devices are made of layers of different materials. In order to solve for the potential or electric field through such an inhomogeneous structure, we need conditions for ϕ and E at the boundary between two materials. Starting with the potential, we observe that an abrupt jump in ϕ would lead to an infinite electric field at the boundary according to Eq. (3.4). Since infinite electric fields are not possible (they would tear the material apart), we conclude that $\phi(x)$ must be continuous

continuity of potential at a boundary

$$\boxed{\phi(x_b^-) = \phi(x_b^+)} \tag{3.9}$$

where the boundary is located at $x = x_b$.

On the other hand, the electric field usually jumps at a boundary. Figure 3.4(a) shows the case of a boundary between materials 1 and 2, where we have located the boundary at $x = 0$ for convenience. The electric field is indicated at two points located a very small distance Δ on either side of the boundary.

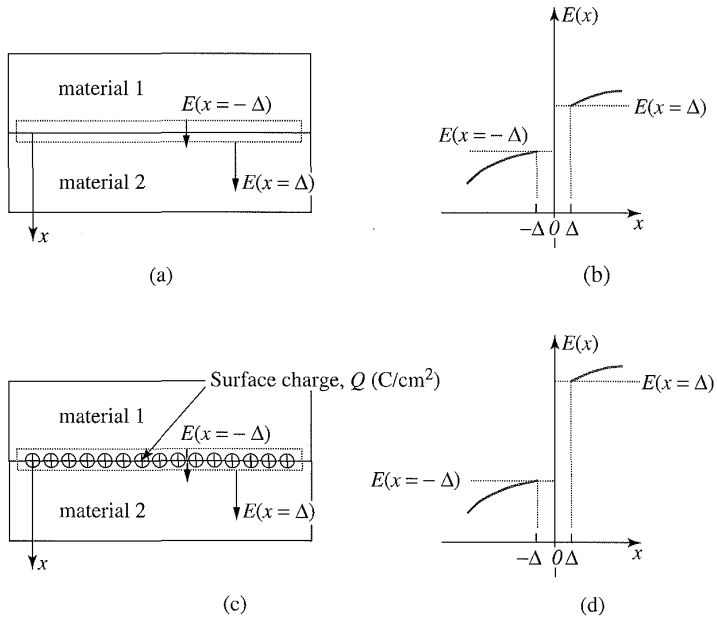
We can use the integral form to Gauss's Law to relate the electric field at $x = -\Delta$ to that at $x = +\Delta$

$$\int_{-\Delta}^{+\Delta} d\varepsilon E(x) = \varepsilon_2 E(x=\Delta) - \varepsilon_1 E(x=-\Delta) = \int_{-\Delta}^{+\Delta} \rho(x)dx = 0 \tag{3.10}$$

The permittivity is left inside the integral on the left hand side since it varies with position through the interval $-\Delta \le x \le \Delta$ The boundary condition is found by letting $\Delta \to 0$

electric field jump for charge-free boundary

$$\boxed{E(0^+) = \left(\frac{\varepsilon_1}{\varepsilon_2}\right) E(0^-)} \tag{3.11}$$



➤ **Figure 3.4** (a) Boundary between materials 1 and 2 with permittivities $\varepsilon_1 > \varepsilon_2$ and (b) resulting jump in the electric field. (c) Boundary, with a surface charge Q located at the interface and (d) resulting jump in the electric field.

in which $E(0^+)$ is the electric field in material 2 at the boundary and $E(0^-)$ is the electric field in material 1 at the boundary. Figure 3.4(b) shows the jump in the electric field with Δ not quite zero, for the case where $\varepsilon_1 > \varepsilon_2$. When $\Delta \to 0$, the plot of electric field becomes discontinuous at the boundary.

A common interface in microelectronic devices is between silicon ($\varepsilon_s = 11.7\ \varepsilon_0$ where $\varepsilon_0 = 8.85 \times 10^{-14}$ F/cm is the permittivity of vacuum) and SiO_2 (silicon dioxide) which has $\varepsilon_{ox} = 3.9\ \varepsilon_0$. For this case, the electric field in the silicon $E_s(0^+)$ is related to the electric field in the oxide $E_{ox}(0^-)$ by

$$E_s(0^+) = \left[\frac{3.9}{11.7}\right] E_{ox}(0^-) = \frac{E_{ox}(0^-)}{3.0} \tag{3.12}$$

The electric field therefore drops by a factor of three from the SiO_2 side to the silicon side of the boundary.

In some cases of practical importance, there can be a sheet of charge Q [units: C/cm^2] at the boundary as shown in Fig. 3.4(c). The presence of this charge sheet changes the integral in Eq. (3.10)

$$\int_{-\Delta}^{+\Delta} d\varepsilon E(x) = \varepsilon_2 E(x{=}\Delta) - \varepsilon_1 E(x{=}-\Delta) = \int_{-\Delta}^{+\Delta} \rho(x)dx = Q \tag{3.13}$$

The charge density $\rho(x)$ [units: C/cm^3] that describes the sheet charge is a "delta function" at the boundary—it integrates to Q even when $\Delta \to 0$. The boundary condition for the case with a sheet charge is

electric field jump for charged boundary

$$E(0^+) = \left(\frac{\varepsilon_1}{\varepsilon_2}\right)E(0^-) + \frac{Q}{\varepsilon_2} \tag{3.14}$$

Figure 3.4(d) shows the jump in E for the case of a positive sheet charge Q and $\varepsilon_1 > \varepsilon_2$. The discontinuity in E is increased by the presence of the positive charge sheet, in comparison to the uncharged boundary in Fig. 3.4(a)-(b).

EXAMPLE 3.1 Metal-Metal Capacitor Electrostatics

In many IC processes, there are two or more levels of metallization separated by deposited oxide films. A planar capacitor can be formed by overlapping metal layers as shown in the layout and cross section in Fig. Ex3.1A below. Assume that a one-dimensional, parallel-plate analysis is adequate.

(**a**) Find the capacitance C in fF/(μm)^2, where 1 fF = 10^{-15} F.

SOLUTION

By definition, capacitance is the derivative of charge with respect to voltage $C = dQ/d\,V$; in this case, C is constant. We are interested in finding the capacitance per unit area, so the charge Q is also interpreted as per unit area. When a positive voltage V is applied to the capacitor, a positive surface charge density $+Q$ accumulates on the top plate of the capacitor, which is mirrored by an equal and opposite negative charge surface charge density $-Q$ on the bottom plate.

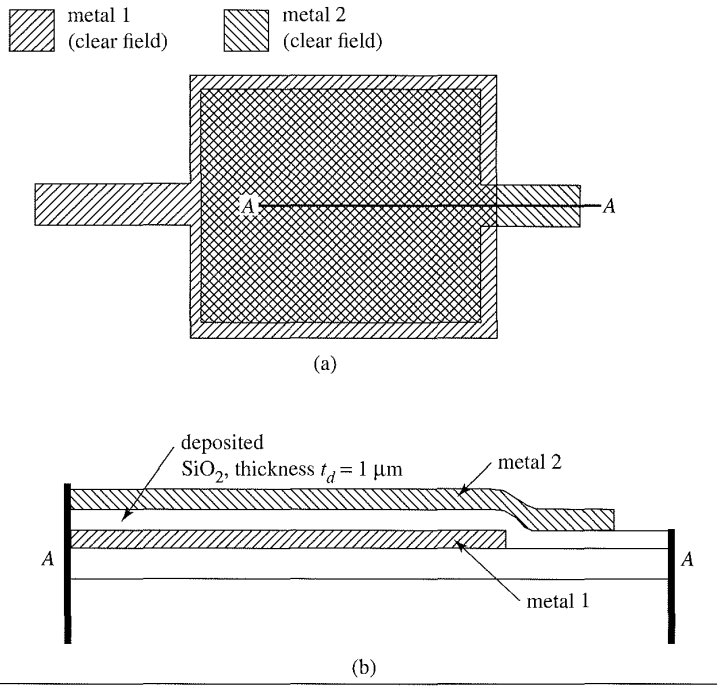
The close-up view of the gap between the plates in Fig. Ex3.1B shows an interior region of the capacitor with all the electric field lines pointing in the $+x$ direction. In this one-dimensional analysis, we will neglect the 2-D effects at the perimeter of the capacitor and assume that all the field lines are parallel to the x axis

As shown in the close-up view of the capacitor, the electric field is constant in the oxide since there is no charge present (see Eq. (3.1), Gauss's Law). Since the electric field is constant and the voltage drop across the gap is V, it follows that $E = V/t_d$. The boundary condition in Eq. (3.14) connects the electric field to the charge density Q [C/cm^2] on the plates

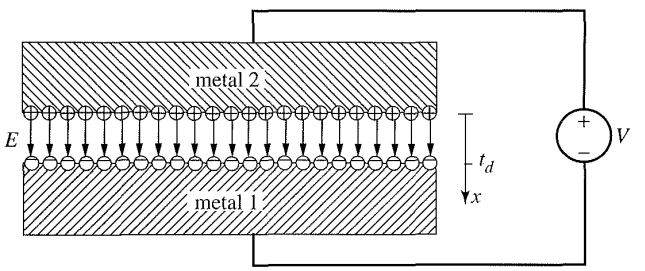
$$E = \left(\frac{\varepsilon_o}{\varepsilon_{ox}}\right)E_{metal} + \frac{Q}{\varepsilon_{ox}} = \frac{Q}{\varepsilon_{ox}}$$

where ε_{ox} is the permittivity of the deposited SiO_2 and $E_{metal} = 0$ (no electric field inside metals). We can substitute $E = V/t_d$ and find the capacitance per unit area

$$\frac{V}{t_d} = \frac{Q}{\varepsilon_{ox}} \to C = \frac{dQ}{dV} = \frac{\varepsilon_{ox}}{t_d}$$



➤ **Figure Ex3.1A** Metal-metal IC capacitor: a) layout and (b) cross section.



➤ **Figure Ex3.1B** Close-up of metal-metal capacitor with applied voltage, showing surface charges on top and bottom metal plates and electric field lines.

Substituting $\varepsilon_{ox} = 3.9\ \varepsilon_o = 3.45 \times 10^{-13}$ F/cm and $t_d = 1\ \mu m = 1 \times 10^{-4}$ cm for the deposited oxide film, we find that

$$C = 3.45 \times 10^{-9} \text{F/cm}^2 = 34.5\ \text{aF}/\mu\text{m}^2$$

where 1 aF = 1 attofarad = 10^{-18} F. For example, a 100 × 100 μm^2 overlap area between metal 1 and metal 2 yields a capacitance of 345 fF, where 1 fF = 1 femtofarad = 10^{-15} F.

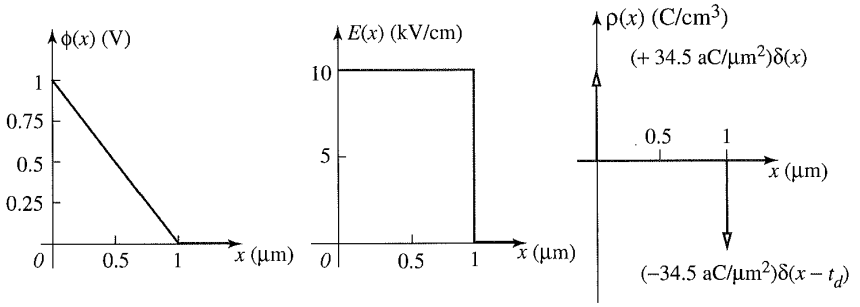
(**b**) Sketch the charge density, electric field, and potential when a voltage $V = 1$ V is applied between the top and bottom metal layers.

SOLUTION

For an applied voltage of $V = 1$ V, the electric field in the gap is

$$E = (1 \text{ V})/(1 \times 10^{-4} \text{ cm}) = 10 \text{ kV/cm}$$

From the definition of potential, the electrostatic potential $\phi(x)$ has a constant slope equal to –10 kV/cm. The bottom electrode is taken as the zero reference in the plot of $\phi(x)$ below. Finally, the charge density ρ [C/cm^3] is zero everywhere except on the inside surfaces of the metal plates. The mathematical function for the surface charge Q is a delta function, which is represented by "spikes" at the two surfaces. Note that the derivative of the electric field is consistent with these spikes at $x = 0$ and $x = t_d$.



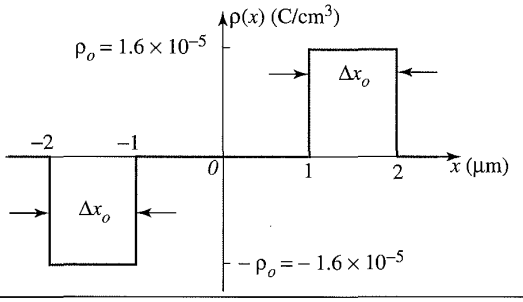
➤ **Figure Ex3.1C** Potential, electric field, and charge density for the capacitor with $t_d = 1$ μm.

EXAMPLE 3.2 Field and Potential from a Distributed Charge Density

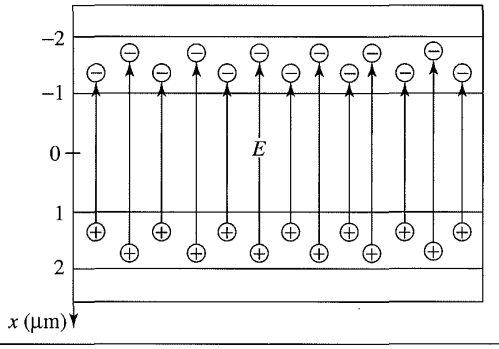
(**a**) Given the following charge distribution, sketch the electric field $E(x)$ and find the numerical value of $E(0)$.

SOLUTION

We should first be sure which way the electric field points.The electric field points from positive charges to negative charges—in the opposite direction as the x axis. Therefore, the electric field is negative. A sketch of the charge distribution with the electric field lines is shown in Fig. Ex3.2B, which indicates that the electric field is constant between the charged layers for -1 μm $< x < 1$ μm, and zero outside the charged layers for $x < -2$ μm or $x > 2$ μm. Our basic intuition from the parallel-plate capacitor in Ex. 3.1 (derived from Gauss's Law) also applies to this "smeared" charge distribution.



➤ **Figure Ex3.2A** Arbitrary charge distribution in silicon.

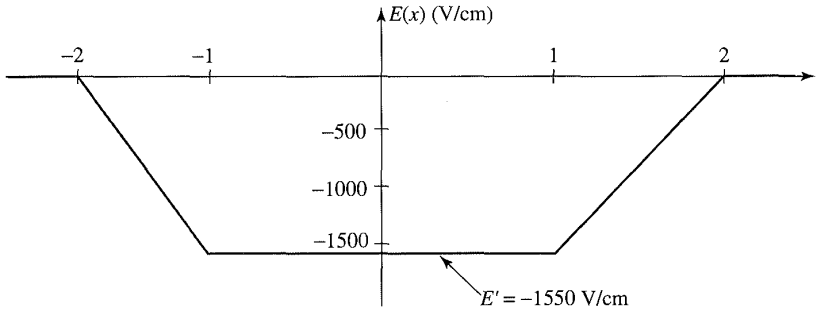


➤ **Figure Ex3.2B** Sketch of electric field lines pointing from positive to negative charges.

We now want to find the field in between the charged layers. Rather than integrate the charge density, we can use the integral form of Gauss's Law to relate the total charge $+Q$ or $-Q$ [C/cm^2] in one of the charged regions to the electric field

$$E(x) - E(-2\mu m) = \frac{-Q}{\varepsilon_s} = \frac{-\rho_o \Delta x_o}{\varepsilon_s} = \frac{-1.6 \times 10^{-9}}{1.036 \times 10^{-12}} = -1.55 \times 10^3 \text{ V/cm}$$

where the boundary at x is anywhere between the charged layers: $-1\ \mu m < x < 1\ \mu m$. Since the field at the other boundary at $x = -2\ \mu m$ is zero, the electric field between the charged layers is $E' = -1550$ V/cm. Within the charged layers, the charge density is constant and the electric field must vary linearly with x according to the differential form of Gauss's Law. This knowledge, combined with the electric field being zero at the boundaries of the charged region at $x = -2\ \mu m$ and $x = 2\ \mu m$, enables us to sketch the trapezoidal shape of $E(x)$ in Fig. Ex3.2C. By "turning the crank" and integrating $\rho(x)/\varepsilon_s$, we could have directly calculated the functional form of $E(x)$ assuming we didn't make algebraic errors along the way. By reasoning through the implications of Gauss's Law and the shape of the charge distribution, we are certain we have the correct answer before doing the math.



➤ **Figure Ex3.2C** Plot of electric field $E(x)$ for the charge distribution in this example.

(**b**) Sketch $\phi(x)$ assuming $\phi(0) = 0$ V. What is $\phi(2\ \mu m)$?

SOLUTION

In order to plot the potential $\phi(x)$, we must graphically integrate $-E(x)$. Starting from $x = 0$, at which point we are given that $\phi(0) = 0$, we can integrate in the $+x$ direction. We know $\phi(2\ \mu m) > 0$ since the positive charge is on the right side of the plot of $\rho(x)$ and also that the electric field points from right to left.

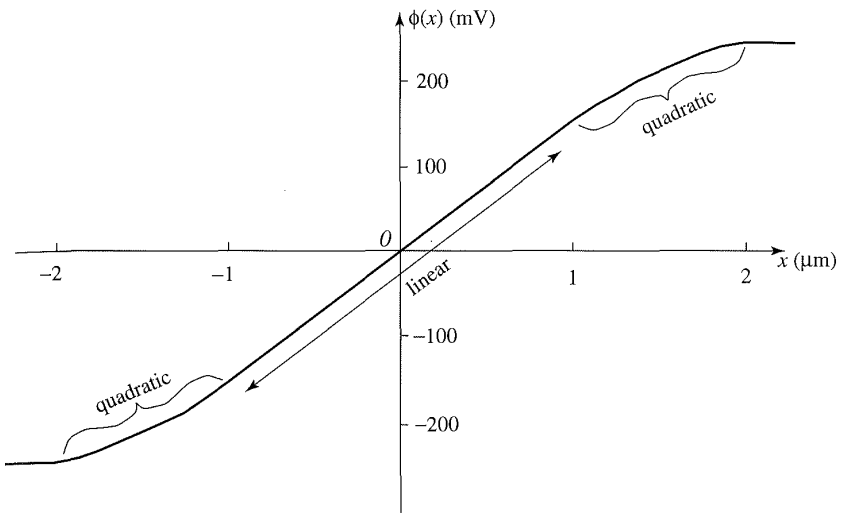
Over the interval $0 < x < 1\ \mu m$, the potential varies linearly since the field is constant. The potential at $x = 1\ \mu m$ is given by

$$\phi(1\mu m) = \phi(0) + \int_0^{1\mu m} -E(x')\,dx' = (1550\ \text{V/cm})(10^{-4}\text{cm}) = 155\ \text{mV}$$

Over the interval $1\ \mu m < x < 2\ \mu m$, the potential varies quadratically since the field is a linear function. The potential change from $x = 1\ \mu m$ to $x = 2\ \mu m$ is given by the area under $-E(x)$

$$\phi(2\mu m) - \phi(1\mu m) = \left(\frac{1}{2}\right)(10^{-4}\text{cm})\ (1550\text{V/cm}) = 77.5\text{mV}$$

Therefore, the potential at $x = 2\ \mu m$ is $\phi(2\ \mu m) = 155 + 77.5\ \text{mV} = 232.5\ \text{mV}$. Since the electric field is symmetric about $x = 0$, the potential for $x < 0$ is $\phi(-x) = -\phi(x)$. The potential $\phi(x)$ is plotted below. If the function for $\phi(x)$ is needed, then the negative of $E(x)$ can be integrated and the solutions in each interval patched together using the continuity boundary condition.



➤ **Figure Ex3.2D** Plot of the potential for the distributed charge density in Example 3.2

EXAMPLE 3.3 Metal-Oxide-Semiconductor (MOS) Capacitor

Another type of common IC capacitor is the Metal-Oxide-Semiconductor (MOS) capacitor. It is formed by a metal layer, an oxide layer, and a layer of silicon. A single mask is needed to create the basic sandwich as shown in the layout and cross section in Fig. Ex3.3A.

(a) Given the qualitative picture of the charge distribution and electric field lines shown in Fig. Ex3.3B, plot the charge density $\rho(x)$. Find an equation relating Q_G and X_d.

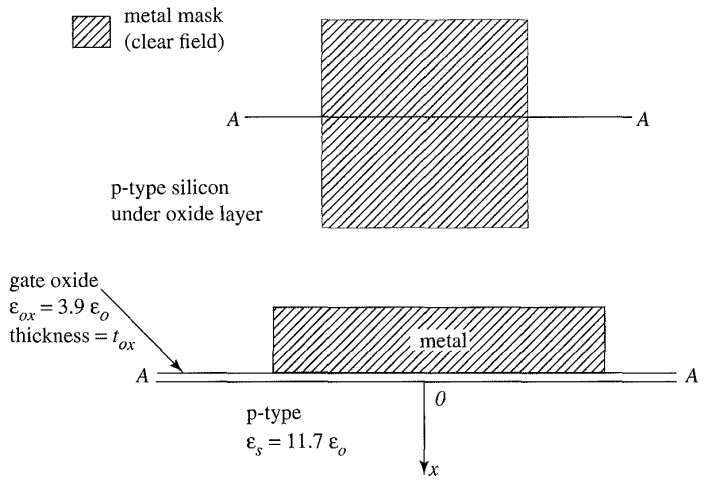
SOLUTION

First, we need to define the coordinate system. The origin is located at the interface between the oxide and p-type silicon. The oxide and metal extend into the $-x$ region and the silicon extends into the $+x$ region. The charge on the metal gate is located at the bottom surface as a sheet charge density, with units C/cm^2. On the plot of charge density in C/cm^3 in Fig. Ex3.3C, the gate charge is plotted as a delta function $Q_G\delta(x + t_{ox})$ that is located at the metal/oxide interface at $x = -t_{ox}$.

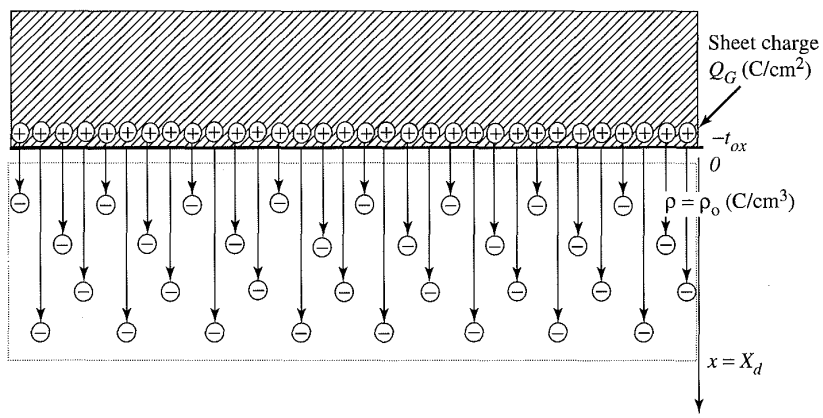
The total charge (gate charge + charge in the silicon substrate) is zero; therefore, the total charge in the silicon per unit area must be equal and opposite to the charge on the gate

$$Q_G = -(\rho_o X_d) = -\rho_o X_d \rightarrow X_d = Q_G/(-\rho_o)$$

(b) Find and sketch the electric field $E(x)$ and find an expression for the field in the oxide, E_{ox}, in terms of ρ_o, X_d, and ε_{ox}.



➤ **Figure Ex3.3A** Metal-oxide-silicon capacitor: (a) layout and (b) cross section.



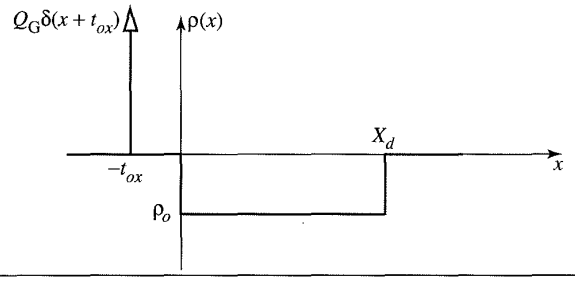
➤ **Figure Ex3.3B** Charge distribution and electric field lines for Example 3.3.

SOLUTION

The electric field points from positive charges to negative charges in the $+x$ direction, and is, therefore, positive. The electric field is confined to the region $-t_{ox} < x < X_d$ because the total charge in this interval is zero. In the oxide $\rho(x) = 0$ and the differential form of Gauss's Law

$$\frac{dE}{dx} = \frac{\rho(x)}{\varepsilon_{ox}} = 0, \text{ for } -t_{ox} < x < 0$$

indicates that the field $E(x) = E_{ox}$ is constant for $-t_{ox} < x < 0$. In the charged region in the silicon, the charge density is constant at $\rho(x) = \rho_o$ and, therefore, the



➤ **Figure Ex3.3C** Charge density $\rho(x)$ corresponding to the qualitative charge distribution in Fig. Ex3.3B.

electric field is a linear function of x there. Since the field is known to be zero at $x = X_d$, we only need the value of the field at $x = 0$ to complete the sketch. The boundary condition at the oxide/silicon interface is

$$\varepsilon_{ox}E_{ox} = \varepsilon_s E(x=0^+) \rightarrow E(x=0^+) = (\varepsilon_{ox}/\varepsilon_s)E_{ox} = E_{ox}/3$$

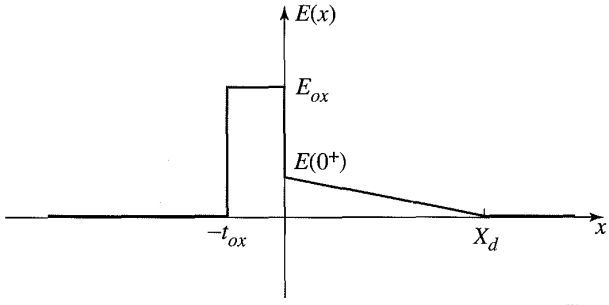
Gauss's Law in integral form is a convenient way to find the value of the electric field at the boundary of a charged region. Using Eq. (3.2) over the interval $0 \le x \le X_d$, we find

$$\varepsilon_s E(X_d) - \varepsilon_{ox}E_{ox} = \int_0^{X_d} \rho_o dx = \rho_o X_d$$

Since $E(X_d) = 0$, the oxide field is

$$E_{ox} = \frac{-\rho_o X_d}{\varepsilon_{ox}}$$

Note that there is the possibility of sign errors in applying Eq. (3.2); however, we are not depending on the math to tell us the sign since we already *know* $E_{ox} > 0$. The electric field $E(x)$ is plotted in Fig. Ex3.3D.



➤ **Figure Ex3.3D** Sketch of electric field for metal-oxide-silicon capacitor with the charge distribution in Fig. Ex3.3B.

(c) Sketch the potential $\phi(x)$ through the structure given that the potential of the metal gate is $\phi(-t_{ox}) = \phi_m$. Find an expression for the surface potential $\phi_s = \phi(0)$ in terms of ϕ_m, ρ_o, X_d, ε_{ox}, and t_{ox}. In addition, find an expression for the potential drop across the charged region in the silicon in terms of ρ_o, X_d, and ε_s.

SOLUTION

As we learned in Example 3.2, there is no need to integrate $-E(x)$ in order to sketch the potential. From the definition of potential and the sketch of $E(x)$ above, the potential is a linear function in the oxide and a quadratic function of x in the charged region of the silicon substrate. Poisson's Equation is useful for finding the concavity of the potential in the charged region

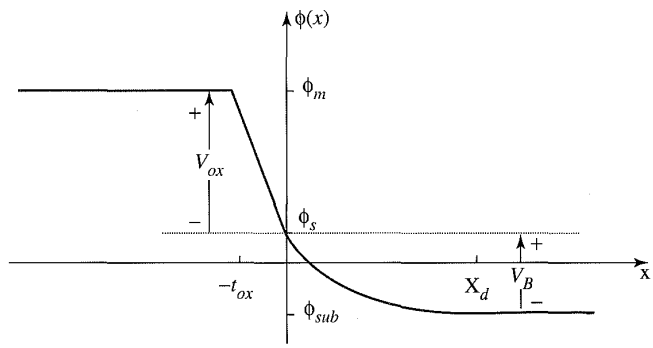
$$\frac{d^2\phi(x)}{dx^2} = -\frac{\rho_o}{\varepsilon_s} > 0$$

Therefore, the potential is "concave up" in the charged region. Finally, we also know that the potential is continuous at the boundary $x = 0$ between the oxide and silicon. We can piece together the linear and quadratic solutions and sketch the potential between the metal and underlying substrate ($x > X_d$), the potential of which is defined as ϕ_{sub}. Figure Ex3.3E is a plot of the potential through the MOS capacitor.

On the sketch of $\phi(x)$ below, the total drop has been divided into two parts, the drop across the oxide V_{ox} and the drop across the charged region in the silicon V_B.

$$\phi_m - \phi_{sub} = V_{ox} + V_B$$

Note that the polarities of V_{ox} and V_B are defined on the potential plot in Fig. Ex3.3E to avoid confusion with signs. The drop across the oxide can be found by integrating the oxide field from $x = -t_{ox}$ to $x = 0$.



➤ **Figure Ex3.3E** Potential plot for metal-oxide-silicon capacitor with the charge distribution in Fig. Ex3.3B.

$$\phi_s - \phi_m = \int_{-t_{ox}}^{0} (-E_{ox})\,dx = -E_{ox}\,(0-(-t_{ox})) = -V_{ox}$$

and therefore, $V_{ox} = E_{ox}t_{ox}$. The surface potential can be found in terms of the given parameters by substituting for the oxide field

$$\phi_s = \phi_m - E_{ox}t_{ox} = \phi_m - \frac{(-\rho_o)\,X_d\,t_{ox}}{\varepsilon_{ox}}$$

This expression can be simplified by defining the charge Q_B [C/cm^2] in the substrate as $Q_B = \rho_o X_d = -Q_G$, and by recognizing that the oxide capacitance per unit area is $C_{ox} = \varepsilon_{ox}/t_{ox}$. The surface potential and drop across the oxide are related to these charges by

$$\phi_s = \phi_m + \frac{Q_B}{C_{ox}} \text{ and } V_{ox} = -\frac{Q_B}{C_{ox}} = \frac{Q_G}{C_{ox}}$$

The last result is not unexpected since the voltage drop V_{ox} must be proportional to the charge $\pm Q_G$ stored on each side of the insulating oxide.
Finally, Poisson's Equation is used to find the drop V_B across the charged region

$$\frac{d^2\phi(x)}{dx^2} = -\frac{\rho_o}{\varepsilon_s} \text{ for } 0 < x < X_d$$

This ordinary differential equation can be solved for $\phi(x)$. The drop across the charged region is

$$\phi(0) - \phi(X_d) = V_B = \frac{1}{2}\left(-\frac{\rho_o}{\varepsilon_s}\right)X_d^2 = \left(-\frac{\rho_o X_d}{2\varepsilon_s}\right)X_d = \frac{-Q_B}{2\varepsilon_s}X_d$$

Note that we have quantified the essential features of the electric field and the potential in the MOS structure without depending on formal mathematical analysis.

➤ 3.2 CARRIER CONCENTRATION AND POTENTIAL IN THERMAL EQUILIBRIUM

In order to find complete solutions for the electrostatic potential in silicon, we must define a potential reference. Although the choice is arbitrary, we will find it very convenient to select a reference that is related to the thermal equilibrium carrier concentration.

Recall from Chapter 2 that thermal equilibrium is equivalent to the absence of *any* stimulus to the device—zero applied voltage, no external light source, etc. In this state, the current density J_o must be zero. Perhaps less obvious is that the electron *and*

hole current densities J_{no} and J_{po} must *each* be equal to zero in thermal equilibrium. The physical reasoning is that the populations of electrons and the holes are each in equilibrium and, therefore, must have zero current densities. For example, $J_{no} \neq 0$ implies that electrons will pile up with time at one end of a structure in thermal equilibrium, which is hardly reasonable.

Setting the electron current density equal to zero in thermal equilibrium:

$$0 = qn_o\mu_n E_o + qD_n\frac{dn_o}{dx} \tag{3.15}$$

Solving for the gradient and using Einstein's Relation (Eq. 2.51) to eliminate the ratio of the mobility to the diffusion coefficient, we find that

$$\frac{dn_o}{dx} = \left(-\frac{\mu_n}{D_n}\right)n_o E_o = \left(-\frac{q}{kT}\right)n_o\left(-\frac{d\phi_o}{dx}\right) \tag{3.16}$$

where the negative gradient of the potential ϕ_o has been substituted for the electric field. Solving for the differential potential $d\phi_o$, we find

$$d\phi_o = \left(\frac{kT}{q}\right)\frac{dn_o}{n_o} = V_{th}\frac{dn_o}{n_o} \tag{3.17}$$

in which we have substituted the thermal voltage $V_{th} = kT/q$.

This differential equation can be solved by direct integration from a reference point x_o to an arbitrary point x

$$\phi_o(x) - \phi_o(x_o) = V_{th}\ln\left(\frac{n_o(x)}{n_o(x_o)}\right) \tag{3.18}$$

To find a numerical value for $\phi_o(x)$, we need to select a reference for the potential. By convention, the reference for potential is chosen to be the point where the electron concentration is the intrinsic concentration

potential reference in thermal equilibrium

$$\phi_o(x_o) = 0 \text{ when } n_o(x_o) = n_i \tag{3.19}$$

Substituting this reference into Eq. (3.18) defines the potential in terms of the electron concentration in thermal equilibrium

$$\phi_o(x) = V_{th}\ln\left(\frac{n_o(x)}{n_i}\right) \tag{3.20}$$

Note that n-type silicon has a positive electrostatic potential since $n_o > n_i$ when electrons are the majority carriers. In order words, a region of n-type silicon is *relatively* positive, when compared with an intrinsic region.

By inverting this equation, we can express the electron concentration in terms of the electrostatic potential

$$n_o(x) = n_i e^{\phi_o(x)/V_{th}} \tag{3.21}$$

A parallel derivation for hole concentration leads to the following results

$$\phi_o(x) = -V_{th}\ln\left(\frac{p_o(x)}{n_i}\right) \qquad p_o(x) = n_i e^{-\phi_o(x)/V_{th}} \tag{3.22}$$

so that p-type silicon ($p_o > n_i$) has a negative electrostatic potential, with respect to intrinsic silicon.

3.2.1 The 60mV Rule

In order to get an easy-to-remember rule, we replace the natural logarithm with the base 10 logarithm in Eq. (3.20) and evaluate the expression at a "warm" room temperature (V_{th} = 26 mV):

60mV rule

$$\boxed{\phi_o(x) = (26\text{mV})\ln(10)\log\left(\frac{n_o(x)}{10^{10}}\right) = (60\text{mV})\log\left(\frac{n_o(x)}{10^{10}}\right)} \tag{3.23}$$

This **60 mV rule** can be evaluated without a calculator for carrier concentrations that are equal to 10^{14}cm^{-3}, 10^{15}cm^{-3}, etc. For example, $n_o = 10^{16}$ cm^{-3}corresponds to the potential

$$\phi_o = (60\text{ mV})\log(10^{16}/10^{10}) = (60\text{ mV})(6) = 360\text{ mV} \tag{3.24}$$

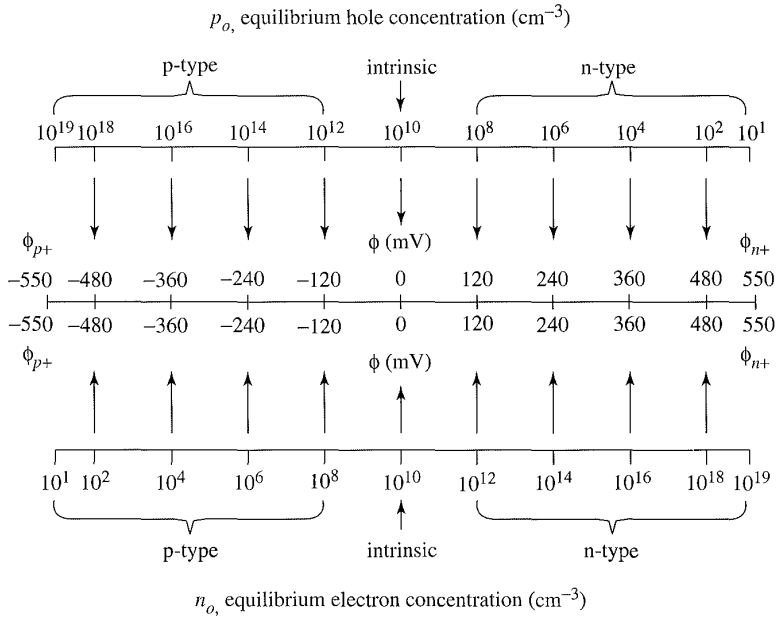
We can also express the potential as a function of the hole concentration in thermal equilibrium by substituting the mass-action law $n_o = n_i^2/p_o$ into Eq. (3.23):

$$\phi_o(x) = -(60\text{mV})\log\left(\frac{p_o(x)}{10^{10}}\right) \tag{3.25}$$

For a hole concentration of $p_o = 10^{16}$ cm^{-3}, the potential is $\phi_o = -360$ mV. The carrier concentrations can range from the intrinsic concentration to about 10^{19} cm^{-3} for electrons or for holes, which corresponds to a potential range in thermal equilibrium of $-540\text{ mV} \le \phi_o \le 540$ mV. The potential of silicon that is extremely heavily doped with donors (called **n^+ silicon**) is $\phi_{n+} = 550$ mV. For **p^+ silicon,** which is heavily doped with acceptors, the potential is $\phi_{p+} = -550$ mV. Figure 3.5 shows the 60 mV rule in graphical form.

n^+ and p^+ silicon

The derivation of Eq. (3.23) only required that the electron and hole currents are zero throughout the sample. In addition to thermal equilibrium, this requirement is also satisfied in capacitors since the insulator prevents carrier transport through the structure.



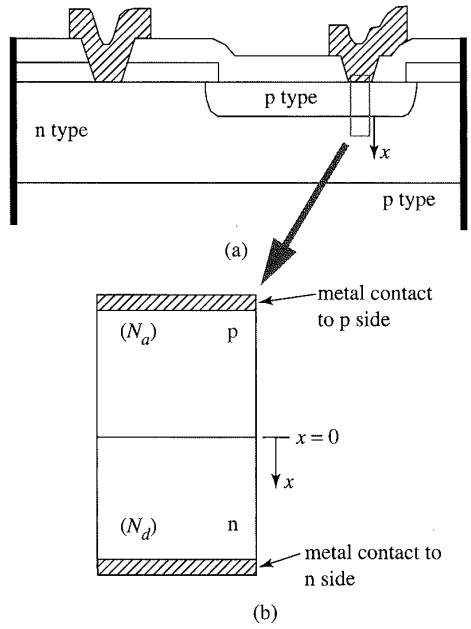
➤ **Figure 3.5** 60 mV rule relating electrostatic potential in thermal equilibrium to hole or electron concentration at room temperature.

➤ 3.3 THE pn JUNCTION: A FIRST PASS

A **pn junction** consists of a p-type silicon region immediately adjacent to an n-type silicon region. As we have seen in Chapter 2, pn junctions are built into nearly all IC device structures—either intentionally or as by-products of the fabrication process. Figure 3.6(a) shows the cross section of a pn juction. The p and n regions are contacted by metal interconnections on the top surface of the chip which are used to apply a voltage across the junction. For now, we will ignore the underlying p-type substrate that forms another junction with the n side of the pn junction. The substrate is electrically contacted and set at a voltage (often ground) to ensure that adjacent device structures have minimum electrical interaction.

The essential features of the pn junction can be understood from the one-dimensional simplification shown in Fig. 3.6(b). For mathematical simplicity, the doping concentrations N_a on the p-side and N_d on the n-side are assumed to be constant. Furthermore, it is assumed that N_a and N_d already incorporate any doping compensation effects and thus represent the net concentrations of acceptors and donors.

How do we develop a qualitative picture of the electrostatics of the pn junction? As a start, we consider the case of thermal equilibrium. We know that the hole and electron current densities $J_{po} = 0$ and $J_{no} = 0$ throughout the structure. Therefore, it follows that the hole drift current density must be equal in magnitude and opposite in direction to the hole diffusion current density throughout the pn junction in thermal equilibrium.



➤ **Figure 3.6** (a) Cross section of a typical pn junction fabricated in an IC process, (b) one-dimensional pn junction consisting of boxed region in (a) with metal contact to n-side added at the bottom. The metallurgical junction (boundary between p-type and n-type regions) is located at $x = 0$.

$$J_{po} = J_{po}^{dr} + J_{po}^{diff} = qp_o\mu_p E_o - qD_p\frac{dp_o}{dx} = 0 \rightarrow qp_o\mu_p E_o = qD_p\frac{dp_o}{dx} \quad (3.26)$$

Similarly, the electron drift and diffusion current densities must cancel

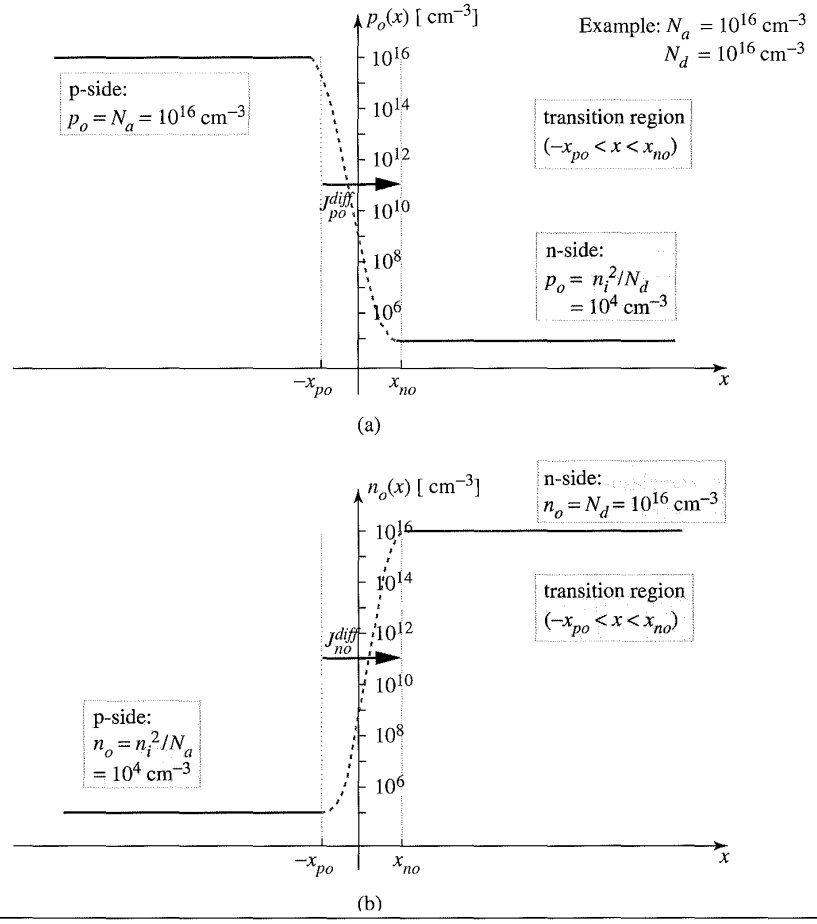
$$J_{no} = J_{no}^{dr} + J_{no}^{diff} = qn_o\mu_n E_o + qD_n\frac{dn_o}{dx} = 0 \rightarrow qn_o\mu_n E_o = -qD_n\frac{dn_o}{dx} \quad (3.27)$$

A rough sketch of the hole concentration $p_o(x)$ is helpful in developing insights into the implications of Eq. (3.26) for the pn junction. Far enough from the metallurgical junction at $x = 0$, the hole concentration on the p-side of the junction must be $p_o(x) = N_a$. In effect, we are asserting that the p-type region can be considered as bulk silicon when it is far enough from the junction since the effects of the pn junction are confined to a region near $x = 0$. We shall quantify the width of the "transition region" when the junction is quantitatively analyzed later in the chapter. However, as a preview of that analysis, the order of magnitude of a typical transition width is 1 μm. On the n-side of the junction, the hole concentration far enough from the junction is $p_o(x) = n_i^2/n_o = n_i^2/N_d$–the hole concentration in bulk n-type silicon. In Fig. 3.7(a), the hole concentration $p_o(x)$ is sketched through the pn junction in thermal equilibrium. The transition region is confined to a region of undetermined width $-x_{po} < x < x_{no}$, where $-x_{po}$ is its border on the p-side and x_{no} is its border on the n-side of the junction in thermal equilibrium. Using similar reasoning, we can sketch the electron concentration $n_o(x)$ in thermal equilibrium through the pn junction.

The concentrations in the transition region are shown as dotted lines connecting the bulk concentrations since we do not know the functional forms.

The most obvious feature of the equilibrium concentrations $p_o(x)$ and $n_o(x)$ is the large gradients across the transition region. The diffusion current densities are proportional to these gradients as described in Eqs. (3.26) and (3.27). Figure 3.7 shows that both J_{po}^{diff} and J_{no}^{diff} are large and positive in the transition region. In order for $J_{po} = 0$ and $J_{no} = 0$ in the transition region, as they must in thermal equilibrium, it is necessary that there be balancing drift current densities $J_{po}^{dr} = -J_{po}^{diff}$ and $J_{no}^{dr} = -J_{no}^{diff}$. Substituting into Eq. (3.26), the required electric field E_o in thermal equilibrium is

$$E_o = \frac{qD_p \dfrac{dp_o}{dx}}{qp_o\mu_p} = \frac{-qD_n \dfrac{dn_o}{dx}}{qn_o\mu_n} \tag{3.28}$$



➤ **Figure 3.7** (a) Qualitative hole concentration $p_o(x)$ and (b) electron concentration $n_o(x)$ in thermal equilibrium. The doping concentrations are 10^{16} cm^{-3} for this example. The equilibrium hole and electron diffusion current densities are both large and positive in the transition region because of the large concentration gradients.

Since we do not know the functional forms for $p_o(x)$ or for $n_o(x)$, we cannot quantify the electric field in the transition region. However, we do know from Eq. (3.28) and from Fig. 3.7 that the electric field in the transition region must be *large* in magnitude and must point in the $-x$ direction so its sign is *negative*. In contrast, outside the transition region p_o and n_o are constants and we conclude that the diffusion current densities and the electric field are zero.

How can a large electric field exist in the transition region? From our understanding of electrostatics, there must be a region of negative and positive charge near $x = 0$. There are four charged species in silicon, two mobile (holes and electrons) and two fixed (ionized donors and ionized acceptors). The charge density in thermal equilibrium $\rho_o(x)$ is given by

$$\rho_o(x) = q\,(p_o - n_o + N_d - N_a) \tag{3.29}$$

In the transition region on the p-side of the junction over the interval $-x_{po} \le x \le 0$, there are very few electrons (see Fig. 3.7(b)) and no donors, so Eq. (3.29) simplifies to

$$\rho_o(x) \cong q\,(p_o - N_a)\ , \text{ for } -x_{po} \le x \le 0 \tag{3.30}$$

As shown in Fig. 3.7(a), the thermal equilibrium hole concentration $p_o \ll N_a$ in the interval $-x_{po} \le x \le 0$, which leads to a *negative* charge density

$$\rho_o(x) < 0\,, \text{ for } -x_{po} \le x \le 0 \tag{3.31}$$

On the n-side of the transition region in the interval $0 \le x \le x_{no}$,there are few holes (see Fig. 3.7(a)) and no acceptors, therefore Eq. (3.29) simplifies to

$$\rho_o(x) \cong q\,(-n_o + N_d)\ , \text{ for } 0 \le x \le x_{no} \tag{3.32}$$

The electron concentration in thermal equilibrium $n_o(x) < N_d$ in this interval (see Fig. 3.7(b)), making the charge density there *positive*

$$\rho_o(x) > 0\,, \text{ for } 0 \le x \le x_{no} \tag{3.33}$$

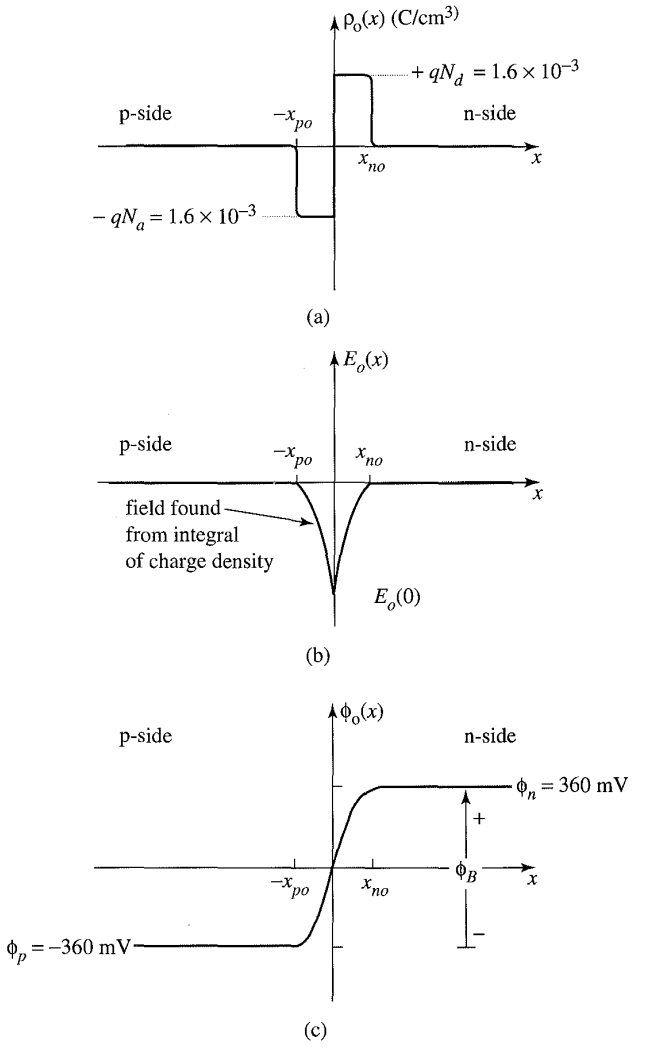
From the sketches of $p_o(x)$ and $n_o(x)$ in Fig. 3.7 and Eqs. (3.30) and (3.32), we can develop the rough sketch of the thermal equilibrium charge density $\rho_o(x)$ in Fig. 3.8(a). Since the doping is symmetrical in Fig. 3.7 ($N_a = N_d = 10^{16}$ cm^{-3}), the transition region is equally split around $x = 0$ in order to maintain charge neutrality. Given the rapid change in concentrations in Fig. 3.7 (note the logarithmic scale), the slopes in $\rho_o(x)$ at $x = -x_{po}$ and $x = x_{no}$ should be greater than depicted in Fig. 3.8(a). Inside the transition region, the mobile carrier concentrations are negligible for finding the charge density. In effect, the mobile carriers have been depleted from the transition region. We will refer to the high-field region at the junction as the **depletion region**.

The sketches for the electric field $E_o(x)$ and the potential $\phi_o(x)$ in thermal equilibrium are given in Figs. 3.8(b) and 3.8(c). The field is found by graphical integration and is approximately triangular in shape. In the regions away from the junction, the potential is given by the 60 mV rule, as shown in Fig. 3.8(c). The potential on the p-side is

defined as ϕ_p and is –360 mV for this diode, since $N_a = 10^{16}$ cm^{-3}. On the n-side, the potential is $\phi_n = 360$ mV since $N_d = 10^{16}$ cm^{-3}.The **built-in potential** ϕ_B is defined as

$$\phi_B = \phi_n - \phi_p = 360 \text{ mV} - (-360 \text{ mV}) = 720 \text{ mV}. \tag{3.34}$$

The sign of the built-in potential is consistent with the direction of the electric field in Fig. 3.7(b) and the fact that the charge density is positive on the n-side of the



➤ **Figure 3.8** Qualitative electrostatics of the pn junction in thermal equilibrium: (a) charge density, (b) electric field, and (c) electrostatic potential. The doping is $N_a = N_d = 10^{16}$ cm^{-3}, as given in Fig. 3.7, which leads to $x_{no} = x_{po}$.

transition region. One way of rationalizing the potential barrier at the junction is that it prevents holes from flooding from the p-side (where they are majority carriers) to the n-side (where they are minority carriers.) Of course, the potential barrier also confines the majority electrons to the n-side of the transition region.

As an electrical device structure, the pn junction has an interesting feature: it has a stored charge with no applied voltage, due to the built-in potential barrier. The charge stored per unit area on the p-side is called the **depletion charge**. In thermal equilibrium, the depletion charge Q_{Jo} [units: C/cm^2] is

$$Q_{Jo} = -qN_a x_{po} \tag{3.35}$$

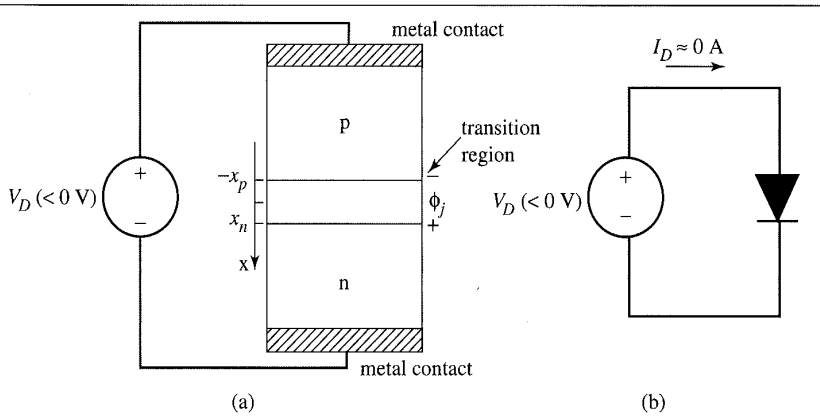
The sign of the depletion charge stored on the p-side of junction is negative, which is consistent with the sign of the potential drop from the p-side to the n-side of the junction in thermal equilibrium:

$$\phi_p - \phi_n = -\phi_B < 0\,. \tag{3.36}$$

Having developed a qualitative picture of the pn junction in thermal equilibrium, we now extend our understanding to the case of an applied voltage. In this chapter, we are considering only the electrostatics of the junction and will only discuss the reverse-biased situation. The reference polarity for the applied voltage V_D is chosen so that a forward bias is positive, as shown in Fig. 3.9(a) and (b), so for reverse bias $V_D < 0$.

In Fig. 3.9(a), we have defined the potential increase from the p to the n side as the **barrier potential** ϕ_j. As we know from basic electronics, the current through the reverse-biased diode is negligible. Since there can be no resistive potential drops across the p and n bulk regions, the applied voltage must be dropped across the depletion region. Accounting for the presence of the built-in potential ϕ_B, the barrier potential under an applied bias is

$$\phi_j = \phi_B - V_D\,. \tag{3.37}$$



➤ **Figure 3.9** (a) pn junction with negative applied voltage V_D (reverse bias) and (b) corresponding circuit schematic.

In thermal equilibrium, $V_D = 0$ and Eq. (3.37) correctly indicates that the barrier potential must equal the built-in potential ϕ_B. Under reverse bias with $V_D < 0$, the barrier at the junction increases, which is consistent with the polarities of V_D and ϕ_j in Fig. 3.9(a).

We have *not* derived Eq. (3.37) from first principles, but rather have established that it correctly accounts for the presence of a built-in barrier at the junction in thermal equilibrium. In Section 3.5, we will investigate in detail how the internal barrier ϕ_j is offset from the externally applied voltage V_D.

What is the effect of the increase in potential barrier on the charge distribution and electric field? More charge is needed on each side of the junction in order to generate enough electric field to yield the larger potential barrier. Since the mobile carriers have been depleted from the depletion region, the charge density is constant at $-qN_a$ on the p-side and $+qN_d$ on the n-side of the junction. In order to get more charge, the width of the depletion region must increase. In Fig. 3.10(a), the edges of the depletion region are shown to have doubled over their thermal-equilibrium values, for the particular reverse bias voltage.

In Fig. 3.10(b), the magnitude of the peak electric field is double that in thermal equilibrium, since the charge on each side of the junction has doubled. Finally, the potential plot in Fig. 3.10(c) shows that the barrier potential under reverse bias has increased over its value in thermal equilibrium: $\phi_j = \phi_B - V_D > \phi_B$. For convenience in plotting $\phi(x)$, we have kept the n-side of the junction fixed at its equilibrium potential and applied the negative diode voltage to the p-side. We can visualize the reverse bias voltage $V_D < 0$ as pulling down the potential on the p-side of the junction, as shown in Fig. 3.10(c).

The effect of reverse bias is to increase the magnitude of the depletion charge q_J, as can be seen by comparing Fig. 3.10(a) with Fig. 3.8(a). As we will derive in Section 3.5, the functional dependence of the depletion charge on the applied bias is:

$$Q_J(V_D) = Q_{Jo}\sqrt{1 - (V_D/\phi_B)} = -qN_a x_{po}\sqrt{1 - (V_D/\phi_B)}\,, \qquad (3.38)$$

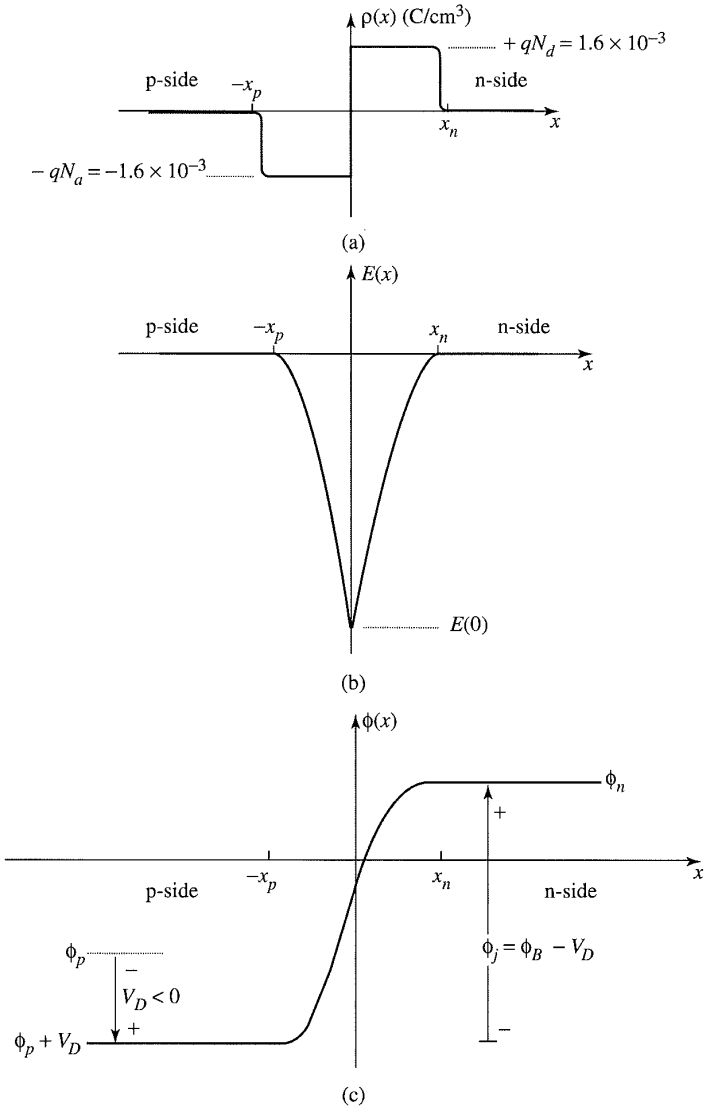
where Q_{Jo} (< 0) is the depletion charge in thermal equilibrium. In summary, the pn junction under reverse bias is a non-linear charge-storage element. Its charge-voltage characteristic in Eq. (3.38) is the starting point for deriving equivalent circuit elements in Section 3.6.

➤ 3.4 THE pn JUNCTION IN THERMAL EQUILIBRIUM

From the previous section, we have gained a "feel" for the pn junction in equilibrium and under reverse bias, which will be valuable in analyzing the electrostatics of the pn junction. Our approach will be to investigate the functional form of the charge density in the transition region and develop an expedient approximation that will simplify the mathematics. We will then solve for the electric field and potential in the transition region and connect these solutions to the known potential in the regions far away from the junction.

3.4.1 The Depletion Approximation

With confidence that we have a correct picture of the electrostatics in Fig. 3.8, we now make a very important approximation that will greatly simplify the mathematics.



➤ **Figure 3.10** Qualitative electrostatics of the pn junction under reverse bias $V_D < 0$ which is sufficient to double the width of the depletion region: (a) charge density, (b) electric field, and (c) electrostatic potential. The doping is $N_a = N_d = 10^{16}$ cm^{-3}, as in Fig. 3.9.

First, we assume that the charge density is zero outside of the transition region near the junction. These regions are "bulk" silicon as defined in Section 2.3. Recall that bulk silicon is electrically neutral and that the majority carrier concentration is equal to the net doping concentration, to an excellent approximation. Inside the transition region, the charge density is given by the general expression

$$\rho_o(x) = q\,(p_o - n_o + N_d - N_a) \cong \begin{cases} q\,(p_o - N_a) & (-x_{po} \le x \le 0) \\ q\,(N_d - n_o) & (0 \le x \le x_{no}) \end{cases} \tag{3.39}$$

where the approximations involve neglecting the minority electrons on the p-side and the minority holes on the n side—both many orders of magnitude less the majority carrier concentrations and doping concentrations. Since the pn junction is in thermal equilibrium, we can substitute for the carrier concentrations in terms of the potential from Eqs. (3.21) and (3.22)

$$\rho_o(x) \cong \begin{cases} q\left(n_i e^{-\phi_o(x)/V_{th}} - N_a\right) & (-x_{po} \le x \le 0) \\ q\left(N_d - n_i e^{\phi_o(x)/V_{th}}\right) & (0 \le x \le x_{no}) \end{cases} \tag{3.40}$$

The exponential functions of the potential present a daunting challenge in solving for $\phi_o(x)$, which leads us to explore whether there is a way to make a further simplification of Eq. (3.40). On the p-side, the hole concentration falls off exponentially with distance for $x > -x_{po}$ as shown qualitatively in Fig. 3.7(a). On the n-side, the electron concentration is similarly negligible for points well inside the transition region. For points very near the edges of the transition region, the majority carrier concentrations are close to the doping concentrations and cannot be neglected. We propose a simplifying approximation to Eq. (3.40) that will be referred to as the **depletion approximation**.

depletion approximation

$$\rho_o(x) \approx \begin{cases} -qN_a & (-x_{po} \le x \le 0) \\ qN_d & (0 \le x \le x_{no}) \end{cases} \quad \text{and} \quad \rho_o(x) = \begin{cases} 0 & (x \le -x_{po}) \\ 0 & (x_{no} \le x) \end{cases} \tag{3.41}$$

In effect, we consider the transition region to be *completely depleted* of mobile holes and electrons. The depletion approximation is slightly inaccurate near the edges of the depletion region since it overestimates the charge density there by neglecting the majority carriers. In Fig. 3.8(a), the charge density $\rho_o(x)$ is shown to taper off at the edges of the junction, which is what Eq. (3.40) predicts and what actually occurs. The depletion approximation in Eq. (3.41) gives a reasonably accurate solution for the depletion region width.

What is the motivation for making this approximation? The answer is simple–the mathematics of solving Poisson's Equation using Eq. (3.40) in the depletion region is too messy. By making the depletion approximation and dividing the pn junction into a depletion region, sandwiched between bulk p and n regions, we can develop relatively simple solutions for the potential, field, and depletion region width. These solutions will be useful for gaining an appreciation of the essential physics of pn junctions.

Several device simulators are available for numerically solving the two- or three-dimensional versions of Poisson's Equation in semiconductors should we desire to study a particular device structure with more accuracy.

3.4.2 Quantitative Analysis Using the Depletion Approximation

In Eq. (3.41), we have an approximate expression for the charge density $\rho_o(x)$ in thermal equilibrium. We now integrate the charge density twice to find first the electric field $E_o(x)$ and then the electrostatic potential $\phi_o(x)$. Gauss's Law relates the charge density to the derivative of the electric field

$$\frac{dE_o}{dx} = \frac{\rho_o(x)}{\varepsilon_s} \tag{3.42}$$

where $\varepsilon_s = 11.7\,\varepsilon_o$ is the electric permittivity of silicon. On the p-side of the depletion region $-x_{po} < x < 0$, we can integrate Eq. (3.42) and find that

$$E_o(x) = \int_{-x_{po}}^{x} \frac{\rho_o(x')}{\varepsilon_s} dx' + E_o(x = -x_{po}) = \int_{-x_{po}}^{x} \frac{-qN_a}{\varepsilon_s} dx' + 0 = -\frac{qN_a}{\varepsilon_s}(x - (-x_{po})) \tag{3.43}$$

where we have used the fact that the electric field is zero at the edge of the depletion region in thermal equilibrium since there is no charge density in the bulk p region. We can integrate the field and find an expression for the electrostatic potential on the p-side, $-x_{po} < x < 0$

$$\phi_o(x) = \int_{-x_{po}}^{x} -E_o(x')\,dx' + \phi_o(x = -x_{po}) = \frac{qN_a}{2\varepsilon_s}(x + x_{po})^2 + \phi_p \tag{3.44}$$

where ϕ_p is the potential of the p-type bulk p region in thermal equilibrium. According to Eq. (3.22), this potential is

$$\phi_p = -V_{th} \ln\left(\frac{N_a}{n_i}\right) \tag{3.45}$$

Using exactly the same approach, we can integrate the charge density in Eq. (3.41) starting from the n-side of the depletion region $0 < x < x_{no}$

$$E_o(x = x_{no}) = E_o(x) + \int_{x}^{x_{no}} \frac{\rho_o(x')}{\varepsilon_s} dx' = E_o(x) + \int_{x}^{x_{no}} \frac{qN_d}{\varepsilon_s} dx' = E_o(x) + \frac{qN_d}{\varepsilon_s}(x_{no} - x) \tag{3.46}$$

in which we have been careful to keep the limits of integration consistent with the positive x direction. Since the electric field $E_o(x = x_{no}) = 0$ at the edge of the depletion region, we can solve Eq. (3.46) for $E_o(x)$ on the n-side of the depletion region $0 < x < x_{no}$

$$E_o(x) = -\frac{qN_d}{\varepsilon_s}(x_{no} - x) \tag{3.47}$$

The potential for $0 \le x \le x_n$ can be found by direct integration of the electric field

$$\phi_o(x = x_{no}) = \phi_o(x) + \int_x^{x_{no}} -E(x')dx' = \phi_o(x) + \int_x^{x_{no}} \frac{qN_d}{\varepsilon_s}(x_{no} - x')\,dx' \tag{3.48}$$

Solving for $\phi_o(x)$ and evaluating the integral, we find

$$\phi_o(x) = \phi_n - \frac{qN_d}{2\varepsilon_s}(x_{no} - x)^2 \qquad (0 \le x \le x_n) \tag{3.49}$$

where ϕ_n is the potential of the bulk n region from Eq. (3.20)

$$\phi_n = V_{th} \ln\left(\frac{N_d}{n_i}\right) \tag{3.50}$$

The two solutions for the electric field must connect at $x = 0$, the metallurgical junction, since there is no sheet of charge at this surface (see the discussion on boundary conditions for the electric field in Section 3.1). By substituting $x = 0$ into Eqs. (3.43) and (3.47), we find that

$$-\frac{qN_a}{\varepsilon_s}(0 + x_{po}) = -\frac{qN_d}{\varepsilon_s}(x_{no} - 0) \tag{3.51}$$

Simplifying, we find

$$N_a x_{po} = N_d x_{no} \tag{3.52}$$

This result states that the negative charge (per unit area) on the p-side of the depletion region $-qN_a x_{po}$, is equal in magnitude and opposite in sign to the positive charge on the n-side: $+qN_d\, x_{no}$. Equation (3.52) is hardly surprising since we are safe in assuming that the total charge in IC device structures is zero, as discussed in Section 3.1.

In order to find expressions for the edges of the depletion region $-x_{po}$ and x_{no}, we first use Eq. (3.52) to substitute for one of them into the two solutions for $\phi_o(x)$, Eqs. (3.44) and (3.49). We then require that they match at the metallurgical junction $x = 0$

$$\frac{qN_a}{2\varepsilon_s}(0 + x_{po})^2 + \phi_p = \phi_n - \frac{qN_d}{2\varepsilon_s}(x_{no} - 0)^2 = \phi_n - \frac{qN_d}{2\varepsilon_s}\left(\frac{N_a}{N_d} x_{po}\right)^2 \tag{3.53}$$

Solving for x_{po} and substituting the built-in potential $\phi_B = \phi_n - \phi_p$, we find that

$$x_{po} = \sqrt{\left(\frac{2\varepsilon_s \phi_B}{qN_a}\right)\left(\frac{N_d}{N_d + N_a}\right)} \tag{3.54}$$

The width on the n-side can now be found directly from Eq. (3.52)

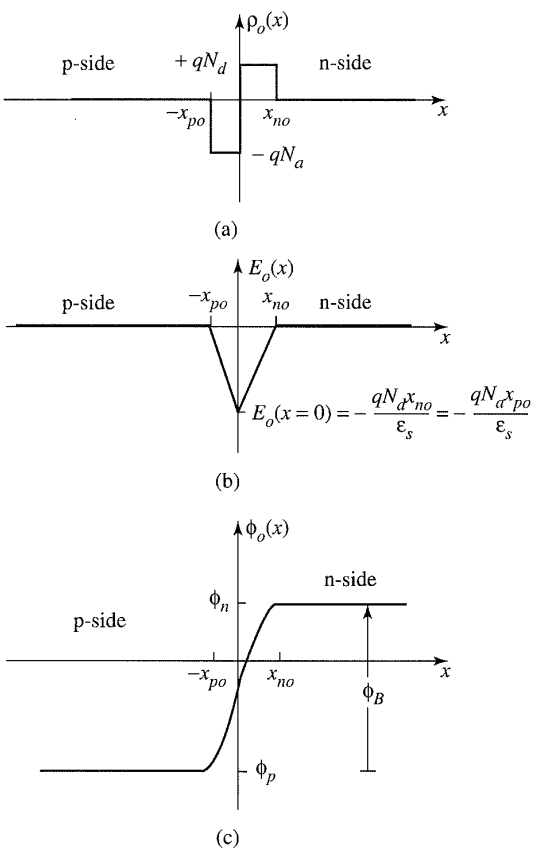
$$x_{no} = \sqrt{\left(\frac{2\varepsilon_s \phi_B}{qN_d}\right)\left(\frac{N_a}{N_d + N_a}\right)} \tag{3.55}$$

The sum of Eqs. (3.54) and (3.55) yields X_{do}, which is the width of the depletion region in thermal equilibrium

$$\boxed{X_{do} = x_{no} + x_{po} = \sqrt{\left(\frac{2\varepsilon_s \phi_B}{q}\right)\left(\frac{1}{N_a} + \frac{1}{N_d}\right)}} \tag{3.56}$$

equilibrium depletion width

Figure 3.11 summarizes the electrostatics of the pn junction in thermal equilibrium using the depletion approximation.



➤ **Figure 3.11** (a) Charge density, (b) electric field, and (c) potential in thermal equilibrium based on the depletion approximation.

EXAMPLE 3.4 Depletion Regions in Thermal Equilibrium

In this example, we will investigate the electrostatics of three pn junctions with progressively more asymmetrical doping. Our goal is to plot $\rho_o(x)$, $E_o(x)$, and $\phi_o(x)$ for junctions 1–3 in thermal equilibrium. The p-side is located on the negative side of the x axis as in Fig. 3.5.

Junction	N_a [cm^{-3}]	N_d [cm^{-3}]
1	5×10^{16}	5×10^{16}
2	5×10^{16}	2×10^{17}
3	5×10^{16}	5×10^{17}

SOLUTION

We start by applying the Eqs. (3.23 and 3.25) to find the potentials on the p and n sides and the built-in voltage.

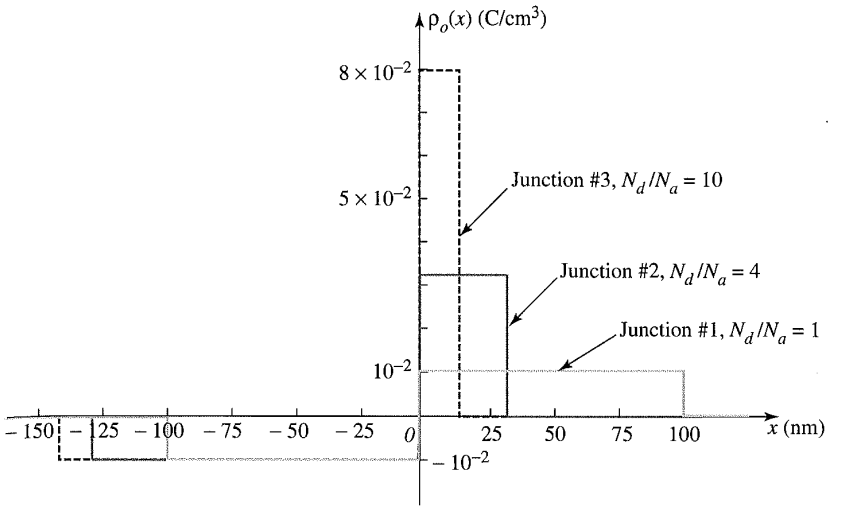
Junction	ϕ_p [mV]	ϕ_n [mV]	ϕ_B [mV]
1	−402	402	804
2	−402	438	840
3	−402	462	864

Note that the built-in voltage is a weak function of the n-side doping concentration, increasing less than 10% between junctions #1 and #3 that have an order-of-magnitude difference in N_d. We substitute into Eqs. (3.54) and (3.55) to find the edges of the depletion region on the p-side and n-side. The results for junction 2 are

$$x_{po} = \sqrt{\left(\frac{(2)\ (11.7)\ (8.85\times10^{-14}\text{Fcm}^{-1})\ (0.840\text{V})}{(1.6\times10^{-19}\text{C})\ (5\times10^{16}\text{cm}^{-3})}\right)\left(\frac{2\times10^{17}}{2\times10^{17}+5\times10^{16}}\right)} = 134\,\text{nm}$$

$$x_{no} = \sqrt{\left(\frac{(2)\ (11.7)\ (8.85\times10^{-14}\text{Fcm}^{-1})\ (0.840\text{V})}{(1.6\times10^{-19}\text{C})\ (2\times10^{17}\text{cm}^{-3})}\right)\left(\frac{5\times10^{16}}{2\times10^{17}+5\times10^{16}}\right)} = 33\,\text{nm}$$

In the plot of the charge density $\rho(x)$ for the three cases in Fig. Ex3.4A, note that the ratios of the depletion widths x_{po}/x_{no} are inversely related to the doping ratios N_a/N_d.



➤ **Figure Ex3.4A** Charge density in thermal equilibrium for the three junctions.

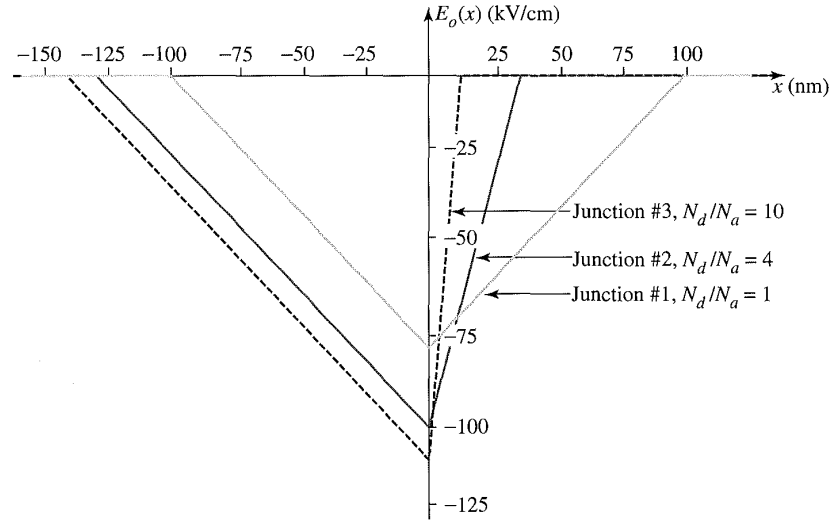
The numerical results for the depletion widths are summarized in this table.

Junction	x_{po} [nm]	x_{no} [nm]
1	102	102
2	134	33
3	143	14.3

To plot the electric field in equilibrium, we need only find its value at $x = 0$ from Eq. (3.47) since we know the shape of $E(x)$ from Fig. 3.11(b). For junction 2, the calculation is

$$E_o(0) = -\frac{qN_d x_{no}}{\varepsilon_s} = \frac{(1.6 \times 10^{-19}\text{C})\ (2 \times 10^{17}\text{cm}^{-3})\ (3.3 \times 10^{-6}\ \text{cm})}{(11.7)\ (8.85 \times 10^{-14}\text{Fcm}^{-1})} = -102\ \text{kV/cm}$$

The peak electric field weakly depends on the doping ratio of the junction when the concentration on one side is fixed. The electric field is plotted in Fig. Ex3.4B, with the numerical results for the peak electric field summarized in the following table.



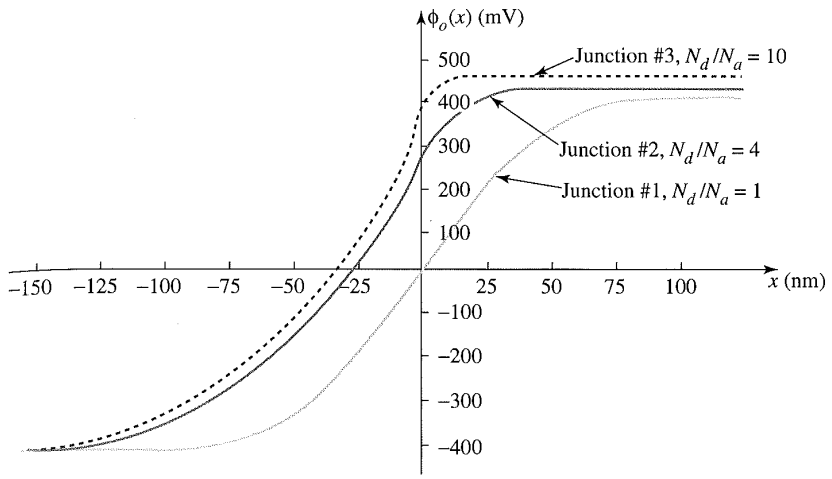
➤ **Figure Ex3.4B** Equilibrium electric field $E_o(x)$ for the three pn junctions in Example 3.4.

Junction	$E_o(0)$ [kV/cm]
1	−79
2	−102
3	−110.5

Poisson's Equation tells us that if the charge distribution is constant, then the potential will behave quadratically as shown in Eqs. (3.44) and (3.49). Evaluating Eq. (3.44) for junction 2, we find that the potential at $x = 0$ is

$$\phi_o(0) = \frac{(1.6\times10^{-19}\text{C})\,(5\times10^{16}\text{cm}^{-3})\,(1.34\times10^{-5}\text{cm})^2}{2\,(11.7)\,(8.85\times10^{-14}\text{F/cm})} - 402\text{mV} = 289\text{ mV}$$

In summary, only a factor of 10 difference in doping is needed for the depletion region to be mostly located on the lightly doped side. Many IC pn junctions have higher doping ratios and can be considered "one-sided" for practical purposes.

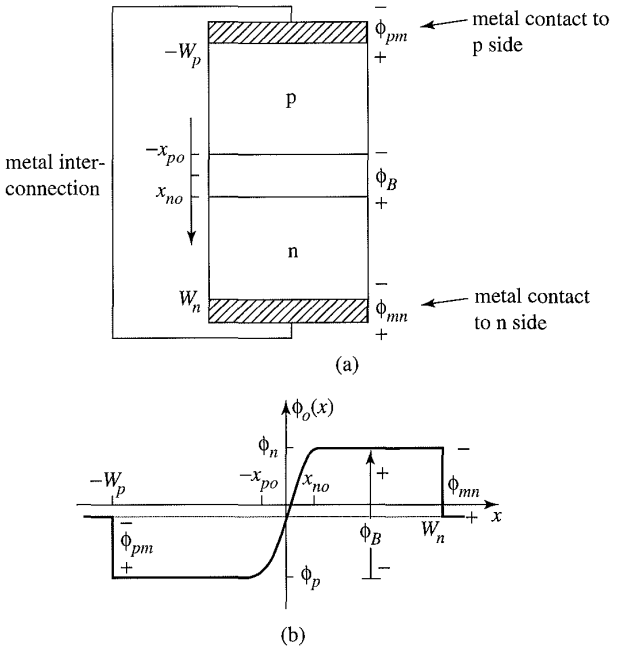


➤ **Figure Ex3.4C** Equilibrium potential $\phi_o(x)$ for the three pn junctions having different doping ratios.

3.4.3 Metal-Semiconductor Contact Potentials

The fact that a pn junction in thermal equilibrium has a built-in potential barrier of approximately 1 V is initially puzzling. For example, the sum of the potential drops around the short-circuited pn junction in Fig. 3.12(a) must be zero since there are no voltage or current sources in the loop and the diode is not illuminated. How then is the built-in potential barrier cancelled around the loop? The metal and bulk silicon regions have constant potentials in equilibrium. Inspection of Fig. 3.12(a) shows the only possibility is that built-in potential drops occur at the metal contacts to the p and n bulk regions. The extended thermal equilibrium potential plot in Fig. 3.12(b) represents these potential changes qualitatively as abrupt jumps at the metal-silicon contacts. We will leave the analysis of the potential variation at the contacts to more advanced courses. In this text, we will ignore that there is a non-zero width to these potential changes.

These contact potentials occur between any dissimilar pair of materials because of the difference in the potential energy of the conduction electrons across the junction. By properly engineering the metal-silicon contact, it is possible to approach an **ohmic contact**. The essential property of an ohmic contact is that it presents no barrier to hole or electron flow in either direction. These specially processed metal-silicon junctions behave like batteries, in that their potential drops are not functions of the forward or reverse current through them. The abrupt changes in potential ϕ_{pm} and ϕ_{nn} shown in Fig. 3.12(b) indicate qualitatively the narrow region of potential change that occurs in ohmic metal-silicon contacts. Note that the metal is *not* at zero potential in Fig. 3.12(b) since the potential reference for the plot is the *internal* silicon reference. The point where $\phi(x) = 0$ occurs when the electron and hole concentrations equal the intrinsic concentration, which is located inside the depletion region on the n-side of the metallurgical junction, from Fig. 3.12(b).



➤ **Figure 3.12** (a) pn diode including ohmic metal-silicon contacts, (b) extended plot of potential in thermal equilibrium including the metal interconnections.

By summing the potential drops around the loop in Fig. 3.12(a), Kirchhoff's voltage law (KVL) around the loop yields

$$0 = \phi_{pm} + \phi_B + \phi_{mn} \tag{3.57}$$

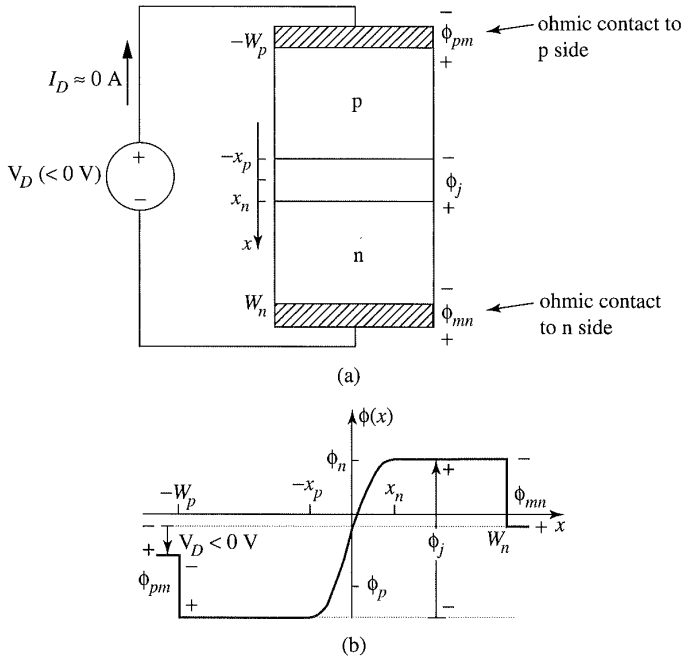
Solving for the sum of the contact potentials, we find that

$$\phi_B = -(\phi_{pm} + \phi_{mn}) \tag{3.58}$$

At first glance, it seems strange that the potential drops at the contacts would have to add up to cancel the built-in potential of the junction. Were this not the case, KVL would imply that current would flow in thermal equilibrium when a resistor is connected across the metal contacts to the p and n regions. This conclusion violates our basic understanding of the meaning of equilibrium; namely, that the net current must be zero for this state.

➤ 3.5 THE pn JUNCTION UNDER REVERSE BIAS

Given our quantitative model for the pn junction in thermal equilibrium, we are in position to analyze how the electrostatics are changed by application of a reverse bias voltage. In Fig. 3.13, we have connected a battery $V_D < 0$ V across the pn junction.



➤ **Figure 3.13** (a) pn junction with negative applied voltage V_D (reverse bias) and (b) plot of the potential across the structure under reverse bias including the ohmic contacts. The "o" subscript has been removed from all variables since the structure is no longer in thermal equilibrium.

Since V_D is negative, it lowers the potential of the p-side of the junction with respect to that of the n-side, which we will consider fixed at ϕ_n. (Alternatively, we could consider the p-side fixed and then $V_D < 0$ V would raise the n-side potential. Recall from Section 3.1 that only potential differences matter.) The effect of the reverse bias is to increase the potential barrier ϕ_j between the p and n bulk regions as shown in Fig. 3.13(b).

Since we are no longer in equilibrium, we must consider whether or not there is a significant current I_D through the pn junction. Raising the potential barrier between the p-side and n-side of the junction is *not* likely to result in a large net carrier transport. In fact, reverse-biased diodes are considered open circuits to a first approximation. We will use this fact and assume $I_D \approx 0$ A for the reverse-biased diode in this chapter. In Chapter 6 we will verify this assumption in a detailed study of the diode under forward and reverse bias.

The potential barrier ϕ_j across the depletion region under reverse bias can be found by applying Kirchhoff's voltage law. Note that the metal-silicon contact potentials at the ohmic contacts are unaffected by the applied voltage. The sum of the potential drops is given by

$$-V_D - \phi_{pm} - \phi_j - \phi_{mn} = 0 \tag{3.59}$$

Solving for the barrier potential, we find that

$$\phi_j = (-\phi_{pm} - \phi_{mn}) - V_D = \phi_B - V_D \tag{3.60}$$

where we have substituted Eq. (3.58) for the built-in potential. We asserted this result for ϕ_j in our first-pass treatment in Section 3.3. Figure 3.13(b) shows the potential variation through the structure. The negative applied voltage increases the barrier height and separates the potential levels of the metal interconnections, which is consistent with our concept of a voltage source.

Now that the barrier height is known as a function of the applied voltage V_D, we can proceed with the analysis of the electrostatics under reverse bias. After applying the depletion approximation, we must solve for the electric field and the potential in the depletion region, assuming it is confined to the interval $-x_p(V_D) < x < x_n(V_D)$. The boundaries of the depletion region are functions of the applied bias since they will change with the barrier height ϕ_j.

Rather than integrate the charge density $\rho(x)$ over the depletion region, we can avoid needless mathematics by recognizing that the form of $\rho(x)$ is *identical* to the charge density $\rho_o(x)$ in thermal equilibrium. The only change is that the boundaries are now $-x_p$ and x_n and the barrier height is ϕ_j. Therefore, the previous solutions for $\phi_o(x)$ and $E_o(x)$ can be adapted to the reverse bias case by simply substituting $\phi_j = \phi_B - V_D$ for ϕ_B. The edges of the depletion region under reverse bias ($V_D < 0$) are given in Eqs. (3.61) and (3.62), and their sum yields the total width in Eq. (3.63)

depletion widths under reverse bias

$$x_p(V_D) = \sqrt{\left(\frac{2\varepsilon_s(\phi_B - V_D)}{qN_a}\right)\left(\frac{N_d}{N_d + N_a}\right)} = x_{po}\sqrt{1 - (V_D/\phi_B)} \tag{3.61}$$

$$x_n(V_D) = \sqrt{\left(\frac{2\varepsilon_s(\phi_B - V_D)}{qN_d}\right)\left(\frac{N_a}{N_d + N_a}\right)} = x_{no}\sqrt{1 - (V_D/\phi_B)} \tag{3.62}$$

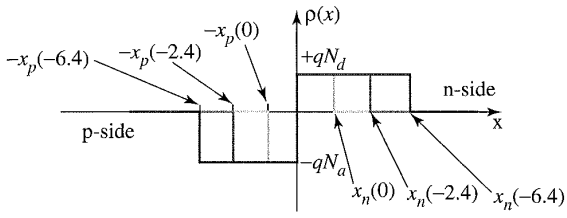
$$X_d(V_D) = \sqrt{\left(\frac{2\varepsilon_s(\phi_B - V_D)}{q}\right)\left(\frac{1}{N_a} + \frac{1}{N_d}\right)} = X_{do}\sqrt{1 - (V_D/\phi_B)} \tag{3.63}$$

where the thermal equilibrium values x_{po}, x_{no}, and X_{do} are defined in Eqs. (3.54)–(3.56). Figure 3.14 shows the effects on the depletion width, electric field, and potential of applying voltages of $V_D = -2.4$ V and $V_D = -6.4$ V across a pn junction with a thermal equilibrium barrier height of $\phi_B = 0.8$ V. These values of reverse bias will double and triple the widths of the depletion region over the thermal equilibrium width since

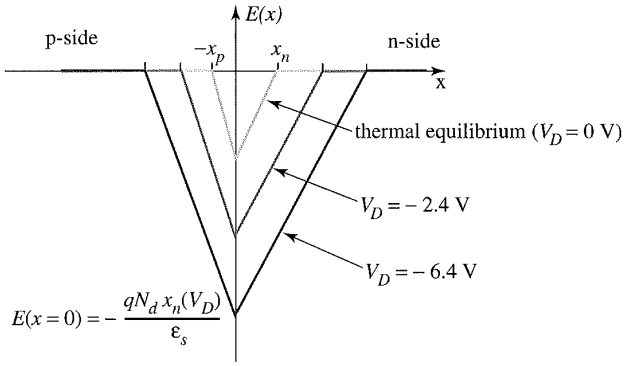
$$\sqrt{1 - (V_D/\phi_B)} = \sqrt{1 - ((-2.4)/0.8)} = \sqrt{4} = 2 \text{ and} \tag{3.64}$$

$$\sqrt{1 - (V_D/\phi_B)} = \sqrt{1 - ((-6.4)/0.8)} = \sqrt{9} = 3 \tag{3.65}$$

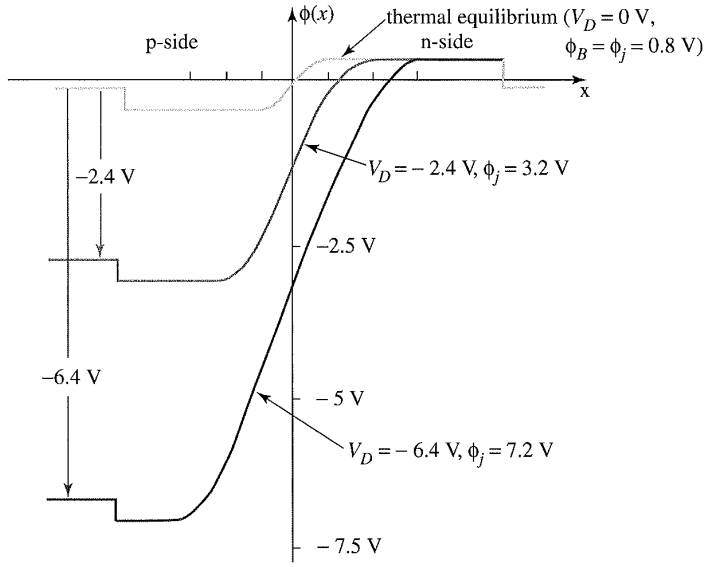
We conclude this section with a numerical example.



(a)



(b)



(c)

➤ **Figure 3.14** pn junction under reverse biases $V_D = -2.4$ V and $V_D = -6.4$ V. (a) charge, (b) electric field, and (c) potential, according to the depletion approximation.

EXAMPLE 3.5 Depletion Regions in Reverse Bias

Consider a pn junction with doping $N_a = 10^{16}\text{cm}^{-3}$ and $N_d = 10^{18}\text{cm}^{-3}$. (**a**) What applied voltage is required for $X_d = 1.5\ \mu\text{m}$? (**b**) What distance x_n does the depletion region penetrate into the n-side of the junction? (**c**) Plot the potential $\phi(x)$ for this applied reverse bias keeping the n-side of the junction fixed and neglecting contact potentials.

SOLUTIONS

(a) In thermal equilibrium, the potential in the bulk regions can be found using Eqs. (3.23) and (3.25): $\phi_n = 480\ \text{mV}$, $\phi_p = -360\ \text{mV}$, and $\phi_B = \phi_n - \phi_p = 840\ \text{mV}$. The thermal equilibrium depletion width is found using Eq. (3.56)

$$X_{do} = \sqrt{\frac{2\,(11.7)\,(8.85\times10^{-14}\text{F cm}^{-1})\,(0.84V)}{1.6\times10^{-19}C}\left(\frac{1}{10^{18}\text{cm}^{-3}}+\frac{1}{10^{16}\text{cm}^{-3}}\right)} = 0.33\mu\text{m}$$

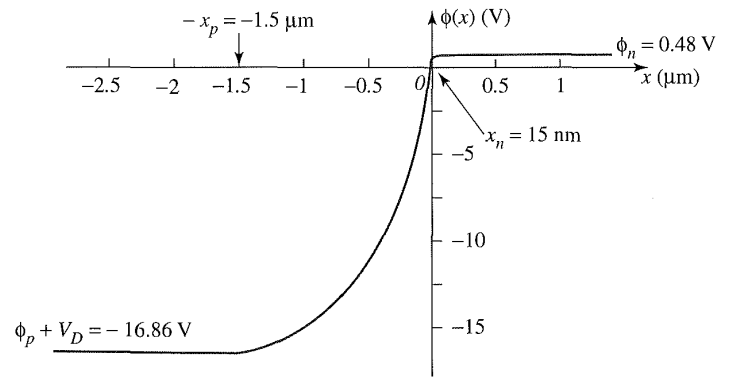
We can now use Eq. (3.63) to solve for the applied voltage V_D necessary for the depletion region to grow to $X_d = 1.5\ \mu\text{m}$

$$X_d(V_D) = X_{do}\sqrt{1-\frac{V_D}{\phi_B}} = (0.33\mu m)\sqrt{1-\frac{V_D}{0.84\text{V}}} = 1.5\mu\text{m}$$

which yields $V_D = -16.5$ V.

(b) In order to find the width on the n-side of the junction, we can substitute into Eq. (3.62) after calculating x_{no} or recall that the charge per unit area on the n-side is equal and opposite to that on the p-side, which implies that $x_n N_d = x_p N_a$. From (a), the total depletion width is $X_d = x_n + x_p = 1.5\ \mu\text{m}$. Eliminating x_p and noting that $N_d = 100\,N_a$, it follows that $x_n = x_p/100 \cong X_d/100 = 15$ nm.

(c) In plotting the potential across the junction, the n-side remains at $\phi_n = 0.48$ V while the p-side potential drops to $\phi_p + V_D = -16.86$ V.



➤ **Figure Ex3.5** Potential plot for Example 3.5 pn junction under applied bias of $V_D = -16.5$ V.

3.6 DEPLETION CAPACITANCE

The pn junction under reverse bias is a two-terminal charge-storage element. It is essential to connect the electrostatics of the reverse-biased pn junction with its circuit models. We will explain modeling of this device structure from several perspectives since the pn junction is universal in ICs and also, since this is our first exposure to circuit modeling.

As we have seen from Sections 3.3–3.5, the charge on the p-side of the depletion region is a function of the applied voltage. Repeating Eq. (3.38), the depletion charge varies with the applied voltage as

$$Q_J(V_D) = -qN_a x_{po}\sqrt{1-(V_D/\phi_B)} = -qN_a x_p(V_D),\ V_D \le 0, \tag{3.66}$$

with the right-hand side coming from Eq. (3.61). Unlike a conventional capacitor, the depletion charge does not change linearly with applied voltage.

To proceed further, we must become more precise in defining symbols for electrical quantities. The convention in this text is that an upper-case symbol with an upper-case subscript (e.g., Q_J or V_D) will refer to a DC or average quantity. A lower-case symbol with an upper-case subscript (e.g., q_J or v_D) will indicate the total quantity, including its time variation.

We have not considered the possibility that the applied voltage or the depletion charge may be functions of time in our analysis thus far. We will model the time-varying depletion charge by simply substituting q_J and v_D into Eq. (3.66). This "quasi-static" approach is an excellent approximation over the range of frequencies for microelectronics. Our starting point for developing circuit models for the reverse-biased pn junction is

$$q_J(v_D) = -qN_a x_{po}\sqrt{1-(v_D/\phi_B)},\ v_D \le 0 \tag{3.67}$$

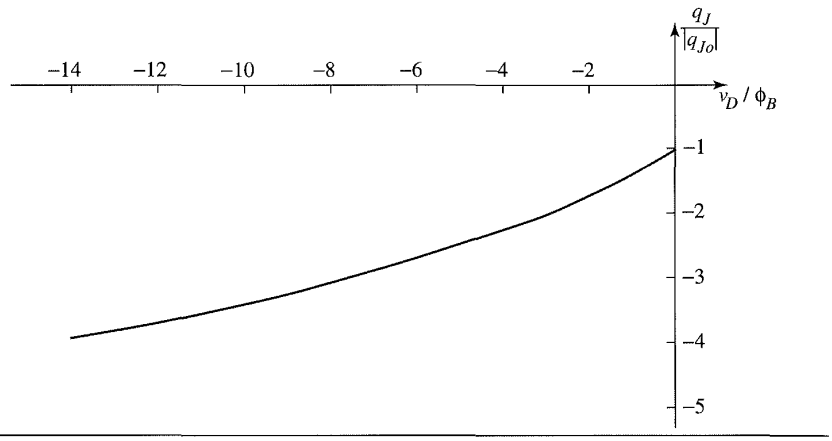
Figure 3.15 is a plot of Eq. (3.67) for negative applied voltages.We have normalized q_J by its magnitude in thermal equilibrium $qN_a x_{po}$ and v_D by the built-in potential ϕ_B.

A numerical circuit simulator such as SPICE incorporates Eq. (3.67) into its model for a reverse-biased pn junction with an abrupt change in doping concentration at the metallurgical junction. For hand analysis, however, it is quite burdensome to have nonlinear models in the node or loop equations. Therefore, we will develop an approach to linearizing the charge-storage characteristics of the pn junction that will be valid under conditions that often occur in practical cases. We will derive an equivalent capacitor that describes the incremental charge storage for small deviations from a specified DC voltage. This technique, termed **small-signal modeling**, is a powerful approach to nonlinear circuit elements and will be used later for transistors.

small-signal modeling

In order to further develop this technique, we introduce the special case where the total depletion charge q_J can be written as the sum of its DC (average) value Q_J and an incremental (i.e., very small) component that is considered time-varying $q_j = q_j(t)$

$$q_J = Q_J + q_j, \text{ where } |q_j| \ll |Q_J| \tag{3.68}$$



➤ **Figure 3.15** Depletion charge q_J, normalized to the magnitude of the equilibrium charge $|q_{Jo}| = qN_a x_{po}$, as a function of the applied voltage v_D, normalized to ϕ_B, the built-in potential barrier.

For this case, we can also express the applied voltage v_D as the sum of its DC value V_D, and a much smaller time-varying component $v_d = v_d(t)$

$$v_D = V_D + v_d \text{ where } |v_d| \ll |V_d| \tag{3.69}$$

The lower-case variables with lower-case subscripts in these equations are reserved for "small signals" which meet the test of being very small compared with the average value of the signal. It is helpful to interpret Eqs. (3.68) and (3.69) on a graph of the depletion charge before pursuing the mathematical analysis further.

3.6.1 Graphical Interpretation

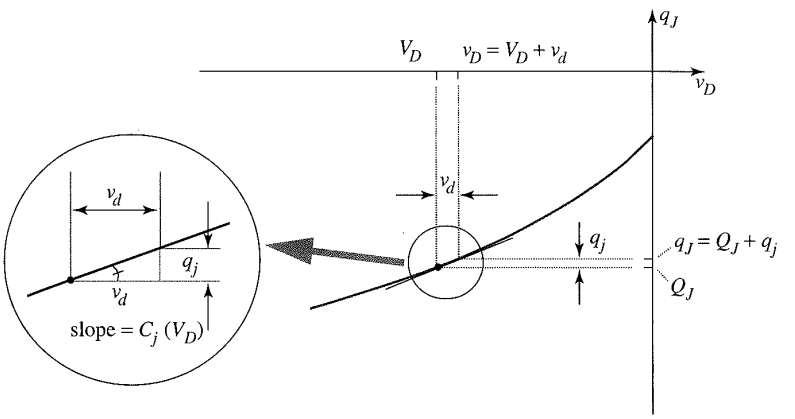
Figure 3.16 shows the depletion charge as a function of the applied voltage with the function $q_J(v_D)$ evaluated at V_D and at $v_D = V_D + v_d$. For illustrative purposes, we have assumed that $v_d \ll |V_D|$ is a constant, positive voltage. From Eq. (3.68), the small-signal depletion charge q_j is given by

$$q_j = q_J - Q_J = q_J(v_D) - q_J(V_D) \tag{3.70}$$

in which we have substituted for the DC depletion charge $Q_J = q_J(V_D)$. In Fig. 3.16, the small-signal depletion charge is the product of the small-signal voltage and the slope of the function $q_J(v_D)$, evaluated at $v_D = V_D$

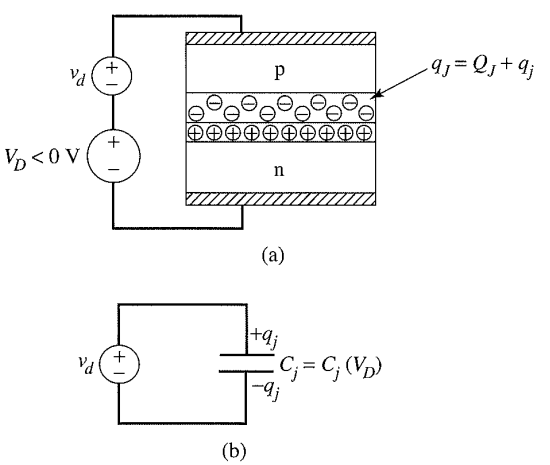
$$q_j = (slope) \times v_d \tag{3.71}$$

The slope has units of $(C/cm^2)/V = F/cm^2$ and is defined as the **depletion capacitance** $C_j(V_D)$. As shown in Fig. 3.16, the slope of $q_J(v_D)$ decreases with increasing reverse bias on the junction. We will consider the physical interpretation of this fact after developing an equation for the depletion capacitance.



➤ **Figure 3.16** Graphical interpretation of the depletion capacitance as the slope of the depletion-charge vs. applied voltage curve evaluated at the point $q_J(V_D) = Q_j$.

The concept of small-signal modeling can be viewed from a circuit perspective as well. In Fig. 3.17(a), the total circuit quantities are broken into their DC and small-signal components. In many cases, we are concerned only with the small-signal behavior of the device structure. Therefore, it is convenient to work with the small-signal equivalent circuit for the reversed-biased pn junction shown in Fig. 3.17(b). Needless to say, this circuit describes only the response of the reverse-biased junction to small changes in the applied voltage. In order to find the numerical value for the capacitance from the slope of $q_J(v_D)$, we must first know the DC bias voltage V_D.



➤ **Figure 3.17** (a) Circuit model with the total voltage $v_D = V_D + v_D$ and the total charge q_J broken into their DC and small-signal components and (b) small-signal equivalent circuit for the reverse-biased pn junction.

3.6.2 Mathematical Formulation

From Fig. 3.16, we can apply calculus to find an expression for the depletion capacitance. Since v_d is a small increment on the average diode voltage V_D, we need to retain only the linear term in the Taylor expansion for the depletion charge at v_D

$$q_J(v_D) = q_J(V_D) + \left[\left. \frac{dq_J}{dv_D} \right|_{V_D} \times (v_D - V_D) \right] \tag{3.72}$$

The total depletion charge $q_J(V_D)$ is the sum of its DC and small-signal components

$$q_J(v_D) = Q_J + q_j = q_J(V_D) + q_j \tag{3.73}$$

Substitution of Eq. (3.73) and the small-signal voltage $v_d = v_D - V_D$ into the Taylor expansion results in cancellation of the DC depletion charge

$$q_J(V_D) + q_j = q_J(V_D) + \left. \frac{dq_J}{dv_D} \right|_{V_D} \times v_d \text{ and} \tag{3.74}$$

$$q_j = \left. \frac{dq_J}{dv_D} \right|_{V_D} \times v_d \tag{3.75}$$

We identify the derivative in Eq. (3.75) with the slope in Eq. (3.71) and therefore, the depletion capacitance is defined as

depletion capacitance

$$\boxed{C_j = C_j(V_D) = \left. \frac{dq_J}{dv_D} \right|_{V_D}} \tag{3.76}$$

in which the second equality is a reminder that C_j is a function of the DC bias voltage.

In order to evaluate the functional form of the depletion capacitance, we substitute the expression for the depletion charge as a function of the diode voltage from Eq. (3.67) into the definition

$$C_j = \frac{d}{dv_D} \{ -qN_a x_{po} \sqrt{1 - v_D/\phi_B} \} \Big|_{V_D} = \frac{qN_a x_{po}}{2\phi_B \sqrt{1 - V_D/\phi_B}} = \frac{C_{jo}}{\sqrt{1 - V_D/\phi_B}} \tag{3.77}$$

in which the various constants are collected and defined as C_{jo}. Closer inspection of the final equality in Eq. (3.77) shows that C_{jo} is the depletion capacitance in thermal equilibrium, or the **zero-bias** capacitance, $C_j(V_D = 0\text{ V}) = C_{jo}$.

There should be a simple physical interpretation for C_{jo}, so we substitute from the definition of x_{po} in Eq. (3.54) to find

$$C_{jo} = \frac{qN_a x_{po}}{2\phi_B} = \frac{qN_a}{2\phi_B}\sqrt{\left(\frac{2\varepsilon_s \phi_B}{qN_a}\right)\left(\frac{N_d}{N_d + N_a}\right)} = \sqrt{\frac{q\varepsilon_s}{2\phi_B}\left(\frac{N_a N_d}{N_d + N_a}\right)} \tag{3.78}$$

The right-hand equality has many common terms with X_{do}, the depletion width in thermal equilibrium (Eq. (3.56)). Factoring out the permittivity of silicon, Eq. (3.78) simplifies to

$$\boxed{C_{jo} = \varepsilon_s\sqrt{\frac{q}{2\varepsilon_s \phi_B}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)^{-1}} = \frac{\varepsilon_s}{X_{do}}} \tag{3.79}$$

zero-bias depletion capacitance

As a final expression for the depletion capacitance that will be useful in the next section, we substitute for C_{jo} in Eq. (3.77)

$$\boxed{C_j(V_D) = \frac{C_{jo}}{\sqrt{1 - V_D/\phi_B}} = \frac{\varepsilon_s}{X_{do}\sqrt{1 - V_D/\phi_B}} = \frac{\varepsilon_s}{X_d(V_D)}} \tag{3.80}$$

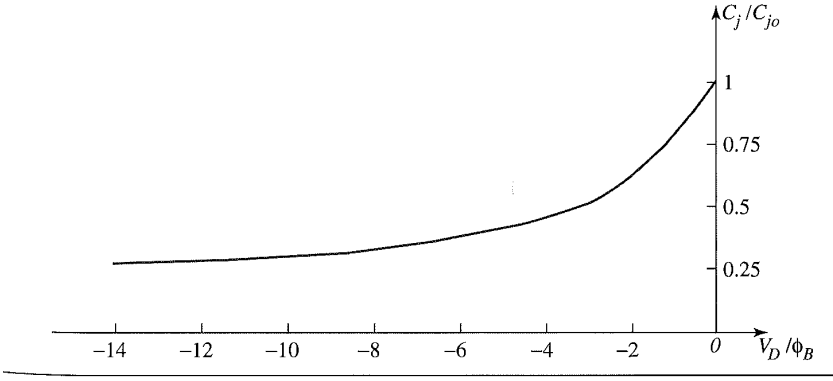
depletion capacitance for reverse bias

where we have substituted for the depletion width under reverse bias in Eq. (3.63).

The normalized depletion capacitance C_j/C_{jo} is plotted in Fig. 3.18. The capacitance decreases with increasing reverse bias, which is consistent with its interpretation as the slope of the depletion charge function plotted in Fig. 3.15. Since a typical equilibrium barrier height $\phi_B \approx 1$ V, the normalized voltage is roughly the same as the DC bias.

3.6.3 Physical Interpretation

Our final perspective on the depletion capacitance will be to develop some insight into the physical meaning of what appears to be a purely mathematical concept. An important clue is provided by Eq. (3.80), which is identical to the capacitance of a par-



➤ **Figure 3.18** Depletion capacitance, normalized to its zero-bias (thermal equilibrium) value, as a function of the applied DC voltage, normalized to the barrier height ϕ_B in thermal equilibrium.

allel-plate capacitor with its electrodes separated by the depletion width $X_d(V_D)$ at the particular DC voltage. What charges are separated by X_d? From Eq. (3.76), the depletion capacitance is the ratio of the small-signal depletion charge q_j to the small-signal voltage v_d

$$C_j = \frac{q_j}{v_d} = \frac{\varepsilon_s}{X_d} \tag{3.81}$$

which indicates that the small-signal charge layers $\pm q_j$ are separated by X_d.

In Fig. 3.19, the charge density in the depletion region is plotted, first for the case where the diode voltage is $V_D < 0$ V in (a), and then where a positive small-signal voltage is added to give $v_D = V_D + v_d$ in (a). The depletion region narrows slightly in the second case since the positive v_d means that $v_D > V_D$. The depletion charge q_J for (b) is less negative than is Q_J in (a) since the depletion region is narrower for v_D.

We graphically subtract (a) from (b) to find the incremental charge density that accounts for the small-signal depletion charge q_j since by definition $q_j = q_J - Q_J$. The result is plotted in Fig. 3.19(c), in which the change in the charge density due to the addition of v_d is confined to the edges of the depletion region. In the limit of $v_d \to 0$, the small-signal charges $\pm q_j$ become sheets that are separated by a gap width of X_d, which explains the physical origin of the parallel-plate formula in Eq. (3.81).

Although we have derived Eq. (3.81) for the case of an abrupt pn junction with constant doping, it is in fact a general result for any pn junction. The functional dependence of the depletion width on the reverse bias will depend on the details of the particular dopant distributions and consequently, the dependence of C_j on the reverse bias may differ slightly from that in Fig. 3.18.

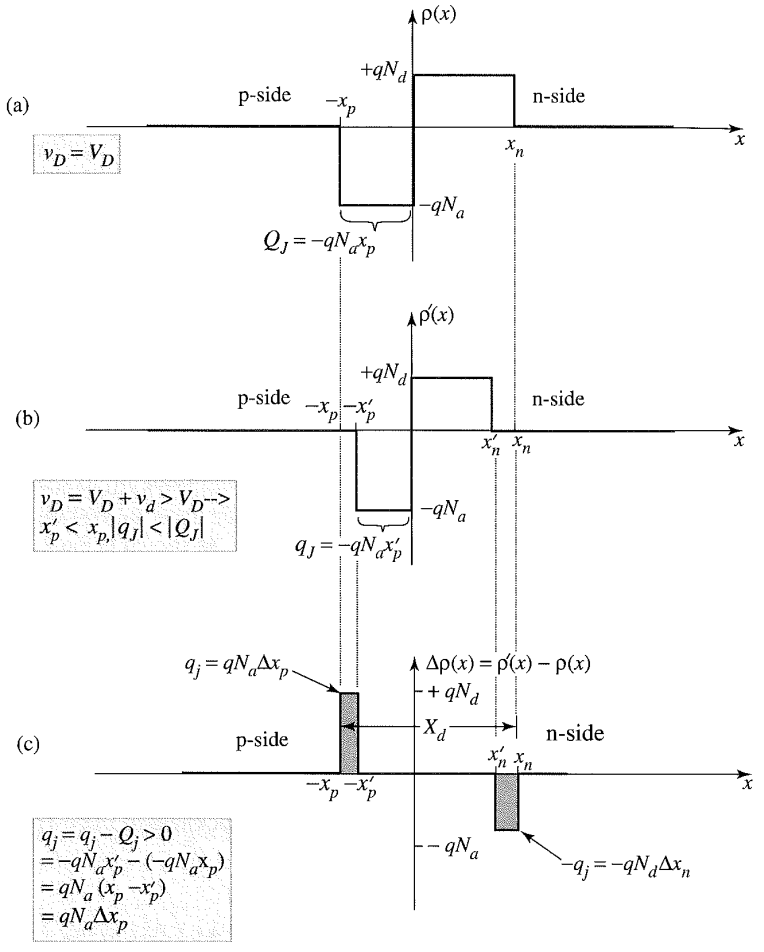
3.6.4 SPICE Models for IC pn Junctions

SPICE is a powerful circuit simulator that is commonly used to simulate integrated circuits. In SPICE, the depletion capacitance of a pn junction under reverse bias is modeled by the following equation

$$C_j = \frac{CJO}{(1 - V_D/VJ)^M} \tag{3.82}$$

in which the SPICE symbols are $CJO = C_{jo}$, the zero-bias (equilibrium) capacitance, $VJ = \phi_B$, the built-in barrier voltage, and M is the grading coefficient of the junction. From Eq. (3.80), $M = 0.5$ for the abrupt pn junction that we have modeled in this chapter. For IC pn junctions, M is usually curve-fit to measurements of $C_j(V_D)$ or is extracted from a numerical device simulator. Typical values for M range from 0.3 to 0.7.

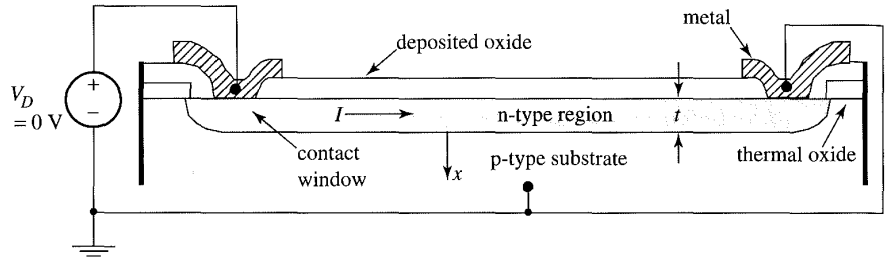
The effect of changing the width of the depletion region is that a displacement current must be supplied through the adjacent bulk silicon regions. The small-signal junction capacitance incorporates this effect into the circuit model, which is illustrated by the following example. This example revisits the ion-implanted resistor and adds charge-storage at the pn junction to the small-signal model.



➤ **Figure 3.19** (a) Charge density $\rho(x)$ in depletion region for a reverse bias of $V_D < 0$, (b) charge density $\rho'(x)$ in depletion region for a perturbed reverse bias $v_D = V_D + v_d$ with $v_d > 0$, and (c) difference $\Delta\rho(x)$ between (b) and (a) showing incremental depletion charge $\pm q_j$ separated by approximately the depletion width X_d. Note that the magnitude of v_d is exaggerated in order to clarify its effect on the depletion region width.

EXAMPLE 3.6 Depletion Capacitance of an IC Resistor

(a) For the ion-implanted resistor described in Ex. 2.5 and shown in cross section below, find the depletion width x_{no} for $V_D = 0$ V.



➤ **Figure Ex3.6A** Cross section of the IC resistor from Example 2.5. The doping concentrations are $N_d = 2.5 \times 10^{16}$ cm^{-3} and $N_a = 2.5 \times 10^{15}$ cm^{-3}. The junction depth is $t = 0.4$ μm = 400 nm, the width of the n-type region is $W = 2.5$ μm, and its length is $L = 20$ μm. The contact regions are each $3W \times 3W$ in area, as shown in Fig.2.22(a).

SOLUTION

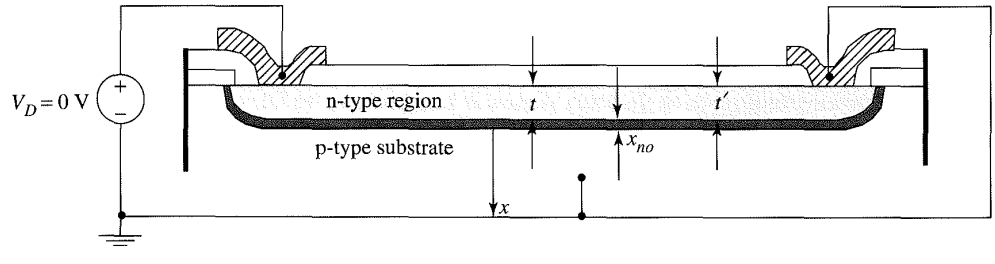
Since N_a is a factor of 10 less than N_d, most of the depletion region will reside on the p-side of the junction. In order to use Eq. (3.55) to find the depletion edge in thermal equilibrium, we first find the potentials of each side under zero bias using Eqs. (3.23) and (3.25): $\phi_n = 384$ mV, $\phi_p = -324$ mV. The depletion edge on the n-side is then $x_{no} = 58$ nm from Eq. (3.55).

(b) Recalculate the sheet resistance by correcting the thickness of the n-type region for the depletion at the junction.

SOLUTION

There are no carriers in the depletion region, so the thickness of the n-type bulk region t' is

$$t' = t - x_{no} = 400\text{ nm} - 58\text{ nm} = 342\text{ nm}$$



➤ **Figure Ex3.6B** IC resistor showing the junction depth t and the narrowing of the n-type region due to depletion width x_{no} on the n-side of the junction.

The sheet resistance increases due to the encroachment of the depletion region into the n-type region as shown in Fig. Ex3.6B.

$$R_{\square} = \frac{1}{qN_d\mu_n t'} = \frac{1}{qN_d\mu_n (t - x_{no})}$$

$$R_{\square} = \frac{1}{(1.6\times10^{-19}\text{C})(1050\,\text{cm}^2\text{V}^{-1}\text{s}^{-1})(2.5\times10^{16}\,\text{cm}^{-3})(400\,\text{nm} - 58\,\text{nm})} = 6.9\text{k}\Omega/\square$$

represents a 15% increase over the value in Ex. 2.5 where the depletion region was neglected. The resistance is $R = (6.9\ \text{k}\Omega/\square)(9.3\square) = 64.2\ \text{k}\Omega$.

(c) Calculate the depletion capacitance C_R in fF between the n region and the p-type substrate assuming the DC voltage $V_D = 0$ V.

SOLUTION

The area of the ion-implanted n-type region from the layout (including the two contact areas) in Example 2.5 is

$$A_R = L\,W + 2(3W)^2 = 162.5\ \mu\text{m}^2 = 1.625\times10^{-6}\,\text{cm},^2$$

where the area of each contact region $(3W)^2$ is included.

The depletion capacitance (per unit area) for $V_D = 0$ is

$$C_{jo} = \frac{\varepsilon_s}{X_{do}} = \frac{(11.7)\ (8.85\times10^{-14}\text{F cm}^{-1})}{0.635\mu\text{m}} = 16.3\times10^{-9}\ \text{F/cm}^2$$

in which we used $X_{do} = x_{no} + x_{po} = x_{no} + (N_d/N_a)\,x_{no} = 11\,x_{no} = 635$ nm. Multiplying by the area A_R, the capacitance is

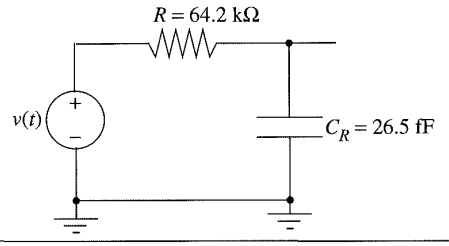
$$C_R = C_{jo}A_R = (16.3\times10^{-9}\text{F/cm}^2)\ (16.25\times10^{-6}\text{cm}^2) = 26.5\ \text{fF}$$

If a DC reverse bias is applied across the pn junction and the depletion region widens, then the resistance will increase, due to further reduction of the bulk n region thickness, and the capacitance will decrease, due to the wider depletion region.

(d) Draw a simple small-signal model of the ion-implanted resistor for the case where $V_D = 0$ V.

SOLUTION

The ion-implanted resistor with depletion capacitance is best modeled as a series of N segments with resistance R/N and capacitance C_R/N. For simplicity, we lump the resistance and capacitance into single elements in the circuit below in Fig. Ex3.6C.



➤ **Figure Ex3.6C** Small-signal circuit for the ion-implanted resistor in Example 2.5 with zero DC bias between the n region and p-type substrate. The resistance includes the effect of encroachment by the depletion region. The capacitance is the depletion capacitance of the entire n region.

➤ 3.7 THE MOS CAPACITOR: A FIRST PASS

We now turn our attention to the more complex, but far more important metal-oxide-semiconductor (MOS) capacitor. Figure 3.20 shows a cross section of the three-layer sandwich in which a heavily doped, n-type polycrystalline silicon layer serves as the "metal" layer. In practice, polysilicon is the most common material used for the "metal" in integrated MOS structures, having supplanted aluminum and other metal films since the 1970s. The oxide film is a thermally grown layer of SiO_2. The silicon substrate could be either p-type or n-type—we will first concentrate on the former. Metal interconnections to the polysilicon gate and to the p-type bulk silicon are shown schematically in Fig. 3.20.

3.7.1 MOS Capacitor in Thermal Equilibrium

As we did with the pn junction, we will start our investigation of MOS electrostatics with the structure in thermal equilibrium. Since the oxide is a near-perfect insulator, we short together the gate and bulk metal layers to enable the charge exchange needed to reach equilibrium as indicated in Fig. 3.20. For the pn junction, $V_D = 0$ V also was a necessary condition for equilibrium.

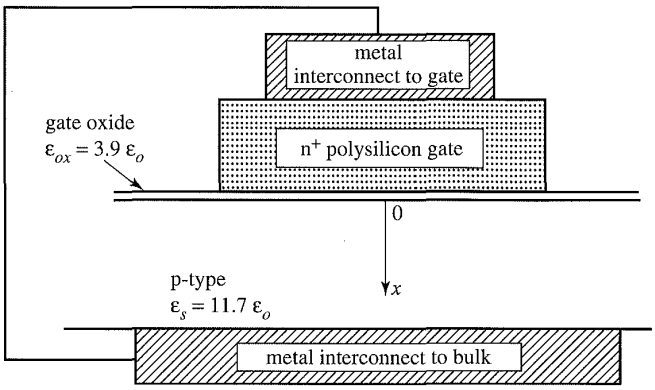
The boundary condition on the potential in the silicon, far from the oxide, is given by Eq. (3.25)

$$\phi_p = -V_{th} \ln\left(\frac{N_a}{n_i}\right) \tag{3.83}$$

For a typical doping concentration of 10^{17} cm^{-3}, the potential of the p-type bulk region far from the surface is $\phi_p = -420$ mV. The n^+ designation on the polysilicon gate indicates that it is very heavily doped with donors, to the point that its potential can be taken as the maximum possible for silicon in thermal equilibrium

$$\phi_{n+} = 550\,\text{mV} \tag{3.84}$$

Based on our intuitive understanding of electrostatics, we can understand the qualitative electrostatics in the MOS structure in thermal equilibrium. The sign of the potential drop (gate positive, p-type silicon negative) means that the electric field E_o



➤ **Figure 3.20** MOS capacitor with p-type substrate. Gate and bulk metal contacts are shorted together for thermal equilibrium.

is positive (points from gate to bulk, in the $+x$ direction as defined in Fig. 3.20). Therefore, a positive charge must be present on the bottom of the polysilicon gate and there must be a balancing negative charge in the p-type silicon underlying the gate oxide. The gate oxide itself will be considered a charge-free perfect insulator in this text. In more advanced courses, the effect of oxide charge on the MOS structure will be investigated in detail.

As we saw with the pn junction in Section 3.2, negative charge density in p-type silicon can occur if a depletion region forms under the gate oxide. Figure 3.21 shows a sketch of the charges in the MOS capacitor with p-type substrate, in the state of thermal equilibrium. Since the gate is highly conductive n^+ polysilicon, the gate charge Q_{Go} (units: C/cm^2) is depicted as a sheet charge located on its bottom surface. An intuitive picture of why the depletion region forms is to view the positive charge on the gate as repelling mobile holes in the silicon under the gate oxide, leaving behind the immobile, negatively charged acceptor ions to a depth X_{do}. The gate charge Q_{Go} is equal and opposite to the bulk charge Q_{Bo} in the silicon substrate

$$Q_{Go} = -Q_{Bo} = -qN_aX_{do} \tag{3.85}$$

Just as we found for the pn junction, there is charge stored on the MOS capacitor in thermal equilibrium. It is helpful to sketch the potential variation through the MOS structure in equilibrium. Example 3.3 analyzed the electrostatics of the MOS capacitor with the charge distribution in Fig. 3.21. Figure 3.22 is a sketch of the potential $\phi(x)$, which is based on the analysis in Ex. 3.3.

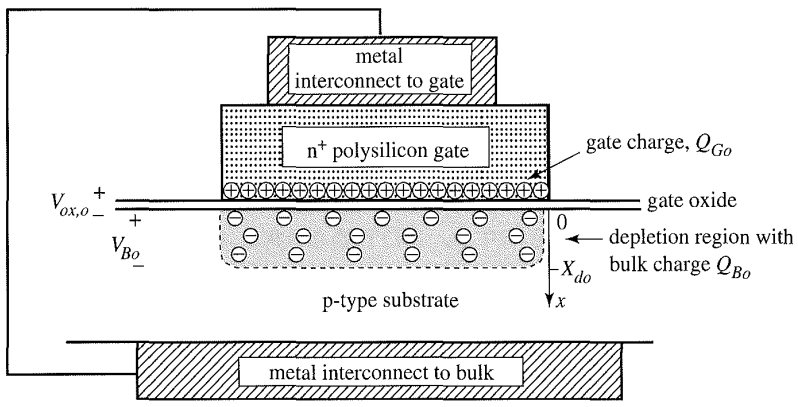
The potential drop across the MOS structure is the difference between the potentials of the n^+ polysilicon gate and the p-type substrate. For the plot in Fig. 3.22, we have used a substrate doping concentration of $N_a = 10^{17}$ cm^{-3} and the total drop is:

$$\phi_{n+} - \phi_p = 550\text{ mV} - (-420\text{ mV}) = 970\text{ mV} \tag{3.86}$$

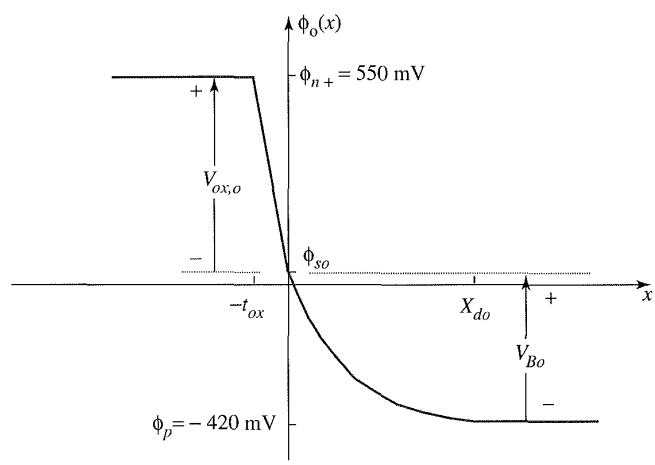
In Figs. 3.21 and 3.22, the potential drop across the capacitor is divided into the sum of the drop across the oxide $V_{ox,o}$ and the drop across the depletion region V_{Bo}

$$\phi_{n+} - \phi_p = V_{ox,o} + V_{Bo} \tag{3.87}$$

Finally, we have defined the potential at $x = 0$ (the SiO_2-silicon interface) as the thermal equilibrium **surface potential**, ϕ_{so}.



➤ **Figure 3.21** Qualitative picture of charge distribution in an MOS capacitor with p-type substrate in thermal equilibrium.



➤ **Figure 3.22** Potential plot for MOS capacitor with p-type substrate in thermal equilibrium. The substrate doping concentration is $N_a = 10^{17}$ cm^{-3}.

3.7.2 MOS Capacitor under Applied Bias

The effect on the gate charge of an applied voltage V_{GB} is rather complicated. The easiest place to start is by applying a voltage that is equal and opposite to the "built-in" voltage across the MOS capacitor in thermal equilibrium. We define this special voltage as the **flatband voltage**, V_{FB}:

$$\boxed{V_{FB} = -(\phi_{n+} - \phi_p)} \tag{3.88}$$

flatband voltage for p-type substrate

For a substrate acceptor concentration of 10^{17} cm^{-3}, the flatband voltage is $V_{FB} = -970$ mV. In plotting the potential variation under applied bias, we will keep the substrate fixed and move the gate. With $V_{GB} = V_{FB} = -970$ mV, the n^+ polysilicon gate's potential is pulled down to 550 mV – 970 mV = –420 mV, as shown in Fig. 3.23 (a). Therefore, there is no internal potential drop across the MOS capacitor when $V_{GB} = V_{FB}$ and as a result, the gate charge is zero:

$$Q_G(V_{GB} = V_{FB}) = 0 \tag{3.89}$$

If we make the applied voltage more negative than the flatband voltage, $V_{GB} < V_{FB}$, then the gate potential drops below that of the substrate. The gate charge therefore becomes negative, as shown in Fig. 3.23(b), and the nature of the charge in the substrate changes. The identity of the positive substrate charge is not hard to determine, since there are plenty of positively charged holes in the p-type silicon that can accumulate under the gate oxide, as shown in Fig. 3.23(b), to cancel the negative gate charge. The MOS capacitor is said to be in **accumulation** for $V_{GB} < V_{FB}$. The potential and charge density in accumulation are shown in Fig. 3.23(b). These plots are identical to those for the parallel-plate capacitor analyzed in Ex. 3.1. Therefore, the gate charge in accumulation is linearly related to the difference between the applied voltage and the flatband voltage:

$$\boxed{Q_G = C_{ox}(V_{GB} - V_{FB}) \text{ for } V_{GB} \leq V_{FB}} \tag{3.90}$$

gate charge in accumulation for p-type substrate

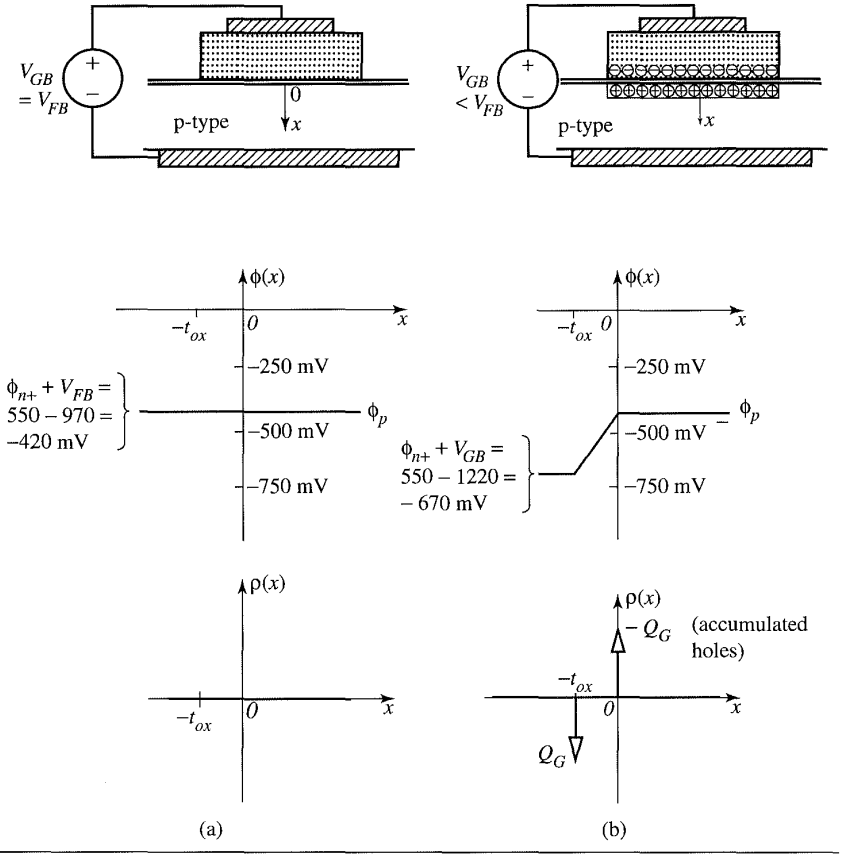
where $C_{ox} = \varepsilon_{ox}/t_{ox}$ is the capacitance per unit area of the gate oxide.

We now apply voltages $V_{GB} > V_{FB}$. Since $V_{GB} = 0\text{ V} > V_{FB}$, we already have examined this case when we considered the MOS capacitor in thermal equilibrium. The gate charge is positive and increases with increasing applied bias

$$Q_G(V_{GB}) = -Q_B = -(-qN_aX_d) = qN_aX_d(V_{GB}) \text{ (depletion)} \tag{3.91}$$

where $X_d(V_{GB})$ indicates the dependence of depletion width on V_{GB}. In order to quantify $X_d(V_{GB})$, we need to find the dependence of the surface potential on the gate-bulk voltage V_{GB}, which we will do in the next section.

Figure 3.24(a) shows the electrostatics of the MOS capacitor biased in **depletion**, with $V_{GB} = 250$ mV. For this bias voltage, the electrostatic analysis in the next section shows that the surface potential is $\phi_s = 185$ mV. The surface potential corresponds to silicon that is slightly n-type—opposite to the p-type substrate. Even though the MOS capacitor is under applied bias, $J_n = 0$ and $J_p = 0$ through the structure due to



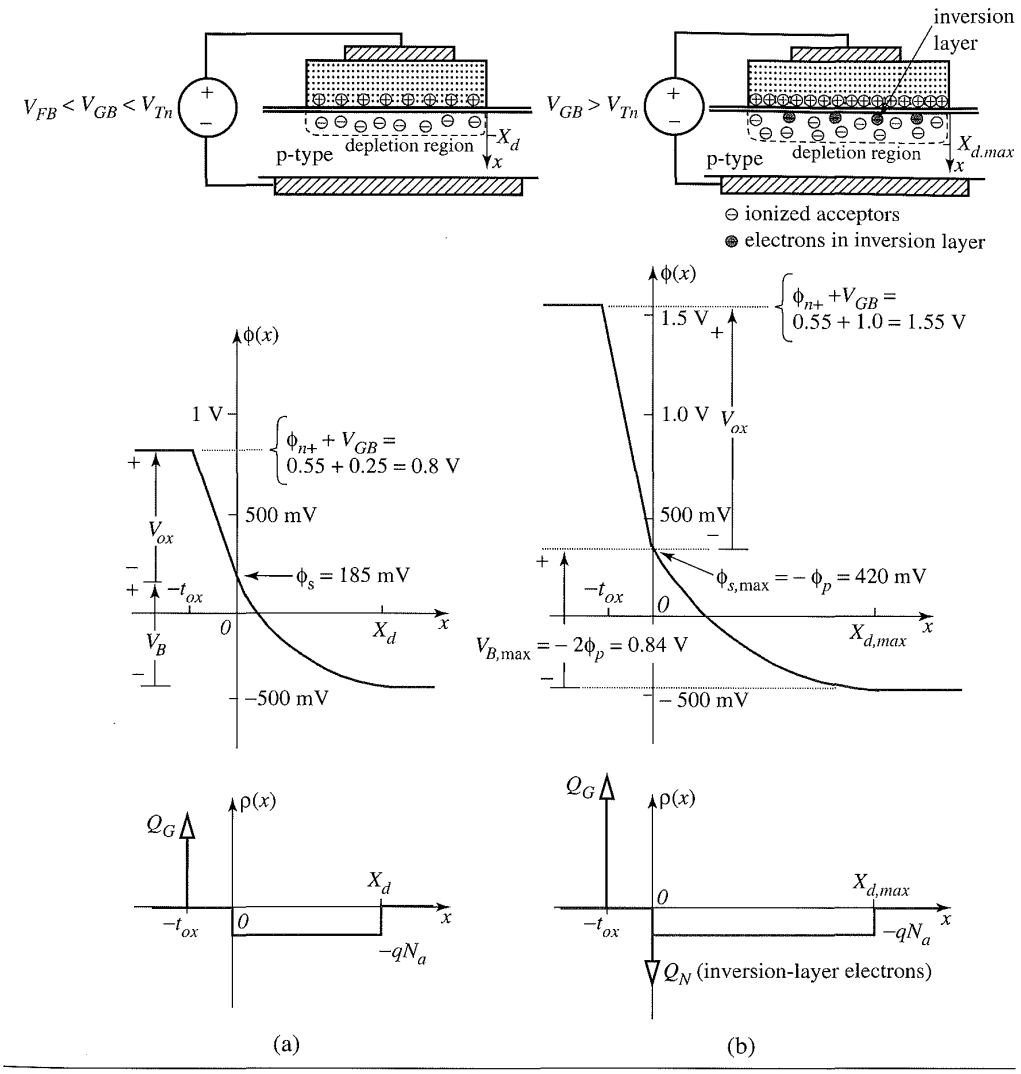
➤ **Figure 3.23** MOS capacitor with p-type substrate under an applied bias. (a) $V_{GB} = V_{FB}$ (flatband) and (b) $V_{GB} = -1.22\ \text{V} < V_{FB}$ (accumulation.) The potential and charge densities are plotted, for the case where the substrate is doped at $N_a = 10^{17}$ cm^{-3}.

the SiO_2 insulating film. Therefore, the carrier concentration is linked to the electrostatic potential by the analysis in Section 3.3.1. Note that this argument is not valid for the reverse-biased pn junction since there is a current, albeit small, through the structure when $V_D < 0$ V.

According to Eq. (3.21), the surface electron concentration at $x = 0$ is

$$n_s = n_i \cdot 10^{\phi_s/(60\ \text{mV})} = 10^{10} \cdot 10^{185/60}\ \text{cm}^{-3} = 1.2 \times 10^{13}\ \text{cm}^{-3} \qquad (3.92)$$

This concentration of electrons is negligible compared to the ionized acceptor concentration in the depletion region. As a result, the additional negative charge at the SiO_2–silicon interface from the electrons at $x = 0$ does not appear in the charge density plot in Fig. 3.24(a).



➤ **Figure 3.24** MOS capacitor with p-type substrate under an applied bias. (a) $V_{FB} < V_{GB} = 250$ mV and $V_{GB} < V_{Tn}$ (depletion) and (b) $V_{GB} = 1.0\text{ V} > V_{Tn}$ (inversion). The potential and charge densities are plotted, for the case where the substrate is doped at $N_a = 10^{17}$ cm^{-3}.

If the gate-bulk voltage V_{GB} is increased further, the surface potential will continue to rise and eventually, there will be a significant electron concentration at the surface. For a first-order analysis of the MOS capacitor, we consider that the surface has **inverted** when the surface potential becomes equal and opposite to the potential of the substrate:

$$\phi_s = -\phi_p \tag{3.93}$$

Note that the surface electron concentration at the onset of inversion is

$$n_s = n_i e^{\phi_s / V_{th}} = n_i e^{-\phi_p / V_{th}} = N_a, \tag{3.94}$$

where the last equality follows from the definition of the potential of p-type silicon in Eq. (3.22). In other words, the surface electron concentration is equal to the hole concentration in the undepleted p-type substrate, when the surface has just inverted.

The applied voltage at which the surface inverts is called the **threshold voltage**, which is given the symbol V_{Tn} for a MOS capacitor on a p-type substrate. Note that the **n** refers to the n-type surface after inversion—not to the p-type substrate. The threshold voltage is the sum of three terms

threshold voltage for p-type substrate

$$\boxed{V_{Tn} = V_{FB} - 2\phi_p + \frac{1}{C_{ox}}\sqrt{2q\varepsilon_s N_a(-2\phi_p)}} \tag{3.95}$$

The derivation of Eq. (3.95) will be left to Section 3.8. A typical range for the threshold voltage of a MOS capacitor on an n-type substrate is $V_{Tn} = 0.5 - 1$ V.

The electron concentration at the surface increases exponentially with increasing ϕ_s as indicated by Eq. (3.94). After inversion occurs, further increases in V_{GB} only slightly increase the surface potential. Note that the maximum possible surface potential is $\phi_{s\,(\max)} = 550$ mV, which is only 130 mV higher than the value of $-\phi_p = -(-420\text{ mV})$, for the case where $N_a = 10^{17}$ cm^{-3}. In order to simplify the electrostatics of the MOS capacitor, we will assume that the surface potential stops increasing after the onset of inversion for $V_{GB} > V_{Tn}$ and remains "pinned" at $-\phi_p$, which we will consider its maximum.

Figure 3.24(b) shows the electrostatics for a MOS capacitor biased with $V_{GB} > V_{Tn}$. The surface potential is $\phi_s = -\phi_p$, which means that the bulk voltage V_B (the drop across the depletion region) has reached its maximum value

$$V_B = V_{B,\max} = \phi_{s,\max} - \phi_p = -\phi_p - \phi_p = -2\phi_p \tag{3.96}$$

Much of the electrostatics in inversion is easily quantified, since the bulk voltage is a known quantity for $V_{GB} \geq V_{Tn}$. For example, the depletion width and bulk charge are constant at their maximum values $X_{d,\max}$ and $Q_{B,\max}$ in inversion. The oxide voltage is

$$V_{ox} = (V_{GB} + \phi_{n+}) - \phi_s = V_{GB} + \phi_{n+} + \phi_p \text{ for } V_{GB} \geq V_{Tn} \tag{3.97}$$

We now consider the important question of how the gate charge varies with applied bias in inversion. The gate charge balances the bulk charge $Q_{B,\max}$ in the depletion region and the electron charge Q_N in the surface inversion layer

$$Q_G(V_{GB}) = -(Q_N + Q_{B,\max}) \text{ for } V_{GB} \geq V_{Tn} \tag{3.98}$$

For $V_{GB} = V_{Tn}$, we consider that the electron charge $Q_N \approx 0$ since the MOS capacitor is at the onset of inversion, which implies that

$$Q_G(V_{Tn}) \approx -Q_{B,\max} \tag{3.99}$$

Taking equalities in these approximations, we find that the additional gate charge from increasing the bias voltage above threshold is

$$Q_G(V_{GB}) - Q_G(V_{Tn}) = -Q_N \tag{3.100}$$

In other words, the additional gate charge is balanced by the inversion-layer electron charge. The added gate charge and the inversion-layer electrons are separated by the gate oxide, from which we conclude that the parallel-plate capacitance applies

$$Q_G(V_{GB}) - Q_G(V_{Tn}) = C_{ox}(V_{GB} - V_{Tn}) \text{ for } V_{GB} \geq V_{Tn} \tag{3.101}$$

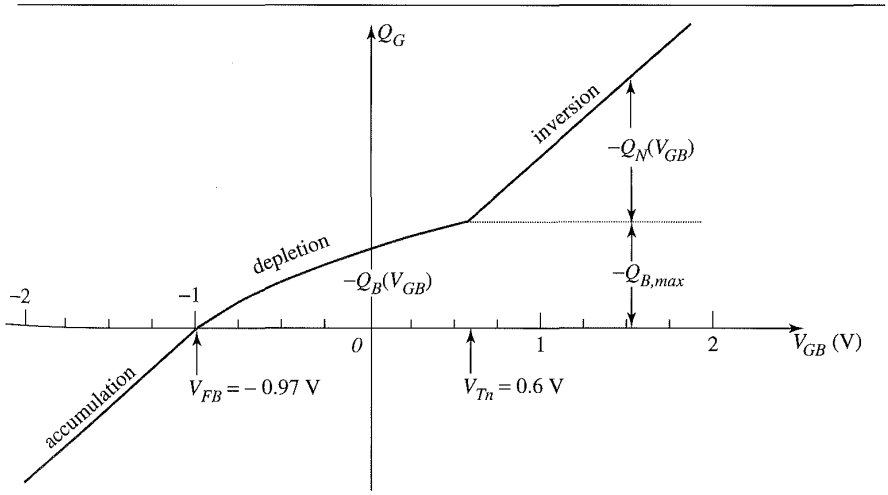
From Eqs. (3.99) and (3.101), the gate charge in inversion is

gate charge in inversion for p-type substrate

$$\boxed{Q_G(V_{GB}) = C_{ox}(V_{GB} - V_{Tn}) - Q_{B,\max}} \tag{3.102}$$

A qualitative picture of the functional dependence of the gate charge on the applied bias for a MOS capacitor with p-type substrate is given in Fig. 3.25, for the case of a substrate doping of $N_a = 10^{17}$ cm^{-3} and a 150 Å-thick oxide. From Eq. (3.88), the flatband voltage is $V_{FB} = -0.97$ V and from Eq. (3.95), the threshold voltage is $V_{Tn} = 0.6$ V. The gate charge varies linearly in accumulation and in inversion, since the additional charge in the substrate due to a change in the gate-bulk voltage is stored at the SiO_2 silicon surface.The slope in these regions is the oxide capacitance per unit area, C_{ox}. In depletion, the substrate charge is smeared over the depletion region. Therefore, Q_G has a non-linear dependence on V_{GB} which has been sketched qualitatively in Fig. 3.25.

In order to quantify the charge storage in the MOS capacitor, we must solve for the charge, field, and potential in the three regions of operation. Prior to tackling the



➤ **Figure 3.25** Gate charge as a function of gate-bulk voltage for an MOS capacitor with a 150 Å-thick gate oxide and a substrate doping $N_a = 10^{17}$ cm^{-3}.

rather involved algebra, it is helpful to summarize the qualitative behavior of the MOS capacitor on an n-type substrate.

3.7.3 MOS Capacitor on an n-type Substrate

Figure 3.26 shows a MOS capacitor with an n^+ polysilicon gate which is fabricated with an n-type substrate. We will use a typical substrate doping of $N_d = 10^{17}$ cm^{-3}, as we did in the previous section. Our qualitative understanding of the MOS capacitor on a p-type substrate is a good foundation for understanding this "complementary" structure.

In thermal equilibrium, the potential drop from the n+ polysilicon gate to the n-type substrate for a typical doping concentration $N_d = 10^{17}$ cm^{-3} is:

$$\phi_{n+} - \phi_n = 550 \text{ mV} - 420 \text{ mV} = 130 \text{ mV} \tag{3.103}$$

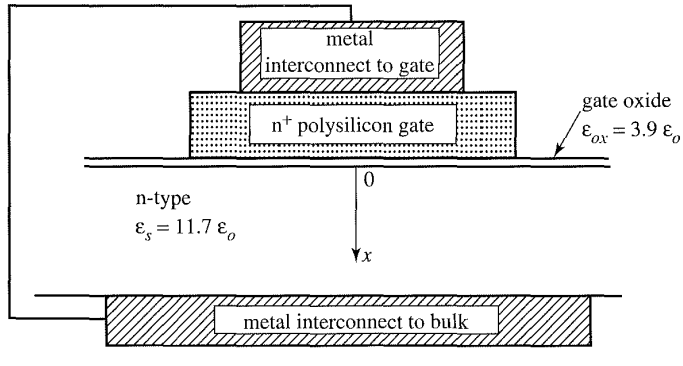
Therefore, the electric field in the oxide is positive (in the same direction as x) in Fig. 3.26 and we conclude that the charge on the gate is positive and that the charge in the n-type substrate is negative. The negative charge consists of accumulated electrons at the SiO_2 silicon interface under the gate. We can visualize the positive charge on the gate attracting the electrons and causing them to accumulate under the gate. In contrast to the situation with a p-type substrate, the MOS capacitor on an n-type substrate is in accumulation when in thermal equilibrium.

Following our procedure from the previous section, we can null the internal voltage drop by applying the flatband voltage $V_{GB} = V_{FB}$

flatband voltage for n-type substrate

$$\boxed{V_{FB} = -(\phi_{n+} - \phi_n) = -130 \text{ mV}} \tag{3.104}$$

For applied voltages more positive than the flatband voltage, the electric field is positive and the MOS capacitor is in accumulation. As we have seen, thermal equilibrium with $V_{GB} = 0 \text{ V} > V_{FB}$ falls into this region. The gate charge in the accumulation region is



➤ **Figure 3.26** MOS capacitor with n-type substrate. $V_{GB} = 0$ for investigating the capacitor in thermal equilibrium.

$$\boxed{Q_G = C_{ox}(V_{GB} - V_{FB}) \text{ for } V_{GB} \geq V_{FB}} \tag{3.105}$$

gate charge in accumulation for n-type substrate

For applied voltages more negative than the flatband voltage, the electric field reverses and a positive charge is needed in the silicon substrate. Since the substrate is n-type, a depletion region forms under the gate that consists of positively ionized donors. The gate charge for depletion is negative for this case

$$Q_G(V_{GB}) = -Q_B = -qN_d X_d(V_{GB}) \tag{3.106}$$

As the gate-bulk voltage becomes more negative, the surface potential is pulled down and the surface hole concentration p_s increases. The gate-bulk voltage when the surface potential is equal and opposite to that of the n-type substrate ($\phi_s = -\phi_n$) is defined as the threshold voltage, which is given by

$$\boxed{V_{Tp} = V_{FB} - 2\phi_n - \frac{1}{C_{ox}}\sqrt{2q\varepsilon_s N_d(2\phi_n)}} \tag{3.107}$$

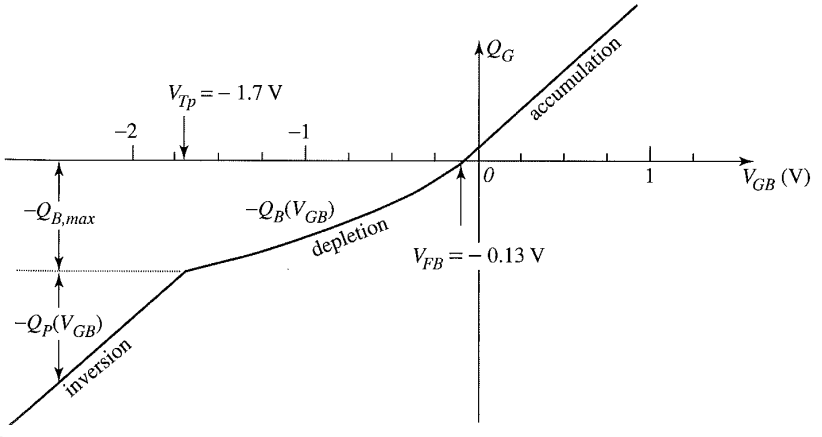
threshold voltage for n-type substrate

The signs of the potential drops across the depletion region and across the oxide are negative for the case of an n-type substrate. For voltages more negative than the threshold voltage, the MOS capacitor is inverted and the gate charge balances holes in the surface inversion layer and the maximum depletion charge:

$$\boxed{Q_G = -Q_P - Q_{B,\max} = C_{ox}(V_{GB} - V_{Tp}) - Q_{B,\max} \text{ for } V_{GB} \leq V_{Tp}} \tag{3.108}$$

gate charge in inversion for n-type substrate

The gate charge as a function of the applied bias voltage is sketched in Fig. 3.27.



➤ **Figure 3.27** Gate charge as a function of gate-bulk voltage for an MOS capacitor on an n-type substrate with doping $N_d = 10^{17}$ cm^{-3} and a 150 Å-thick gate oxide.

EXAMPLE 3.7 MOS Electrostatics in Accumulation and in Inversion

In this example, we consider a MOS capacitor with $t_{ox} = 200$ Å on a p-type substrate with acceptor concentration $N_a = 5 \times 10^{16}$ cm^{-3}.

(a) Find the electric field in the oxide and the charge per unit area in the silicon substrate (units C/cm^2) for $V_{GB} = -2.5$ V.

SOLUTION

The first step is to determine whether the capacitor is in the accumulation, depletion, or inversion regions. The flatband voltage for an MOS capacitor on a p-type substrate with these parameters is given by Eq. (3.88)

$$V_{FB} = -(\phi_{n+} - \phi_p) = -(550 - (-402)) = -0.95\text{V}$$

and the threshold voltage is found from Eq. (3.95)

$$V_{Tn} = -0.95\text{V} - 2(-0.40) + \frac{\sqrt{2(1.6\times10^{-19})(1.04\times10^{-12})(5\times10^{16})(-2)(-0.40)}}{(3.45\times10^{-13})/(2\times10^{-6})}$$

$$V_{Tn} = -0.95\text{V} - 2(-0.40) + 0.67 = 0.52\text{ V}$$

In evaluating the final term, we have substituted the permittivities of silicon and SiO_2

$$\varepsilon_s = 1.04\times10^{-12}\text{ F/cm and } \varepsilon_{ox} = 3.45\times10^{-13}\text{ F/cm}.$$

The applied bias is $V_{GB} = -2.5$ V $< V_{FB} = -0.95$ V and we conclude that the p-type substrate is accumulated and the electrostatics of Fig. 3.23(b) apply. We can find the electric field from the information on the potential plot in accumulation:

$$E_{ox} = \frac{V_{ox}}{t_{ox}} = \frac{(V_{GB} + \phi_{n+}) - \phi_p}{t_{ox}} = \frac{-2.5 + 0.55 - (-0.40)}{2\times10^{-6}} = -7.8\times10^5\text{ V/cm}$$

The charge in the substrate consists of accumulated holes. It is opposite in sign to the gate charge, which is given by Eq. (3.90)

$$Q_P = -Q_G = -C_{ox}(V_{GB} - V_{FB}) = -1.72\times10^{-7}(-2.5 - (-0.95)) = 2.67\times10^{-7}\text{ C/cm}^2$$

(b) Find the numerical value of the depletion width and depletion charge when the capacitor is biased in the inversion region.

SOLUTION

In inversion, the potential drop across the depletion region is

$$V_{B,\max} = \phi_s - \phi_p = -\phi_p - \phi_p = -2\,\phi_p = -2(-0.40) = 0.8\text{ V}.$$

Adapting the results of Ex. 3.3, the relationship between the depletion region width $X_{d,\max}$ and the potential drop $V_{B,\max}$ is

$$V_{B,\max} = \frac{1}{2}\left(\frac{qN_a}{\varepsilon_s}\right)X_{d,\max}^2$$

Solving for the depletion region width, we find that

$$X_{d,\max} = \sqrt{\frac{2\varepsilon_s V_{B\max}}{qN_a}} = \sqrt{\frac{2\,(1.035\times10^{-12})\,(0.8)}{(1.6\times10^{-19})\,(5\times10^{16})}} = 1440\ \text{Å}.$$

The bulk charge in the depletion region is

$$Q_{B,\max} = -qN_a X_{d,\max} = -(1.6\times10^{-19})(5\times10^{16})(1.44\times10^{-5}) = -1.15\times10^{-7}\ \text{C/cm}^2.$$

(c) Find the electric field in the oxide and the inversion-layer electron charge, for $V_{GB} = 2.5$ V.

SOLUTION

We know that the MOS capacitor is inverted, since $V_{GB} = 2.5\ \text{V} > V_{Tn} = 0.52$ V. The electrostatics in Fig. 3.24(b) apply for this region. From the potential plot, we can evaluate the electric field in the oxide:

$$E_{ox} = \frac{V_{ox}}{t_{ox}} = \frac{(V_{GB} + \phi_{n+}) - (-\phi_p)}{t_{ox}} = \frac{2.5 + 0.55 - 0.40}{2\times10^{-6}} = 1.32\times10^6\ \text{V/cm}.$$

Electrons in the inversion layer are one component of the total charge in the silicon substrate. Using Gauss's Law, we can relate the electric field in the oxide to the substrate charge:

$$E_{ox} = \frac{-(Q_{B,\max} + Q_N)}{\varepsilon_{ox}}$$

Solving for the inversion-layer charge, we find that

$$Q_N = -\varepsilon_{ox}E_{ox} - Q_{B,\max} = -4.55\times10^{-7} - (-1.15\times10^{-7}) = -3.4\times10^{-7}\ \text{C/cm}^2$$

We could have found the electron charge directly from Eq. (3.101)

$$Q_N = -C_{ox}(V_{GB} - V_{Tn}) = -\left(\frac{3.45\times10^{-13}}{2\times10^{-6}}\right)(2.5 - 0.52) = -3.4\times10^{-7}\ \text{C/cm}^2$$

(d) Electric fields with magnitudes greater than $E_{ox,\max} = 5\times10^6$ V/cm will cause irreversible damage to the gate oxide. Find the permissible range of gate-bulk voltages.

SOLUTION

In accumulation, the electric field in the oxide is

$$E_{ox} = \frac{V_{ox}}{t_{ox}} = \frac{(V_{GB} + \phi_{n+}) - \phi_p}{t_{ox}}$$

Solving for the most negative gate-bulk voltage using $E_{ox} = -5 \times 10^6$ V/cm, we find that

$$V_{GB,\min} = E_{ox} t_{ox} - \phi_{n+} + \phi_p = (-5 \times 10^6)(2 \times 10^{-6}) - 0.55 + (-0.4) = -10.95 \text{ V}$$

In inversion, the electric field in the oxide is

$$E_{ox} = \frac{V_{ox}}{t_{ox}} = \frac{(V_{GB} + \phi_{n+}) - (-\phi_p)}{t_{ox}}$$

Substituting the maximum positive field $E_{ox} = 5$ x 10^6 V/cm, we find that

$$V_{GB,\max} = E_{ox} t_{ox} - \phi_{n+} - \phi_p = (5 \times 10^6)(2 \times 10^{-6}) - 0.55 - 0.4 = 9.05 \text{ V}$$

In this example, we have found numerical values for several properties of the MOS capacitor when biased in accumulation and in inversion. A clear understanding of the MOS structure and its regions of operation, combined with one-dimensional electrostatics, can prove adequate for solving practical problems—without the need for complicated analysis.

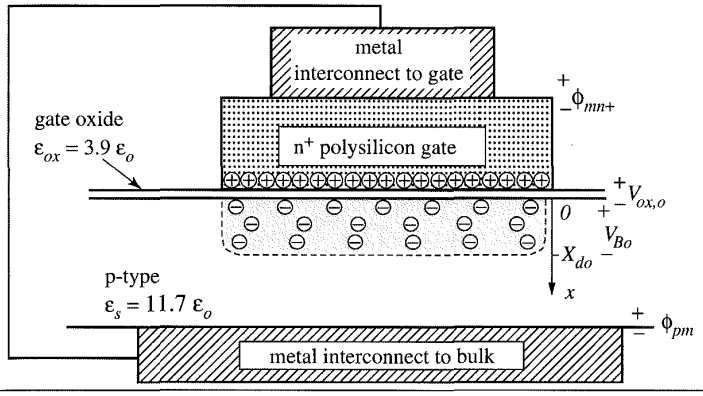
➤ 3.8 THE ELECTROSTATICS OF THE MOS CAPACITOR

In the previous section, the basic features of the MOS capacitor were outlined without deriving the results from a thorough analysis of the electrostatics. In this section, the MOS capacitor is considered more systematically, beginning with the MOS capacitor on a p-type substrate in thermal equilibrium. We will rely on the previous section to provide the context for the electrostatics.

3.8.1 MOS Electrostatics in Thermal Equilibrium

Figure 3.28 shows the MOS capacitor with $V_{GB} = 0$ V and indicates the locations and symbols for the internal potential drops. In particular, the contact potentials ϕ_{mn+} between the metal and n^+ polysilicon gate, and ϕ_{pm} between the p-type bulk and its metal contact are included in this section. These battery-like potential drops were described in Section 3.4.3 in the analysis of the pn junction under reverse bias. We define the potential drop across the oxide is defined as $V_{ox,o}$ and the drop across the depletion region as V_{Bo}, where the subscript **o** indicates thermal equilibrium.

The charge density $\rho_o(x)$ through the MOS capacitor on a p-type substrate is sketched in Fig. 3.29(a), for the case of thermal equilibrium. We would like to quantify the width of the depletion region X_{do} as a function of the material properties and oxide thickness. The bulk charge Q_{Bo} in the depletion region must be equal in magnitude and opposite in sign to the gate charge



➤ **Figure 3.28** MOS capacitor with p-type substrate, with the charge distribution taken from Fig. 3.21. Gate and bulk metal contacts shorted together for the thermal equilibrium analysis.

$$Q_{Bo} = -qN_a X_{do} = -Q_{Go} \tag{3.109}$$

where N_a is the doping concentration in the bulk. As in Section 3.7, numerical values will be calculated for a substrate acceptor concentration of $N_a = 10^{17}$ cm^{-3} and an oxide thickness $t_{ox} = 150$ Å. The charge density in Fig. 3.29(a) is the same as in Ex. 3.3. Following the same steps as in that analysis, we graphically integrate $\rho_o(x)$ and find that the electric field in the gate oxide is constant and that $E_o(x)$ varies linearly in the depletion region as shown in Fig. 3.29(b). The boundary condition on electric field

$$\varepsilon_{ox} E_o(x = 0^-) = \varepsilon_s E_o(x = 0^+) \tag{3.110}$$

is used to connect the electric field solutions across the oxide/silicon interface. The electric field drops a factor of three across $x = 0$ since the permittivity of silicon is a factor of three larger than that of oxide.

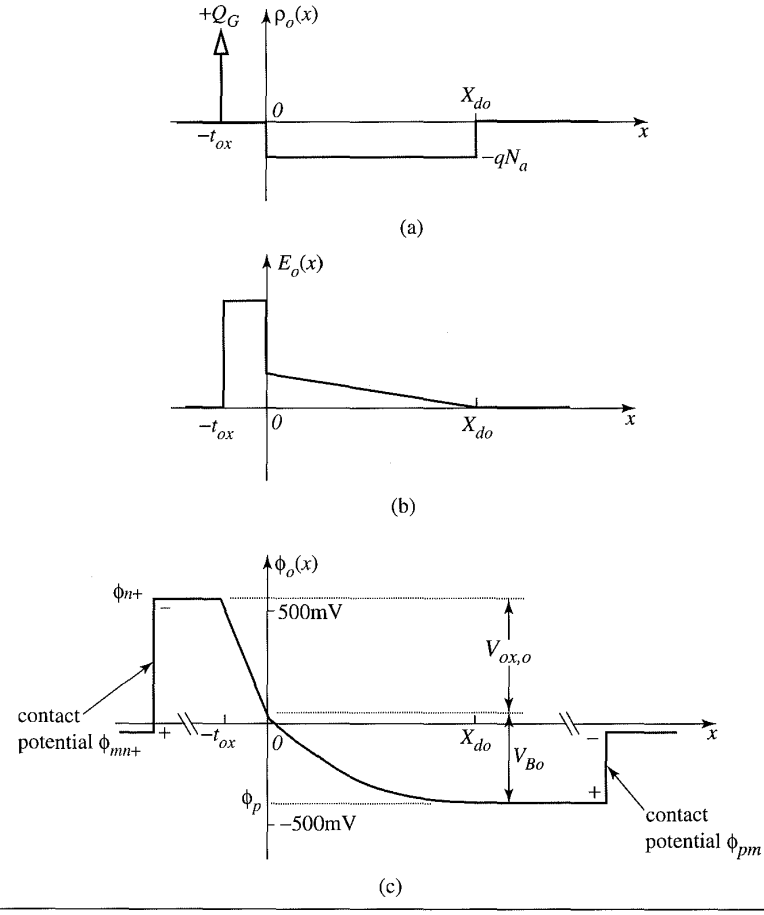
The first step in plotting the potential is to apply Kirchhoff's voltage law to the MOS capacitor in thermal equilibrium. From Fig. 3.28,

$$-\phi_{mn^+} - V_{ox,o} - V_{Bo} - \phi_{pm} = 0 \tag{3.111}$$

Solving for the internal drop from the n^+ polysilicon gate to the p-type bulk, we find that

$$V_{ox,o} + V_{Bo} = -(\phi_{mn} + \phi_{pm}) \,. \tag{3.112}$$

For the MOS capacitor, the internal potential drop is known to be the difference in the thermal equilibrium potentials of the n^+ gate and the p-type bulk:



➤ **Figure 3.29** MOS capacitor in thermal equilibrium: (a) charge density $\rho_o(x)$, (b) electric field $E_o(x)$, and (c) electrostatic potential $\phi_o(x)$. The potential jumps at the metal contacts to the n^+ polysilicon gate and to the p-type bulk are both negative quantities in (c).

$$V_{ox,o} + V_{Bo} = \phi_{n+} - \phi_p \tag{3.113}$$

By equating Eqs. (3.113) and (3.112), the sum of the contact potentials is equal to the difference in the potentials of the gate and the bulk

$$\phi_{n+} - \phi_p = -(\phi_{mn+} + \phi_{pm}) \tag{3.114}$$

which is similar to Eq. (3.58) for the pn junction in thermal equilibrium.

In order to plot the potential $\phi_o(x)$ in thermal equilibrium, we graphically integrate the negative of the electric field across the depletion region and the oxide. The potential varies linearly in the oxide since E_o is constant there, and quadratically in

the depletion region. As a sign check, we verify that the potential of the gate comes out higher than that of the underlying bulk silicon. The contact potentials ϕ_{mn+} between the metal and n^+ polysilicon gate, and ϕ_{pm} between the p-type bulk and its metal contact are shown schematically in Fig. 3.29(c) as abrupt jumps. For the case of thermal equilibrium, note that the gate and bulk metal contacts are at the same potential in Fig. 3.29(c).

In order to quantify the potential variation $\phi_o(x)$ and the width of the depletion region X_{do}, it is necessary to wade through considerable algebra. This is an unfortunate feature of MOS device physics, as will be seen in this chapter and Chapter 4. Nevertheless, it is worthwhile to work out the detailed solutions for $E_o(x)$ and $\phi_o(x)$.

To start, Gauss's Law in integral form connects the electric field just inside the silicon to the depletion charge Q_{Bo} [C/cm^2]

$$E_o(0^+) = \frac{-Q_{Bo}}{\varepsilon_s} = \frac{qN_aX_{do}}{\varepsilon_s} \tag{3.115}$$

Since there is no charge layer at the oxide/silicon interface, the boundary condition in Eq. (3.11) connects the silicon field to the oxide field on the other side of the interface

$$E_o(0^-) = \frac{\varepsilon_s}{\varepsilon_{ox}}\left(\frac{qN_aX_{do}}{\varepsilon_s}\right) = \frac{qN_aX_{do}}{\varepsilon_{ox}} = E_{ox} \tag{3.116}$$

From the definition of potential, it follows that the potential in the oxide where $-t_{ox} \le x \le 0$ is given by

$$\phi_o(x) - \phi_{n+} = \int_{-t_{ox}}^{x} -E_{ox}\,dx = -E_{ox}(x + t_{ox}) = -\frac{qN_aX_{do}}{\varepsilon_{ox}}(x + t_{ox}) \tag{3.117}$$

It is not difficult to make a sign error, but we can check our algebra since we know the potential decreases through the oxide. The potential at the oxide/silicon interface can be found by evaluating Eq. (3.117) at the origin

$$\phi_o(0) = \phi_{n+} - \frac{qN_aX_{do}t_{ox}}{\varepsilon_{ox}} \tag{3.118}$$

This result can be expressed in terms of $V_{ox,o}$ which is the drop across the oxide

$$V_{ox,o} = \phi_{n+} - \phi_o(0) = \frac{qN_aX_{do}}{C_{ox}} = \frac{Q_{Go}}{C_{ox}} \tag{3.119}$$

where we have substituted the oxide capacitance $C_{ox} = \varepsilon_{ox}/t_{ox}$ [F/cm^2]. In Eq. (3.119), we have also substituted the equality between the charge (per unit area) on the gate Q_{Go} and the negative of the charge (per unit area) in the silicon $-Q_{Bo} = qN_aX_{do}$.

We now turn our attention to the depletion region in the silicon substrate. From the integral form of Gauss's Law, the electric field $E_o(x)$ in the charged region $(0 \le x \le X_{do})$ is given by

$$\varepsilon_s E_o(x) - \varepsilon_s E_o(0^+) = \int_{0^+}^{x} -qN_a\,dx = -qN_a x \tag{3.120}$$

Solving for $E_o(x)$ and substituting Eq. (3.115) for $E_o(0^+)$, the electric field at the interface, we find the linear equation sketched in Fig. 3.29(b)

$$E_o(x) = \frac{qN_a(X_{do} - x)}{\varepsilon_s} \tag{3.121}$$

We then integrate the electric field to find $\phi_o(x)$ in the charged region $0 \le x \le X_{do}$

$$\phi_o(x) - \phi_o(0) = \int_{0^+}^{x} -E_o(x)dx = \left(-\frac{qN_a}{\varepsilon_s}\right)\left(X_{do}x - \frac{x^2}{2}\right) \tag{3.122}$$

Substituting $\phi_o(0)$ from Eq. (3.118), the potential in the charged region is given by

$$\phi_o(x) = \left(\phi_{n+} - \frac{qN_a X_{do}}{C_{ox}}\right) - \frac{qN_a}{\varepsilon_s}\left(X_{do}x - \frac{x^2}{2}\right) \tag{3.123}$$

Since the potential at the bottom of the charged region is $\phi_o(X_{do}) = \phi_p$, we can solve for the thermal equilibrium built-in potential across the polysilicon-oxide-silicon sandwich

$$\phi_{n+} - \phi_p = \frac{qN_a X_{do}}{C_{ox}} + \frac{qN_a X_{do}^2}{2\varepsilon_s} \tag{3.124}$$

The first term in on the right-hand side of Eq. (3.124) is $V_{ox,o}$, the equilibrium drop across the oxide (see Eq. (3.119)), therefore the second term is the potential drop across the charged region V_{Bo}, as can also be shown from Eq. (3.122). This result is represented graphically in Fig. 3.29(c).

The width of the depletion region can be found by solving for X_{do}

$$X_{do} = t_{ox}\left(\frac{\varepsilon_s}{\varepsilon_{ox}}\right)\left(\sqrt{1 + \frac{2C_{ox}^2(\phi_{n+} - \phi_p)}{q\varepsilon_s N_a}} - 1\right) \tag{3.125}$$

Evaluating Eq. (3.123) at $x = 0$, we find that the thermal equilibrium surface potential is

$$\phi_{so} = \phi_o(x{=}0) = \phi_{n+} - \frac{qN_a X_{do}}{C_{ox}} \tag{3.126}$$

where X_{do} is given in Eq. (3.125).

EXAMPLE 3.8 MOS Capacitor in Thermal Equilibrium

It is always helpful to substitute typical dimensions and material properties into complicated expressions such as Eqs. (3.125) and (3.126) in order to develop a feel for the magnitudes of the quantities. For a typical gate oxide thickness of $t_{ox} = 15$ nm and a typical bulk doping of $N_a = 10^{17}$ cm^{-3}, we find that the depletion layer is

$$X_{do} = 76 \text{ nm}$$

and that the surface potential at the oxide/silicon interface is

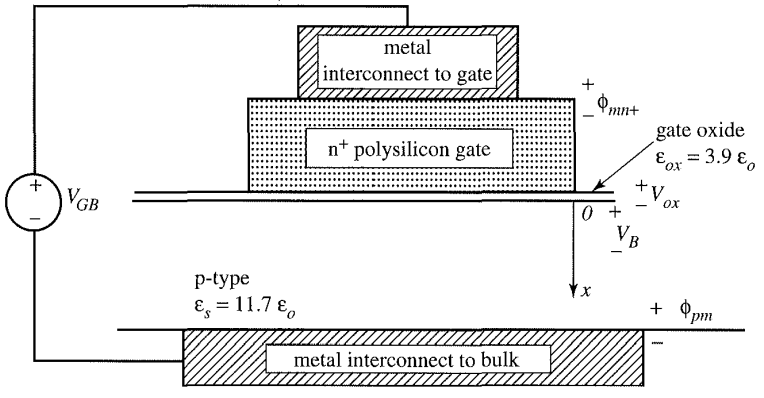
$$\phi_{so} = \phi(x = 0) = 550 \text{ mV} - 529 \text{ mV} = 21 \text{ mV}$$

For this particular case, the potential drop across the oxide $V_{ox,o} = 529$ mV is slightly larger than the drop across the depletion region $V_{Bo} = 21 - (-420) = 441$ mV, with the depletion region width equal to about five times the oxide thickness.

3.8.2 MOS Electrostatics under Applied Bias

From the numerical results in Ex. 3.8, we note that the surface potential for this example corresponds to silicon that is nearly intrinsic. In other words, the "built-in" voltage drop across the MOS capacitor, $\phi_{n+} - \phi_p$, has set up electric fields that have affected the electrical properties of the silicon underlying the gate oxide. This phenomenon, termed the **field effect**, is controllable through an applied voltage on the gate. We will see in the next chapter that it forms the basic principle of MOS transistors.

Having done a thorough analysis of the electrostatics of the MOS capacitor in thermal equilibrium, we now turn to the case where an external bias V_{GB} is applied between the metal interconnections to the polysilicon gate and p-type bulk silicon. Figure 3.30 shows the MOS capacitor with a voltage source V_{GB} connected from the



➤ **Figure 3.30** MOS capacitor on a p-type substrate for one-dimensional analysis of charge distribution ρ(x), electric field E(x), and potential ϕ(x) for the case of an applied bias V_{GB}.

gate to the bulk. We would like to investigate the effect of V_{GB} on the potential $\phi(x)$ through the structure. For convenience, the metal interconnection to the bulk will be kept fixed in plotting the change in potential under applied bias. As we have seen in Section 3.6, the electrostatics of the MOS capacitor are very different depending on the sign and magnitude of the applied voltage. We now quantify the charge, electric field, and potential for the MOS capacitor for the different regions of applied bias.

FLATBAND

A good starting point is to apply a gate voltage that is *opposite* to the built-in internal potential drop $\phi_{n+} - \phi_p$, which we define as the flatband voltage V_{FB}. For example, the built-in drop is 970 mV for the case where the substrate doping is $N_a = 10^{17}$ cm^{-3}. Applying $V_{GB} = V_{FB} = -970$ mV shifts the gate interconnect lower by 970 mV with respect to the bulk interconnect. The contact potential ϕ_{mn+} is unchanged by the applied bias, so the polysilicon gate potential is lowered to

$$\phi_{n+} - V_{FB} = 550\,\text{mV} - 970\,\text{mV} = -420\,\text{mV} \tag{3.127}$$

The bulk metal contact is considered fixed under applied bias and its contact potential ϕ_{pm} is also unchanged, so the p-type bulk remains at $\phi_p = -420$ mV. Therefore, there is no potential drop between the n^+ polysilicon and p-type bulk in flatband. The potential is therefore constant at $\phi(x) = -420$ mV throughout the MOS structure when biased at flatband. There is no electric field or charge density in the MOS capacitor for this case either, as shown in Fig. 3.31.

ACCUMULATION

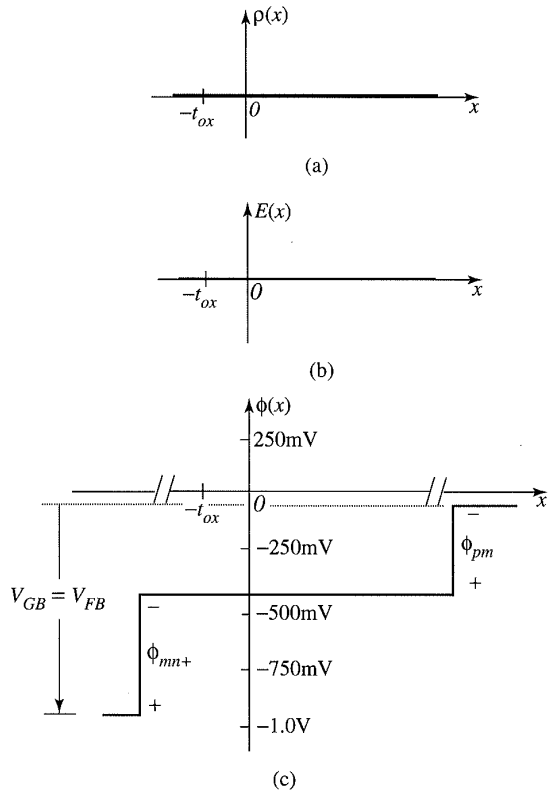
The electrostatics of the MOS capacitor are very different for voltages $V_{GB} < V_{FB}$ compared with $V_{GB} > V_{FB}$. In the former case, the potential of the n^+ polysilicon gate is pulled less than that of the p-type bulk, leading to a negative charge on the gate and a positive charge in the silicon underneath the gate oxide. As we have seen in Section 3.7, the p-type substrate has a high concentration of mobile holes that accumulate at the SiO_2–silicon interface due to attraction by the negative gate charge. An excess of holes over the acceptor concentration results in a net positive charge at the silicon surface. The surface potential is pulled lower, due to the surface hole concentration p_s exceeding the bulk doping concentration:

$$\phi_s = -V_{th} \ln(p_s/n_i) < -V_{th} \ln(N_a/n_i) \tag{3.128}$$

However, the logarithmic function is weak and it is a reasonable approximation that $\phi_s \approx \phi_p$ in accumulation. In Section 3.7, we were able to quantify the electrostatics in accumulation since the parallel-plate analysis in Ex. 3.1 could be adapted to this case.

DEPLETION

We have already worked through an example of the MOS capacitor with $V_{GB} > V_{FB}$ in the case of thermal equilibrium since $V_{GB} = 0$ V and $V_{FB} = -0.97$ V. As shown in Fig. 3.29(a), the positive sheet charge on the gate in thermal equilibrium is mirrored in a negatively charged depletion region in the silicon substrate. In this section, we will establish the range of voltages V_{GB} for which the MOS capacitor is in the depletion region. Since the form of the charge density for a depleted MOS capacitor is identical to that in thermal equilibrium, the results of the equilibrium electrostatic analysis can be adapted and applied here.



➤ **Figure 3.31** MOS capacitor on a p-type substrate biased in flatband with $V_{GB} = V_{FB} = -0.97$ V. (a) Charge density, (b) electric field, and (c) potential. The arrow on the potential plot graphically indicates that the flatband voltage pulls the gate metallization lower relative to the bulk metallization.

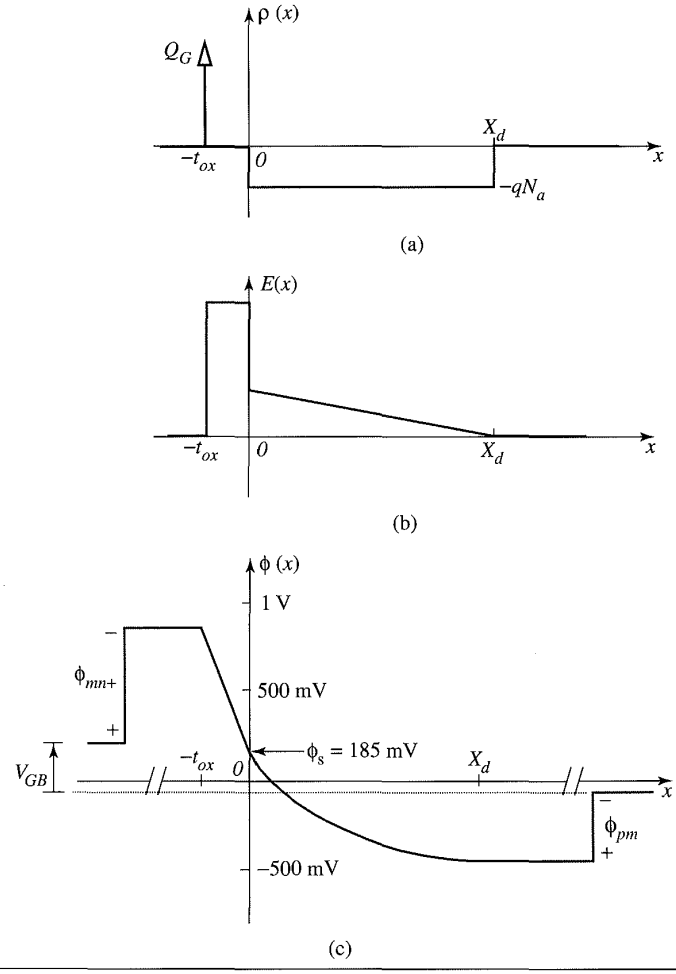
Figure 3.32 shows the electrostatics of a MOS capacitor with $V_{GB} = 250$mV and $V_{FB} = -970$ mV. Later in this section, we will justify that the capacitor is depleted for this particular gate-bulk bias. The only difference from equilibrium is that the internal voltage drop across the oxide and depletion region is now

$$V_{GB} - V_{FB} = V_{ox} + V_B = 250 - (-970) = 1220\,\text{mV} = 1.22\,\text{V} \qquad (3.129)$$

rather than $-V_{FB} = \phi_{n+} - \phi_p = 0.97$ V. Making the substitution $V_{GB} - V_{FB}$ for $-V_{FB}$ in the equilibrium depletion width in Eq. (3.125) and $V_{GB} + \phi_{n+}$ for ϕ_{n+} in the equilibrium surface potential in (3.126), we find that

$$X_d(V_{GB}) = t_{ox}\left(\frac{\varepsilon_s}{\varepsilon_{ox}}\right)\left(\sqrt{1 + \frac{2C_{ox}^2\,(V_{GB} - V_{FB})}{q\varepsilon_s N_a}} - 1\right) \qquad (3.130)$$

$$\phi_s(V_{GB}) = (V_{GB} + \phi_{n+}) - \frac{qN_a X_d(V_{GB})}{C_{ox}} \qquad (3.131)$$



➤ **Figure 3.32** MOS capacitor on a p-type substrate biased in depletion with an applied bias V_{GB} = 250 mV for V_{FB} = –970 mV: (a) charge density, (b) electric field, and (c) potential.

For V_{GB} = 250 mV, the depletion region width increases to X_d = 88.5 nm and the surface potential is ϕ_s = (250 mV + 550 mV) – 615 mV = 185 mV.

We discussed the implications of the surface becoming n-type in Section 3.7. Here, we will quantify the onset of inversion and derive the expression for the threshold voltage.

The Threshold Voltage

As we increase the applied bias, the surface potential rises according to Eq. (3.131). At some point the assumptions underlying the depletion charge distribution in Fig.

3.32(a) become invalid. For example, if we substitute $V_{GB} = 1$ V into Eqs. (3.130) and (3.131), we find that the surface potential $\phi_s = 740$ mV, which is greater than the maximum potential possible in silicon ($\phi_{n+} = 550$ mV).

The electron concentration increases exponentially as the surface potential increases, according to Eq. (3.92), to the point where it eventually becomes the dominant component of the negative charge in the silicon substrate. After the surface becomes strongly n-type, the charge density in the silicon substrate must be modified to include the electron contribution

$$\rho(x) = -q(N_a + n(x)) = -q(N_a + n_i e^{\phi(x)/V_{th}}) \text{ for } 0 \le x \le X_d \qquad (3.132)$$

Substitution of Eq. (3.132) into Poisson's Equation results in a nonlinear differential equation for the potential $\phi(x)$. The solution will be left for an advanced device physics course. As usual, our goal is to find an approximation that allows us to capture the essential results in relatively simple equations.

The exponential increase in the electron concentration with increasing surface potential turns out to be the key to making a reasonable approximation to the charge density. We are able to set a critical surface potential ϕ_s' below which $n(x)$ can be neglected in Eq. (3.132) and the MOS capacitor considered depleted. The **onset of inversion** is defined to be when the surface potential is equal and opposite to the potential of the bulk p-type substrate.

$$\boxed{\phi_s' = -\phi_p} \qquad (3.133)$$

onset of inversion for p-type substrate

In simplifying the charge density in Eq. (3.132), we will neglect the surface electron concentration for surface potentials less than the critical surface potential at the onset of inversion

$$n_s \approx 0 \text{ for } \phi_s \le \phi_s' = -\phi_p \qquad (3.134)$$

Note that the surface electron concentration when $\phi_s' = -\phi_p$ is

$$n_s = n_i e^{\phi_s'/V_{th}} = n_i e^{-\phi_p/V_{th}} = N_a = p(x > X_d) \qquad (3.135)$$

In other words, Eq. (3.135) states that at the critical surface potential, the electron concentration is equal to the acceptor concentration, or that the surface is "as much n-type as the bulk is p-type." Although the electron concentration $n_s = N_a$ at this point, we will nevertheless neglect n_s in solving the electrostatics for $\phi_s' = -\phi_p$ according to the approximation in Eq. (3.134). This approach seems likely to lead to large errors, but the very rapid (exponential) change in n_s with surface potential makes the results reasonably accurate and very useful for understanding the essentials of MOS electrostatics.

The applied gate bias V_{GB} at which the onset of inversion occurs is obviously a very important quantity for the MOS capacitor. We define it as the **threshold voltage** V_{Tn}[1] for capacitors on p-type substrates. In order to find an expression for V_{Tn}, we can

1. Not to be confused with the thermal voltage, $V_{th} = kT/q$.

directly apply our understanding of the electrostatics of the depleted MOS capacitor. According to Eq. (3.134), $n_s \approx 0$ for $\phi_s \leq -\phi_p$ and so the charge density from Fig. 3.32(a) remains valid. We can solve for V_{Tn} by substituting $\phi_s = -\phi_p$ into Eqs. (3.130) and (3.131). However, there is little insight gained. Instead, we will begin from the plot of potential at the onset of inversion in Fig. 3.33.

At the onset of inversion, the potential drop V_B' across the depletion region is a known quantity

$$V_B' = \phi_s' - \phi_p = -\phi_p - \phi_p = -2\phi_p \tag{3.136}$$

The charge density in the depletion region is $\rho(x) = -qN_a$ and can be integrated twice using Poisson's Equation, as we did in Ex. 3.3. At the onset of inversion, the depletion width has increased to its maximum value which we call $X_{d,\max}$. Integration of the constant charge density in the depletion region leads to

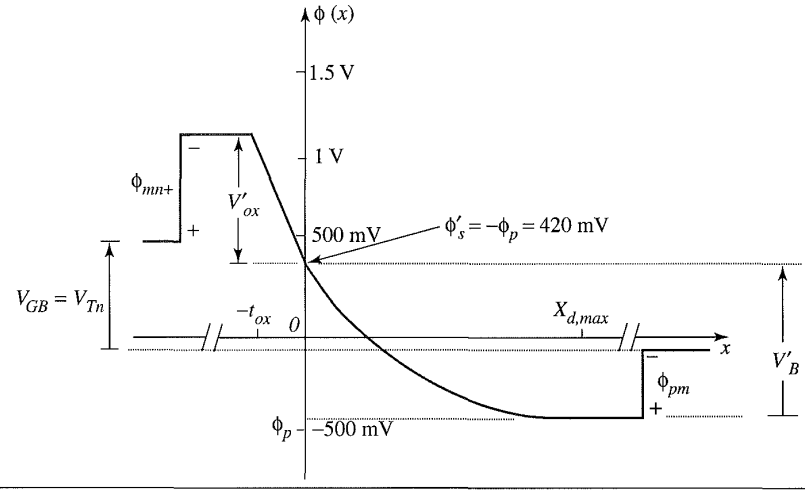
$$V_B' = \left(\frac{qN_a}{2\varepsilon_s}\right) X_{d,\max}^2 = -2\phi_p \tag{3.137}$$

Solving for $X_{d,\max}$ we find it is related to the drop across the depletion region by

$$X_{d,\max} = \sqrt{2\frac{\varepsilon_s}{qN_a}(-2\phi_p)} \tag{3.138}$$

The charge in the depletion region at the onset of inversion $Q_B' = Q_{B,\max}$ is the product of its charge density and the depletion width

$$Q_{B,\max} = -qN_a X_{d,\max} = -\sqrt{2q\varepsilon_s N_a(-2\phi_p)} \tag{3.139}$$



➤ **Figure 3.33** Potential for gate biased at the threshold voltage $V_{GB} = V_{Tn}$, the onset of inversion. The surface potential is equal and opposite to the bulk potential.

Gauss's integral law can be applied to find the electric field in the oxide, which when multiplied by the oxide thickness yields V_{ox}', the potential drop across the oxide in Fig. 3.33

$$V_{ox}' = E_{ox}' t_{ox} = \frac{-Q_{B,\max} t_{ox}}{\varepsilon_{ox}} = \left(\frac{t_{ox}}{\varepsilon_{ox}}\right)\sqrt{2q\varepsilon_s N_a(-2\phi_p)} \tag{3.140}$$

The term in parentheses in Eq. (3.140) is the inverse of the oxide capacitance C_{ox}. Since the internal voltage drop in the MOS capacitor is the sum of the drops across the depletion region and across the oxide as shown in Fig. 3.33

$$V_{GB} - V_{FB} = V_{Tn} - V_{FB} = V_B' + V_{ox}' \tag{3.141}$$

Substituting from Eqs. (3.137) and (3.140), we can solve for the threshold voltage on a p-type substrate, which was given previously in Eq. (3.95) in Section 3.7

$$V_{Tn} = V_{FB} + V_B' + V_{ox}' = V_{FB} - 2\phi_p + \frac{1}{C_{ox}}\sqrt{2q\varepsilon_s N_a(-2\phi_p)} \tag{3.142}$$

This expression hardly seems like a "neat" final result. However, V_{Tn} is the sum of three simple terms. First, the flatband voltage is the built-in offset across the MOS structure. The second term is the drop across the depletion region at the onset of inversion. The final term is the magnitude of the depletion charge at inversion divided by the oxide capacitance, which turns out to be the potential drop across the gate oxide. For the typical case where $t_{ox} = 15$ nm and $N_a = 10^{17}$ cm^{-3}, the threshold voltage from Eq. (3.142) is $V_{Tn} = 600$ mV.

INVERSION

What happens when the applied gate-bulk voltage V_{GB} is increased further? From Eq. (3.135), small changes in the surface potential lead to large increases in the surface electron concentration. As we know from Eq. (3.23), a 60 mV increase in ϕ_s over ϕ_s' corresponds to an order-of-magnitude increase in n_s. Therefore, we conclude that the surface potential only slowly increases as V_{GB} is raised, since the electron charge will increase linearly with V_{GB}. This insight allows us to greatly simplify the electrostatics in inversion, by making the **delta-depletion approximation**

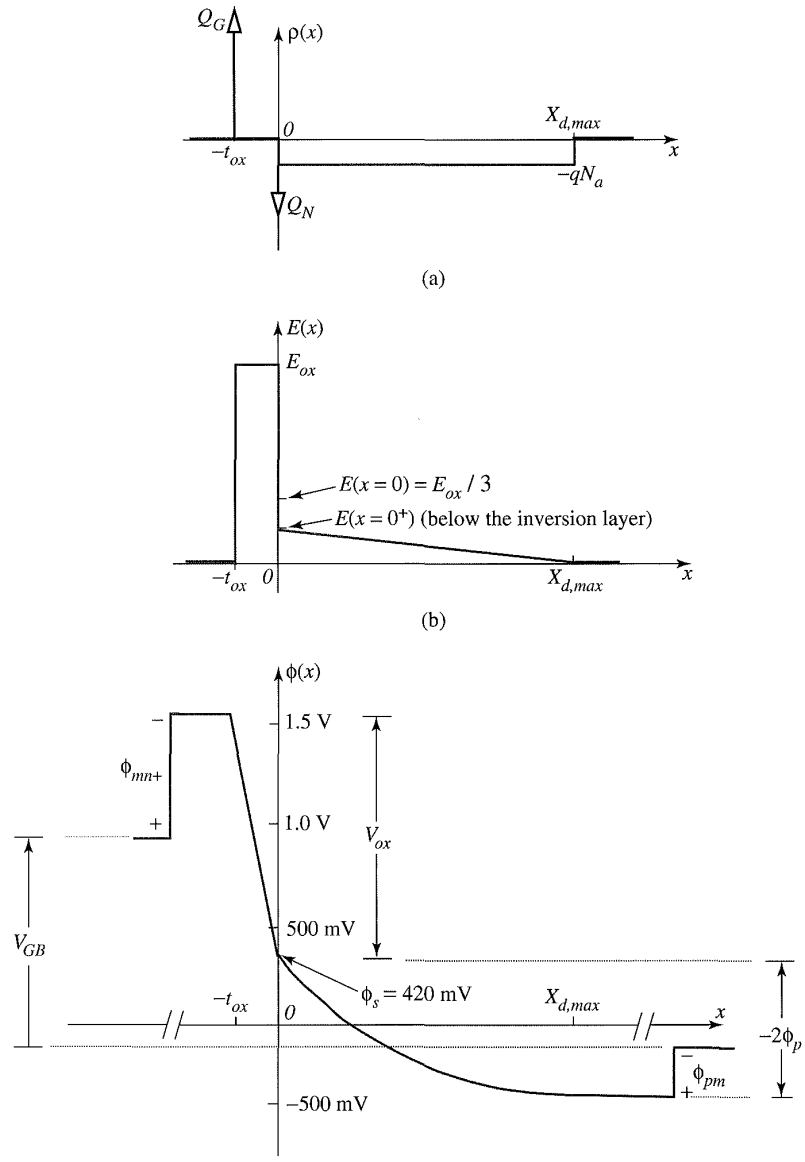
delta-depletion approximation

$$\boxed{\phi_s \approx \phi_s' = -\phi_p \text{ for } V_{GB} \geq V_{Tn}} \tag{3.143}$$

In other words, we consider the surface potential **pinned** at $-\phi_p$ in inversion. The approximation that ϕ_s is pinned means that we cannot use Eq. (3.21) to find the electron concentration at the surface in the inversion region. However, we will be able to find the inversion-layer electron charge Q_N [C/cm^2] by other means.

Now that the boundary between depletion and inversion has been established using the delta-depletion approximation, we can analyze the electrostatics of the inverted MOS capacitor. Our primary goal is to find the relationship between the electron charge in the inversion layer and the applied voltage V_{GB}.

Figure 3.34(a) shows the charge density for the case where $V_{GB} = 1.0$ V when $V_{Tn} = 600$ mV. Electrons at the SiO_2–silicon surface constitute a sheet charge Q_N



➤ **Figure 3.34** MOS capacitor on a p-type substrate biased in inversion with $V_{GB} = 1$ V, $V_{Tn} = 600$ mV: (a) charge density, (b) electric field, and (c) potential.

[C/cm^2] that is represented by the delta function in Fig. 3.34(a). The depletion region width is $X_{d,\max}$ after inversion because the drop across it remains fixed at $-2\phi_p$. The potential drop across the oxide increases since $V_{GB} > V_{Tn}$ and

$$V_{GB} - V_{FB} = V_{ox} - 2\phi_p \tag{3.144}$$

Once again, the integral form of Gauss's Law is useful for connecting V_{ox} to the charge in the silicon that consists of the sum of the electron charge Q_N, and the (maximum) depletion charge $Q_B' = Q_{B,\max}$

$$V_{ox} = E_{ox}\, t_{ox} = \left(\frac{t_{ox}}{\varepsilon_{ox}}\right)(-Q_{B,\max} - Q_N) \tag{3.145}$$

In Fig. 3.34(b), the electric field drops a factor of three from the oxide to just inside the silicon at $x = 0$

$$E(0) = \frac{-Q_N - Q_{B,\max}}{\varepsilon_s} = E_{ox}\left(\frac{\varepsilon_{ox}}{\varepsilon_s}\right) = \frac{E_{ox}}{3} \tag{3.146}$$

In inversion, there is a second discontinuity in the electric field, as shown in Fig. 3.34(b). The electrons in the inversion layer form a charge sheet located at the SiO_2–silicon interface. Therefore, the electric field just below the inversion layer at $x = 0^+$ does not include the contribution of the inversion charge

$$E(0^+) = -\frac{Q_{B,\max}}{\varepsilon_s} = E(0) - \frac{(-Q_N)}{\varepsilon_s} \tag{3.147}$$

The width of the inversion layer is determined in an advanced course to be on the order of 50 Å, so the drop in electric field is not quite as abrupt as shown in Fig. 3.34(b).

Our goal is to express the inversion-layer charge Q_N as a function of the applied voltage, so we substitute Eq. (3.145) into Eq. (3.144)

$$V_{GB} - V_{FB} = V_{ox} - 2\phi_p = \frac{(-Q_{B,\max} - Q_N)}{C_{ox}} - 2\phi_p \tag{3.148}$$

in which we have also substituted $C_{ox} = \varepsilon_{ox}/t_{ox}$. Solving Eq. (3.148) for the inversion-layer charge, we find that

inversion-layer electron charge for p-type substrate

$$\boxed{Q_N = -C_{ox}\left(V_{GB} - V_{FB} + 2\phi_p - \frac{-Q_{B,\max}}{C_{ox}}\right) = -C_{ox}\,(V_{GB} - V_{Tn}) \quad \text{for } V_{GB} \geq V_{Tn}} \tag{3.149}$$

where we have identified the threshold voltage from Eq. (3.142). Note that there is *no* inversion layer charge unless the gate voltage is greater than the threshold voltage. This result is central to the model of the MOS field-effect transistor that will be developed in Chapter 4.

The results of this section for the MOS capacitor on a p-type substrate can be summarized by writing the gate charge as a function of the applied gate-bulk bias in accumulation, depletion, and inversion.

Gate charge for: accumulation depletion inversion

$$Q_G = C_{ox}(V_{GB} - V_{FB}) \quad \text{for } V_{GB} \le V_{FB} \tag{3.150}$$

$$Q_G = -Q_B(V_{GB}) = \frac{q\varepsilon_s N_a}{C_{ox}}\left(\sqrt{1 + \frac{2C_{ox}^2(V_{GB} - V_{FB})}{q\varepsilon_s N_a}} - 1\right) (V_{FB} \le V_{GB} \le V_{Tn}) \tag{3.151}$$

$$Q_G = C_{ox}(V_{GB} - V_{Tn}) + \frac{q\varepsilon_s N_a}{C_{ox}}\left(\sqrt{1 + \frac{2C_{ox}^2(V_{Tn} - V_{FB})}{q\varepsilon_s N_a}} - 1\right) \text{ for } V_{GB} \ge V_{Tn} \tag{3.152}$$

The depletion charge in Eq. (3.151) is found from Eq. (3.130) which gives the depletion width as a function of bias $X_d(V_{GB})$. The results in Eqs. (3.150)–(3.152) quantify the sketch of Q_G versus V_{GB} in Fig. 3.25 and will be useful for finding the MOS capacitance in the next section.

We conclude this section with an example that adapts the quantitative analysis of the MOS capacitor to the case of an n-type substrate.

EXAMPLE 3.9 MOS Electrostatics with an N-type Substrate

Given a MOS capacitor on an n-type substrate with an n^+ poly gate (ϕ_{n+}= 550mV), a donor concentration $N_d = 10^{16}$ cm^{-3}, and an oxide thickness t_{ox} = 500 Å. The contact potentials have values $\phi_{mn+} = -400$ mV and ϕ_{nm}= 210 mV.

(a) Find the flatband voltage V_{FB}.

SOLUTION

The flatband voltage is equal and opposite to the built-in voltage in the MOS structure

$$V_{FB} = -(\phi_{n+} - \phi_n) = -(550 \text{ mV} - 360 \text{ mV}) = -190 \text{ mV}$$

(b) Plot the potential $\phi_o(x)$ in thermal equilibrium

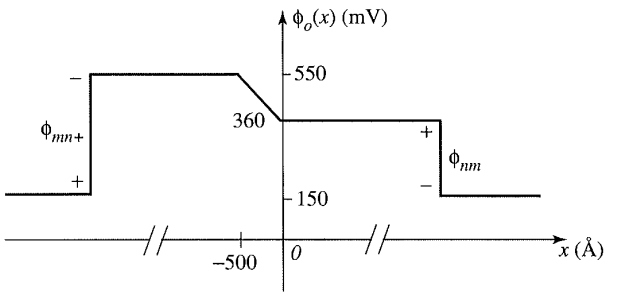
SOLUTION

The MOS capacitor is accumulated in thermal equilibrium, since accumulated electrons under the gate balance the positive gate charge. We can sketch the potential since the substrate charge is a delta function at the SiO_2 silicon interface with value $Q_{No} = -Q_{Go}$, as shown in Fig. Ex3.9A.

(c) Find the threshold voltage V_{Tp}.

SOLUTION

In order to invert the surface, we must first apply V_{FB} to reach flatband. To deplete the substrate, we must apply $V_{GB} < V_{FB}$ to repel the mobile electrons from the surface and leave the positively charged ionized donors. At the onset of inversion, the surface potential will be lowered to the point where it is equal and opposite to that of the n-type bulk: $\phi_s' = -\phi_n$. Adding the flatband voltage, the



➤ **Figure Ex3.9A** Potential of the MOS structure with n-type substrate in thermal equilibrium.

voltage drop across the depletion region $V_B' = V_{B,\max} = -2\phi_n$ and the voltage drop across the oxide V_{ox}' the threshold voltage V_{Tp} is

$$V_{Tp} = V_{FB} - 2\phi_n + V_{ox}' = V_{FB} - 2\phi_n - \frac{Q_{B,\max}}{C_{ox}}$$

Note that the drop across the oxide is negative since the substrate has a positive depletion charge $Q_{B,\max} > 0$ and the gate charge is negative. Substituting for the maximum depletion charge $Q_{B,\max}$, we find

$$V_{Tp} = V_{FB} - 2\phi_n - \frac{\sqrt{2q\varepsilon_s N_d (2\phi_n)}}{C_{ox}} = -0.19 - 0.72 - \frac{4.88 \times 10^{-8}\,\text{C/cm}^2}{6.9 \times 10^{-8}\,\text{F/cm}} = -1.62\,\text{V}$$

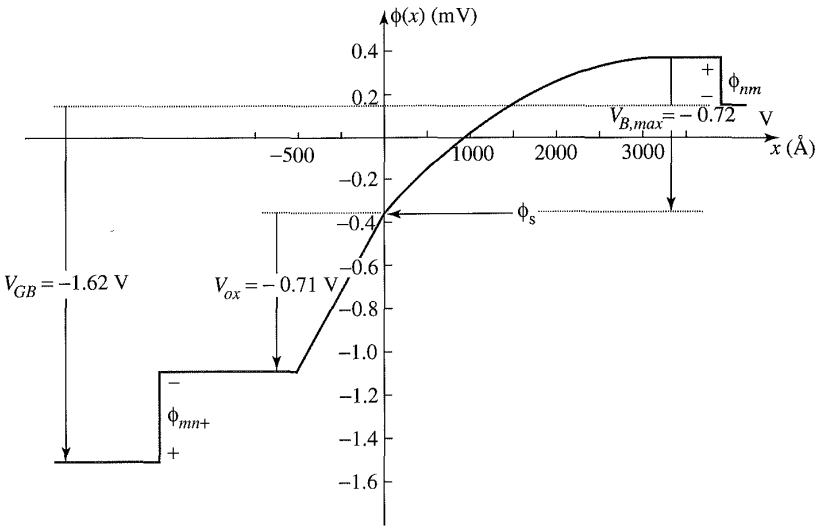
(d) Plot the potential $\phi(x)$ for $V_{GB} = V_{Tp}$.

SOLUTION

When $V_{GB} = V_{Tp} = -1.62$ V, the n^+ polysilicon gate has a potential of 0.55 – 1.62 = –1.07 V. The potential in the n-type bulk remains fixed at $\phi_n = 0.36$ V with the surface potential at the onset of inversion $\phi_s = -\phi_n = -0.36$ V. The potential varies quadratically in the depletion region, which has a width

$$X_{d,\max} = \sqrt{\frac{2\varepsilon_s (2\phi_n)}{qN_a}} = \sqrt{\frac{2 \cdot 11.7 \cdot 8.85 \times 10^{-14} \cdot 0.72}{1.6 \times 10^{-19} \cdot 10^{16}}} = 0.31\,\mu\text{m}$$

and linearly in the oxide as shown in Fig. Ex3.9B.



➤ **Figure Ex3.9B** Potential of MOS structure with n-type substrate for the case where $V_{GB} = V_{Tp}$.

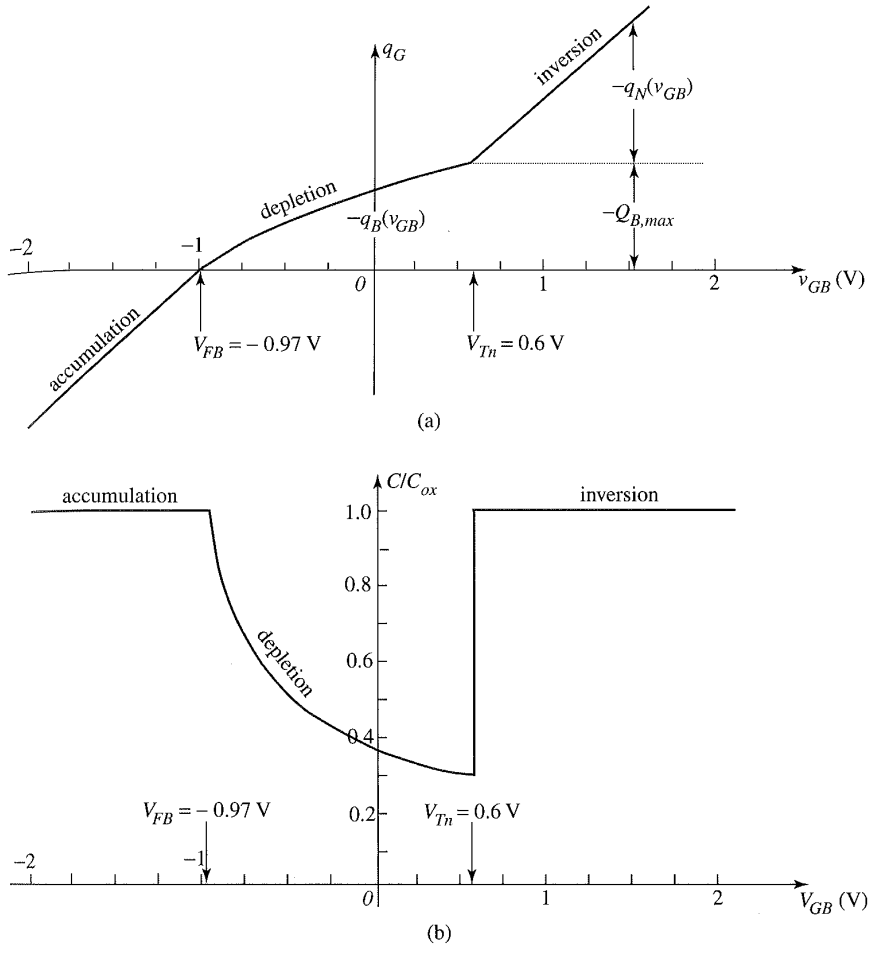
➤ 3.9 CAPACITANCE OF THE MOS STRUCTURE

In this section, we will investigate the small-signal modeling of the MOS structure. Our approach will parallel Section 3.6 for the pn junction under reverse bias. We begin by viewing the capacitance as a graphical derivative of the charge-voltage characteristics of the capacitor sketched in Fig. 3.25 (for p-type substrate) and Fig. 3.27 (for n-type substrate). By differentiating the expressions for gate charge as a function of gate-bulk bias, we can find mathematical results for the MOS capacitance. Finally, we can gain insight into the physics of the MOS capacitor by determining where the small-signal increment in gate charge is mirrored in the underlying silicon substrate, for the three regions of gate-bulk bias.

3.9.1 Graphical Interpretation

Compared to the pn junction, the MOS structure has a much more complicated charge-storage characteristic. For case of a p-type substrate, we can apply the definition of the capacitance as the slope of $q_G(V_{GB})$ to Fig. 3.25 (repeated for convenience in Fig. 3.35(a)) to derive the plot of $C(V_{GB})$ in Fig. 3.35(b). We have used symbols for the total variables (DC + small-signal) in Fig. 3.35(a), since we are assuming that our analysis in Section 3.7 and Section 3.8 can be extended to time-varying signals.

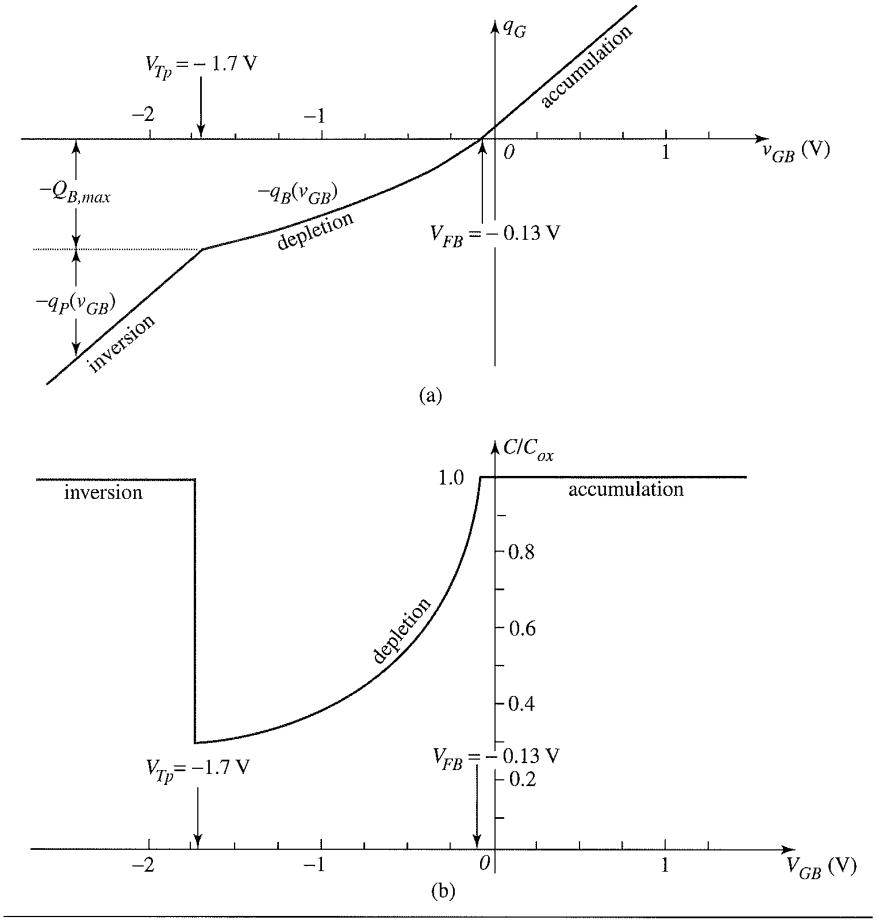
By definition, the capacitance of the MOS structure is the slope of the $q_G(v_{GB})$ curve in Fig. 3.35, (a) evaluated at a particular DC bias voltage V_{GB}. In accumulation and inversion, the gate charge has a term that is proportional to v_{GB}, with a slope equal to C_{ox}. Therefore, the capacitance is independent of the DC bias and is equal



➤ **Figure 3.35** (a) Gate charge as a function of gate-bulk voltage for an MOS capacitor on a p-type substrate with a 150 Å-thick gate oxide and a substrate doping $N_a = 10^{17}$ cm^{-3} and (b) capacitance as a function of gate-bulk voltage, found by graphically differentiating (a).

to the oxide capacitance $C_{ox} = \varepsilon_{ox}/t_{ox}$. In Fig. 3.35(b), the capacitance is plotted normalized to C_{ox}. Later in this section, we will explain the reason for the variation of the MOS capacitance with gate bias in the depletion region. The abrupt jump in capacitance at the threshold voltage is a convenient way to measure V_{Tn}.

For an MOS capacitor on an n-type substrate, the gate charge versus gate-bulk voltage is plotted in Fig. 3.27, which is repeated as Fig. 3.36(a). Graphical differentiation yields the capacitance as a function of the DC gate-bulk voltage V_{GB} in Fig. 3.36(b). Note that substrate type can be determined from the capacitance plots in Figs. 3.35(b) and 3.36(b).



➤ **Figure 3.36** (a) Gate charge as a function of gate-bulk voltage for an MOS capacitor on an n-type substrate, with a 150 Å-thick gate oxide and a substrate doping $N_d = 10^{17}$ cm^{-3} and (b) capacitance as a function of gate-bulk voltage, found by graphically differentiating (a).

3.9.2 Mathematical Formulation

The expressions for gate charge as a function of gate-bulk bias in Eqs. (3.150)–(3.152) can be differentiated to find the capacitance in accumulation, depletion, and inversion. The capacitance of the metal-oxide-semiconductor structure is defined as

$$C = \left.\frac{dq_G}{dv_{GB}}\right|_{V_{GB}} \tag{3.153}$$

In accumulation, the MOS capacitance for a p-type substrate is

$$C = \frac{d}{dv_{GB}}[C_{ox}(v_{GB} - V_{FB})]\Big|_{V_{GB} \le V_{FB}} = C_{ox} \text{ for } V_{GB} \le V_{FB} \tag{3.154}$$

For the case of inversion, the MOS capacitance for a p-type substrate is the same

$$C = \frac{d}{dv_{GB}}[(C_{ox}(v_{GB} - V_{Tn}) - Q_{B,\max})]\Big|_{V_{GB} \ge V_{Tn}} = C_{ox} \tag{3.155}$$

since the bulk charge is fixed at $Q_{B,\max}$ in inversion. The situation in depletion requires differentiation of Eq. (3.151), which is valid for gate-bulk biases $V_{FB} \le V_{GB} \le V_{Tn}$

$$C = \frac{dq_G}{dv_{GB}}\Big|_{V_{FB} \le V_{GB} \le V_{Tn}} = \frac{C_{ox}}{\sqrt{1 + \dfrac{2C_{ox}^2(V_{GB} - V_{FB})}{q\varepsilon_s N_a}}} \tag{3.156}$$

This result allows the capacitance versus gate-bulk bias to be quantified in the depletion region. The minimum capacitance occurs just prior to inversion, which we can evaluate by substituting $V_{GB} = V_{Tn}$ into Eq. (3.156)

$$C(V_{GB} = V_{Tn}^-) = \frac{C_{ox}}{\sqrt{1 + \dfrac{2C_{ox}^2(V_{Tn} - V_{FB})}{q\varepsilon_s N_a}}} \tag{3.157}$$

In order to interpret the meaning Eq. (3.156), we will express it in terms of the depletion width $X_d(V_{GB})$ from Eq. (3.130), which we repeat here for convenience

$$X_d(V_{GB}) = \left(\frac{\varepsilon_s}{C_{ox}}\right)\left(\sqrt{1 + \frac{2C_{ox}^2(V_{GB} - V_{FB})}{q\varepsilon_s N_a}} - 1\right) \tag{3.158}$$

After some algebra, the capacitance in the depletion region can be written as

$$C = \frac{C_{ox}}{\left(\dfrac{C_{ox}}{\varepsilon_s}\right)X_d(V_{GB}) + 1} = \frac{C_{ox}\left(\dfrac{\varepsilon_s}{X_d(V_{GB})}\right)}{C_{ox} + \left(\dfrac{\varepsilon_s}{X_d(V_{GB})}\right)} \text{ for } V_{FB} \le V_{GB} \le V_{Tn} \tag{3.159}$$

From Section 3.6, we can identify the capacitance C_b of the depletion region as

$$C_b = \frac{\varepsilon_s}{X_d(V_{GB})} \tag{3.160}$$

which enables Eq. (3.159) to be rewritten as the capacitances C_{ox} and C_b in series

$$C = \frac{C_{ox}C_b}{C_{ox}+C_b} \text{ for } V_{FB} \leq V_{GB} \leq V_{Tn} \tag{3.161}$$

Later in this section, we will see why the oxide and depletion capacitances are in series combination when the MOS capacitor is biased in depletion.

EXAMPLE 3.10 MOS Capacitance-Voltage Curves

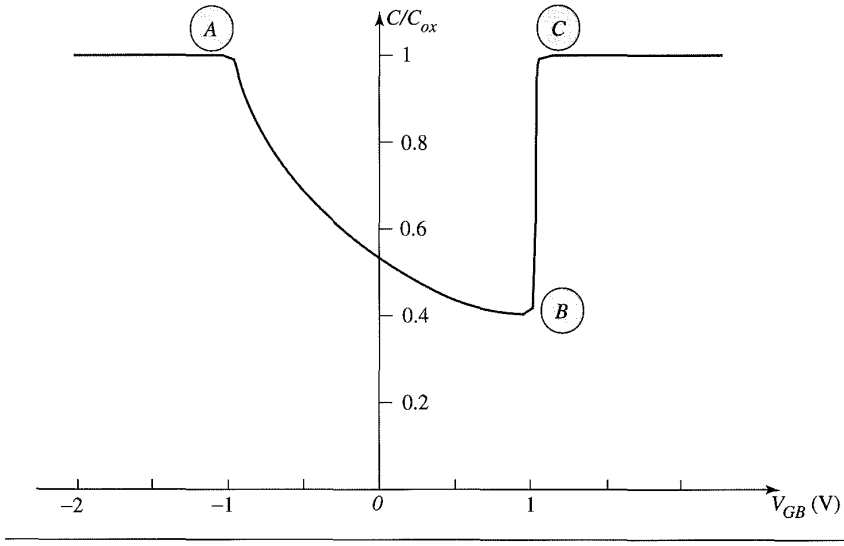
Shown below is a measured *"C-V"* curve for an MOS capacitor. The gate is n^+ polysilicon with $\phi_{n+} = 550$ mV.

(a) From the normalized capacitance-voltage characteristic, determine the substrate type and doping level.

SOLUTION

To make sense of the *C-V* curve, we must determine whether point *A* or *C* corresponds to flatband. Going from point *A* to point *B*, the capacitance decreases with increasing voltage, which corresponds to a widening depletion region. Also, the transition from *B* to *C* is much more abrupt than the transition from *A* to *B*. These observations are consistent with *A* being the flatband condition and *C* being inversion. From the characteristic, we estimate that

$V_{FB} = -1\text{V}$ $\quad$ $V_T = 1\text{V}$



➤ **Figure Ex3.10** Measured capacitance-voltage curve for an MOS structure with the capacitance normalized to its maximum value—the oxide capacitance.

Using the definition of the flatband voltage

$$V_{FB} = -1\text{V} = -(0.55\text{V} - \phi_{sub}) \rightarrow \phi_{sub} = -450\,\text{mV}$$

where ϕ_{sub} is the substrate potential. Since ϕ_{sub} is negative, the substrate is p-type. To find the acceptor doping level in the substrate, we apply the 60mV rule.

$$\phi_p = (-60\text{mV})\log\left(\frac{N_a}{1\times10^{10}\,\text{cm}^{-3}}\right) = -450\,\text{mV} \rightarrow N_a = 3.16\times10^{17}\,\text{cm}^{-3}$$

(b) Determine the oxide thickness t_{ox}.

SOLUTION

To find the oxide thickness t_{ox}, we must first find the oxide capacitance. Since the capacitance has been normalized, it is not possible to use the measurements in accumulation or in inversion to find C_{ox} directly. Instead, we use the depletion region curve where the capacitance is the series combination of C_{ox} and the capacitance of the depletion region C_b. From Eq. (3.157), the minimum normalized capacitance is given by

$$C_{\min}/C_{ox} = C(V_{GB} = V_{Tn}^-)/C_{ox} = \frac{1}{\sqrt{1+\dfrac{2C_{ox}^2(V_{Tn}-V_{FB})}{q\varepsilon_s N_a}}}$$

Solving for the oxide capacitance, we find that

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \sqrt{\frac{q\varepsilon_s N_a}{2(V_{Tn}-V_{FB})}\left[\left(\frac{C_{ox}}{C_{\min}}\right)^2 - 1\right]}$$

Substituting $C_{\min}/C_{ox} = 0.4$ and the other parameters from part (a) into the expression for the oxide thickness, we find that

$$t_{ox} = \frac{\varepsilon_{ox}}{\sqrt{\dfrac{q\varepsilon_s N_a}{2(V_{Tn}-V_{FB})}\left[\left(\dfrac{C_{ox}}{C_{\min}}\right)^2 - 1\right]}}$$

$$t_{ox} = \frac{(3.9)(8.85\times10^{-14})}{\sqrt{\dfrac{1.6\times10^{19}(1.035\times10^{-12})(3.16\times10^{17})}{2(1-(-1))}\left(\left(\dfrac{1}{0.4}\right)^2 - 1\right)}} = 132\,\text{Å}$$

An alternative approach is to use the threshold voltage to calculate C_{ox} from Eq. (3.95). Using this approach, we find that $t_{ox} = 125$ Å which is about 5% lower. Uncertainties in reading the threshold voltage from the *C-V* curve and limitations of the delta-depletion approximation account for the discrepancy.

(c) From your results from part (a) and (b), find the gate charge Q_G for $V_{GB} = -3$ V and for $V_{GB} = +3$ V

SOLUTION

For $V_{GB} = -3\text{ V} < V_{FB}$, the MOS capacitor is in accumulation. The gate charge is

$$Q_G = C_{ox}(V_{GB} - V_{FB}) = 2.55 \times 10^{-7}\text{F/cm}^2(-3\text{V} - (-1\text{V})) = -5.1 \times 10^{-7}\text{ C/cm}^2$$

For $V_{GB} = 3\text{ V} > V_{Tn}$, the MOS capacitor is in inversion. The gate charge is equal and opposite to the sum of the magnitudes of the inversion electron charge and the magnitude of the maximum depletion charge. The latter is found from Eq. (3.140) and Eq. (3.142)

$$Q_G = C_{ox}(V_{GB} - V_{Tn}) - Q_{B,\max} = 5.1 \times 10^{-7} + 2.8 \times 10^{-7}\text{ C/cm}^2 = 7.9 \times 10^{-7}\text{ C/cm}^2$$

3.9.3 Physical Interpretation

In order to make physical sense of this rather complicated capacitance-voltage curve, we will investigate where the increment in gate charge q_g is mirrored for each of the three different regions of operation: accumulation, depletion, and inversion. Starting with a MOS capacitor biased in accumulation, Fig. 3.37 illustrates the charge density before and after a positive increment q_g is added to the DC gate charge Q_G. The increment in gate charge is due to a positive increment in the gate voltage v_{gb}. In Fig. 3.37(a), we note that $Q_G < 0$ for accumulation. The charge increment on the gate is mirrored in the accumulated holes under the gate oxide as shown in the small-signal charge density plotted in Fig. 3.37(c). The incremental charge distribution is identical to that of the parallel-plate capacitor in Ex. 3.1. Therefore, the capacitance of the MOS structure in accumulation is C_{ox}, the oxide capacitance.

Figure 3.38 shows the more complicated situation for a MOS capacitor biased in depletion. In this case, the gate charge is positive and the gate-bulk bias voltage is in the range $V_{FB} \leq V_{GB} \leq V_{Tn}$. The gate charge increment is mirrored at the bottom of the depletion region as shown in Fig. 3.38(c). The incremental charges $+q_g$ and $-q_g$ are separated by an oxide layer of thickness t_{ox} with capacitance $C_{ox} = \varepsilon_{ox}/t_{ox}$, and a depletion region of thickness $X_d(V_{GB})$ with capacitance $C_b = \varepsilon_s/X_d(V_{GB})$. The equivalent circuit is the series combination of C_{ox} and C_b as shown in Fig. 3.38(c). Therefore, the MOS capacitance when biased in depletion is

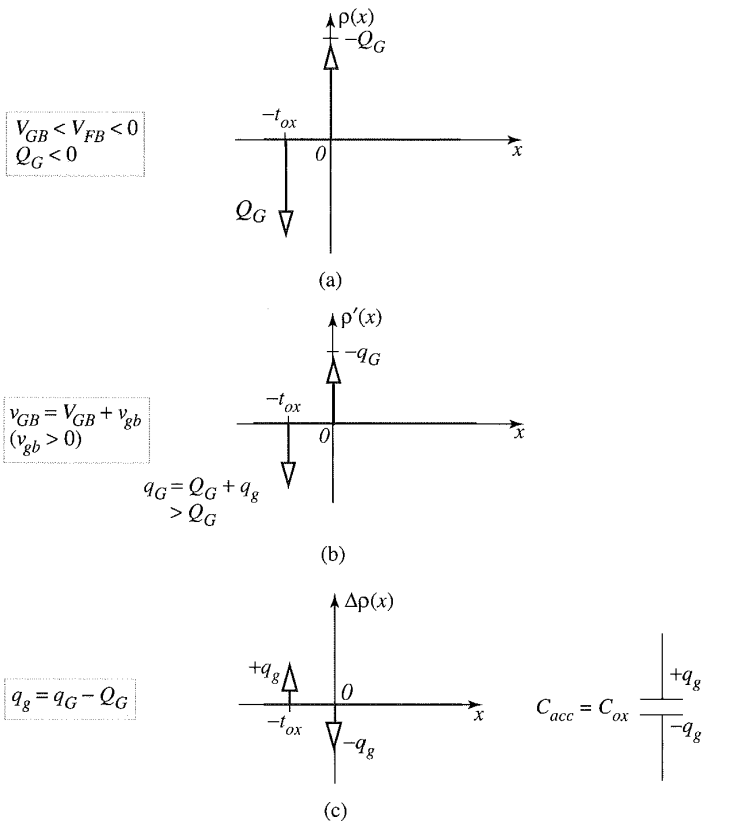
$$C = \frac{C_{ox}C_b}{C_{ox} + C_b} \text{ for } V_{FB} \leq V_{GB} \leq V_{Tn}. \tag{3.162}$$

which explains the physical basis for Eq. (3.161).

For an inverted MOS capacitor, an increment in gate voltage is mirrored in an increase electron charge in the inversion layer as shown in Fig. 3.39. The depletion charge is fixed and does not vary with gate voltage as is reflected in the constant depletion width $X_{d,\max}$ in inversion. As a result, the capacitance in inversion is again the oxide capacitance

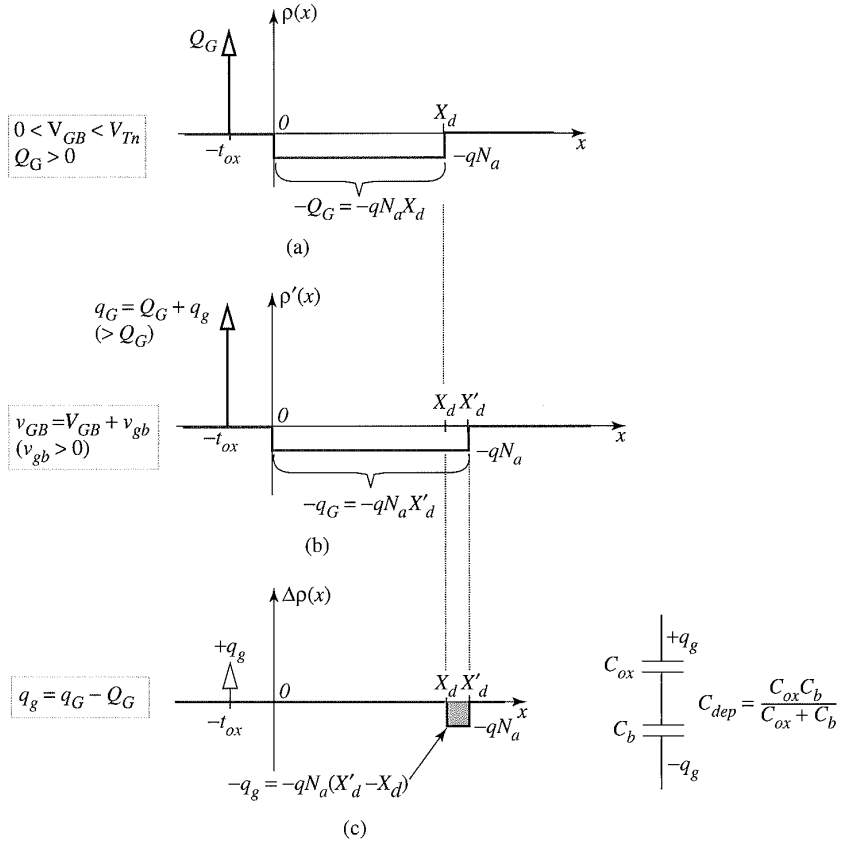
$$C = C_{ox} \text{ for } V_{GB} \geq V_{Tn} \tag{3.163}$$

Our analysis of the MOS structure has not considered the time variation of voltages or charges. In fact, our use of the logarithmic relationship between potential and carrier concentration for finding the carrier concentrations in the silicon substrate means we have implicitly assumed that sufficient time has elapsed for transient currents to die out. Measurement of the capacitance-voltage curves shown in this chapter requires use of a slow ramp of the bias voltage V_{GB} and a low-frequency AC

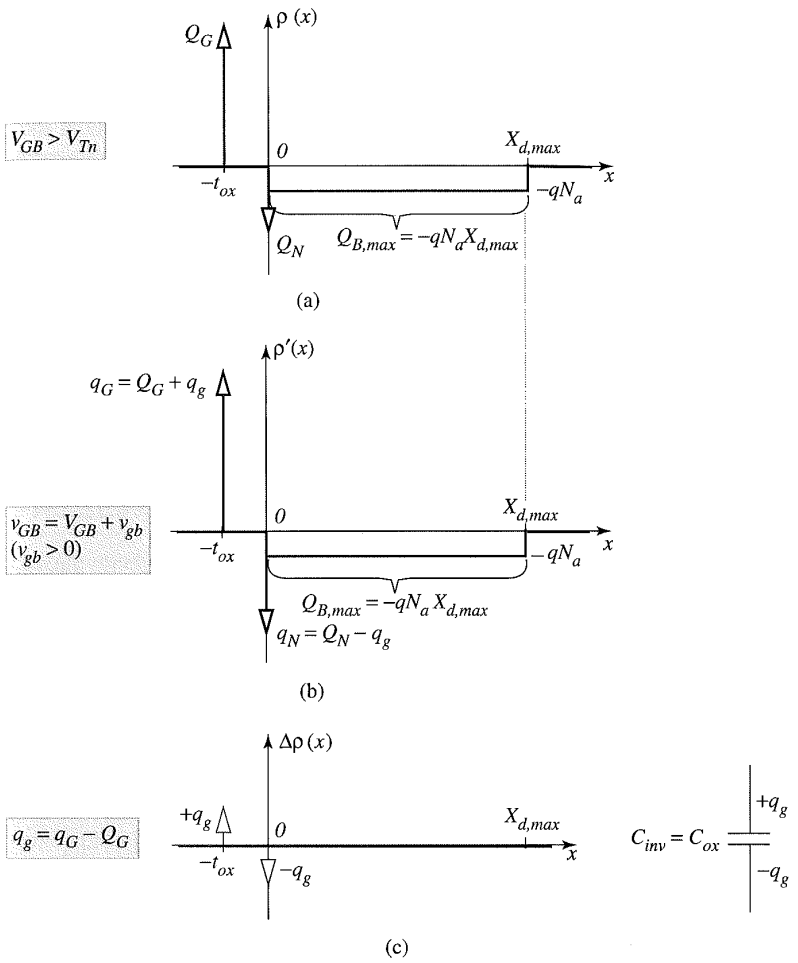


➤ **Figure 3.37** Charge density ρ(x) for a MOS structure on a p-type substrate in accumulation: (a) bias voltage $V_{GB} < V_{FB}$, (b) perturbed charge density $\rho'(x)$ for $v_{GB} = V_{GB} + v_{gb}$, where the incremental voltage $v_{gb} > 0$, and (c) incremental charge density $\Delta\rho(x) = \rho'(x) - \rho(x)$ showing that the incremental charges $\pm q_g$ are separated by the gate oxide. The small-signal model in accumulation is the oxide capacitance.

incremental voltage $v_{gb}(t)$. We will leave further discussion of the various ways to measure MOS *C-V* curves to more advanced texts in device physics or IC fabrication technology.



➤ **Figure 3.38** Charge density ρ(*x*) for a MOS structure on a p-type substrate in depletion: (a) bias voltage $V_{FB} < V_{GB} < V_{Tn}$, (b) perturbed charge density ρ′(*x*) for $v_{GB} = V_{GB} + v_{gb}$, where the incremental voltage $v_{gb} > 0$, and (c) incremental charge density $\Delta\rho(x) = \rho'(x) - \rho(x)$ showing that the incremental charges $\pm q_g$ are separated by the gate oxide and the depletion region. The small-signal model in depletion is two capacitors in series.



➤ **Figure 3.39** Charge density $\rho(x)$ for a MOS capacitor on a p-type substrate in inversion: (a) bias voltage $V_{GB} > V_{Tn}$, (b) perturbed charge density $\rho'(x)$ for $v_{GB} = V_{GB} + v_{gb}$, where the incremental voltage $v_{gb} > 0$, and (c) incremental charge density $\Delta\rho(x) = \rho'(x) - \rho(x)$ showing that the incremental gate charge is mirrored by an increment in the electron inversion charge. Therefore, the small-signal model in the inversion region is again the oxide capacitance.

➤ SUMMARY

This chapter began with a review of basic electrostatic theory and its application to one-dimensional device structures. These concepts were applied to the pn junction, first in thermal equilibrium and then under a reverse bias $V_D < 0$. The charge density in this structure was modeled using the depletion approximation, in which the electron and hole concentrations in the transition region at the junction were assumed to be negligible. In order to pin down the potential in silicon in thermal equilibrium, we selected a potential reference such that intrinsic silicon is at zero volts. The relationship of potential to the logarithm of carrier concentration followed from this reference, with n-type silicon having a positive potential and p-type silicon a negative potential. The very important concept of small-signal modeling was introduced by finding the capacitance of the reverse-biased pn junction.

The MOS capacitor was the second important device structure analyzed in this chapter. Its nonlinear charge-storage characteristics were described by finding the charge stored on the gate as a function of the applied voltage using the delta-depletion approximation. An unusual and very useful feature of the MOS structure is the possibility of inverting the type of the substrate, which means creating a surface layer that is of opposite type to the substrate. Finally, the small-signal capacitance of the MOS capacitor was found as a function of the applied voltage.

The following important concepts and techniques should have been mastered in Chapter 3.

- Interrelationships between charge density, electric field, and electrostatic potential in one dimension from Gauss's Law and the definition of potential.
- One-dimensional boundary conditions on electric field and potential.
- Techniques for graphically sketching the charge density, electric field, and potential without doing the formal mathematics.
- Equilibrium in the pn junction as a balance of drift and diffusion current densities.
- The logarithmic relationship between potential and electron and hole concentration in thermal equilibrium.
- Charge density in pn junctions using the depletion approximation.
- Electrostatics of the pn junction under reverse bias.
- Contact potentials and the use of Kirchhoff's voltage law to show the current is zero in a short-circuited pn junction in equilibrium.
- Capacitance of the pn junction and its interpretation from the small-signal model of charge storage.
- The MOS structure and its electrostatics in thermal equilibrium.
- Accumulation, depletion, and inversion of the MOS capacitor and the corresponding charge densities
- The delta-depletion approximation and the quantitative analysis of MOS electrostatics
- Capacitance of the MOS structure in accumulation, depletion, and inversion.

FURTHER READING

Reverse-Biased pn Junction

1. G. N. Neudeck, *The PN Junction Diode*, Modular Series on Solid State Devices, 2nd ed. Vol. II, G. W. Neudeck and R. F. Pierret, eds., Addison-Wesley, 1989, Chapter 2. About the same level.
2. R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed., Wiley, 1986, Chapter 4. More advanced discussion of the depletion approximation; includes Fermi level concept.

MOS Capacitor

1. C. G. Fonstad, *Microelectronic Devices and Circuits*. McGraw-Hill, 1994, Chapter 9. About the same level; includes a discussion of the effects of oxide charge.
2. R. F. Pierret, *Field Effect Devices*, Modular Series on Solid State Devices, Vol. IV, G. W. Neudeck and R. F. Pierret, eds., Addison-Wesley, 1983, Chapters 2–4. More advanced discussion, including the concept of the Fermi level.

PROBLEMS

EXERCISES

Permittivities: crystalline silicon: $\varepsilon_s = 1.035 \times 10^{-12}$ F/cm
silicon dioxide (SiO_2): $\varepsilon_{ox} = 3.45 \times 10^{-13}$ F/cm
silicon nitride (Si_3N_4): $\varepsilon_n = 8.63 \times 10^{-13}$ F/cm

E3.1 The charge distribution in a sample of silicon is given in Fig. E3.1.

(a) Determine the sign of the electric field at $x = -750$ nm and $x = 250$ nm.

(b) Determine the numerical value of the electric field at $x = -250$ nm.

(c) Plot the electric field as a function of x.

E3.2 The charge distribution in an IC structure is shown in Fig. E3.2.

(a) Find the numerical value of the surface charges Q so that the total charge is zero in the structure. Note that the units of Q are C/cm^2.

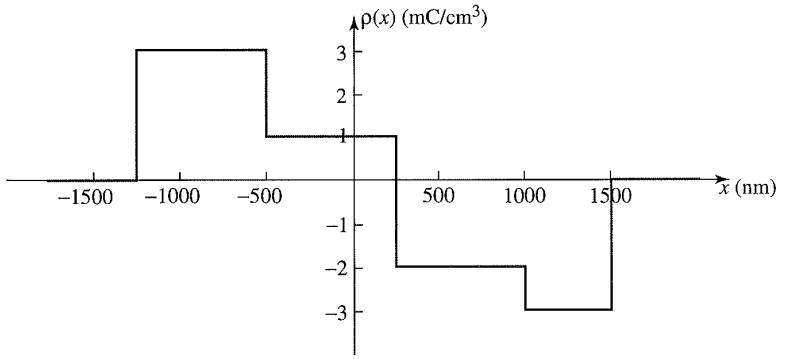
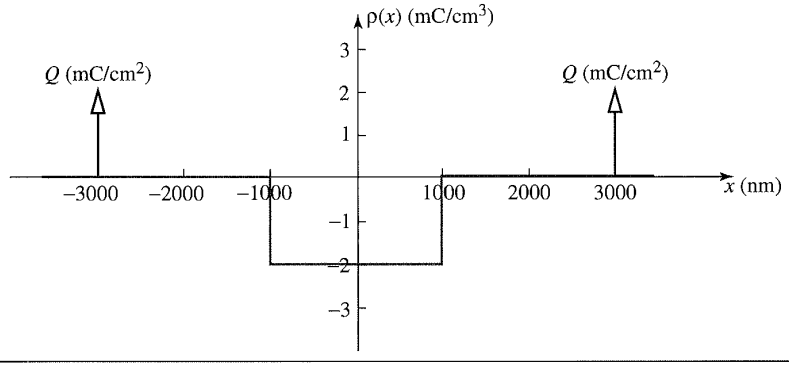


Figure E3.1



➤ **Figure E3.2**

(b) Plot the electric field $E(x)$ through the structure.

(c) Sketch the potential $\phi(x)$, given that the potential $\phi(\pm 3000) = 0$.

E3.3 The charge distribution in a silicon region is plotted in Fig. E3.3.

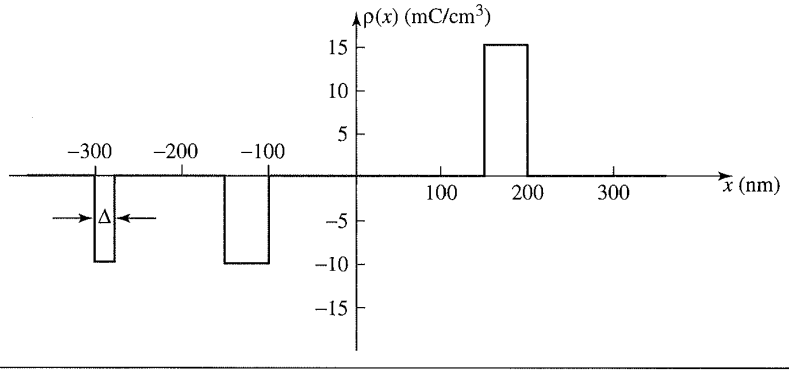
(a) Find the width Δ so that the silicon region is electrically neutral. The boundary at $x = -300$ nm is fixed.

(b) Find the numerical value of the electric field at $x = -250$ nm and at $x = 150$ nm

(c) Plot the electric field $E(x)$

(d) If the potential $\phi(-300 \text{ nm}) = -500$ mV, find the potential at $x = 350$ nm.

E3.4 The charge distribution in a silicon region is given in Fig. E3.4. There is no need to find $E(x)$ to solve (a)–(c).



➤ **Figure E3.3**

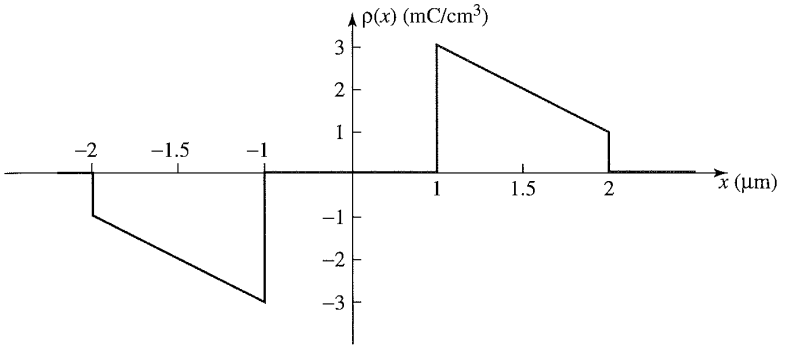
(a) Find the numerical value of the electric field E at $x = -1.5$ μm.

(b) Find the numerical value of the electric field E at $x = 0$.

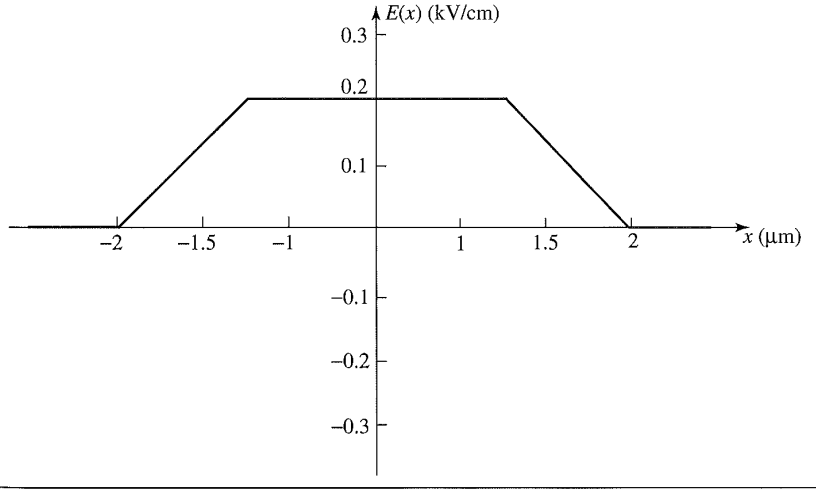
(c) Find the numerical value of the electric field E at $x = 1.5$ μm.

E3.5 The electric field in a silicon structure is given in Fig. E3.5.

(a) Plot the charge density $\rho(x)$.



➤ **Figure E3.4**



➤ **Figure E3.5**

(b) If the potential $\phi(-2\ \mu m) = -500$ mV, what is the potential at $x = 0$? What is the potential at $x = 2\ \mu m$. Note that you don't need to solve for the potential function in order to answer this question.

(c) Sketch the potential $\phi(x)$, given that the potential $\phi(-2\ \mu m) = -500$ mV.

E3.6 In the plot of the electric field $E(x)$ in Fig. E3.6, the materials consist of a metal film ($x < t_i$), an insulator with unknown permittivity ($t_i < x < 0$) and silicon ($x \geq 0$).

(a) Given that there is no sheet charge located at $x = 0$, what is the numerical value of the permittivity of the insulating film?

(b) Sketch the charge density $\rho(x)$ through the structure.

(c) How much charge [C/cm^2] is on the metal film at $x = t_i$?

(d) Given that the potential of the metal is $\phi_m = 350$ mV, sketch $\phi(x)$.

E3.7 The electric field in a metal-SiO_2-silicon sandwich is shown in Fig. E3.7. The metal is located at $x < -100$ Å, the oxide film is located between -100 Å $< x < 0$, and the silicon substrate ranges over $x > 0$. There is a sheet charge at $x = 0$ with units of C/cm^2.

(a) From the plot of electric field in Fig. E3.7, determine the magnitude and sign of the sheet charge at $x = 0$.

(b) What is the charge on the metal film at $x = -100$ Å?

(c) Sketch the potential $\phi(x)$, given that $\phi(-100\ Å) = -850$ mV.

E3.8 The electric field $E(x)$ in a silicon region is given in Fig. E3.8.

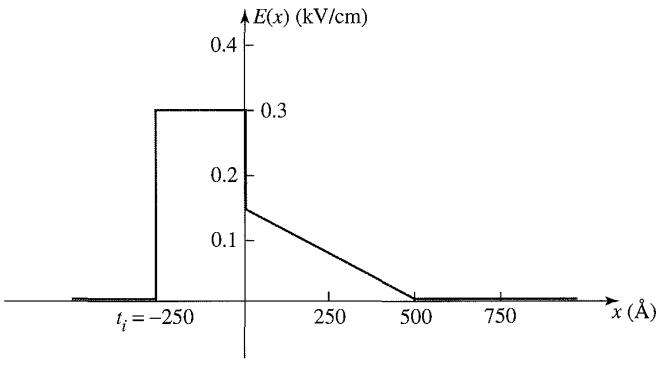
(a) Plot the charge density $\rho(x)$.

(b) Given that the potential $\phi(0) = 0$, sketch the potential $\phi(x)$.

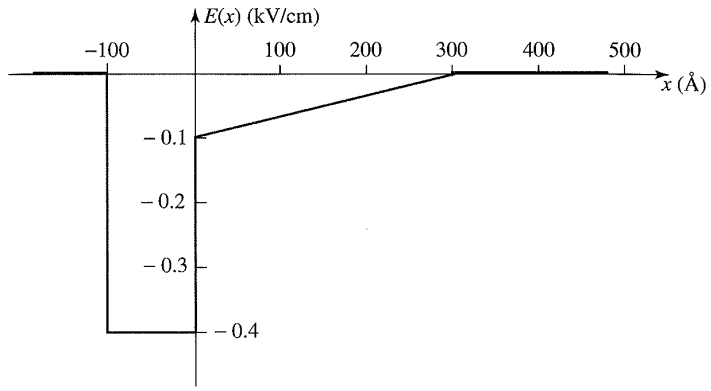
E3.9 The potential $E(x)$ through a silicon sample is graphed in Fig. E3.9.

(a) Sketch the potential $\phi(x)$, given that $\phi(0) = 0.5$ V.

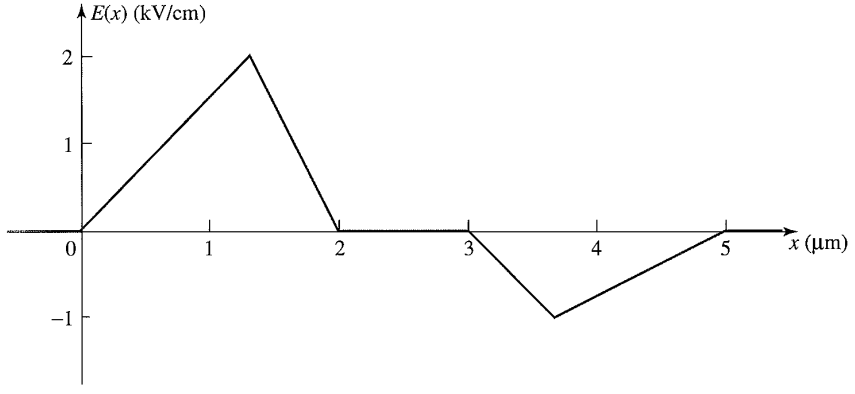
(b) Sketch the charge density $\rho(x)$.



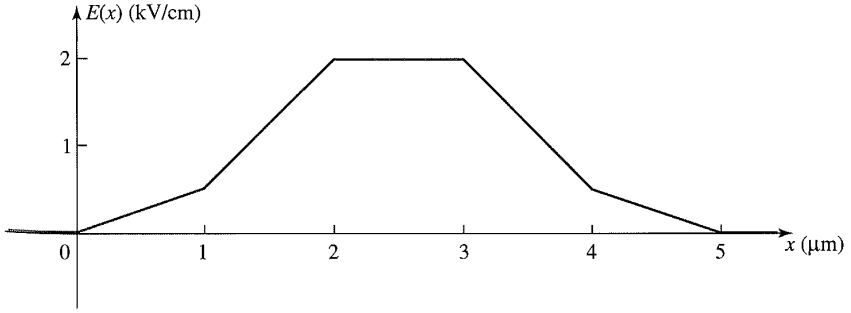
➤ **Figure E3.6**



➤ **Figure E3.7**



➤ **Figure E3.8**



➤ **Figure E3.9**

E3.10 This exercise investigates the depletion approximation, for a pn junction diode with symmetrical doping $N_a = N_d = 10^{15}$ cm^{-3}.

(a) Find the thermal equilibrium potential $\phi_o(x)$ according to the depletion approximation.

(b) Plot on a linear scale the electron and hole concentrations through the depletion region, in thermal equilibrium.

(c) Plot on a linear scale the charge density $\rho(x)$ in the depletion region, including the contributions of the electrons and holes from your answer in part (b). Estimate the percentage error in the charge on one side of the junction, if we neglect the electron and hole charge.

E3.11 Consider a pn junction diode with symmetrical doping $N_a = N_d = 10^{17}$ cm^{-3}.

(a) What is the magnitude of the electric field $|E_o(x = 0)|$ (where the metallurgical junction is located at $x = 0$) in thermal equilibrium?

(b) For what value of bias voltage V_D is the electric field at $x = 0$ increased by a factor of 3?

(c) Plot $|E(0)|$ as a function of V_D for -7.5 V $< V_D < 0$.

E3.12 Consider a pn junction with p-side doping $N_a = 10^{19}$ cm^{-3} and n-side doping $N_d = 10^{17}$ cm^{-3}.

(a) Sketch the potential in thermal equilibrium, using the approximation that the junction is "one-sided" as discussed in Ex. 3.5.

(b) What is the electric field $|E_o(x = 0)|$ (where the metallurgical junction is located at $x = 0$) in thermal equilibrium?

(c) If the maximum electric field allowable in the junction is $E_{max} = 10^4$ V/cm what is the maximum reverse bias permitted?

E3.13 Low doped silicon is used in particle detectors, in order to obtain a wide depletion region. Consider a pn junction with p-side doping $N_a = 5 \times 10^{12}$ cm^{-3} and n-side doping $N_d = 10^{17}$ cm^{-3}.

(a) What is the thermal equilibrium depletion region width?

(b) What is the reverse bias needed to increase the depletion width to 50 μm?

(c) If the maximum electric field allowable in the junction is $E_{max} = 5 \times 10^5$ V/cm what is the maximum depletion width? What is the bias voltage corresponding to the maximum depletion width?

E3.14 We would like to investigate the charge distribution in a pn junction with doping levels $N_a = 10^{16}$ cm^{-3} and $N_d = 5 \times 10^{16}$ cm^{-3}.

(a) For $V_D = -2.5$ V, plot the charge density $\rho(x)$.

(b) We add to V_D a time-varying voltage $v_d(t) = (250 \text{ mV})\cos(\omega t)$. Plot the charge density when $v_D(t) = V_D + v_d(t)$ is at its maximum ($\rho_{max}(x)$) and its minimum ($\rho_{min}(x)$).

(c) Plot the difference between $\rho_{max}(x)$ and $\rho(x)$ and the difference between $\rho_{min}(x)$ and $\rho(x)$ on the same graph. Find the small-signal charge per unit area q_j in units C/cm^2 for each case. Are they equal and opposite? Why or why not?

(d) Find the depletion capacitance $C(V_D)$ for $V_D = -2.5$ V and the small-signal depletion charge q_j, assuming that $v_d(t)$ can be considered a small signal. Compare with your answer to (c).

E3.15 In order to evaluate the junction diode as a circuit element, we would like to know the capacitance per unit area and the tuning range.

(a) Plot the capacitance per unit area in thermal equilibrium versus the p-side doping concentration from $N_a = 10^{16}$ cm^{-3} to 10^{17} cm^{-3}. Assume that the n-side doping concentration is $N_d = 10^{19}$ cm^{-3}.

(b) If the maximum electric field allowable in the junction is $E_{max} = 10^4$ V/cm, plot the maximum reverse bias $V_{D,max}$ on the capacitor as a function of the p-side doping concentration.

(c) From your results in part (b), plot the ratio of $C(0)/C(V_{D,max})$ as a function of p-side doping concentration.

E3.16 We are interested in using a reverse-biased pn junction as a small-signal capacitor. The doping levels are $N_a = 10^{16}$ cm^{-3} and $N_d = 10^{18}$ cm^{-3}. We would like to have a capacitor with $C = 500$ fF ± 150 fF.

(a) What is the capacitance per unit area [units: fF/μm^2] for $V_D = 0$ V.

(b) What is the area needed for the capacitor and the voltage range needed for tuning? Note that $V_D = 0$ V must correspond to $C = 650$ fF, the maximum capacitance.

E3.17 An MOS capacitor is fabricated on a p-type silicon substrate with doping concentration $N_a = 10^{16}$ cm^{-3}, using an n$^+$ polysilicon gate.

(a) What is the flatband voltage, V_{FB}?

(b) What is the surface potential in inversion?

(c) What is the maximum width of the depletion region?

(d) For $t_{ox} = 200$ Å, what is the threshold voltage V_{Tn}?

E3.18 An MOS capacitor is fabricated on an n-type silicon substrate with doping concentration $N_d = 5 \times 10^{15}$ cm^{-3}, using an n$^+$ polysilicon gate.

(a) What is the flatband voltage, V_{FB}?

(b) What is the surface potential in inversion?

(c) What is the maximum width of the depletion region?

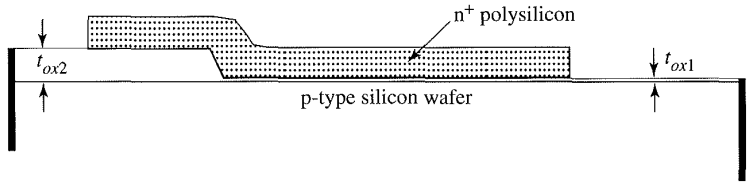
(d) For $t_{ox} = 150$ Å, what is the threshold voltage V_{Tp}?

E3.19 In an IC containing MOS structures, the n$^+$ polysilicon gate is often used as an interconnect layer. In the cross section in Fig. E3.19, the polysilicon layer forms a parasitic MOS capacitor over the "field" region in between devices. The oxide thickness $t_{ox1} = 200$ Å and the substrate doping is $N_a = 5 \times 10^{16}$ cm^{-3}.

(a) Find the threshold voltage V_{Tn1} of the MOS capacitor formed over the thin oxide of thickness t_{ox1}.

(b) What thickness t_{ox2} of field oxide is required if the threshold voltage V_{Tn2} of this MOS capacitor must be a factor of 5 higher than V_{Tn1}?

(c) If the substrate doping under the field regions is increased by a factor of 10 and the doping in the thin-oxide region kept at 5×10^{16} cm^{-3}, what is the thickness t_{ox2} such that $V_{Tn2} = 5\ V_{Tn1}$?



➤ **Figure E3.19**

E3.20 It is occasionally useful to design a MOS capacitor with a threshold voltage $V_T = 0$ V. The oxide thickness is $t_{ox} = 250$ Å and we use an n^+ polysilicon gate.

(a) What p-type substrate doping concentration is required for $V_{Tn} = 0$ V? An iterative approach may be helpful, since the substrate potential ϕ_p is a weak function of the doping concentration.

(b) Sketch the potential $\phi_0(x)$ in thermal equilibrium.

E3.21 This exercise concerns an MOS capacitor with an n^+ polysilicon gate, a 500 Å-thick gate oxide, and a p-type substrate with acceptor concentration $N_a = 7.5 \times 10^{15}$ cm^{-3}.

(a) If the gate charge is $Q_G = -1 \times 10^{-7}$ C/cm^2, sketch the charge density through the MOS capacitor.

(b) For the gate charge in part (a), what is the gate-bulk bias V_{GB}?

(c) If the gate charge is $Q_G = +2.5 \times 10^{-7}$ C/cm^2, sketch the charge density through the MOS capacitor.

(d) For the gate charge in part (c), what is the inversion charge Q_N in C/cm^2?

(e) For the gate charge in part (c), what is the gate-bulk bias V_{GB}?

E3.22 For this exercise, the MOS capacitor has an n^+ polysilicon gate, a 450 Å-thick gate oxide, and an n-type substrate with donor concentration $N_d = 7.5 \times 10^{16}$ cm^{-3}.

(a) If the gate charge is $Q_G = 2 \times 10^{-7}$ C/cm^2, sketch the charge density through the MOS capacitor.

(b) For the gate charge in part (a), what is the gate-bulk bias V_{GB}?

(c) If the gate charge is $Q_G = -5 \times 10^{-7}$ C/cm^2, sketch the charge density through the MOS capacitor.

(d) For the gate charge in part (c), what is the inversion charge Q_P in C/cm^2?

E3.23 The MOS capacitor is sometimes used for storing charge. Consider an MOS capacitor with n^+ polysilicon gate, $t_{ox} = 100$ Å, and a p-type silicon substrate with an acceptor concentration $N_a = 10^{17}$ cm^{-3}.

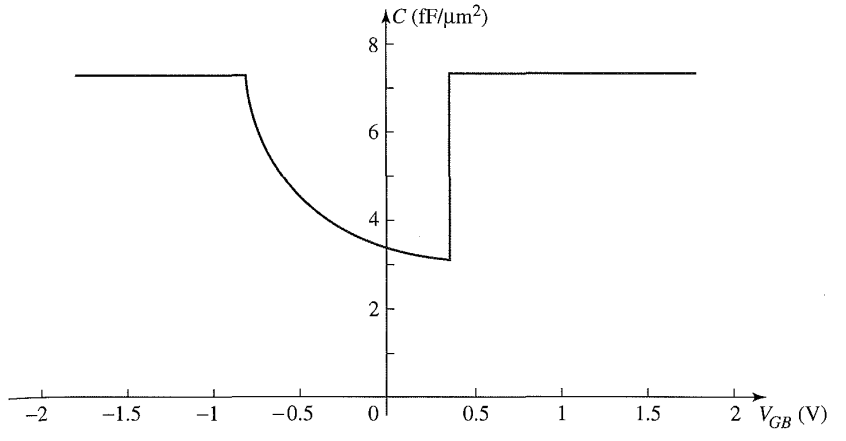
(a) What is the threshold voltage, V_{Tn}?

(b) For a capacitor with area 1 μm × 5 μm and a bias voltage $V_{GB} = 2.9$ V, what is the inversion layer electron charge, Q_N in C? What is the number of electrons stored in the capacitor?

E3.24 The capacitance-voltage curve for an MOS capacitor with an n^+ polysilicon gate is given in Fig. E3.24.

(a) What is the flatband voltage, V_{FB}? What is the threshold voltage, V_T?

(b) What is the oxide thickness, t_{ox}?



➤ **Figure E3.24**

(c) What is the substrate type?

(d) What is the substrate doping?

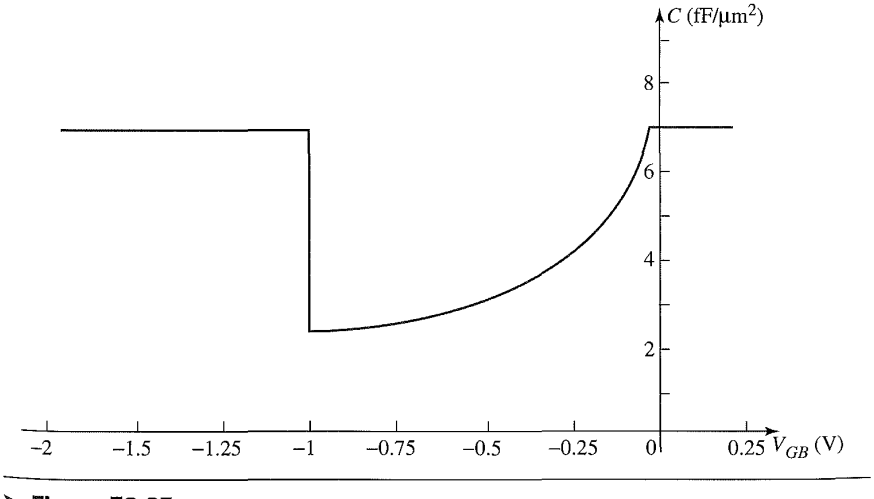
E3.25 The capacitance-voltage curve for an MOS capacitor with an n^+ polysilicon gate is given in Fig. E3.25.

(a) What is the flatband voltage, V_{FB}? What is the threshold voltage, V_T?

(b) What is the oxide thickness, t_{ox}?

(c) What is the substrate type?

(d) What is the substrate doping?



➤ **Figure E3.25**

PROBLEMS

P3.1 An aluminum-aluminum capacitor has a composite dielectric of Si_3N_4 (silicon nitride) and SiO_2 with thicknesses indicated in Fig. P3.1.

(a) For an applied voltage $V_{21} = 1$ V between the metal layers, plot the electric field $E(x)$ in the composite dielectric ($0 < x < 500$ Å).

(b) Plot the potential $\phi(x)$, with the bottom plate (metal 2) being considered the reference ($\phi(500\text{ Å}) = 0$).

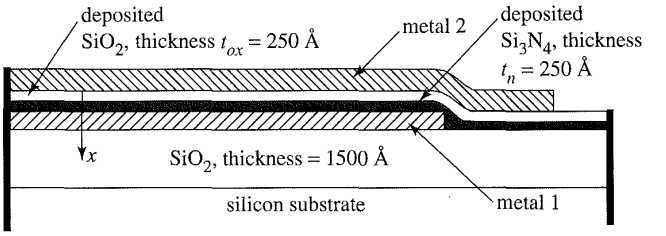
(c) What is the magnitude and sign of the sheet charge on the top plate (metal 2), for $V_{21} = 1$ V? A circuit approach to this part might be helpful.

P3.2 A region of silicon has the charge density shown in Fig. P3.2.

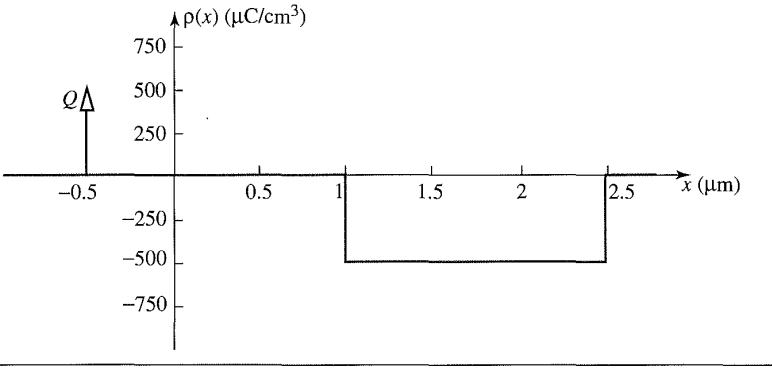
(a) What is the sheet charge Q [units: C/cm^2] needed to balance the negative charge density over the interval $1\ \mu m < x < 2.5\ \mu m$.

(b) Find the electric field at $x = 0.5\ \mu m$.

(c) Find the potential at $x = 0$, given that $\phi(2.5\ \mu m) = -0.4$ V.



➤ **Figure P3.1**



➤ **Figure P3.2**

P3.3 A region of silicon has the charge density given in Fig. P3.3. The sheet charge $Q = -2.5 \times 10^{-7}$ C/cm^2.

(a) Find the sheet charges Q_a at $x = 0$ and $x = 1.75$ μm needed to have the region of silicon be electrically neutral.

(b) Sketch the electric field $E(x)$ through the structure.

(c) Sketch the potential $\phi(x)$ given that $\phi(-1\ \mu m) = 0$.

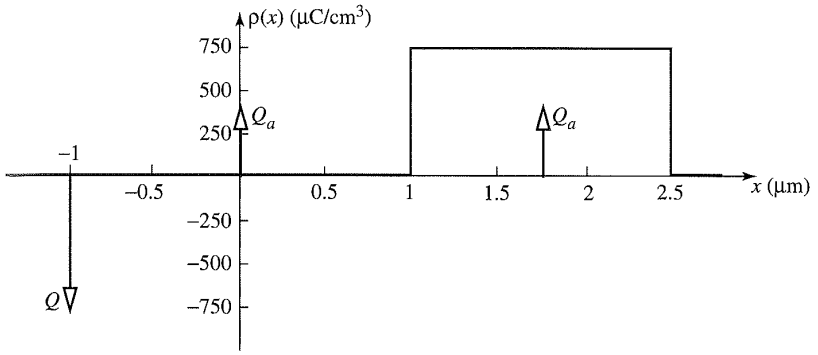
P3.4 In the aluminum-aluminum capacitor shown in cross section in Fig. P3.1, we now add a sheet charge $Q_I = 5 \times 10^{-10}$ C/cm^2 at the interface between the SiO_2 and Si_3N_4 at $x = 250$ Å.

(a) For $V_1 = V_2 = 0$, plot the electric field $E(x)$ through the dielectrics.

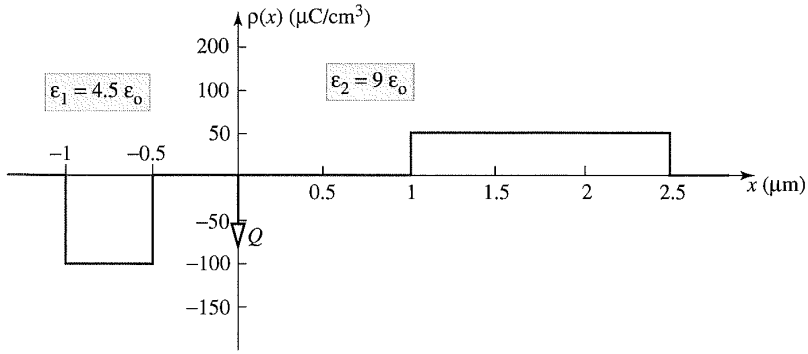
(b) Repeat part (a) for $V_1 = 1$ V and $V_2 = 0$ V.

(c) Plot the potential $\phi(x)$ for part (b).

P3.5 In Fig. P3.5, there is a material interface at $x = 0$ at which a sheet charge is also located, with value $Q = -2.5$ nC/cm^2.



➤ **Figure P3.3**



➤ **Figure P3.5**

(a) What is the relationship between $E(0^-)$ and $E(0^+)$?

(b) Sketch the electric field through the structure.

P3.6 The silicon structure in Fig. P3.6 consists of two pn junctions that are separated by 1 μm. The n-type to metal contact potential $\phi_{nm} = 400\ \text{mV} = \phi_n - \phi_m = -\phi_{mn}$.

(a) Find the depletion widths for each junction in thermal equilibrium.

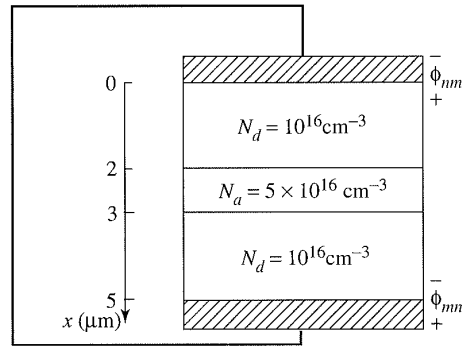
(b) Plot the electric field $E_o(x)$ through the structure in thermal equilibrium.

(c) Plot the potential $\phi_o(x)$ in thermal equilibrium, including the contact potentials.

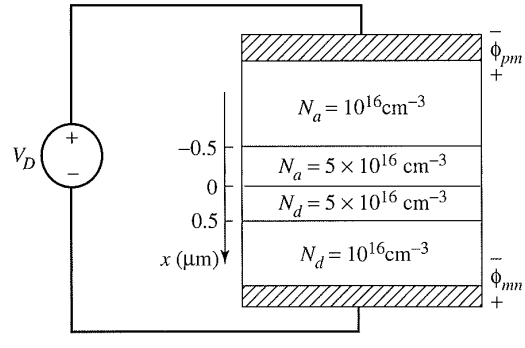
P3.7 The pn junction in Fig. P3.7 has a stepped doping concentration on each side of the metallurgical junction.

(a) Find the depletion width X_{do} in thermal equilibrium.

(b) If the n and p sides of the junction are each depleted by 1 μm, sketch the charge density, the electric field, and the potential.



➤ **Figure P3.6**



➤ **Figure P3.7**

(c) Find the applied voltage V_D (< 0 V) that corresponds to the depletion width in part (b).

P3.8 The pn junction diode in Fig. P3.8 has a stepped doping concentration.

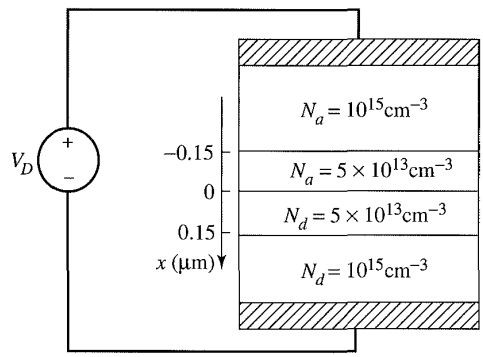
(a) Find the depletion width X_{do} in thermal equilibrium.

(b) If the n and p sides of the junction are depleted to ±1.0 μm, sketch the charge density, the electric field, and the potential.

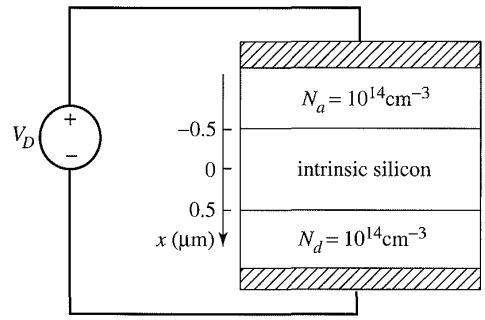
(c) Find the applied voltage V_D (< 0 V) that corresponds to the depletion width in part (b).

P3.9 Figure P3.9 shows a p-i-n diode. The *i* stands for intrinsic silicon, which in practice means silicon that is compensated with $N_a = N_d$.

(a) Sketch the charge density, electric field, and potential in thermal equilibrium.



➤ **Figure P3.8**



➤ **Figure P3.9**

(b) For an applied bias $V_D = -3$ V, find the widths of the depletion regions in the p and the n silicon.

P3.10 The IC structure shown in cross section in Fig. P3.10 is part of a voltage-controlled resistor.

(a) What is the thickness t of the undepleted n-type layer for $V_C = 0$. The distance between the metallurgical junctions between the n-type layer and the p^+ layers is $t_o = 1$ μm.

(b) For what value of V_C does $t = 0$? This condition is called pinch-off.

(c) Sketch the potential $\phi(x)$ at pinch-off.

P3.11 This problem continues the voltage-controlled resistor from Fig. P3.10.

(a) Plot the sheet resistance of the n-type layer as a function of V_C.

(b) Plot the junction capacitance per unit area (fF/μm^2) of the n-type layer as a function of V_C.

P3.12 Sketch the depletion capacitance of the pn junction in Fig. P3.8 as a function of the applied bias V_D. The capacitance at any breakpoints (where the slope of $C(V_D)$ changes) should be accurate.

P3.13 Sketch the depletion capacitance of the pn junction in Fig. P3.9 as a function of the applied bias V_D. The capacitance at any breakpoints (where the slope of $C(V_D)$ changes) should be accurate.

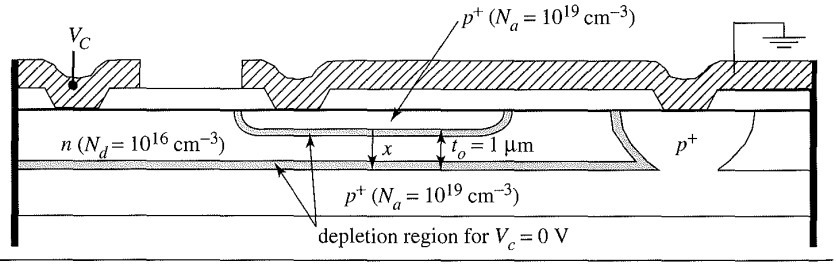
P3.14 The analysis of the electrostatics of the pn junction remains valid for small forward biases $V_D < 250$ mV. What is the percentage increase in junction capacitance per unit area over that of thermal equilibrium is found for the voltage-controlled resistor in Fig. P3.10 if we use $V_C = -250$ mV?

P3.15 The fabrication process used to make the voltage-controlled resistor in Fig. P3.10 has the following normalized uncertainties for each junction depth x_j and each doping concentration N.

$$x_j = \bar{x}_j(1 \pm 0.02) \quad \text{and} \quad N = \bar{N}(1 \pm 0.04)$$

(a) Find the uncertainty in the pinch-off voltage, assuming that all of the contributions are independent.

(b) Find the uncertainty in the zero-bias junction capacitance per unit area.



➤ **Figure P3.10**

P3.16 Figure P3.16 is the cross section of an ion-implanted region. In this problem, we will develop a very simple model for the extra capacitance contributed by the edges (or sidewalls) of the region. The bottom of the region has area $W \times W$ and the edges are modeled as half-circles in cross section, with a radius equal to the junction depth x_j. The doping concentrations for the bottom and sidewalls are approximated by:

$$\text{bottom: } N_a = 10^{17}\ \text{cm}^{-3},\ N_d = 10^{16}\ \text{cm}^{-3}$$

$$\text{sidewall: } N_a = 5 \times 10^{17}\ \text{cm}^{-3},\ N_d = 10^{17}\ \text{cm}^{-3}$$

(a) Find the sidewall capacitance per unit edge length in thermal equilibrium by applying the one-dimensional depletion model to the semi-circular edge of the region. Use $x_j = 0.75\ \mu\text{m}$.

(b) Plot the ratio of the sidewall capacitance to the total junction capacitance in thermal equilibrium, as a function of the lateral dimension W over the range $W = 2\ \mu\text{m} - 100\ \mu\text{m}$. Use $x_j = 0.75\ \mu\text{m}$.

P3.17 Heavily boron-doped polysilicon (referred to as p^+ polysilicon) is sometimes used as a gate material. Its thermal equilibrium potential is $\phi_{p+} = -550$ mV. The gate oxide is $t_{ox} = 200$ Å and the substrate doping is $N_a = 7.5 \times 10^{16}\ \text{cm}^{-3}$ for this problem.

(a) Find the flatband voltage V_{FB}. What state is the capacitor in for the case of thermal equilibrium?

(b) Find the threshold voltage V_{Tn}.

(c) Sketch the charge density, electric field, and potential through the MOS capacitor in thermal equilibrium.

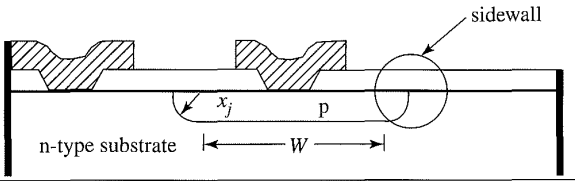
(d) Sketch the charge density, electric field, and potential through the MOS capacitor for $V_{GB} = V_{Tn}$.

P3.18 Using the p^+ polysilicon gate introduced in P3.17, we fabricate an MOS capacitor with a gate oxide is $t_{ox} = 250$ Å and a substrate doping is $N_d = 5 \times 10^{16}\ \text{cm}^{-3}$ of

(a) Find the flatband voltage V_{FB}. What state is the capacitor in for the case of thermal equilibrium?

(b) Find the threshold voltage V_{Tp}.

(c) Sketch the charge density, electric field, and potential through the MOS capacitor in thermal equilibrium.



➤ **Figure P3.16**

(d) Sketch the charge density, electric field, and potential through the MOS capacitor for $V_{GB} = V_{Tp}$.

P3.19 A composite dielectric is sometimes useful in an MOS capacitor. For this problem, the gate insulator consists of 50 Å of oxide (on the substrate), over which is 450 Å of *vacuum*. The gate is n^+ polysilicon and the substrate is doped p-type with $N_a = 5 \times 10^{15}$ cm^{-3}.

(a) What is the threshold voltage?

(b) Sketch the electric field through the MOS capacitor for $V_{GB} = -2.5$ V.

(c) Sketch the electric field through the MOS capacitor for $V_{GB} = +2.5$ V.

(d) For $V_{GB} = 2.5$ V, what is the change in the gate charge if the vacuum gap is reduced by 10%? A pressure transducer has been made using this principle.

(e) What is the electrostatic force f_E per unit area on the polysilicon gate for a gate-bulk bias $V_{GB} = 2.5$ V? This effect must be accounted for in designing a pressure sensor.

P3.20 This problem considers the effect of fixed charge at the SiO_2–silicon interface on the parameters of an MOS capacitor. For a MOS capacitor with n^+ polysilicon gate, $t_{ox} = 150$ Å, and $N_a = 10^{17}$ cm^{-3}, we have $Q_F = +2.5 \times 10^{-8}$ C/cm^2 located at $x = 0$.

(a) Sketch the charge density in thermal equilibrium for this MOS capacitor. What is the shift in the flatband voltage V_{FB} due to the fixed charge?

(b) What is the shift in the threshold voltage V_{Tn} due to the fixed charge?

(c) Sketch normalized *C-V* curve with and without the fixed charge.

P3.21 This problem considers the effect of fixed charge at the SiO_2–silicon interface on the parameters of an MOS capacitor. For a MOS capacitor with n^+ polysilicon gate, $t_{ox} = 150$ Å, and $N_d = 10^{17}$ cm^{-3}, we have $Q_F = +2.5 \times 10^{-8}$ C/cm^2 located at $x = 0$.

(a) Sketch the charge density in thermal equilibrium for this MOS capacitor. What is the shift in the flatband voltage V_{FB} due to the fixed charge?

(b) What is the shift in the threshold voltage V_{Tp} due to the fixed charge?

(c) Sketch normalized *C-V* curve with and without the fixed charge.

P3.22 The control of the threshold voltage is critical for VLSI CMOS processes. In this problem, we explore the implications for the control of gate oxide thickness. We would like the threshold voltage to be $V_{Tn} = 500$ mV ±50 mV in a particular process.

(a) Find the required average gate oxide thickness is $\overline{t_{ox}}$ in Å, given $N_a = 10^{17}$ cm^{-3}.

(b) Assuming that the substrate doping is $N_a = 10^{17}$ cm^{-3} and is perfectly controlled, what is the allowable variation in the gate oxide thickness?

(c) Now consider that $N_a = 10^{17} \text{ cm}^{-3} \pm 2 \times 10^{15}$ cm^{-3}. How much does the substrate doping variation add to the variation in V_{Tn}?

P3.23 Due to a processing error, the n^+ polysilicon gate of an MOS capacitor is merely n-type with a doping concentration $N_d = 10^{17}$ cm^{-3}. In this problem, assume you can treat the electrostatics of the n-type polycrystalline silicon gate exactly as you would treat n-type crystalline silicon. The gate oxide thickness is 200 Å and the substrate doping concentration is $N_a = 10^{17}$ cm^{-3}.

(a) What is the flatband voltage V_{FB}?

(b) For $V_{GB} = 2$V, what is the width of the depletion region in the n-type gate polysilicon? Sketch the electric field and the potential.

(c) Sketch the normalized *C-V* curve for this capacitor.

P3.24 In this problem, the substrate is doped at $N_a = 5 \times 10^{16}$ cm^{-3} for $0 < x < 500$ Å and $N_a = 5 \times 10^{17}$ cm^{-3} for $x > 500$ Å. The gate oxide is 250 Å thick and the gate is n^+ polysilicon.

(a) Sketch the charge density in thermal equilibrium.

(b) What is the flatband voltage V_{FB}?

(c) What is the threshold voltage V_{Tn}?

P3.25 Due to an error in implantation, the substrate doping for an MOS capacitor is $N_a = 10^{16}$ cm^{-3} rather than 5×10^{16} cm^{-3}. The gate oxide is 250 Å thick and the gate is n^+ polysilicon. We would like to remedy this error by doing another boron implant.

(a) What are the minimum dose and depth of the implant required to obtain the desired threshold voltage?

(b) Sketch the normalized *C-V* curve before and after the corrective implant.

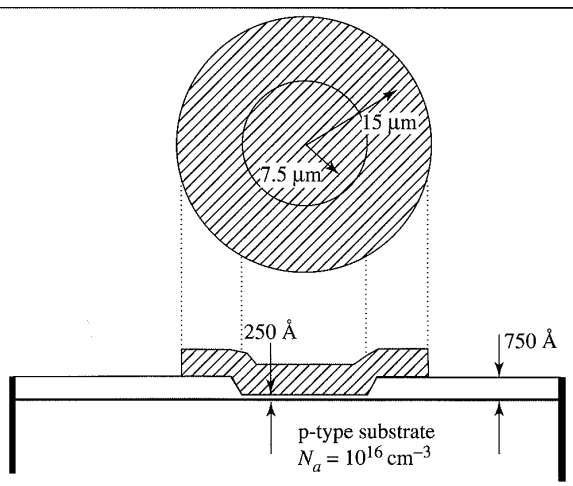
P3.26 Sketch the normalized *C-V* curve for the MOS capacitor with the stepped substrate doping in P3.24.

P3.27 The layout and cross section of a MOS structure is shown in Fig. P3.27.

(a) What are the threshold voltages of the thin oxide and thick oxide portions of this MOS capacitor?

(b) For $V_{GB} = 5$ V, what is the ratio of the inversion layer charge stored on the thick-oxide to the thin-oxide portions?

(c) Sketch the capacitance (in fF) versus V_{GB} for this structure.



➤ **Figure P3.27**

Design Problems

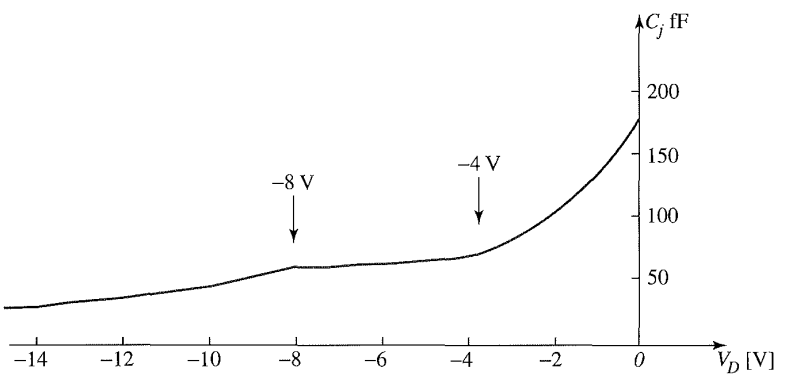
D3.1 The junction capacitance versus voltage curve is given below for a one-sided pn-junction. Determine the area and the (nonuniform) doping concentration that are required to have this capacitance curve. Note the breakpoints at $V_D = -4$ V and at $V_D = -8$ V.

D3.2 The requirement for a particular CMOS process is that the threshold voltage variation is $V_{Tn} = 0.7 \pm 0.1$ V. The oxide thickness is controlled to $t_{ox} = \overline{t_{ox}}(1 \pm 0.1)$ and the doping concentration is controlled to $N_a = \overline{N_a}(1 \pm 0.02)$. The possible range for the oxide thickness is limited to $100\text{Å} \leq t_{ox} \leq 500$ Å and the doping concentration is limited to $10^{15} \leq N_a \leq 10^{17}$ cm^{-3}.

(a) Choose an average doping concentration and an average oxide thickness that yield $\overline{V_{Tn}} = 0.7$ V.

(b) Find the variation in the threshold voltage, given the uncertainties in the oxide thickness and the doping concentration. Iterate to find a solution for $\overline{t_{ox}}$ and $\overline{N_a}$ that satisfies the 100 mV variation in the threshold voltage.

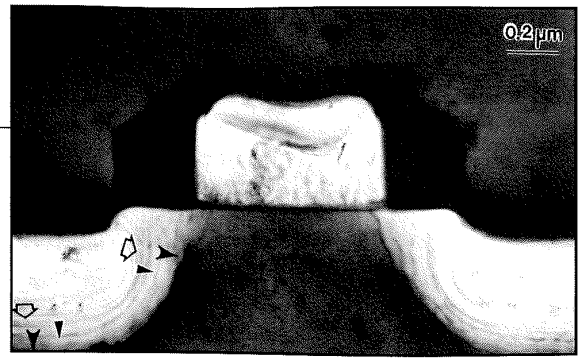
(c) Rather than use the statistical model for the uncertainties, it is more conservative to use worst-case analysis in which the signs of the variations in t_{ox} and in N_a are chosen to largest magnitude shift in V_{Tn}. Repeat (b) using this approach and compare your answers with the statistical model.



➤ **Figure D3.1**

chapter 4

The MOS Field-Effect Transistor



Transmission electron micrograph (TEM) of a 0.6 μm channel length MOSFET. (Courtesy of Prof. R. Gronsky, Dept. of MSME, Univ. of California at Berkeley)

This chapter introduces our first multi-terminal microelectronic device—the MOS field-effect transistor (MOSFET). Your understanding of the MOS capacitor will be the foundation for seeing how MOSFETs work and developing more accurate models. MOSFETs are the dominant transistor type for digital ICs and also have important applications in analog signal processing circuits.

Chapter Objectives

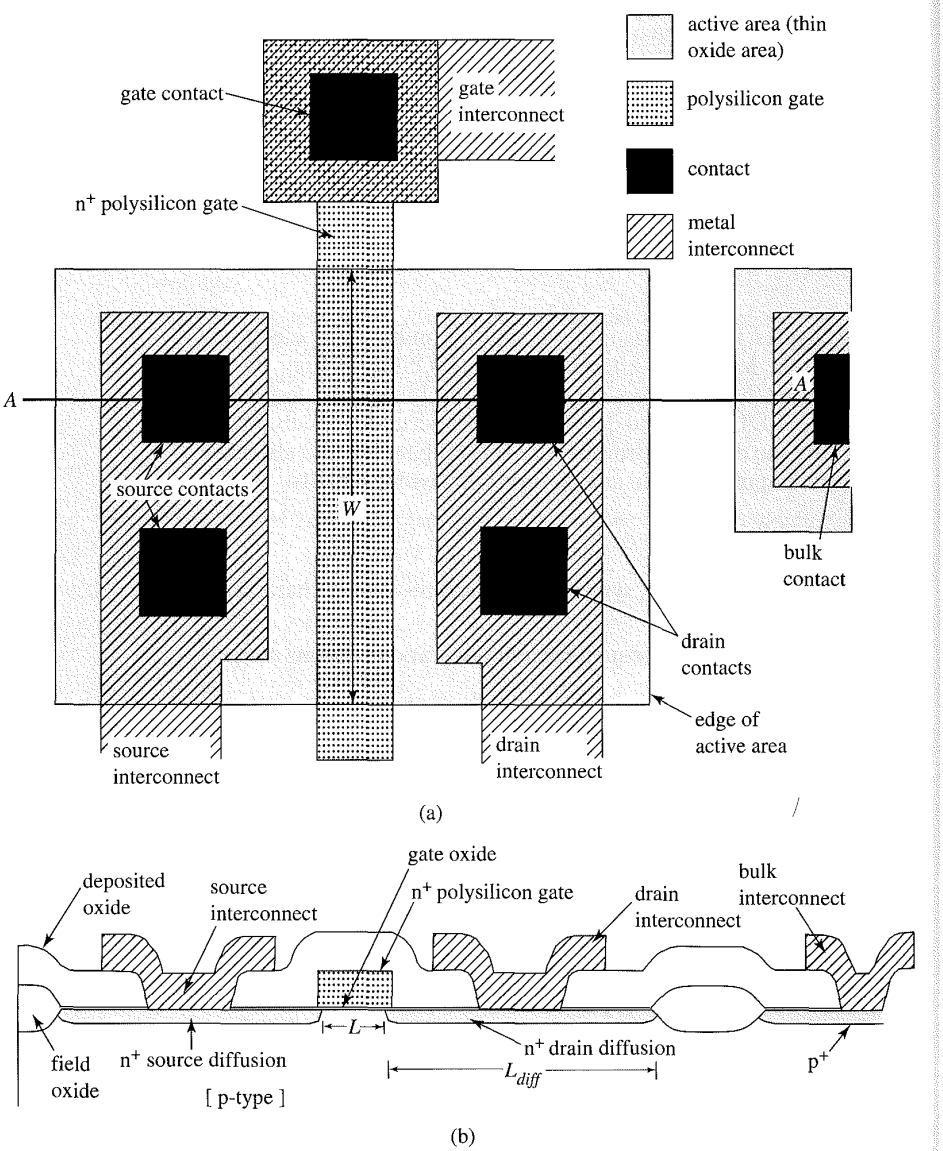
- MOSFET structures, symbols, and current-voltage characteristics.
- Cutoff, triode, and constant current (saturation) regions of operation.
- The gradual channel approximation and its use to model the current-voltage characteristics.
- Large- and small-signal models of n-channel and p-channel MOSFETs for circuit design.
- Level 1 SPICE model for the MOSFET.

➤ 4.1 INTRODUCTION

4.1.1 MOSFET Physical Structure

The **MOS field-effect transistor (MOSFET)** is the electronic device behind the explosive growth in digital electronics since 1970. The complexity and capability of MOS integrated circuits have increased many orders of magnitude over the past 25 years. To begin your study of this device, the layout and cross section of a

modern n-channel MOSFET is shown in Fig. 4.1. The basic structure is identical to that of the simple MOSFET whose fabrication process was described in detail in Chapter 2. The **gate** of the transistor is a film of n^+ polycrystalline silicon (polysilicon). Underlying the gate is the **gate oxide** which is a layer of thermally grown silicon dioxide (SiO_2).



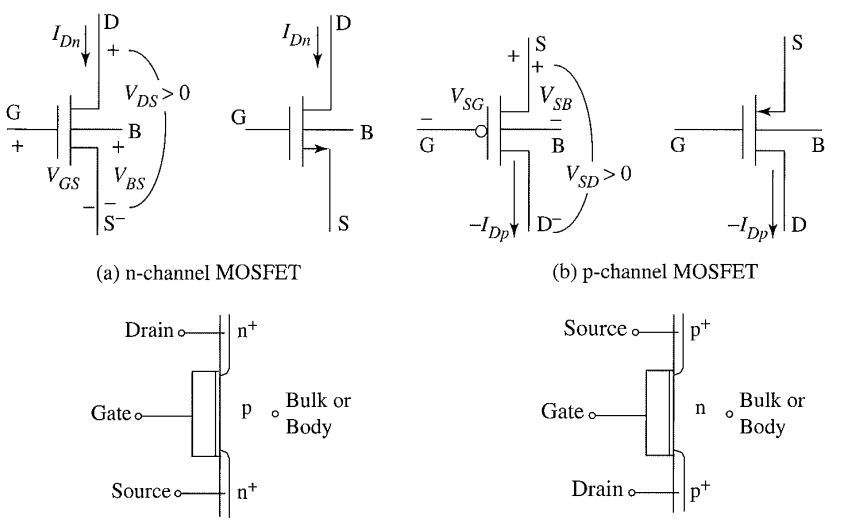
➤ **Figure 4.1** (a) Four-mask layout and (b) cross section of an integrated n-channel MOSFET.

The **channel** is formed under the polysilicon gate between the two n^+ regions as shown in Fig 4.1 (b). The **channel length** L of the MOSFET as is defined in Fig. 4.1 (b) as the gap between the n^+ diffusions. On the layout in Fig. 4.1(a), the **channel width** W of the MOSFET is defined as the width of the active area. One n^+ diffusion is called the **source**, the other is called the **drain**. The potentials on the diffusions will be the basis for differentiating the source from the drain, as discussed in the next section. The source and drain diffusions have a length L_{diff} and the same width W as the channel. Metal interconnections make ohmic contacts to the gate, source, and drain. In addition, the underlying p-type bulk region (the substrate or a deep p-type diffusion) is contacted with a fourth metal interconnection called the **bulk** or **body**.

4.1.2 MOSFET Circuit Symbol and Terminal Characteristics

Before introducing the current-voltage characteristics of the MOSFET, it is helpful to have a symbol for this device. Unfortunately, there are two common symbols for each MOSFET type. Figure 4.2 shows the circuit symbols and basic structure for an **n-channel MOSFET** (shown in Fig. 4.1) and for the complementary **p-channel MOSFET** that has p^+ source and drain diffusions in an n-type bulk region. In this text we will use the symbols without arrows for digital circuits and the symbols with arrows for analog circuits. The cross sections for the two MOSFETS have been rotated 90° in order to illustrate the close correspondence between the symbols and the physical structure.

The MOSFETs are oriented as they often appear in circuit schematics: the drain for the n-channel MOSFET and the source for the p-channel MOSFET are on top—toward the positive voltage supply. Voltages between terminals are labeled by an ordered pair of subscripts such as $V_{GS} = V_G - V_S$. By IEEE convention, the reference



➤ **Figure 4.2** (a) n-channel MOSFET symbols and structure and (b) p-channel symbols and structure.

directions for currents are defined as positive *into* the device terminals. Since the drain current for the p-channel device is negative, it is convenient to use $-I_{Dp} > 0$, which is positive leaving the drain terminal. For the n-channel MOSFET in Fig. 4.2(a), the drain and source terminals are selected so that $V_{DS}>0$, which is equivalent to defining the source as the one with the *lower* potential. As shown in Fig. 4.2(b), the labeling convention is opposite for the p-channel MOSFET. Its source and drain are identified so that $V_{SD}>0$, which means its source is defined as the one with the *higher* potential.

The n-channel MOSFET's static (DC) terminal characteristics can be measured with voltage sources and ammeters. Since the MOSFET is a four terminal device, we need to experiment and discover which voltages are most important. For our initial investigation, we short the bulk to the source to set $V_{BS} = 0$ V. Since the voltage *differences* between the drain and source and gate and source are most important, we ground the source ($V_S = 0$ V) for convenience. Since the gate terminal is insulated, the only current of interest is that into the drain terminal I_{Dn}.

Figure 4.3(a) is a schematic of the circuit used to find the drain current's functional dependence on the two voltages under our control: the gate-source voltage V_{GS} and the drain-source voltage V_{DS}. The conventional way to graph the measured data is to plot a family of curves $I_{Dn}(V_{GS}, V_{DS})$, with V_{GS} as a parameter. The drain current versus drain-source voltage curves for a selected set of gate-source voltages are called the MOSFET **drain characteristics**.

Figure 4.3(b) shows measured drain characteristics for an n-channel MOSFET in a digital CMOS process. Although the voltage and current ranges can vary with the technology and device layout, the ranges in Fig. 4.3(b) are typical. In this graph, the drain current I_D is measured as a function of drain-source voltage V_{DS} for eight gate-source voltages

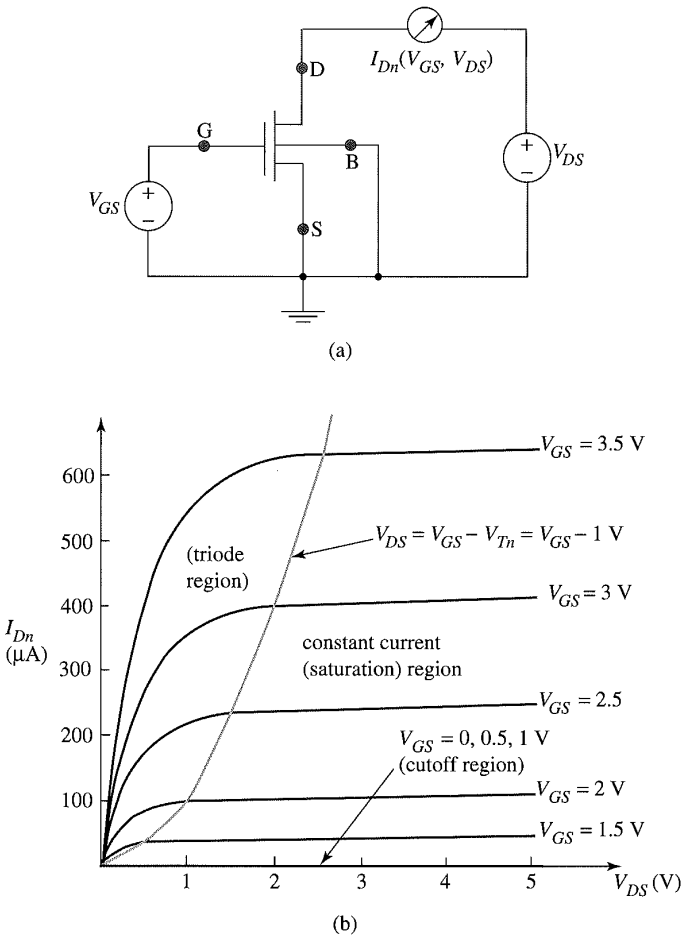
$$I_{Dn}\ (V_{GS} = 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5\ \text{V},\ V_{DS}),\ \text{for}\ 0\ \text{V} < V_{DS} < 5\text{V} \qquad (4.1)$$

Several observations can be made from studying the drain characteristics in Fig. 4.3(b).

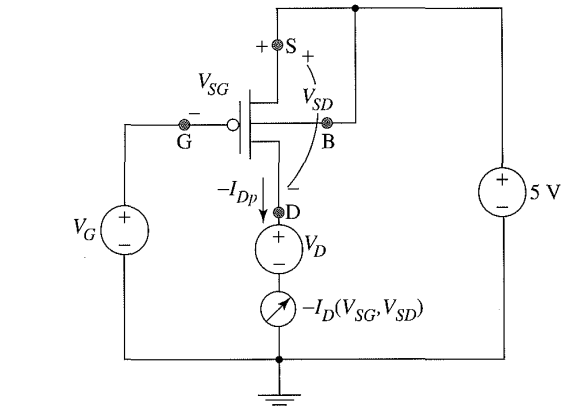
1. The MOSFET is **cutoff** ($I_{Dn} = 0$ A) for gate-source voltages that are less than some critical voltage, which is defined as the **threshold voltage** V_{Tn} of the n-channel device. A typical value for the n-channel threshold voltage is $V_{Tn} = 1$ V. The n^+ source and drain are electrically isolated until sufficient voltage is applied to the gate to create an n-type inversion layer or **channel** between them.
2. The drain current is nearly independent of the drain-source voltage once $V_{DS} > V_{GS} - V_{Tn}$. In this region of operation, called **saturation**, the MOSFET behaves like a current source. Further study reveals that the drain current in saturation is proportional to the square of the gate-source voltage above threshold $(V_{GS} - V_{Tn})^2$. This can be seen from the ratios of drain current for $V_{GS} = 3$ V and for $V_{GS} = 2$ V. This constant-current behavior will be exploited in both analog and digital circuit design.
3. The region where I_{Dn} depends on both V_{GS} and V_{DS} is termed the **triode** region.

The p-channel MOSFET's drain characteristics are shown in Fig. 4.4. In this case, the source and n-type bulk terminals are both connected to a 5 V supply. The effect of varying the drain voltage and gate voltage on the drain current can be investigated. The shape of the drain characteristics are identical to those of the n-channel MOSFET

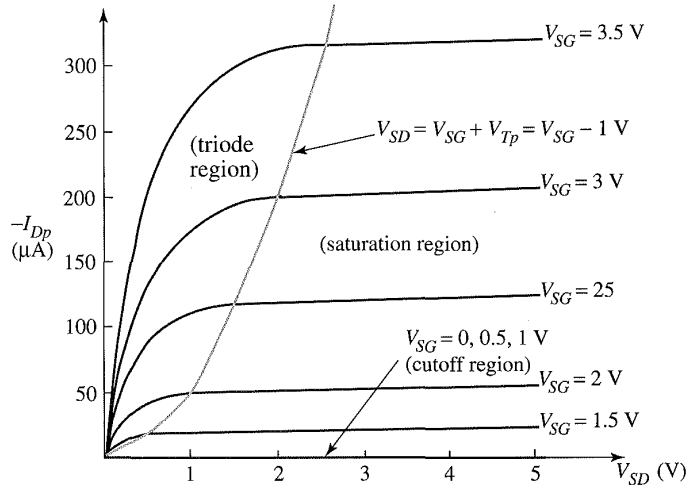
if $-I_{Dp}$ (which is positive) is plotted against the source-drain drop $V_{SD} = 5\text{ V} - V_D$ as a function of the source-gate voltage $V_{SG} = 5\text{ V} - V_G$. The drain current is zero unless the source-gate voltage $V_{SG} > -V_{Tp}$, where V_{Tp} is the p-channel threshold voltage. A typical value for the p-channel threshold voltage is $V_{Tp} = -1\text{ V}$.



➤ **Figure 4.3** n-channel MOSFET drain characteristics: (a) test circuit and (b) measurements of drain current as function of drain-source voltage, with gate-source voltage as a parameter for a typical integrated device.



(a)

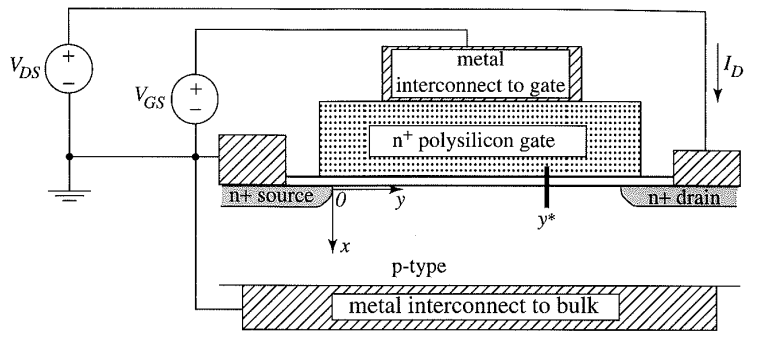


(b)

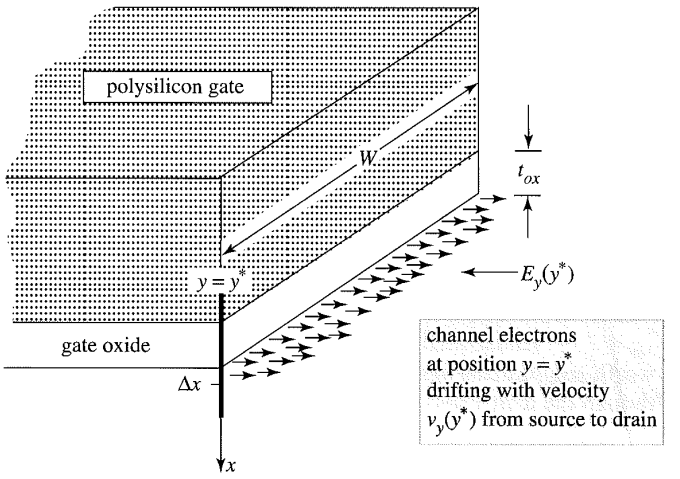
➤ **Figure 4.4** (a) Test circuit and (b) measured drain characteristics of a p-channel MOSFET with the same geometric dimensions as the n-channel device used for Fig. 4.3. The negative of the drain current $-I_{Dp}$ (> 0) is plotted against the source-drain drop V_{SD}, with the source-gate drop V_{SG} as a parameter. The drain current for the p-channel is about half that of the n-channel device for identical geometries and complementary threshold voltages $V_{Tp} = -V_{Tn} = -1$ V.

➤ 4.2 DEVICE PHYSICS OF THE MOSFET: DRAIN CURRENT AND CHANNEL CHARGE

We are now prepared to quantify how the drain current varies with the terminal voltages for the n-channel MOSFET. As a starting point, we will build upon the electrostatics of the MOS capacitor that were discussed in Chapter 3. Once the channel forms, an inversion layer of electrons exists between source and drain. An applied



(a)



(b)

➤ **Figure 4.5** (a) MOSFET biased with $V_{GS} > V_{Tn}$ and $V_{DS} > 0$. (b) Cutaway view of drifting electrons in the channel at position y^* (see (a)) between the source ($y = 0$) and drain ($y = L$).

voltage between the drain and source will impose an electric field across the channel that results in a drift current density. A good first step is to relate this drift current density, which is the origin of the drain current I_D, to the charge in the inversion layer.

Figure 4.5(a) illustrates an n-channel MOSFET with applied voltages $V_{GS} > V_{Tn}$ and $V_{DS} > 0$. For simplicity, we again short the bulk to the source so $V_{BS} = 0$. Furthermore, we assume that a channel extends from the source to the drain. The drain current I_D consists of electrons drifting from source to drain along the channel. At any point along the channel, the electron drift current must equal the drain current I_D since there is no current to the substrate or through the insulating oxide to the gate.

Figure 4.5(b) shows a perspective view of a portion of the channel at position y^* between the source and drain. From basic semiconductor physics, the drift current density J_y at any point (x,y) in the channel is given by

$$J_y = -qn(x, y)v_y(y) \tag{4.2}$$

where $n(x,y)$ is the electron concentration (cm^{-3}) in the channel and $v_y(y)$ is the drift velocity. The channel is a very thin region under the gate oxide—there are negligible electrons for $x > \Delta x \approx 50$ Å $= 0.005$ μm. The drain current I_D is equal to the integral of the drift current density J_y over the cross section of the channel in Fig. 4.5(b), which has an area of $W\Delta x$. Adding a negative sign to account for the fact that the reference direction for I_D in Fig. 4.4(a) is opposite to the positive direction of the y coordinate, we find

$$I_D = -W\int_0^{\Delta x} J_y(x, y)dx = Wv_y(y)\left(\int_0^{\Delta x} qn(x, y)dx\right) \tag{4.3}$$

where we have used the fact that v_y is not a function of the position x in the thin channel. The integral in Eq. (4.3) is the negative of the electron charge $Q_N(y)$ (units: C/cm^2) in the channel at point y between source and drain, since

$$Q_N(y) = -\int_0^{\Delta x} qn(x, y)dx \tag{4.4}$$

As we will investigate shortly, the electron charge will vary along the channel between the source and drain. Substituting Eq. (4.4) into Eq. (4.3), we relate the drain current to the channel charge

drain current = product of channel charge, width, and drift velocity

$$\boxed{I_D = -Wv_y(y)Q_N(y)} \tag{4.5}$$

It is worth emphasizing again that the drain current does *not* vary from source to drain, although the drift velocity $v_y(y)$ and channel electron charge $Q_N(y)$ generally do.

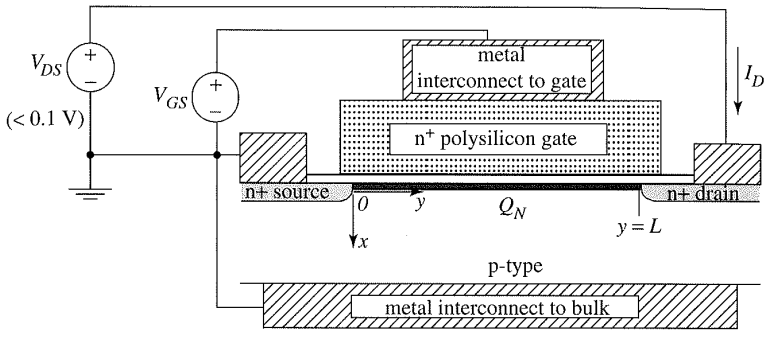
The drain current equation in Eq. (4.5) has been derived from basic physics. Next, we must determine how the channel electron charge and drift velocity vary with position in the channel in order to proceed toward finding the functional dependence of drain current on V_{GS} and V_{DS}. Since the full analysis is rather involved, we will make a "first pass" to gain an understanding of the important features of MOSFET operation.

➤ 4.3 MOSFET DEVICE PHYSICS: A FIRST PASS

From Chapter 3, we have a quantitative result for the inversion charge under a MOS capacitor

$$Q_N = -C_{ox}(V_{GB} - V_{Tn}) \tag{4.6}$$

In order to apply this one-dimensional result to the two-dimensional channel of a MOSFET, first consider the special case where the drain-source voltage V_{DS} is very small—on the order of 100 mV = 0.1 V or less. As shown in Fig. 4.6, the channel is uniform from source to drain for small drain-source biases. Note that for $V_{DS} = 0$ V,



➤ **Figure 4.6** MOSFET biased with $V_{GS} > V_{Tn}$ and very small V_{DS}, so the channel can be assumed to be uniform from source to drain. The channel thickness is greatly exaggerated in order to be visible.

the structure degenerates to a capacitor and Eq. (4.6) is an *exact* expression for the channel charge.

Substituting Eq. (4.6) into the drain current expression from the previous section and noting from Fig. 4.6 that $V_{GB} = V_{GS}$, we find

$$I_D = -Wv_y Q_N = Wv_y C_{ox}(V_{GS} - V_{Tn}), \text{ where } V_{GS} > V_{Tn} \text{ and } V_{DS} < 0.1 \text{ V} \quad (4.7)$$

The remaining unknown in Eq. (4.7) is the drift velocity in the channel $v_y(y)$. Since the channel is nearly uniform, the voltage in the channel must vary linearly from source to drain with a constant electric field in the channel

$$E_y = -\left(\frac{V_{DS}}{L}\right) \quad (4.8)$$

Given the small V_{DS}, the channel electric field is also small in magnitude, making it a good approximation that the drift velocity is proportional to the electric field

$$v_y = -\mu_n E_y = -\mu_n(-V_{DS}/L) = \mu_n (V_{DS}/L) \quad (4.9)$$

Substitution of Eq. (4.9) into (4.7) results in a equation for the drain current in terms of V_{GS} and V_{DS}, for the special case of $V_{DS} < 0.1$ V:

$$I_D = \left(\frac{W}{L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn})V_{DS} \text{ where } V_{GS} > V_{Tn} \text{ and } V_{DS} < 0.1 \text{ V} \quad (4.10)$$

For a fixed gate-source bias and small drain-source bias, Eq. (4.10) indicates that the MOSFET behaves like a voltage-controlled resistor, with resistance given by

$$R = \frac{V_{DS}}{I_D} = \left(\frac{1}{\mu_n C_{ox}(V_{GS} - V_T)}\right)\left(\frac{L}{W}\right) = [R_\square (V_{GS})]\left(\frac{L}{W}\right). \quad (4.11)$$

In order to model the MOSFET for $V_{DS} > 0.1$ V without solving differential equations, first approximate Eq. (4.5) by taking the averages of the channel charge and drift velocity over the channel

$$I_D \approx -W\overline{v_y}\,\overline{Q_N} \tag{4.12}$$

Since we do not know the functional dependence of Q_N, we estimate it as the mean of the channel charge at the source and drain ends of the channel. On the source side, the channel is adjacent to an n^+ diffusion that is connected to the p-type bulk region so $V_{GS} = V_{GB}$. The n^+ source diffusion supplies electrons to the inversion layer as soon as the surface inverts. Therefore, the channel charge stored on the source side of the circuit is

$$Q_N(y{=}0) = -C_{ox}(V_{GS} - V_{Tn}) \quad \text{where } V_{GS} > V_{Tn} \tag{4.13}$$

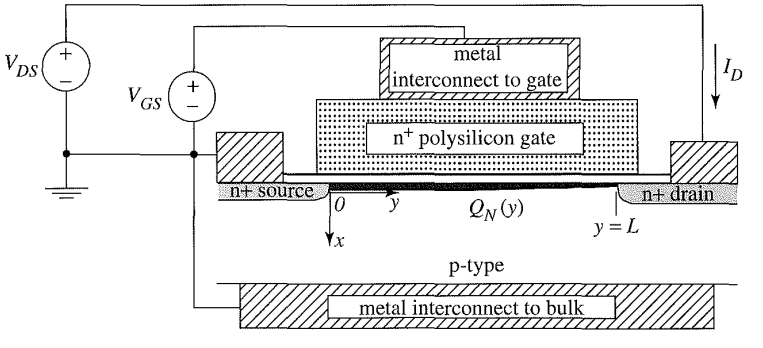
On the drain side of the circuit, the results from the MOS capacitor must be modified since the voltage of the n^+ diffusion adjacent to the drain end of the channel is V_{DS} and the electrostatics of the problem are changed. What is the effect of the higher potential at the drain end? The elevated drain potential lowers the voltage drop from the gate to the channel at the drain end and *reduces* the channel charge. This phenomenon can be approximated by using the gate-drain voltage drop $V_{GD} = V_{GS} - V_{DS}$ rather than the gate-source voltage drop in Eq. (4.13). The channel charge at the drain end is

$$Q_N(y{=}L) = -C_{ox}(V_{GD} - V_{Tn}) \text{ , where } V_{GD} = V_{GS} - V_{DS} > V_{Tn} \tag{4.14}$$

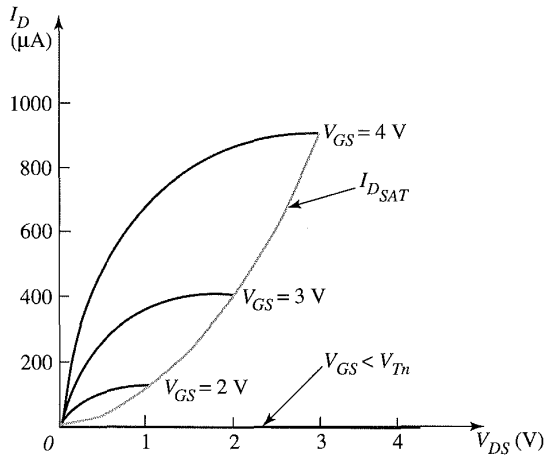
The conditions on V_{GS} and V_{DS} in Eq. (4.14) ensure that the channel charge is greater than zero at the drain end. The reduced charge at the drain end of the channel is shown qualitatively in Fig. 4.7. An estimate of the average channel charge from its values at the ends of the channel is

$$\overline{Q_N} \approx \frac{Q_N(y{=}0) + Q_N(y{=}L)}{2} = \frac{-C_{ox}(V_{GS} - V_{Tn}) - C_{ox}(V_{GS} - V_{DS} - V_{Tn})}{2} \tag{4.15}$$

Now that the channel charge varies from source to drain, it is no longer true that the channel voltage will increase linearly along the channel, or equivalently, that the



➤ **Figure 4.7** MOSFET biased in the triode region with $V_{GS} - V_{DS} > V_{Tn}$. The channel charge decreases from source to drain. The thickness of the channel is greatly exaggerated.



➤ **Figure 4.8** Drain characteristics for $V_{DS} \leq V_{GS} - V_{Tn}$, according to Eq. (4.17), assuming $V_{Tn} = 1$ V, $\mu_n C_{ox} = 50\ \mu A/V^2$, and $(W/L) = 4$.

channel electric field is constant. However, in the spirit of this approximate analysis, we will estimate the average electric field by

$$\overline{E_y} = -\overline{\left(\frac{dV_C}{dy}\right)} \approx \frac{-V_{DS}}{L} \tag{4.16}$$

where $V_C(y)$ is the channel voltage at position y between source and drain. The average drift velocity is then $\overline{v_y} = -\mu_n \overline{E_y} = \mu_n V_{DS}/L$. Substituting the estimates of average channel charge from Eq. (4.15) and average drift velocity into Eq. (4.12), we find

$$I_D = \left(\frac{W}{L}\right)\mu_n C_{ox}\left[V_{GS} - V_{Tn} - (V_{DS}/2)\right]V_{DS} \text{ where } V_{GS} - V_{DS} > V_{Tn} \tag{4.17}$$

As a check, Eq. (4.17) reduces to Eq. (4.10) for small V_{DS}. Figure 4.8 is a plot of Eq. (4.17) for three values of gate-source voltage. The curves have the general shape of the measured MOSFET drain characteristics in the triode region from Fig. 4.3. The inverted parabolas reach their maxima when the drain-source voltage is equal to $V_{DS_{SAT}}$, which is defined by

$$V_{DS_{SAT}} = V_{GS} - V_{Tn}\ . \tag{4.18}$$

This voltage is also the largest drain-source voltage for which Eq. (4.17) is valid. The maximum drain current, called the **saturation current** $I_{D_{SAT}}$, is found by substituting $V_{DS_{SAT}}$ into Eq. (4.17)

$$I_{D_{SAT}} = \left(\frac{W}{2L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn})^2 = \left(\frac{W}{2L}\right)\mu_n C_{ox} V_{DS_{SAT}}^2 \tag{4.19}$$

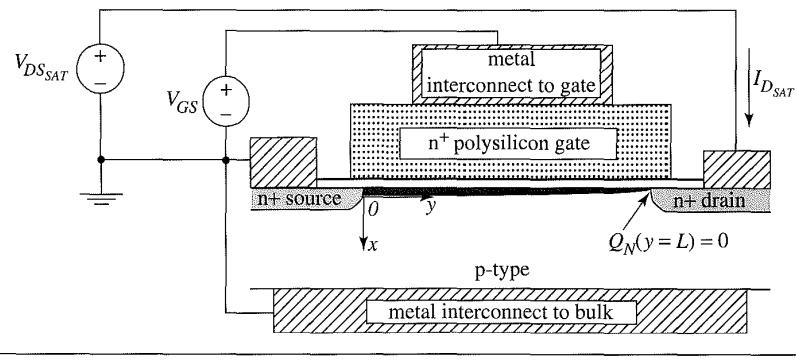
From the measured drain characteristics in Fig. 4.3, we know the drain current remains nearly constant at $I_{D_{SAT}}$ when the drain-source voltage increases above the saturation voltage $V_{DS} > V_{DS_{SAT}}$. How can this fact be explained? The behavior of the channel charge at the drain end with increasing V_{DS} is an important clue. For the case where $V_{DS} = V_{DS_{SAT}} = V_{GS} - V_{Tn}$, the channel charge vanishes at the drain end

$$Q_N(y=L)\big|_{V_{DS_{SAT}}} = -C_{ox}(V_{GD} - V_{Tn}) = -C_{ox}(V_{GS} - V_{DS_{SAT}} - V_{Tn}) = 0 \quad (4.20)$$

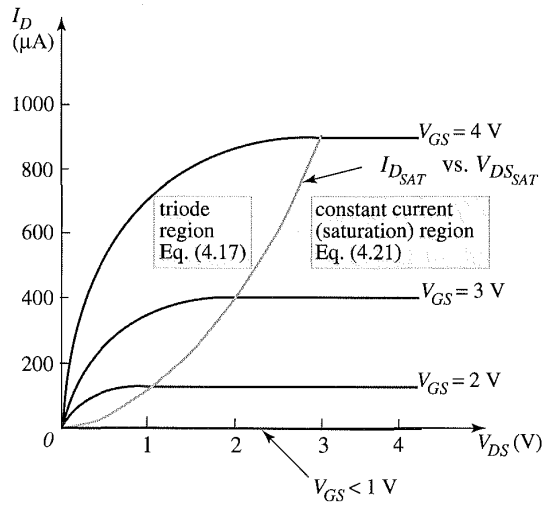
In Fig. 4.9, the channel charge in a MOSFET biased with $V_{DS} = V_{GS} - V_{Tn}$ is shown qualitatively. How can $Q_N(y = L) = 0$ and the drain current not also vanish? The answer is contained in Eq. (4.5)—the drain current is proportional to the product of the channel charge and drift velocity. For the drain current to remain constant in saturation, the drift velocity must become infinite at the drain end. In reality, the channel charge becomes very small and the drift velocity reaches its limiting value. Further increases in the drain-source voltage essentially have no effect on the channel charge at the drain end $Q_N(y = L)$ since it has already been reduced to nearly zero. As a result, the drain current remains constant, to first order, at its maximum value

$$I_D = I_{D_{SAT}} = \left(\frac{W}{2L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn})^2 \text{ for } V_{DS} \geq V_{DS_{SAT}} \quad (4.21)$$

The very simple models constructed in this section for the triode region (Eq. (4.17)) and the saturation region (Eq. (4.21)) reproduce the essential features of the MOSFET drain characteristics as shown in Fig. 4.10. After some minor refinements in Section 4.6, we will find their simplicity makes them very useful for hand analysis and design of digital and analog circuits.



➤ **Figure 4.9** MOSFET biased with $V_{DS} = V_{DS_{SAT}} = V_{GS} - V_{Tn}$ with the channel charge vanishing at the drain end. As in Figs. 4.6 and 4.7, the thickness of the channel is greatly exaggerated in order to be visible.

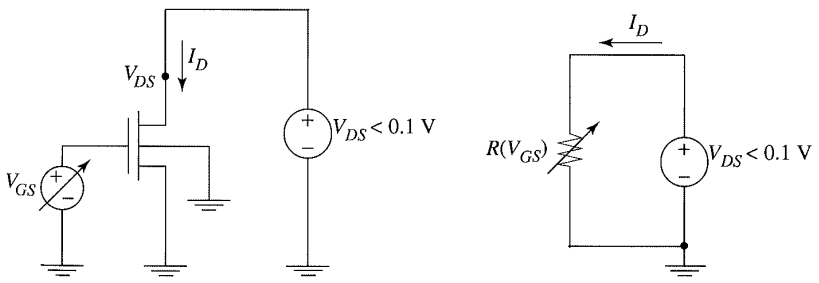


➤ **Figure 4.10** MOSFET drain characteristics generated from Eq. (4.17) for the triode region and Eq. (4.21) for the saturation region. Device parameters are the same as for Fig. 4.8.

➤ EXAMPLE 4.1 MOSFET as a Voltage Controlled Resistor

The circuit below shows an n-channel MOSFET that is used as a voltage-controlled resistor. A MOSFET in this configuration is found in circuits such as variable gain amplifiers.

(a) Find the sheet resistance of the MOSFET over the range $V_{GS} = 1.5$ V to $V_{GS} = 4$ V using $\mu_n = 215\ \text{cm}^2\text{V}^{-1}\text{s}^{-1}$, $C_{ox} = 2.3\ \text{fF}/\mu\text{m}^2$ and $V_{Tn} = 1$ V.



➤ **Figure Ex4.1A** An n-channel MOSFET used as a voltage controlled resistor.

SOLUTION

Note that for a gate-source voltage $V_{GS} > V_{Tn} + 0.1\ \text{V} = 1.1\ \text{V}$, the MOSFET will be operating in the **triode** region. Since the drain-source voltage is small, Eq. (4.17) reduces to Eq. (4.10)

$$I_D = \left(\frac{W}{L}\right)\mu_n C_{ox}\,(V_{GS} - V_{Tn})\,V_{DS} = \frac{1}{R}V_{DS}$$

For a particular value of V_{GS}, I_D is a *linear* function of V_{DS} and the circuit model for the MOSFET is a resistor as shown in the figure. Now we relate R to the sheet resistance.

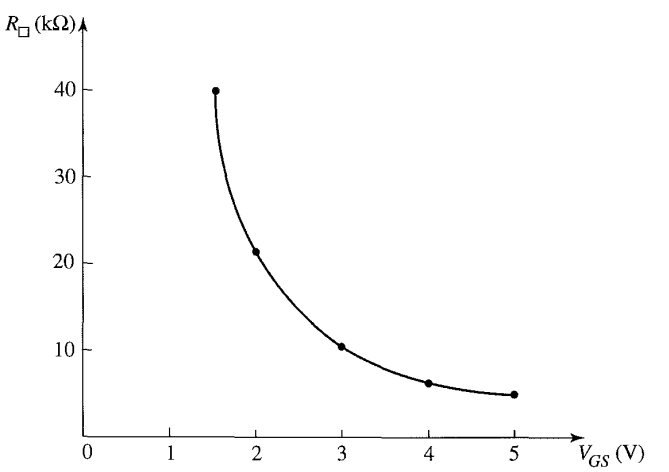
$$R = \frac{1}{\mu_n C_{ox}\left(\frac{W}{L}\right)(V_{GS} - V_{Tn})} = \frac{1}{\mu_n C_{ox}(V_{GS} - V_{Tn})}\left(\frac{L}{W}\right) = R_{\square}(L/W)$$

$R_{\square}$ as a function of V_{GS} is

$$R_{\square} = \frac{1}{(215\text{cm}^2\text{V}^{-1}\text{s}^{-1})\,(2.3\times 10^{-7}\text{F/cm}^2)}\,\frac{1}{(V_{GS} - 1V)} = \frac{20\,\text{k}\Omega\text{-V}}{(V_{GS} - 1\text{V})}$$

The plot of as $R_{\square}$ as a function of V_{GS} is shown below

(b) For a particular application, we need to control the resistor between 200 Ω and 1 kΩ for $V_{GS} = 1.5$ V to 4 V. How wide should the MOSFET be if the channel length $L = 1.5$ μm?



➤ **Figure Ex4.1B** Plot of sheet resistance vs. V_{GS} for a n-channel MOSFET operating with very small V_{DS}.

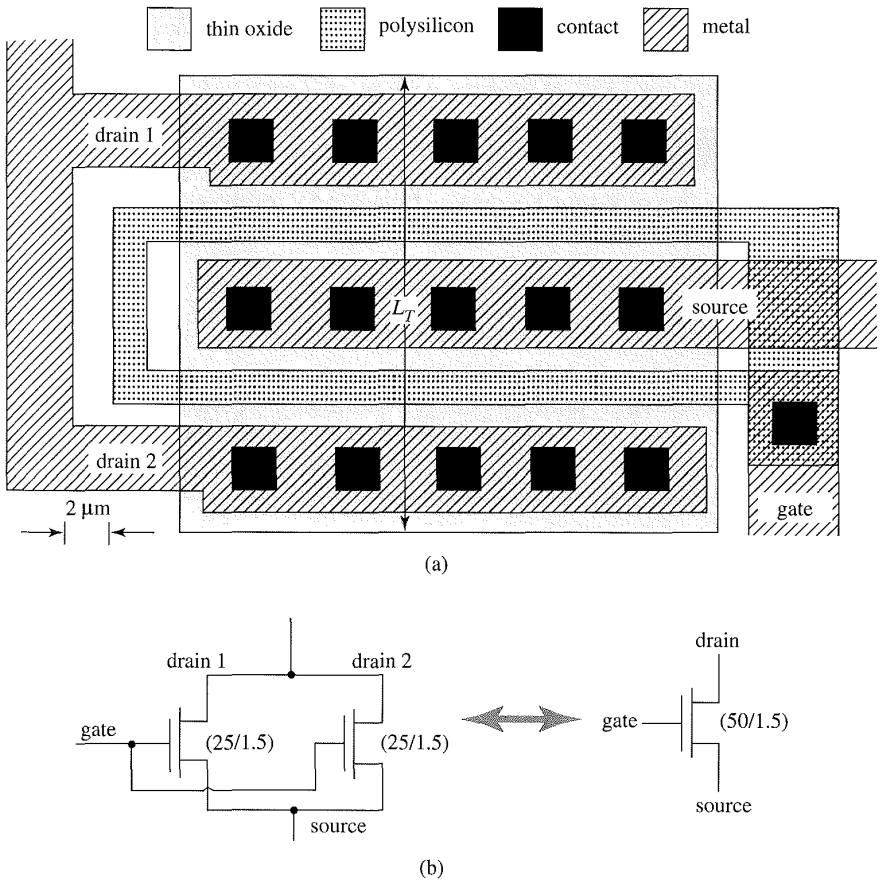
SOLUTION

We already solved for the sheet resistance in part (a), so we can find the range of sheet resistances for $V_{GS} = 1.5$ V to 4 V

$$R_{\square min} = \frac{20\,\text{k}\Omega\text{-V}}{(4\text{V} - 1\text{V})} = 6666.7\Omega \text{ and } R_{\square max} = \frac{20\,\text{k}\Omega\text{-V}}{(1.5\text{V} - 1\text{V})} = 40\text{k}\Omega$$

Solving for (W/L) to obtain $R_{min} = 200\ \Omega$ and $R_{max} = 1\ \text{k}\Omega$ yields

$$\left(\frac{W}{L}\right)_{min} = \frac{6666.7\Omega}{200\Omega} = 33.3 \text{ and } \left(\frac{W}{L}\right)_{max} = \frac{40000\Omega}{1000\Omega} = 40$$



➤ **Figure Ex4.1C** (a) Layout of folded n-channel MOSFET with L=1.5μm and W = 50 μm and (b) schematic showing the equivalence of the two MOSFETs with W = 25 μm in parallel with a single MOSFET having L = 50 μm

$(W/L) = 33.3$ will satisfy the requirements for both R_{min} and R_{max} so the width of the MOSFET should be

$$W = 1.5\mu m\,(33.3) = 50\mu m$$

(c) Design the layout for this MOS resistor so it occupies a minimum area. The length of the source/drain diffusions is $L_{diff} = 6$ μm with contacts that are 2 μm × 2 μm.

SOLUTION

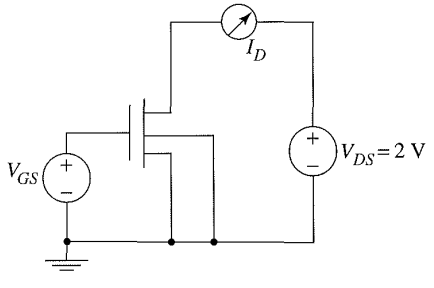
Given the high ratio of width to length, it is desirable to fold the MOSFET, which is equivalent to putting two MOSFETs of half the width in parallel. Since the diffusions are 6 μm long, the total length is $L_T = 3\ L_{diff} + 2\ L = 3 \times 6\ \mu m + 2 \times 1.5\ \mu m = 21\ \mu m$, which is nearly equal to half the width $W/2 = 25$ μm. The layout, drawn to scale, and the equivalent circuit are shown in Fig. Ex 4.1C.

➤ EXAMPLE 4.2 MOSFET operating regions

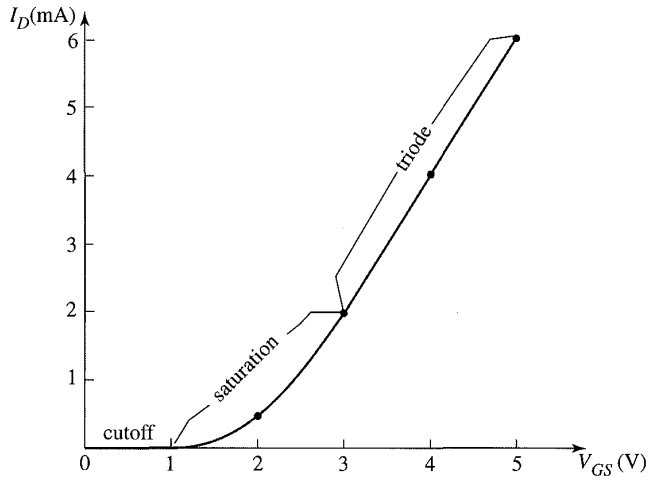
Plot I_D as a function of V_{GS} for V_{GS} ranging from 0 to 5V for the n-channel MOSFET shown in the circuit below with $\mu_n = 215\ cm^2V^{-1}s^{-1}$, $t_{ox} = 150$ Å, $V_{Tn} = 1$ V, $L = 1.5$ μm, and $W = 30\mu m$.

SOLUTION

For $V_{GS} < 1\ V = V_{Tn}$, the MOSFET is in the *cutoff* region and I_D is zero. Once the gate-source voltage is just high enough to turn on the device, the drain-source voltage $V_{DS} = 2\ V > V_{GS} - 1$ and the MOSFET is operating in the *saturation* region. The drain current in saturation is given by Eq. (4.21)



➤ **Figure Ex4.2A** n-channel MOSFET circuit.



➤ **Figure Ex4.2B** Plot of the I_D vs. V_{GS} for a n-channel MOSFET with fixed V_{DS} at 2 V.

$$I_D = \left(\frac{20}{2}\right)\left(50\frac{\mu A}{V^2}\right)(V_{GS} - 1\text{ V})^2 = 500\frac{\mu A}{V^2}(V_{GS} - 1\text{ V})^2$$

Once the gate-source voltage is high enough so $V_{GS} - V_{Tn} = V_{GS} - 1\text{ V} = V_{DS} = 2\text{ V}$, the MOSFET operates in the *triode* region with Eq. (4.17) describing the drain current

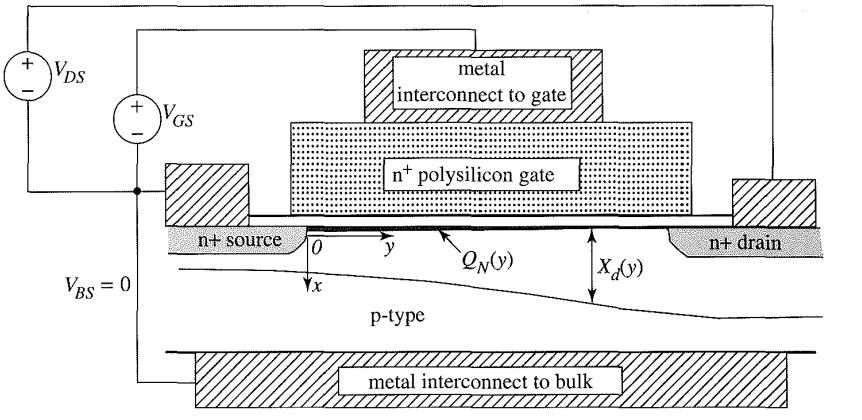
$$I_D = (20)\left(50\frac{\mu A}{V^2}\right)\left(V_{GS} - 1\text{V} - \frac{V_{DS}}{2}\right)V_{DS} = 2\frac{\text{m A}}{V^2}(V_{GS} - 2\text{V}) \quad \text{for } V_{GS} \geq 3\text{ V}$$

➤ 4.4 MOSFET DEVICE PHYSICS: THE GRADUAL CHANNEL APPROXIMATION

The starting point for developing a physical model for the current-voltage characteristics of the n-channel MOSFET is Eq. (4.5), which is repeated here.

$$I_D = -Wv_y(y)Q_N(y) \tag{4.22}$$

Our first step will be to find the functional dependence of the channel electron charge $Q_N(y)$ on the position y between the source and drain. Fig. 4.11 shows the cross section of a MOSFET that has bias voltages V_{GS} and V_{DS}. The bulk terminal is shorted to the source to make $V_{BS} = 0$ to simplify the initial analysis. Later in this section we will consider the case where the bulk is biased at a voltage different from the source.



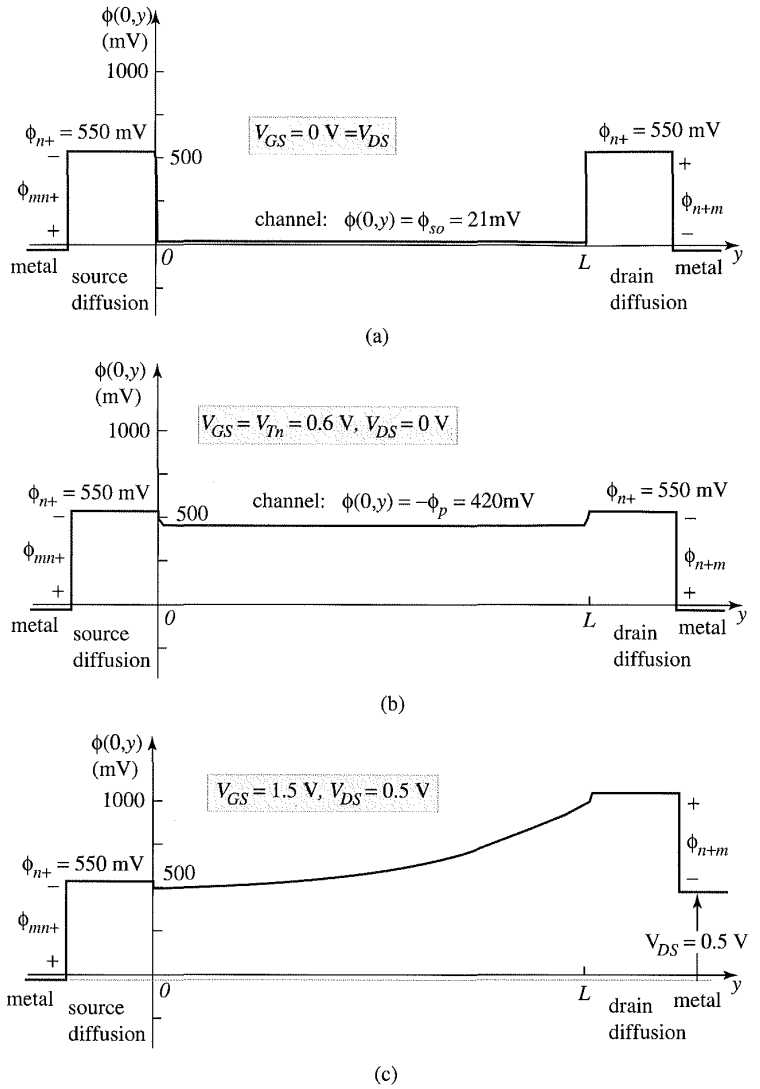
➤ **Figure 4.11** Cross section of n-channel MOSFET operating in the triode region with the applied voltages satisfying $V_{GS} > V_{Tn}$ and $V_{DS} < V_{GS} - V_{Tn}$. The depletion width increases and the channel charge decreases in magnitude from source to drain.

The coordinate system in Fig. 4.11 has its origin at the intersection of the boundary of the n^+ source diffusion with the oxide/silicon interface. In Chapter 3 we saw that the surface potential $\phi_s = \phi(0, y)$ is important in calculating the charge in the bulk depletion region and the charge in the inversion layer (the channel electron charge). Starting with thermal equilibrium, Fig. 4.12(a) plots the surface potential from source to drain. The example structure has a substrate doping of $N_a = 10^{17}$ cm^{-3} and an oxide thickness $t_{ox} = 150$ Å. The potential reference is the internal reference for silicon with intrinsic silicon being defined as zero. Note that the metal source and drain interconnections have the same potential since $V_{DS} = 0$ for equilibrium. The surface potential $\phi(0, y) = 21$ mV as calculated in Ex. 3.8, which corresponds to the surface being very slightly n-type. When a gate-source voltage is applied that is equal to the threshold voltage $V_{GS} = V_{Tn} = 0.6$ V and we maintain $V_{DS} = 0$, the surface potential increases to $\phi(0, y) = -\phi_p = 420$ mV as plotted in Fig. 4.12(b). The surface is inverted and an n-type channel has formed between the source and drain. Further increasing the gate-source voltage to $V_{GS} = V_{Tn} + 0.9$ V $= 1.5$ V and then applying a drain-source voltage of $V_{DS} = 0.5$ V, the sketch in Fig. 4.12(c) shows that the potential $\phi(0, y)$ increases by V_{DS} from the source to the drain. The metal contact to the n^+ source has been kept fixed in Fig. 4.12(c).

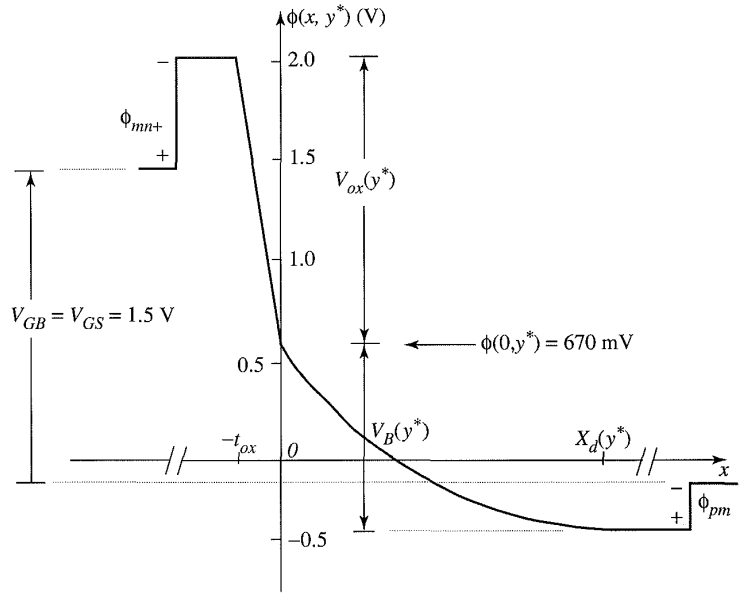
In Fig. 4.12(b), note that the potential in the channel is slightly lower than that of the n^+ source and drain regions when the gate is biased at the onset of inversion. For higher gate biases, the surface potential increases somewhat and reaches a maximum value of $\phi_{n+} = 550$ mV, an effect that is ignored in the delta-depletion approximation. In fact, the potential jump at the source becomes negligible for normal gate biases.

gradual channel approximation

We now consider the potential variation perpendicular to the channel (in the x direction in Fig. 4.11). In order to use the one-dimensional analysis from Chapter 3, we assume the lateral electric field (in the y direction) caused by V_{DS} is much smaller in magnitude than the vertical electric field caused by V_{GS}. Overall, this assumption is plausible since the gate oxide thickness (150Å) is much smaller than the channel length (on the order of 1 μm or longer), while the drain and gate voltages are each on



➤ **Figure 4.12** Surface potential along the channel of a MOSFET with the substrate doping concentration $N_a = 10^{17}$ cm^{-3} and oxide thickness $t_{ox} = 150$ Å. (a) Thermal equilibrium, (b) gate biased at the threshold voltage with no drain-source bias, and (c) gate bias increased to well above threshold $V_{GS} = 1.5$ V with a drain-source bias $V_{DS} = 0.5$ V that places the MOSFET in the triode region.



➤ **Figure 4.13** Potential $\phi(x, y^*)$, where the potential at point $y = y^*$ is 250 mV above its potential at the onset of inversion. The MOSFET is biased at the operating point used for Fig. 4.12(c).

the order of a few volts. This standard approach to MOSFET modeling is known as the **gradual channel approximation**.

Applying the one-dimensional results from Chapter 3, we equate the vertical potential drop through the MOSFET at any position in the channel to the sum of the drop across the oxide and the drop across the depletion region. In Fig. 4.13, we plot the potential as a function of x for a point y^* in the channel of a MOSFET biased with $V_{GS} = V_{GB} = 1.5$ V and $V_{DS} = 0.5$ V. In order to plot numerical values, we choose y^* so

$$\phi(0, y^*) = -\phi_p + 250\,\text{mV} = 670\,\text{mV} \tag{4.23}$$

At a general point y in the channel, we can equate

$$V_{GB} - V_{FB} = V_{GS} - V_{FB} = V_{ox}(y) + V_B(y) \tag{4.24}$$

since $V_{GB} = V_{GS}$ from Fig. 4.10. The drop across the depletion region is

$$V_B(y) = \phi(0, y) - \phi_p \tag{4.25}$$

At the source end of the channel ($y = 0$), the drop across the depletion region is

$$V_B(0) = \phi(0, 0) - \phi_p = -2\phi_p \tag{4.26}$$

which is the familiar result for an MOS capacitor from Chapter 3. As shown in Fig. 4.12(c), the applied drain-source bias does not affect the potential at the source end of the channel.

Since we need to find the variation of the channel electron charge $Q_N(y)$, we apply Gauss's Law to relate the oxide field to the charge in the silicon

$$V_{ox}(y) = E_{ox}(y)t_{ox} = \left(\frac{t_{ox}}{\varepsilon_{ox}}\right)(-Q_N(y) - Q_B(y)) = \frac{-Q_N(y) - Q_B(y)}{C_{ox}} \tag{4.27}$$

Substituting for $V_{ox}(y)$ from Eq. (4.24), we can solve for $Q_N(y)$

$$Q_N(y) = -C_{ox}\left(V_{ox}(y) + \frac{Q_B(y)}{C_{ox}}\right) = -C_{ox}\left(V_{GS} - V_{FB} - V_B(y) + \frac{Q_B(y)}{C_{ox}}\right) \tag{4.28}$$

The drop across the depletion region is known in terms of the surface potential from Eq. (4.25), which leads to

$$Q_N(y) = -C_{ox}\left(V_{GS} - V_{FB} - \phi(0, y) + \phi_p + \frac{Q_B(y)}{C_{ox}}\right) \tag{4.29}$$

This result can be compacted by defining the threshold voltage $V_{Tn}(y)$ to account for the variation in the bulk charge $Q_B(y)$ along the channel

$$V_{Tn}(y) = V_{FB} - 2\phi_p - \frac{Q_B(y)}{C_{ox}} = V_{FB} - 2\phi_p + \frac{1}{C_{ox}}\sqrt{2q\varepsilon_s N_a(\phi(0, y) - \phi_p)} \tag{4.30}$$

in which we substitute from Eq. (4.25) for the drop across the depletion region. Note from Eq. (4.30) and Fig. 4.12(c) that the threshold at the drain end is higher than at the source end of the channel because of the larger depletion charge resulting from $V_B(L) > V_B(0) = -2\,\phi_p$. This varying of threshold voltage along the channel is referred to as the **body effect**. Substituting Eq. (4.30) into Eq. (4.29), we find

$$Q_N(y) = -C_{ox}(V_{GS} - V_{Tn}(y) - \phi(0, y) - \phi_p) \tag{4.31}$$

Another convenient variable is the **channel voltage** $V_C(y)$, which is defined as

$$V_C(y) = \phi(0, y) + \phi_p \tag{4.32}$$

and is the component of the surface potential in the channel due to the drain-source voltage. Note from Fig. 4.12(c) that $V_C(0) = 0$ at the source end and $V_C(L) = V_{DS}$ at the drain end. Substituting for $V_C(y)$ into Eq. (4.31), the electron charge in the channel at point y is given by

$$Q_N(y) = -C_{ox}(V_{GS} - V_{Tn}(y) - V_C(y)) \tag{4.33}$$

in which the threshold voltage in terms of $V_C(y)$ is found by substituting into Eq. (4.30)

$$V_{Tn}(y) = V_{FB} - 2\phi_p + \frac{1}{C_{ox}}\sqrt{2q\varepsilon_s N_a(V_C(y) - 2\phi_p)} \tag{4.34}$$

Having expressed the electron charge in terms of the channel voltage in Eqs. (4.33) and (4.34), we now turn to the drift velocity term in Eq. (4.22). Assuming that the drift velocity is proportional to the electric field component along the channel E_y

$$v_y(y) = -\mu_n E_y(y) = \mu_n \frac{dV_C(y)}{dy} \tag{4.35}$$

Note that this assumption will be invalid for high electric fields, due to velocity saturation. The channel electron mobility in Eq. (4.35) is around one-third the bulk electron mobility from Fig. 2.8 due to excess scattering off the gate oxide in the thin channel. Substituting for the drift velocity and channel electron charge in Eq. (4.22), we find a differential equation in the channel voltage

$$I_D = -WQ_N(y)v_y(y) = WC_{ox}[V_{GS} - V_{Tn}(y) - V_C(y)]\mu_n \frac{dV_C(y)}{dy} \tag{4.36}$$

This equation can be integrated along the channel by separation of variables to eliminate V_C, resulting in an analytical expression for drain current as a function of V_{GS} and V_{DS} in the triode region. The final expression for $I_D(V_{GS},V_{DS})$ is too complicated to be useful for hand design or analysis. Note that the square-root term in $V_{Tn}(y)$ leads to three-halves powers of the terminal voltages.

To find a simpler expression for the drain current, we neglect the complication of a varying threshold voltage along the channel. We can see the nature of this approximation by first evaluating Eq. (4.34) at the source end of the channel, which is defined as the MOSFET threshold voltage V_{Tn}

n-channel MOSFET threshold voltage

$$\boxed{V_{Tn} = V_{Tn}(0) = V_{FB} - 2\phi_p + \frac{\sqrt{2q\varepsilon_s N_a(-2\phi_p)}}{C_{ox}}} \tag{4.37}$$

Substituting V_{Tn} into Eq. (4.30), the threshold voltage variation $V_{Tn}(y)$ due to the body effect is

$$V_{Tn}(y) = V_{Tn} + \frac{-Q_B(y) + Q_B(0)}{C_{ox}} = V_{Tn} + \frac{qN_a}{C_{ox}}(X_d(y) - X_d(0)) \tag{4.38}$$

Neglecting the body effect, we approximate $V_{Tn}(y) \approx V_{Tn}$ along the channel. The error in Eq. (4.38) due to body effect is small near the source and only depends on the square root of the channel voltage. This implies that we can model the essential features of the channel charge $Q_N(y)$ by approximating Eq. (4.33) with

$$Q_N(y) \approx -C_{ox}(V_{GS} - V_{Tn} - V_C(y)) \tag{4.39}$$

The differential equation for drain current Eq. (4.36) simplifies to

$$I_D = WC_{ox}(V_{GS} - V_{Tn} - V_C(y))\mu_n \frac{dV_C(y)}{dy} \tag{4.40}$$

We can eliminate the channel voltage by separating variables and integrating

$$\int_0^L I_D dy = \int_0^{V_{DS}} W\mu_n C_{ox}(V_{GS} - V_{Tn} - V_C(y))\, dV_C \tag{4.41}$$

which leads to a closed-form expression for the drain current

$$I_D = \left(\frac{W}{L}\right)\mu_n C_{ox}\,(V_{GS} - V_{Tn} - V_{DS}/2)\,V_{DS} \tag{4.42}$$

Surprisingly, this expression is identical to Eq. (4.17) that we derived using very crude approximations for the drift velocity and channel electron charge. As mentioned in the previous section, this equation for drain current is valid in the triode region when the inversion layer extends from source to drain. This implies from Eq. (4.39) that

$$V_{GS} \geq V_{Tn} + V_C(L) = V_{Tn} + V_{DS} \tag{4.43}$$

or equivalently, that $V_{DS} \leq V_{GS} - V_{Tn}$

4.4.1 Backgate Effect

In some circuit configurations, the bulk terminal of the MOSFET is biased at a voltage V_{BS} with respect to the source, thereby making it a four-terminal device. The voltage V_{BS} in Fig. 4.14 is called the **backgate bias** since the bulk potential affects the electron charge in the channel and functions as a second gate. Since the source and drain form pn diodes with the bulk terminal, the voltage V_{BS} is negative (V_{SB} positive) in order to ensure that these diodes are reverse biased.

The effect of the backgate bias is to shift the potential of the substrate lower, thereby increasing the potential drop across the depletion region as shown in Fig. 4.15. The MOSFET is under the same gate-source and drain-source bias ($V_{GS} = 1.5$ V and $V_{DS} = 0.5$ V) and the potential is plotted where $\phi(0, y^*) = 670$ mV as in Fig. 4.13.

From Fig. 4.15, the potential drop across the depletion region at a general point y between the source and drain is increased by the backgate bias to

$$V_B(y) = \phi(0, y) - (\phi_p + V_{BS}) = V_C(y) - 2\phi_p - V_{BS} \tag{4.44}$$

in which we substituted the definition of the channel voltage from Eq. (4.32). The larger $V_B(y)$ increases the bulk charge $Q_B(y)$, resulting in a larger threshold voltage. At the source end of the channel, the channel voltage is zero and $V_B(0) = -2\phi_p - V_{BS}$. Substitution of the increased drop at the source end of the channel into Eq. (4.37) yields the dependence of the MOSFET threshold on the backgate bias

$$V_{Tn} = V_{FB} - 2\phi_p + \frac{1}{C_{ox}}\sqrt{(2qN_a\varepsilon_s)\,(-2\phi_p - V_{BS})} \tag{4.45}$$

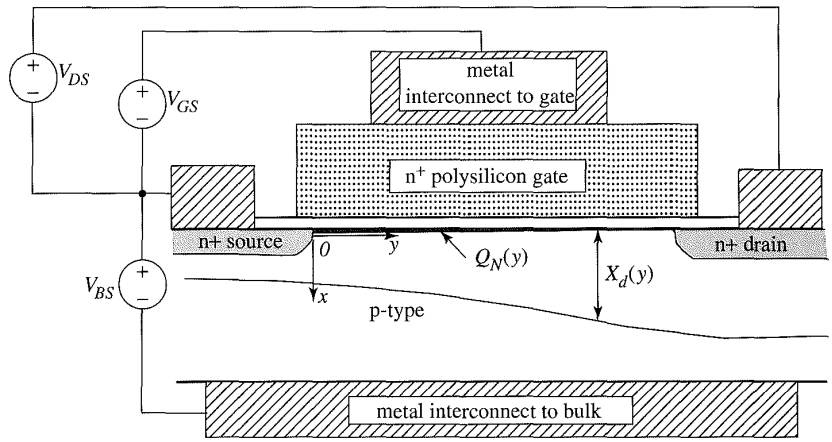
Note that both ϕ_p and V_{BS} are negative numbers. The conventional form for Eq. (4.45) is

n-channel MOSFET threshold voltage with backgate bias

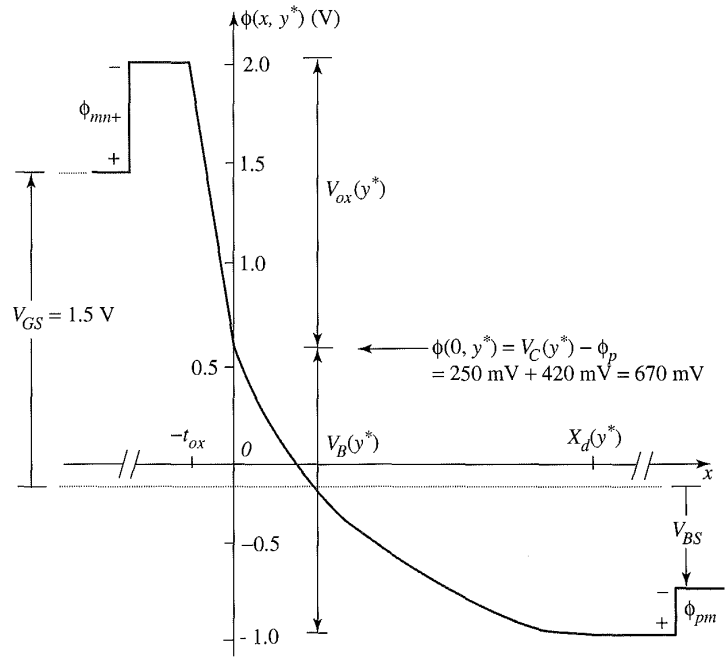
$$\boxed{V_{Tn} = V_{TOn} + \gamma_n\left(\sqrt{-2\phi_p - V_{BS}} - \sqrt{-2\phi_p}\right)} \tag{4.46}$$

where the n-channel MOSFET **backgate effect parameter** is defined as

$$\boxed{\gamma_n = \left(\sqrt{2qN_a\varepsilon_s}\right)/C_{ox}} \tag{4.47}$$



➤ **Figure 4.14** n-channel MOSFET with backgate bias $V_{BS}<0$. The depletion region has widened in comparison with Fig. 4.11 due to the larger potential drop across it.



➤ **Figure 4.15** Potential variation in x direction through the MOSFET at y^* where the channel voltage is $V_C(y^*) = 250$ mV. The bias voltages are $V_{GS} = 1.5$ V, $V_{DS} = 0.5$ V, and $V_{BS} = -0.5$ V.

and V_{TOn} is the threshold voltage without backgate effect (Eq. (4.37)). Therefore, the drain current is a function of the backgate bias through its effect on V_{Tn}. In Chapters 8 and 9 we will see that the backgate bias can have significant consequences for circuit design with MOSFETs.

n-channel backgate effect parameter

➤ EXAMPLE 4.3 Measuring the Backgate Effect Parameter

The test circuit from Example 4.1 can be modified to find an experimental value for the backgate parameter, γ_n. Note that a negative voltage V_{BS} is applied from the bulk to the source of the MOSFET. The circuit varies V_{GS} continuously from 0 to 5V, for $V_{BS} = 0$ and $V_{BS} = -5$ V. The drain-source voltage is $V_{DS} = 100$ mV.

(a) From the drain current measurements plotted below, find the backgate effect parameter. Assume that the device parameters are the same as in Example 4.1 with the substrate doping $N_a = 10^{17}$ cm^{-3}.

SOLUTION

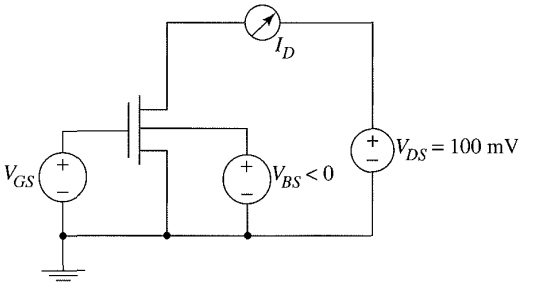
Since the drain voltage is only 100 mV, the MOSFET operates in its triode region once V_{GS} exceeds the threshold voltage. The drain current I_D is given by Eq. (4.42)

$$I_D = \left(\frac{W}{L}\right)\mu_n C_{ox}\left[V_{GS} - V_{Tn}(V_{BS}) - V_{DS}/2\right] V_{DS}$$

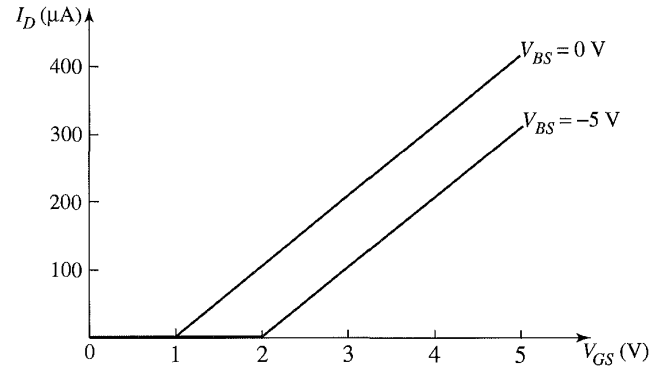
where the functional dependence of the threshold voltage on the backgate bias is indicated explicitly. The triode equation can be approximated with little error by neglecting $V_{DS}/2 = 50$ mV in the parentheses

$$I_D \cong \left(\frac{W}{L}\right)\mu_n C_{ox}\left[V_{GS} - V_{Tn}(V_{BS})\right] V_{DS}$$

This equation is linear in V_{GS}, with $V_{Tn}(V_{BS})$ as the intercept. From the graph, we can see $V_{TOn} = 1.0$ V and for $V_{BS} = -5$ V, $V_{Tn} = 2$V.



➤ **Figure Ex4.3A** Circuit to find the backgate effect parameter.



➤ **Figure Ex4.3B** Experimental data for circuit in Example 4.3

Applying Eq. (4.46) with $\phi_p = -0.42$ V, we can extract the backgate effect parameter

$$2\,\text{V} = 1\,\text{V} + \gamma_n\left(\sqrt{(0.84\,\text{V} + 5\,\text{V})} - \sqrt{0.84\,\text{V}}\right) \rightarrow \gamma_n = 0.67\ \text{V}^{1/2}$$

4.4.2 The MOSFET in Saturation

We will now consider the device physics of the MOSFET in saturation where the drain-source voltage increases up to and beyond the point where the channel charge vanishes at the drain end of the channel, according to our simple one-dimensional analysis. The variation of channel voltage $V_C(y)$ from source to drain can be derived from Eqs. (4.40) and provides insight into this phenomenon. By integrating Eq. (4.40) to a point y along the channel, we find that the channel voltage is given by

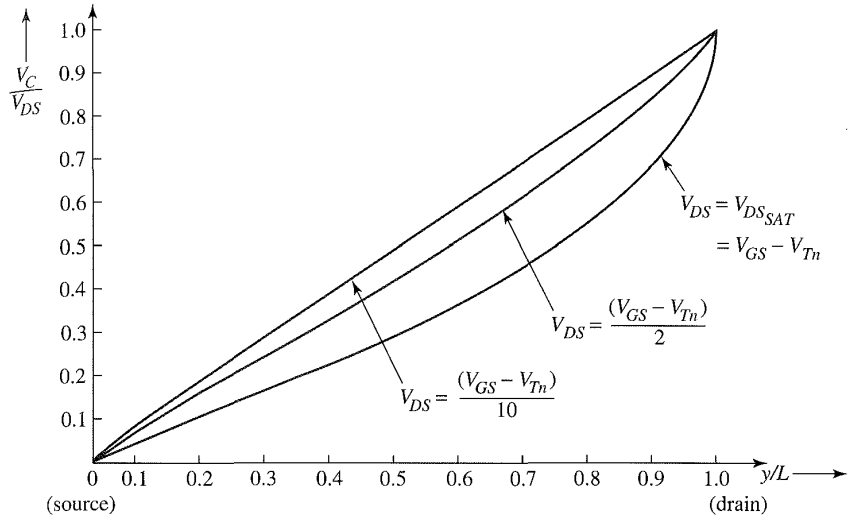
$$\int_0^y \frac{I_D}{\mu_n W C_{ox}}\,dy = \int_0^{V_C} (V_{GS} - V_{Tn} - V_C)\,dV_C = (V_{GS} - V_{Tn})\,V_C - V_C^2/2 \qquad (4.48)$$

After integrating the left hand side (remembering that I_D is constant along the channel), a quadratic equation yields the channel voltage

$$V_C(y) = V_{GS} - V_{Tn} - \sqrt{(V_{GS} - V_{Tn})^2 - 2\,(V_{GS} - V_{Tn} - V_{DS}/2)\,V_{DS}\,(y/L)} \qquad (4.49)$$

where we have substituted Eq. (4.42) for the drain current.

Figure 4.16 is a plot of the channel voltage, normalized to V_{DS}, from source to drain for three values of V_{DS}. For the smallest drain-source voltage, one-tenth of $V_{GS} - V_{Tn}$, the channel voltage increases nearly linearly from source to drain. With V_{DS} equal to one-half of $V_{GS} - V_{Tn}$, more of the channel voltage is dropped near the drain end. Finally, for the saturation drain voltage, $V_{DS_{SAT}} = V_{GS} - V_{Tn}$, most of the



➤ **Figure 4.16** Normalized channel voltage $V_C(y)/V_{DS}$ along the channel for three values of drain-source voltage.

voltage drop occurs at the drain end with the channel voltage becoming vertical at the drain.

The channel electric field E_y is the negative derivative of $V_C(y)$:

$$E_y(y) = -\frac{dV_C}{dy} = \frac{-(V_{GS} - V_{Tn} - V_{DS}/2)\,V_{DS}}{L\sqrt{(V_{GS} - V_{Tn})^2 - 2(V_{GS} - V_{Tn} - V_{DS}/2)\,V_{DS}\,(y/L)}} \tag{4.50}$$

which becomes infinite at the drain ($y = L$) when $V_{DS} = V_{DS_{SAT}} = V_{GS} - V_{Tn}$. However, an infinite electric field is not reasonable and we must reexamine the assumptions underlying the gradual channel approximation. The reason why E_y blows up at the drain is seen from the basic equation for drain current, evaluated at $y = L$

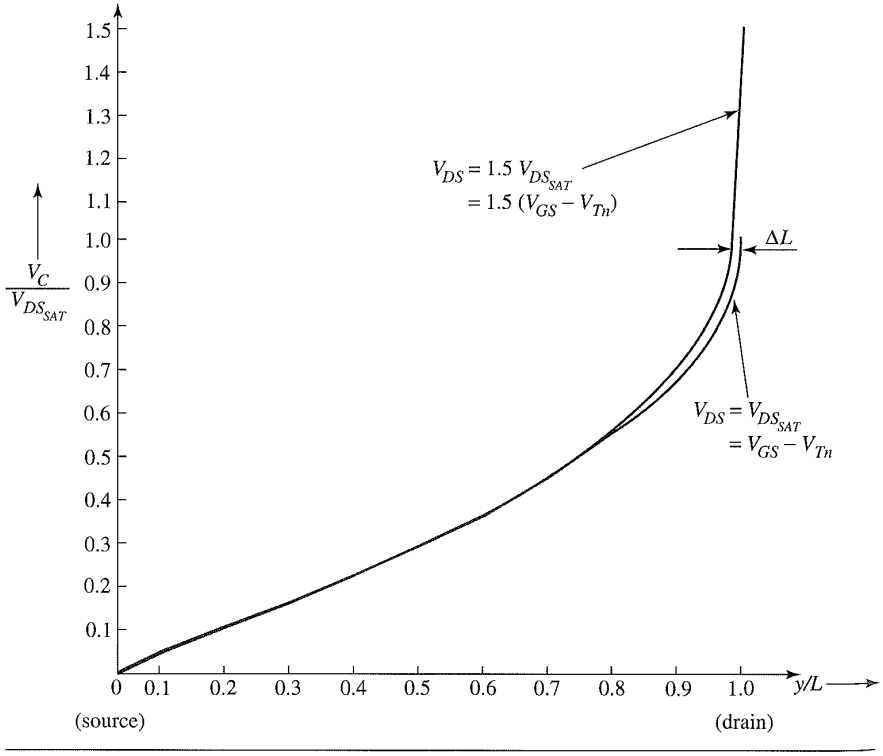
$$I_D = -WQ_N(L)v_y(L) = WC_{ox}(V_{GS} - V_{Tn} - V_C(L))\,\mu_n \left.\frac{dV_C}{dy}\right|_{y=L} \tag{4.51}$$

When the drain voltage is $V_{DS} = V_{GS} - V_{Tn}$, the channel charge vanishes at the drain end of the channel. Consequently, the channel is said to be "pinched off" at the drain. In order for the drain current to remain constant, the drift velocity must be infinite at $y = L$. Since we assumed that the drift velocity is proportional to the electric field, this condition can be satisfied mathematically by the channel voltage having an infinite slope at the drain as shown in Fig. 4.16. In reality, the drift velocity saturates and consequently, the channel electron charge is very small, but does not vanish, at the drain end of the channel.

We know from the measured drain characteristics in Fig. 4.3 that the drain current I_D **saturates** or remains nearly constant as the drain voltage increases beyond $V_{DS_{SAT}}$. The explanation for this phenomenon lies in the drain current equation (Eq. (4.41) that is repeated here for convenience.

$$\int_0^L I_D\,dy = \int_0^{V_{DS}} W\mu_n C_{ox}\,(V_{GS} - V_{Tn} - V_C(y))\,dV_C \tag{4.52}$$

When $V_{DS} < V_{DS_{SAT}}$, increasing the drain voltage increases the upper limit on the integral on the right-hand side, but also modifies the integrand by changing $V_C(y)$ as shown in Fig. 4.16. Once the channel pinches off for $V_{DS} = V_{DS_{SAT}}$ the channel charge nearly vanishes at the drain end and the lateral electric field becomes very large. An increase in drain voltage beyond $V_{DS_{SAT}}$ causes the high-field region at the drain to widen by a small distance ΔL—enough to accommodate the additional drop $V_{DS} - V_{DS_{SAT}}$. Figure 4.17 shows a qualitative picture of this effect that is called **channel length modulation**. To first order, the drain current I_D is unchanged by increases in V_{DS} above saturation. The right-hand integral in Eq. (4.52) is over the channel (where $0 < V_C(y) < V_{DS_{SAT}}$) and is unaffected. Electrons drifting along the channel enter the high-field region near the drain and are simply swept into the n^+ drain—at their saturation velocity.



➤ **Figure 4.17** Channel voltage, normalized to the saturation drain-source voltage $V_{DS_{SAT}}$, for $V_{DS} = V_{DS_{SAT}}$ and for $V_{DS} = 1.5\ V_{DS_{SAT}}$.

Upon closer inspection, the drain current is perturbed slightly by an increase in drain voltage above $V_{GS} - V_{Tn}$. The point y in the channel at which $V_C(y) = V_{DS_{SAT}}$ and therefore $Q_N(y) = 0$ moves a small distance ΔL toward the source as the drain voltage increases beyond saturation. The channel length is shortened to $L - \Delta L$, which changes the limit of integration on the left-hand integral in Eq. (4.52)

$$\int_0^{(L-\Delta L)} I_D\, dy = I_D(L-\Delta L) = \mu_n C_{ox} W \int_0^{V_{DS_{SAT}}} (V_{GS} - V_{Tn} - V_C)\, dV_C \qquad (4.53)$$

where the channel extends over $0 \le y \le L - \Delta L$ and the channel voltage ranges from $0 \le V_C \le V_{DS_{SAT}}$. Carrying out the integration and substituting $V_{DS_{SAT}} = V_{GS} - V_{Tn}$, we find

$$I_{D_{SAT}} = \frac{1}{2}\left(\frac{W}{L-\Delta L}\right)\mu_n C_{ox} (V_{GS} - V_{Tn})^2 \text{ for } V_{DS} \ge V_{DS_{SAT}} = V_{GS} - V_{Tn} \qquad (4.54)$$

An exact calculation of the channel-length modulation ΔL as a function of the terminal voltages involves solving the two-dimensional Poisson's Equation in the drain region. However, a useful approximation can be made by noting that $\Delta L \ll L$

$$\frac{1}{L-\Delta L} \approx \frac{1}{L}\left(1 + \frac{\Delta L}{L}\right) \qquad (4.55)$$

For simplicity, we assume that the fractional change in channel length is proportional to the drain voltage

$$\frac{\Delta L}{L} = \lambda_n V_{DS} \qquad (4.56)$$

where λ_n is the **channel-length modulation parameter**. Device measurements and simulations indicate that λ_n varies roughly with the inverse of channel length. For the MOSFETs in this text we will use

$$\lambda_n = \frac{0.1\,\mu\text{m}\text{V}^{-1}}{L} \qquad (4.57)$$

where L is in μm. Finally, we substitute Eqs. (4.55) and (4.56) into Eq. (4.54) and find a very useful approximation to the drain current in saturation

$$I_D = \frac{W}{2L}\mu_n C_{ox} (V_{GS} - V_{Tn})^2 (1 + \lambda_n V_{DS}) \qquad (4.58)$$

where $V_{DS} \ge V_{DS_{SAT}} = V_{GS} - V_{Tn}$.

Considering the several simplifying assumptions that have gone into the drain current equations, it is remarkable that they are quite adequate models for hand analysis and design. In practice, the channel electron mobility μ_n is adjusted to fit Eq. (4.42) to the measured drain current. A typical value for channel electron mobility in modern MOSFETs is 200 $\text{cm}^2/(\text{Vs})$, substantially less than the electron mobility in bulk silicon.

➤ EXAMPLE 4.4 Vertical and lateral electric fields in the channel

For an n-channel MOSFET with $L = 1.5\ \mu m$, $t_{ox} = 150$ Å, $N_a = 10^{17} cm^{-3}$, and $V_{Tn} = 0.6$ V, evaluate $E_x(0, y)$ and $E_y(0, y)$ to verify the gradual channel approximation, which assumes $|E_y| \ll E_x$.

(a) For $V_{GS} = 1.5$ V, $V_{DS} = V_{DS_{SAT}} = 0.9$ V, $V_{BS} = 0$ V, find the vertical and lateral electric fields at $y = 0.75\ \mu m$.

SOLUTION

The vertical electric field in the silicon at $x = 0^+$ can be found from the oxide field by applying the boundary condition at the oxide/silicon interface

$$E_x(0^+, y = 0.75\mu m) = \left(\frac{\varepsilon_{ox}}{\varepsilon_s}\right)\frac{V_{ox}(y = 0.75\mu m)}{t_{ox}}$$

The potential drop across the oxide is given by (see Fig. 4.13)

$$V_{ox}(y) = (V_{GS} + \phi_{n+}) - \phi(0, y) = V_{GS} + \phi_{n+} - V_C(y) + \phi_p$$

where we substituted the channel voltage $V_C(y)$ using Eq. (4.32). We can calculate the channel voltage at $y = 0.75\ \mu m$ from Eq. (4.49)

$$V_C = (1.5 - 0.6) - \sqrt{(1.5 - 0.6)^2 - 2(1.5 - 0.6 - 0.45)(0.9)(0.75/1.5)} = 0.26\ V$$

Note that the channel voltage is much less than $V_{DS}/2 = 0.45$ V at the halfway point in the channel, which is consistent with the normalized plots in Fig. 4.16. The vertical electric field at $y = 0.75\ \mu m$ is

$$E_x(0^+, y = 0.75\mu m) = \left(\frac{1}{3}\right)\frac{(1.5 + 0.55 - (0.26 - 0.42))}{1.5 \times 10^{-6} cm} = 3.3 \times 10^5\ V/cm$$

The lateral electric field at the halfway point in the channel can be found directly from Eq. (4.50)

$$E_y = \frac{-(1.5 - 0.6 - 0.45)(0.9)}{(1.5)\sqrt{(1.5 - 0.6)^2 - 2(1.5 - 0.6 - 0.45)(0.9)(0.75/1.5)}} = -4250\ V/cm$$

The ratio of the lateral field magnitude to the vertical field is $0.013 \ll 1$. The results of the gradual channel approximation justify the use of one-dimensional results from the MOS capacitor in Chapter 3.

(b) Find the vertical and lateral fields at $y = 1.45\ \mu m$.

SOLUTION

Applying the same analysis near the drain end of the channel, we find that the channel voltage is $V_C(y = 1.45\ \mu m) = 0.74$ V. The vertical electric field is

$$E_x(0^+, y = 1.4\mu m) = \left(\frac{1}{3}\right)\frac{(1.5 + 0.55 - (0.74 - 0.42))}{1.5 \times 10^{-6} cm} = 2.0 \times 10^5\ V/cm$$

which is lower than at mid-channel because the channel potential is higher near the drain. The lateral electric field magnitude increases dramatically near the drain, as can be seen from Eq. (4.50)

$$E_y = \frac{-(1.5-0.6-0.45)\,(0.9)}{(1.5)\sqrt{(1.5-0.6)^2-2\,(1.5-0.6-0.45)\,(0.9)\,(1.45/1.5)}} = -3.3\times 10^4 \text{ V/cm}$$

The ratio $|E_y|/E_x$ at $y = 1.45$ μm is about 0.16—the lateral field is much more significant at the drain end of the channel. In order to model the MOSFET accurately near the drain in saturation, it is necessary to solve numerically the two-dimensional Poisson's Equation

4.5 MOSFET CIRCUIT MODELS

In the previous section, we developed the groundwork for circuit models for the MOSFET by deriving the steady-state equations for drain current as a function of the terminal voltages. These equations are useful for finding the steady-state behavior of digital circuits and for biasing analog amplifiers. For analog applications, MOSFETs are usually biased in the constant-current (saturation) region and the circuit is designed to process incremental signals. Therefore, we will derive a small-signal model for the n-channel MOSFET in the saturation region. In order to determine the response to time-varying signals, we must include capacitances between the terminals of the saturated MOSFET. We will find that the small-signal analysis of the pn junction and MOS capacitor from Chapter 3 can be applied here to find the capacitances in the MOSFET.

After deriving the n-channel MOSFET circuit models, we can adapt the analysis to the p-channel MOSFET. In a CMOS technology, both n-channel and p-channel MOSFETs are fabricated on the same substrate. One of the transistors is placed in a well of opposite type to the substrate, which affects the device models and the options for the circuit designer.

4.5.1 Steady-State Models

The drain current equations for cutoff, triode, and saturation for an n-channel MOSFET were derived in sections 4.3 and 4.4. The addition of the factor $(1+\lambda_n V_{DS})$ in Eq. (4.58) to account for channel-length modulation in the saturation region creates a discontinuity in the drain current when $V_{DS} = V_{DS_{SAT}} = V_{GS} - V_{Tn}$. To avoid mathematical difficulties, this factor is added into the current equation for the triode region. The steady-state model for the n-channel MOSFET is

n-channel MOSFET steady-state drain current

$$\begin{aligned}
&I_D = 0 \text{ A} && (V_{GS} \le V_{Tn}) \\
&I_D = (W/L)\mu_n C_{ox}\left[V_{GS} - V_{Tn} - (V_{DS}/2)\right](1+\lambda_n V_{DS})\,V_{DS} && (V_{GS} \ge V_{Tn},\ V_{DS} \le V_{GS} - V_{Tn}) \\
&I_{D_{SAT}} = (W/2L)\,\mu_n C_{ox}\,(V_{GS} - V_{Tn})^2\,(1+\lambda_n V_{DS}) && (V_{GS} \ge V_{Tn},\ V_{DS} \ge V_{GS} - V_{Tn})
\end{aligned} \tag{4.59}$$

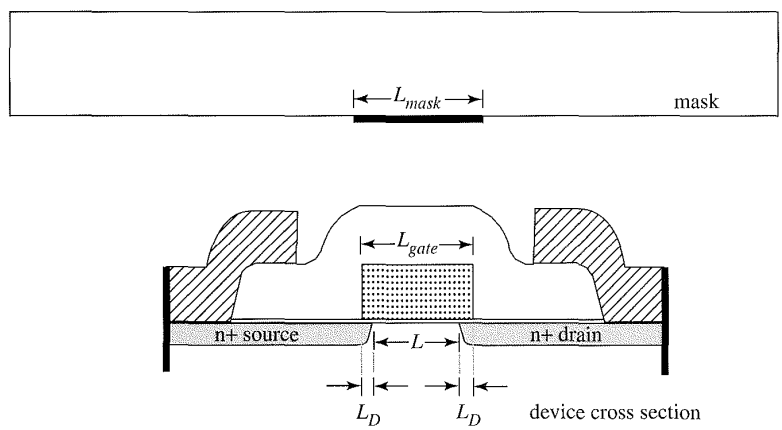
Channel-length modulation can be neglected in the triode region for hand calculations, since the drain voltage V_{DS} is relatively small and the effect has no practical significance. It is easy to forget that the drain current is also a function of the bulk voltage through the threshold voltage. Therefore, Eq. (4.46) is repeated here

$$V_{Tn} = V_{TOn} + \gamma_n \left(\sqrt{-V_{BS} - 2\phi_p} - \sqrt{-2\phi_p}\right) \tag{4.60}$$

where $\gamma_n = \left(\sqrt{2q\varepsilon_s N_a}\right) / C_{ox}$ is the backgate effect parameter.

Of the various terms in Eqs. (4.59) and (4.60), the gate oxide capacitance C_{ox} can be found directly from measurements of the gate oxide thickness t_{ox}. The MOSFET threshold voltage expression in Eq. (4.37) is very useful for indicating functional dependencies on device dimensions and parameters, although without additional terms it is not sufficiently accurate to predict V_{Tn}. These additional terms account for the contributions of oxide charges and the threshold-adjustment implant and are discussed in advanced device physics texts. In practice, experimental values for V_{TOn}, μ_n, λ_n, and γ_n are found by "curve-fitting" Eqs. (4.59) and (4.60) to the measured drain characteristics.

At first glance, the ratio W/L in Eqs. (4.59) can be read directly from the device layout. However, the channel length is often designed as the minimum allowed dimension in the fabrication technology. Since the channel length is in the denominator, small errors in L can lead to substantial errors in W/L. Figure 4.18 illustrates the need to be precise in specifying the channel length for the MOSFET. The "as drawn" channel length L_{mask} on the layout becomes the "as etched" length L_{gate} of the polysilicon gate, due to systematic lithography and etching effects. Finally, the ion-implanted source and drain regions diffuse under the polysilicon gate by a dis-



➤ **Figure 4.18** Definitions of the as-drawn gate length L_{mask} on the mask, the "as-etched" polysilicon gate length L_{gate}, and the channel length $L = L_{gate} - 2L_D$.

tance L_D, as shown in Fig. 4.18. The actual channel length, as defined earlier in Fig. 4.1, is given by

$$L = L_{gate} - 2L_D . \tag{4.61}$$

For this reason and because of the limitations of the gradual channel approximation, the ratio W/L is determined by curve-fitting to the measured drain characteristics.

4.5.2 Small-Signal Model Elements

In analog applications, a DC operating point, specified by the bias voltages V_{GS}, V_{DS}, and V_{BS}, is established so $V_{DS} > V_{DS_{SAT}} = V_{GS} - V_{Tn}$. Incremental voltages v_{gs}, v_{ds}, and v_{bs} that are much smaller in magnitude than the bias voltages perturb the operating point. The small-signal model is a circuit representation of the response of the drain current of the MOSFET to these perturbations.

The total drain current $i_D = I_D + i_d$ is a function of the gate, drain, and bulk voltages in saturation. By summing the contributions from the perturbations v_{gs}, v_{bs}, and v_{ds}, we can write a first-order expansion for the drain current

$$I_D + i_d = I_D + \left.\frac{\partial i_D}{\partial v_{GS}}\right|_Q v_{gs} + \left.\frac{\partial i_D}{\partial v_{BS}}\right|_Q v_{bs} + \left.\frac{\partial i_D}{\partial v_{DS}}\right|_Q v_{ds} \tag{4.62}$$

where Q is the DC operating point of the MOSFET, which is defined by V_{GS}, V_{BS}, and V_{DS}. The partial derivatives in Eq. (4.62) represent small-signal circuit elements and are given the following symbols

$$i_d = g_m v_{gs} + g_{mb} v_{bs} + g_o v_{ds} \tag{4.63}$$

where

$$g_m = \left.\frac{\partial i_D}{\partial v_{GS}}\right|_Q , \quad g_{mb} = \left.\frac{\partial i_D}{\partial v_{BS}}\right|_Q , \text{ and } \quad g_o = \left.\frac{\partial i_D}{\partial v_{DS}}\right|_Q \tag{4.64}$$

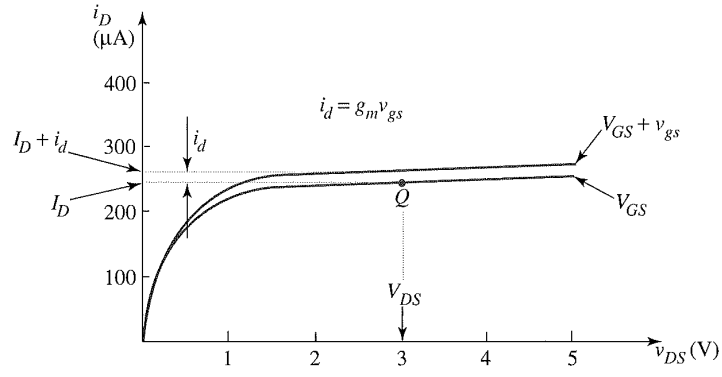
The MOSFET **transconductance** g_m connects the small-signal gate-source voltage v_{gs} to the resulting incremental drain current. The **backgate transconductance** g_{mb} represents the perturbation of the drain current by an incremental change in the bulk-source voltage. Finally, the **output conductance** g_o relates the incremental change in drain current due to a small-signal, drain-source voltage.

For an operating point Q in the saturation region, Fig. 4.19 is a graphical interpretation of the transconductance. Substituting the saturation drain current expression from Eq. (4.59), we find

$$g_m = \left.\frac{\partial}{\partial v_{GS}}\right|_Q \left\{\left(\frac{W}{2L}\right)\mu_n C_{ox} (v_{GS} - V_{Tn})^2 (1 + \lambda_n v_{DS})\right\} \tag{4.65}$$

Evaluating the partial derivative at the operating point, we find

$$g_m = \left(\frac{W}{L}\right)\mu_n C_{ox} (V_{GS} - V_{Tn}) (1 + \lambda_n V_{DS}) . \tag{4.66}$$



➤ **Figure 4.19** Graphical interpretation of the perturbation of the saturated drain current by the small-signal, gate-source voltage v_{gs} at bias point $Q = (V_{GS}, V_{DS}, V_{BS})$. Note that the drain-source and bulk-source voltages are kept constant so $v_{ds} = 0$ V, $v_{bs} = 0$ V,

A useful approximation to this equation can be made by neglecting the contribution from the channel length modulation, which will be a small error if $\lambda_n V_{DS} \ll 1$

n-channel MOSFET transconductance g_m

$$g_m \cong \left(\frac{W}{L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn}) = \sqrt{2\left(\frac{W}{L}\right)\mu_n C_{ox} I_D} \tag{4.67}$$

These results indicate that g_m is equally sensitive to the bias drain current, the channel width-to-length ratio, and $\mu_n C_{ox}$. In Chapter 7 we will see that the bipolar junction transistor has a transconductance that is linearly related to the bias current, in contrast to the square-root dependence on bias current for the MOSFET. For a typical channel mobility 215 $cm^2V^{-1}s^{-1}$, and oxide capacitance 2.3 $fF\mu m^{-2}$, bias drain current $I_D = 100$ μA, and $(W/L) = 10$, we find that the transconductance $g_m = 316$ μS (S = A/V).

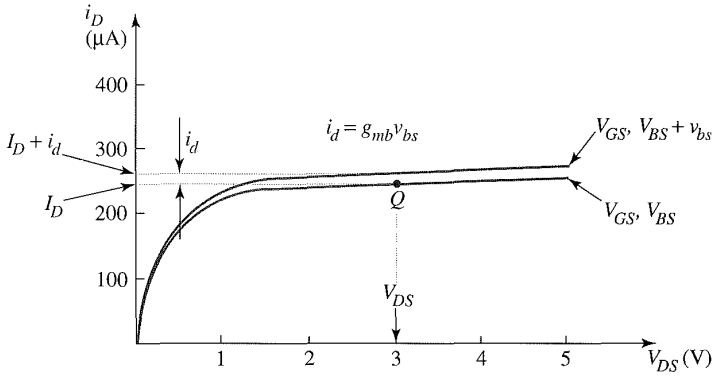
Continuing with the backgate transconductance g_{mb}, it is convenient to expand the partial derivative in Eq. (4.64) using the chain rule

$$g_{mb} = \left.\frac{\partial i_D}{\partial v_{BS}}\right|_Q = \left.\frac{\partial i_D}{\partial V_{Tn}}\right|_Q \left.\frac{\partial V_{Tn}}{\partial v_{BS}}\right|_Q \tag{4.68}$$

The first partial derivative can be evaluated by substitution of Eq. (4.59)

$$\left.\frac{\partial i_D}{\partial V_{Tn}}\right|_Q = \frac{\partial}{\partial V_{Tn}}\left(\frac{W}{2L}\right)\mu_n C_{ox}(v_{GS} - V_{Tn})^2(1 + \lambda_n v_{DS})) = -\left.\frac{\partial i_D}{\partial v_{GS}}\right|_Q = -g_m \tag{4.69}$$

where we recognize that the result is the negative of the transconductance. Partial differentiation of the threshold voltage (Eq. (4.60)) yields an expression for the backgate transconductance



➤ **Figure 4.20** Perturbation of the saturated drain current by the small-signal, bulk-source voltage v_{bs} at bias point $Q = (V_{GS}, V_{DS}, V_{BS})$. Note that the gate-source and drain-source voltages are kept constant so $v_{gs} = 0$ V, $v_{ds} = 0$ V.

$$g_{mb} = (-g_m)\left.\frac{\partial V_{Tn}}{\partial v_{BS}}\right|_Q = (-g_m)\left(\frac{-\gamma_n}{2\sqrt{-2\phi_p - V_{BS}}}\right) = \frac{\gamma_n g_m}{2\sqrt{-2\phi_p - V_{BS}}} \tag{4.70}$$

n-channel MOSFET backgate transconductance g_{mb}

where γ_n is the n-channel backgate effect parameter defined in Eq. (4.47). Figure 4.20 shows how the backgate transconductance can be measured on the drain characteristics.

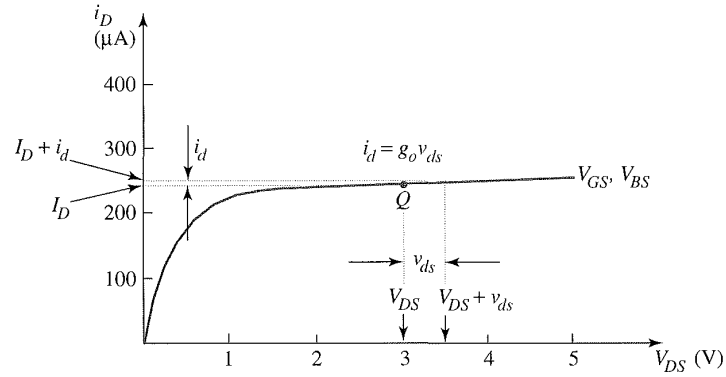
The physical meaning of the backgate transconductance is that incremental changes in the bulk-source voltage alter the electrostatics of the MOS structure and affect the channel charge exactly as do small-signal gate-source voltages. The bulk thus acts as a **backgate** (actually, "bottom gate" would be more descriptive) that is separated from the channel by a depletion layer. Using Eq. (4.70), we can solve for the ratio of the two transconductances

$$\frac{g_{mb}}{g_m} = \frac{\sqrt{2q\varepsilon_s N_a}}{2C_{ox}\sqrt{-2\phi_p - V_{BS}}} = \frac{1}{C_{ox}}\sqrt{\frac{q\varepsilon_s N_a}{2(-2\phi_p - V_{BS})}} = \frac{C_b(y=0)}{C_{ox}} \tag{4.71}$$

in which we identified the depletion capacitance at the source end of the channel, $C_b(y = 0)$ from

$$C_b(y=0) = \frac{\varepsilon_s}{X_d(y=0)} = \frac{\varepsilon_s}{\sqrt{2\dfrac{\varepsilon_s V_B(y=0)}{qN_a}}} \quad \text{and } V_B(y=0) = -2\phi_p - V_{BS}. \tag{4.72}$$

Evaluating Eq. (4.71) with the bulk shorted to the source ($v_{BS} = 0$ V) and $N_a = 10^{17}$ cm^{-3}, the backgate transconductance is about 40% of the "front gate" transconductance g_m. However, the bulk-source signal $v_{bs} = 0$ V for this case, making $g_{mb} v_{bs} = 0$ A. For a MOSFET with $V_{BS} = -1.5$ V, $g_{mb}/g_m \approx 0.25$. In order to minimize



➤ **Figure 4.21** Perturbation of the saturated drain current by the small-signal, drain-source voltage v_{ds} at bias point $Q = (V_{GS}, V_{DS}, V_{BS})$. Note that the gate-source and bulk-source voltages are kept constant, so that $v_{gs} = 0$ V, $v_{bs} = 0$ V.

the relative magnitude of the backgate transconductance, Eq. (4.71) shows it is necessary to increase the gate capacitance (by reducing the gate oxide) or reduce the depletion capacitance between the channel and the bulk. The latter can be achieved by reducing the bulk doping or increasing the magnitude of the bulk-source bias voltage.

The final small-signal element is the drain conductance g_o that represents the dependence of the drain current on the drain-source voltage

n-channel MOSFET output conductance g_o

$$g_o = \left.\frac{\partial i_D}{\partial v_{DS}}\right|_Q = \left(\frac{W}{2L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn})^2 \lambda_n \cong \lambda_n I_D \tag{4.73}$$

where we again make the approximation that $\lambda_n V_{DS} \ll 1$ at the operating point to obtain a simpler result. The graphical interpretation of the output conductance is shown in Fig. 4.21.

The channel-length modulation parameter λ_n has been found to vary inversely with the channel length

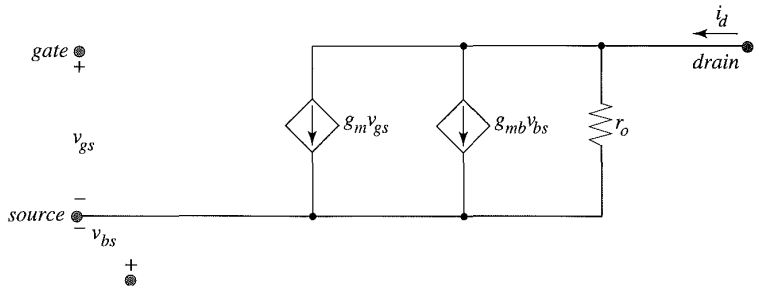
$$\lambda_n = \propto \frac{1}{L}, \tag{4.74}$$

For a MOSFET with $L = 1.5$ μm, $\lambda_n = 0.06\ \text{V}^{-1}$. At a DC bias current $I_D = 100$ μA, the inverse of the drain conductance, called the **output resistance**, is

$$(g_o)^{-1} = r_o = 150\ \text{k}\Omega \tag{4.75}$$

4.5.3 Small-Signal Circuit

We can assemble the small-signal model for steady-state perturbations on the operating point by properly connecting these three elements between the four terminals of the MOSFET. Kirchhoff's current law indicates that the three elements are con-



➤ **Figure 4.22** Steady-state small-signal model for the n-channel MOSFET in saturation.

nected in parallel at the drain node as shown in Fig. 4.22, since their contributions sum to give the small-signal drain current

$$i_d = g_m v_{gs} + g_{mb} v_{bs} + g_o v_{ds} \tag{4.76}$$

Note that the small-signal, steady-state gate current of the MOSFET is zero. In addition, the reverse-biased source-bulk junction is an open circuit in steady-state as shown in Fig. 4.22.

4.5.4 MOSFET Capacitances in Saturation

The MOS capacitor formed between the gate and substrate is the central structure of the MOSFET. The inversion-layer electron charge under the gate plays a fundamental role in the MOSFET by enabling current between the drain and source. In order to model the small-signal response to time-varying small signals, we must model the MOS capacitor and other capacitors in the transistor, when the MOSFET is operated in the saturation region.

In saturation, the channel charge is a function only of the gate-source voltage. As discussed in Sections 4.3 and 4.4, changes in the drain-source voltage have negligible effect on the channel for $V_{GS} > V_{DS_{SAT}}$. The total electron charge in the channel in saturation is given by

$$q_N(v_{GS}) = -W\int_0^L C_{ox}\left(v_{GS} - V_{Tn} - v_C(y)\right) dy \tag{4.77}$$

where we integrate the channel charge per unit area from Eq. (4.33) over the area of the channel. Symbols for total channel charge q_N, gate-source voltage v_{GS}, and channel voltage v_C are used since incremental (small) signals are present. We can change variables in the integral using Eq. (4.40)

$$dy = \left(\frac{W\mu_n C_{ox}}{i_D}\right)\left(v_{GS} - V_{Tn} - v_C\right) dv_C \tag{4.78}$$

and then integrate from the source ($y = 0$ and $v_C = 0$ V) to the drain ($y = L$ and $v_C = v_{DS} = v_{GS} - V_{Tn}$ for saturation)

$$q_N(v_{GS}) = -\left(\frac{W^2 \mu_n C_{ox}^2}{i_D}\right) \int_0^{v_{GS}-V_{Tn}} (v_{GS} - V_{Tn} - v_C^2)\, dv_C \tag{4.79}$$

After evaluation of the integral and substitution of the drain current equation in saturation, we can express the channel charge as a function of the gate-source voltage

$$q_N(v_{GS}) = -\frac{2}{3} WLC_{ox} (v_{GS} - V_{Tn}) \tag{4.80}$$

The gate charge is equal and opposite to the sum of the channel charge and the maximum bulk (depletion region) charge

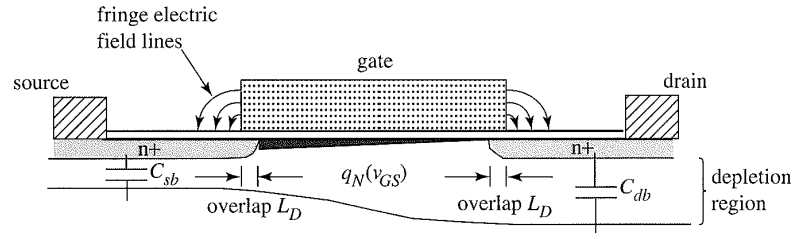
$$q_G(v_{GS}) = -q_N(v_{GS}) - Q_{B,max} = \frac{2}{3} WLC_{ox} (v_{GS} - V_{Tn}) - Q_{B,max} \tag{4.81}$$

The capacitance in saturation is the derivative of Eq. (4.81)

$$\left.\frac{dq_G}{dv_{GS}}\right|_{V_{GS} \geq V_{DS_{SAT}}} = \frac{2}{3} WLC_{ox} \tag{4.82}$$

where we have used the fact that the bulk charge is constant in saturation.

Figure 4.23 is a cross section of the MOSFET biased in saturation. Due to overlap of the source and drain diffusions and the polysilicon gate and the contribution of fringe electric fields from the gate, there is significant additional capacitance. The overlap capacitance C_{ov} is quantified as a linear capacitance proportional to the gate width, with units of fF/μm. The total gate-source capacitance in saturation is the sum of Eq. (4.82) and the overlap capacitance:



➤ **Figure 4.23** MOSFET cross section in saturation showing the channel charge $q_N(v_{GS})$ and the overlap and fringe contributions to C_{ov}. The source-bulk and drain-bulk depletion capacitances are also shown qualitatively.

$$C_{gs} = \frac{2}{3}WLC_{ox} + WC_{ov} \tag{4.83}$$

gate-source capacitance in saturation

Since the drain has no influence on the channel charge, the only contribution to the gate-drain capacitance is C_{ov}:

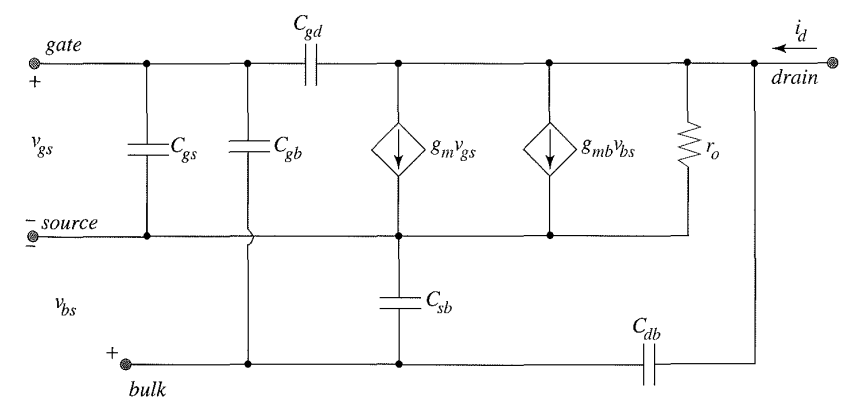
$$C_{gd} = WC_{ov} \tag{4.84}$$

gate-drain capacitance in saturation

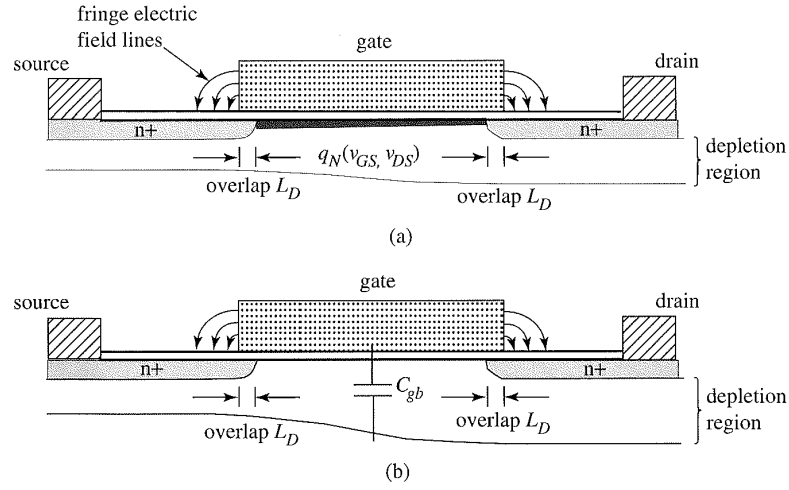
The drain-bulk and source-bulk capacitances C_{db} and C_{sb} originate from charge storage in the depletion regions between the drain and source n^+ regions and the p-type bulk. Given the area of these reverse-biased junctions $A_{diff} = L_{diff}W$ from the layout in Fig. 4.1, C_{db} and C_{sb} can be found as functions of the junction voltages V_{DB} and V_{SB} using the results from Section 3.6. Finally, the capacitance C_{gb} between gate and bulk is from the overlap of the polysilicon gate onto the field oxide region as shown in Fig. 4.1. This MOS capacitor is in depletion for normal gate voltages and its capacitance can be found from the MOS capacitor analysis in Section 3.9. It is often ignored as negligible, however, and is included only for completeness.

To predict the frequency response of the MOSFET, we add the capacitances between the terminals at the operating point. Figure 4.24 shows the schematic for the small-signal model of the intrinsic n-channel MOSFET with the capacitances added to the steady-state model in Fig. 4.22.

One final, but very important point. We have only modeled the *intrinsic* MOSFET, that is, the device without interconnections. The parasitic capacitances of the interconnections between MOSFETs are often the limiting factor in switching speed and *must* be estimated from the layout and cross section for accurate analysis of a design. Off-chip wiring and package capacitances are also critical for evaluating the performance of any integrated circuit.



➤ **Figure 4.24** Complete small-signal model for the n-channel MOSFET in saturation.



➤ **Figure 4.25** MOSFET cross section showing components of C_{gs}, C_{gd}, and C_{gb} in (a) triode and (b) cutoff.

4.5.5 MOSFET Capacitances in Triode and Cutoff

A basic understanding of the terminal capacitances in triode and in cutoff is useful for digital IC design. Figure 4.25 (a) shows the channel charge in triode, which is a function of both v_{GS} and v_{DS}. Rather than derive models for C_{gs} or C_{gd} from $q_N(v_{GS}, v_{DS})$, we split the total channel charge equally between source and drain. After adding in the overlap capacitances, the triode-region capacitances are

C_{gs} and C_{gd} in triode

$$C_{gs} = C_{gd} = \frac{WLC_{ox}}{2} + WC_{ov} = C_{gd} \tag{4.85}$$

In triode, C_{gd} is much larger than its value in saturation, whereas the gate-source capacitance C_{gs} decreases from its value in saturation.

In cutoff, there is no channel present as shown in Fig. 4.25(b). The depletion region under the gate is under gate control and the MOS capacitor theory from Section 3.9 can be applied to find C_{gb}. The gate-bulk capacitance is a sensitive function of the gate-bulk bias voltage, as seen in Fig. 3.35. Since no channel is present in cutoff, the gate-source and gate-drain capacitances are each equal to the overlap capacitance.

C_{gs} and C_{gd} in cutoff

$$C_{gs} = C_{gd} = WC_{ov} \tag{4.86}$$

This result completes our overview of charge storage in the MOSFET. The terminal capacitances are very different in the three regions of operation, as can be seen from Eqs. (4.83)-(4.86). In the case of large transients such as in digital logic, the complexities of charge storage in the MOSFET are a challenge even for numerical simulators. For small-signal analysis in the saturation region, which is the usual case in analog design, the capacitances in Eqs. (4.83) and (4.84) are reasonably accurate for hand analysis.

EXAMPLE 4.5 MOSFET small-signal model

For the n-channel MOSFET in Fig. 4.1, the channel length is $L = 2\ \mu m$, the width is $W = 30\ \mu m$, the length of the source-drain diffusions is $L_{diff} = 6\ \mu m$, the channel mobility is $\mu_n = 215\ cm^2V^{-1}s^{-1}$, the oxide thickness is $t_{ox} = 150$ Å, and the substrate doping is $N_a = 10^{17}\ cm^{-3}$, and the channel length modulation parameter is $\lambda_n = 0.05\ V^{-1}$. The overlap capacitance is $C_{ov} = 0.5$ fF/μm. The threshold voltage without backgate bias is $V_{TOn} = 1$ V. Note that this measured threshold voltage is greater than predicted by the simple MOS theory in Chapter 3 (due to the presence of a shallow threshold-adjustment implant and charge in the oxide film.)

(a) For the DC operating point Q ($V_{GS} = 1.5$ V, $V_{DS} = 1.5$ V, $V_{BS} = 0$ V), find all the small-signal model parameters in Fig. 4.24, except the gate-bulk capacitance C_{gb} that we will consider negligible.

SOLUTION

Saturation drain current:
The first step is to find the DC bias current from Eq. (4.59)

$$I_D = \left(\frac{W}{2L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn})^2(1 + \lambda_n V_{DS})$$

in which
$\mu_n C_{ox} = (215\ cm^2V^{-1}s^{-1})(3.9 \times 8.85 \times 10^{-14}\ F/cm)(1.5 \times 10^{-6}\ cm)^{-1} = 49.5\ \mu AV^{-2}$
The threshold voltage is

$$V_{Tn} = V_{TOn} + \gamma_n(\sqrt{-2\phi_p - V_{BS}} - \sqrt{-2\phi_p}) = V_{TOn} = 1\ V \text{ since } V_{BS} = 0\ V$$

The error in neglecting the channel-length modulation term in finding I_D is $(0.05)(1.5) = 0.075$, less than 10%. Therefore, we neglect it for hand calculations. Substituting $\mu_n C_{ox}$, V_{Tn}, (W/L), and V_{GS} into the saturation drain current, we find

$$I_D = \frac{1}{2}\left(\frac{30}{2}\right)(49.5\mu AV^{-2})(1.5V - 1V)^2 = 91\ \mu A$$

Steady-State Parameters:
From Eq. (4.67), the transconductance is

$$g_m = \left(\frac{30}{2}\right)(49.5\mu AV^{-2})(1.5V - 1V) = 371\ \mu S$$

The backgate effect parameter is found from Eq. (4.47)

$$\gamma_n = \frac{\sqrt{2 \cdot 1.6 \times 10^{-19} \cdot 11.7 \cdot 8.85 \times 10^{-14} F\,cm^{-1} \cdot 10^{17} cm^{-3}}}{(3.9 \cdot 8.85 \times 10^{-14} F\,cm^{-1})/(1.5 \times 10^{-6} cm)} = 0.79\ V^{1/2}$$

Substitution of γ_n and $\phi_p = -0.42$ V (from the known substrate doping) into Eq. (4.70) yields the backgate transconductance

$$g_{mb} = \frac{(0.79\ V^{1/2}) \cdot (371\mu S)}{2\sqrt{-2(-0.42V) - 0}} = 160\ \mu S$$

The output resistance is

$$r_o = \frac{1}{g_o} = \frac{1}{\frac{1}{2}\left(\frac{30}{2}\right)(49.5\mu AV^{-2})\,(1.5V - 1V)^2\,(0.05V^{-1})} = 220\ k\Omega$$

Capacitances:
The gate oxide capacitance (per unit area) is

$$C_{ox} = \frac{(3.9 \cdot 8.85 \times 10^{-14} F\,cm^{-1})}{(1.5 \times 10^{-6} cm)} = 2.3\ fF/\mu m^2$$

Substituting C_{ox} and $C_{ov} = 0.5$ fF/μm into Eq. (4.83), together with the MOSFET dimensions, we find

$$C_{gs} = \left(\frac{2}{3}\right)(30 \cdot 2\ \mu m^2)\,(2.3 fF\mu m^{-2}) + (30\ \mu m)\,(0.5 fF\mu m^{-1}) = 107\ fF$$

and the gate-drain capacitance: $C_{gd} = (0.5\ fF/\mu m)(30\ \mu m) = 15$ fF.
The remaining capacitances are the pn junction depletion capacitances C_{db} between the n^+ drain and the substrate and C_{sb} between the n^+ source and the substrate. The zero-bias depletion capacitance per unit area for both junctions is

$$C_{jo} = \varepsilon_s \sqrt{\frac{q}{2\varepsilon_s \phi_B}\left(\frac{1}{N_a} + \frac{1}{N_d}\right)^{-1}} \cong \sqrt{\frac{q\varepsilon_s N_a}{2\phi_B}}$$

from Eq. (3.79). Since the source and drain are heavily doped, the junction can be considered one-sided and the contribution of the depletion region on the n-side of the junction can be neglected. The built-in voltage is

$$\phi_B = \phi_{n+} - \phi_p = 0.55\ V - (-0.42\ V) = 0.97\ V$$

The zero-bias junction capacitance is

$$C_{jo} = \sqrt{\frac{(1.6 \times 10^{-19} C)\,(11.7)\,(8.85 \times 10^{-14} F/cm)\,(10^{17} cm^{-3})}{2\,(0.97V)}} = 0.92 fF\mu m^{-2}$$

The source junction has a bias voltage of $V_{BS} = 0$ V, which implies

$$C_{sb} = C_{jo}(WL_{diff}) = (0.92 fF\mu m^{-2})\,(30\mu m \cdot 6\mu m) = 166\ fF$$

The drain junction has a bias voltage of $V_{BD} = -V_{DS} = -1.5$ V, so its depletion capacitance is reduced (Eq. 3.80)

$$C_{db} = \frac{(WL_{diff})\,C_{jo}}{\sqrt{1 - V_{BD}/\phi_B}} = \frac{(30\mu m \cdot 6\mu m)\,(0.92 fF\mu m^{-2})}{\sqrt{1 - (-1.5V)/(0.97V)}} = 104\ fF$$

We have found the complete small-signal model for the n-channel MOSFET in Fig. 4.1 evaluated at the given operating point.

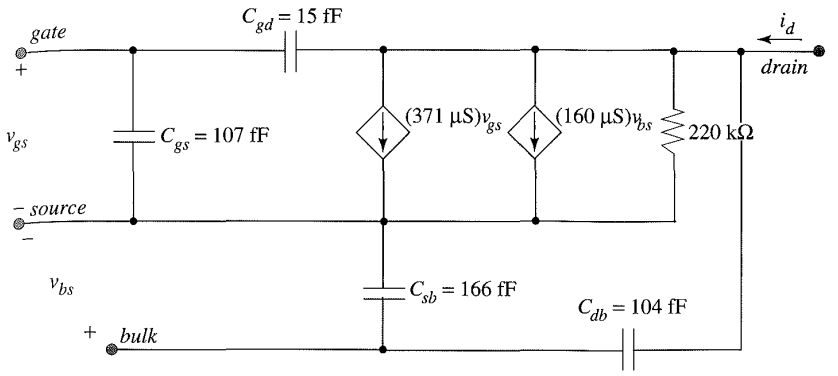


Figure Ex4.5 Small-signal model for the MOSFET in Fig. 4.1 as calculated in Example 4.5.

As we will see in Section 4.6 on the MOSFET SPICE model, the contribution of the perimeters of the source and drain diffusions can be important. The one-dimensional junction capacitance analysis in Chapter 3 is appropriate for the planar bottom of the junction. In order to estimate the perimeter capacitance, a two-dimensional or even three-dimensional analysis is needed.

(b) The product of the transconductance and the output resistance is an important quantity in small-signal amplifiers. For the MOSFET in part (a), this quantity is

$$g_m \cdot r_o = (371\mu\text{S}) \cdot (220\ \text{k}\Omega) = 82$$

For many applications, $g_m\, r_o$ must be much larger. Given the same operating point as in part (a) and the same width-to-length ratio, redesign the MOSFET so that $g_m\, r_o = 150$. What penalty is paid for improving the product?

SOLUTION

It is worthwhile to examine the functional dependence of $g_m\, r_o$ by multiplying Eq. (4.67) and the inverse of Eq. (4.73)

$$g_m \cdot r_o = \frac{\left(\frac{W}{L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn})}{\left(\frac{W}{2L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn})^2 \lambda_n} = \frac{2}{(V_{GS} - V_{Tn})\lambda_n}$$

Since the operating point is unchanged, the only term that can be adjusted is the channel length modulation parameter λ_n. From Eq. (4.74), λ_n varies inversely with channel length. Given that $\lambda_n = 0.05\ \text{V}^{-1}$ for $L = 2\ \mu\text{m}$, we find the constant of proportionality to be $0.1\ \mu\text{mV}^{-1}$. Solving for λ_n.

$$\lambda_n = \frac{2}{(g_m \cdot r_o)(V_{GS} - V_{Tn})} = \frac{0.1\ \mu\text{mV}^{-1}}{L}.$$

from which we can solve for L

$$L = \left(\frac{1}{2}\right)[150 \cdot (1.5\text{V} - 1\text{V})]\,(0.1\,\mu\text{mV}^{-1}) = 3.75\,\mu\text{m}$$

Since the width-to-length ratio has been held constant, the channel width is

$$W = (30/2)(3.75)\,\mu\text{m} = 56.25\,\mu\text{m}$$

The penalty for improving the $g_m r_o$ product is an increase in the area of the MOSFET. In addition, C_{gs} and C_{gd} are increased due to larger W and L. The total area (not counting any backgate contact) is now

$$A = W(L + 2L_{diff}) = (56.25)(3.75 + 2(6))\,\mu\text{m}^2 = 886\,\mu\text{m}^2$$

By comparison, the original device area was $(30)(2 + 2(6)) = 420\,\mu\text{m}^2$, so the area has more than doubled.

(c) An alternative approach to increasing the $g_m r_o$ product is to keep the device geometry the same and modify the operating point. We will also specify that $V_{DS} = 1.5$ V and that $V_{BS} = 0$ V. Find V_{GS} and I_D for $g_m r_o = 150$.

SOLUTION

We can solve for the required gate-source bias from

$$g_m \cdot r_o = \frac{2}{(V_{GS} - V_{Tn})\lambda_n}$$

The channel length remains $L = 2$ μm and so the channel-length modulation parameter remains $\lambda_n = 0.05\ \text{V}^{-1}$. Solving for V_{GS}, we find

$$V_{GS} = V_{Tn} + \frac{2}{\lambda_n (g_m \cdot r_o)} = 1\text{V} + \frac{2}{(0.05\text{V}^{-1})\,(150)} = 1.27\,\text{V}$$

The drain current for this gate-source bias is

$$I_D = \left(\frac{30}{2}\right)\left(\frac{49.5\,\mu\text{AV}^{-2}}{2}\right)(1.27\text{V} - 1\text{V})^2 = 27\,\mu\text{A}$$

What is the trade-off for lowering the drain current? The transconductance g_m is reduced by more than a factor of two, which degrades the frequency response of the MOSFET, as we will see in Chapter 10. The increased capacitance in (b) also has the same affect, so increasing $g_m r_o$ comes at the cost of diminished bandwidth. This example is an introduction to the trade-offs in device and circuit design in ICs.

4.5.6 p-Channel MOSFET Circuit Models

The drain current equations for a p-channel MOSFET can be derived using exactly the same approach as for the n-channel device since the basic physics is the same. For a p-channel device, the gate must be made negative with respect to the n-type substrate in order to form an inversion layer of holes. Since holes drift across the channel from source to drain in the p-channel MOSFET, the drain voltage must be negative with respect to the source, and the drain current is negative.

A practical problem for the circuit designer is to keep track of the minus signs and negative quantities in the p-channel equations. The common orientation for the p-channel MOSFET on a schematic is with the source "up"—meaning near the positive supply as shown in Fig. 4.2. By taking the source potential as a positive reference, the source-drain potential difference V_{SD}, the source-gate potential V_{SG}, and the bulk-source potential V_{BS} are all positive. By using these quantities in the p-channel drain current equations, only the drain current and the threshold voltage are negative.

p-channel MOSFET steady-state drain current

$$\begin{aligned} -I_D &= 0 \text{ A} && (V_{SG} \le -V_{Tp}) \\ -I_D &= (W/L)\,\mu_p C_{ox}[V_{SG} + V_{Tp} - (V_{SD}/2)]\,(1 + \lambda_p V_{SD})\,V_{SD} && (V_{SG} \ge -V_{Tp},\ V_{SD} \le V_{SG} + V_{Tp}) \\ -I_D &= (W/2L)\,\mu_p C_{ox}(V_{SG} + V_{Tp})^2\,(1 + \lambda_p V_{SD}) && (V_{SG} \ge -V_{Tp},\ V_{SD} \ge V_{SG} + V_{Tp}) \end{aligned} \tag{4.87}$$

The p-channel MOSFET threshold voltage is

p-channel MOSFET threshold voltage

$$V_{Tp} = V_{TOp} - \gamma_p\left(\sqrt{2\phi_n - V_{SB}} - \sqrt{2\phi_n}\right) \tag{4.88}$$

The small-signal parameters can be found from the drain current in the saturation region. Rather than carry out the derivations again, we adapt the n-channel results. The p-channel MOSFET transconductance is

p-channel MOSFET transconductance

$$g_m = \left(\frac{W}{L}\right)\mu_p C_{ox}(V_{SG} + V_{Tp}) \cong \sqrt{2\left(\frac{W}{L}\right)\mu_p C_{ox}(-I_D)} \tag{4.89}$$

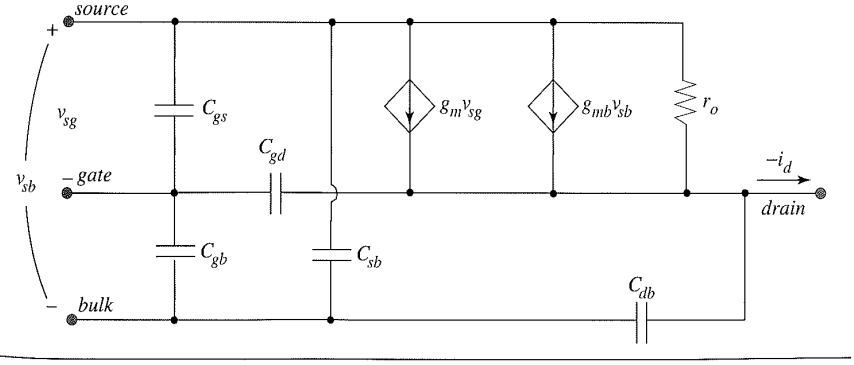
where the transconductance is defined as

$$g_m = \frac{\partial}{\partial v_{SG}}(-i_D)\bigg|_Q \tag{4.90}$$

The backgate transconductance is

p-channel MOSFET backgate transconductance

$$g_{mb} = g_m \frac{\partial V_{Tp}}{\partial v_{SB}}\bigg|_Q = \frac{\gamma_p g_m}{2\sqrt{2\phi_n - V_{SB}}} \tag{4.91}$$



➤ **Figure 4.26** Small-signal model for the p-channel MOSFET in saturation.

and the output conductance is

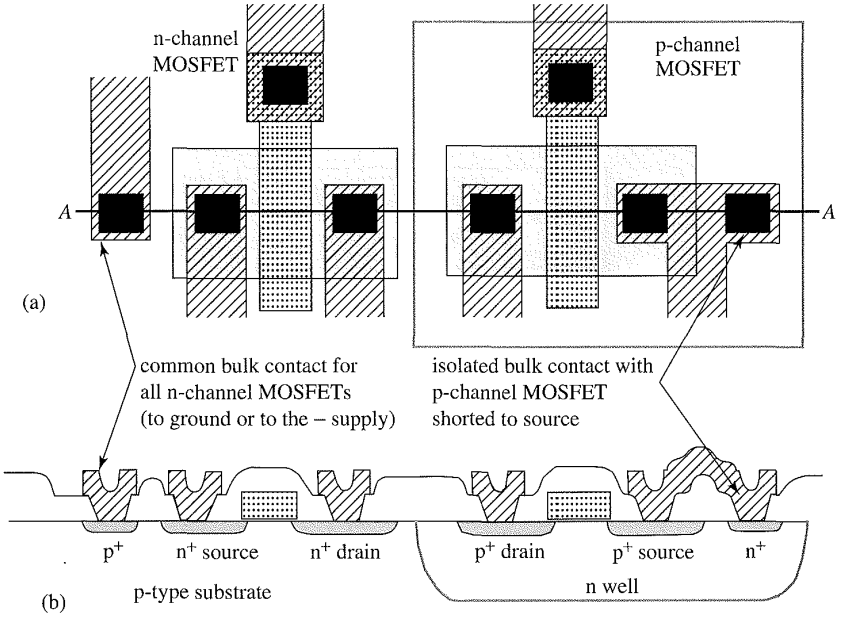
p-channel MOSFET output conductance

$$g_o = \left.\frac{\partial(-i_D)}{\partial v_{SD}}\right|_Q = \left(\frac{W}{2L}\right)\mu_p C_{ox}(V_{SG}+V_{Tp})^2\lambda_p \cong \lambda_p(-I_D) \tag{4.92}$$

Finally, the capacitances associated with the p-channel MOSFET are identical to those of its complementary device. Figure 4.26 is the small-signal model for the p-channel MOSFET with the source located at the top, which is its most common orientation.

4.5.7 MOSFET Structures in CMOS Processes

The n- and p-channel MOSFETs can be made together in a Complementary MOS (CMOS) process. Having both types of MOSFETs available has important advantages for both digital and analog circuit design. Since the two devices need a bulk region with the opposite doping type, a CMOS process includes a deep well diffusion. Figure 4.27 shows the layout and a schematic cross section of an n-well CMOS process, in which the p-channel MOSFETs are fabricated in an n-type well in a p-type substrate. The opposite scheme, the p-well CMOS process, is less common.



➤ **Figure 4.27** (a) layout and (b) cross section of an n-well CMOS technology with isolated bulks regions for p-channel MOSFETs.

All n-channel MOSFETs share a common bulk region (the p-type substrate) in an n-well technology. Multiple contacts are made to the substrate, one of which is shown in Fig. 4.27, to ensure a low-resistance connection to ground. The IC designer has the option of placing a p-channel MOSFET in its own well so that its bulk (the n-well) can be shorted to its source, as shown in Fig. 4.27. Alternatively, a large number of p-channel devices can share a single n-well.

As we will see in Chapter 8, there are circuit performance advantages if the bulk and source contacts can be shorted. For this configuration, the backgate transconductance generator does not add to the small-signal drain current since $v_{BS} = 0$ V. In an n-well process, $V_{BS} = 0$ V only for those n-channel MOSFETs that have their sources grounded, since their shared bulk terminal must be grounded. Other n-channel devices will have their thresholds shifted by the backgate effect through Eq. (4.60) and must have the backgate generator included in their small-signal models. For the p-channel MOSFETs, on the other hand, we can arrange for $v_{SB} = 0$ for a particular device by placing it in its own n-well. There is an area penalty for placing a device in an isolated well, rather than having it share a well with many other p-channel MOSFETs. As a result, an IC designer will only isolate a device when it is necessary for performance reasons.

4.6 SPICE MODELS FOR THE MOSFET

MOSFETs can be simulated using models with varying levels of complexity in SPICE. In this text, we will use the simplest model, which is referred to as the *"Level 1"* model. This model corresponds closely with the large-signal and small-signal models that have been developed in this chapter. In this section, we will introduce the Level 1 MOSFET model and define the notation used in SPICE.

4.6.1 Level 1 DC Model

SPICE models the drain-source current I_{DS} of an n-channel MOSFET in the cutoff, triode, and saturation regions by:

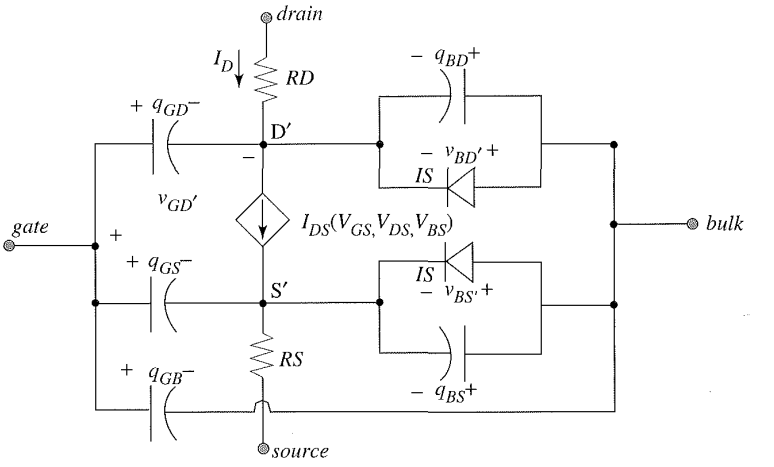
$$\begin{aligned} I_{DS} &= 0 && (V_{GS} \le -V_{TH}) \\ I_{DS} &= \frac{KP}{2}(W/L_{eff})\, V_{DS}\,[2\,(V_{GS} - V_{TH}) - V_{DS}]\,(1 + LAMBDA \cdot V_{DS}) && (0 \le V_{DS} \le V_{GS} - V_{TH}) \\ I_{DS} &= \frac{KP}{2}(W/L_{eff})\,(V_{GS} - V_{TH})^2\,(1 + LAMBDA \cdot V_{DS}) && (0 \le V_{GS} - V_{TH} \le V_{DS}) \end{aligned} \tag{4.93}$$

where the threshold voltage is given by

$$V_{TH} = V_{TO} + GAMMA\,(\sqrt{2 \cdot PHI - V_{BS}} - \sqrt{2 \cdot PHI})\,. \tag{4.94}$$

The SPICE parameter names correspond closely with the variables used in the analytical model developed in this chapter. In most cases, the SPICE parameter can be deduced from the symbol. However, the SPICE symbols for the channel length unfortunately conflict with the definitions in Fig. 4.18 and Eq. (4.61). The channel length in Eq. (4.93) is called L_{eff} in SPICE and is given by

$$L_{eff} = L - 2 \cdot LD\,, \tag{4.95}$$



➤ **Figure 4.28** SPICE large-signal model for an n-channel MOSFET, which includes nonlinear charge-storage between the gate and the other terminals and also in the drain and source depletion regions.

where L in SPICE is the length of the polysilicon gate and LD is the gate overlap of the source and drain.

Table 4.1 summarizes the DC SPICE parameter names and their equivalent analytical symbols and units, for the n-channel MOSFET. The elements in the large-signal MOSFET model are shown in Fig. 4.28.

4.6.2 Specifying MOSFET Geometry in SPICE

An n-channel MOSFET is defined by the following statement in SPICE:

M*name D G S B MODname* **L= _ W=_ AD= _ AS=_ PD=_ PS=_ NRD=_ NRS=_**

where *name* identifies the MOSFET with up to seven characters in SPICE2; *D, G, S,* and *B* are the drain, gate, source, and bulk node numbers, *MODname* is the model for the transistor, and the eight geometric parameters are defined in Fig. 4.29. The MOSFET in Fig 4.29 has different diffusion widths for the source and the drain, which though unusual can be modeled in SPICE. The geometry parameters are:

L = Polysilicon gate length,
W = Polysilicon gate width,
AD = Drain area,
AS = Source area,
PD = Perimeter of drain diffusion (not including the edge under the gate),
PS = Perimeter of source diffusion (not including the edge under the gate),
NRD = Equivalent number of squares in the drain diffusion, and
NRS = Equivalent number of squares in the source diffusion.

Table 4.1 n-Channel DC MOSFET SPICE parameters

Parameter Name (SPICE/this text)	SPICE Symbol Eqs. (4.93), (4.94)	Analytical Symbol Eqs. (4.59), (4.60)	Units
Channel length	L_{eff}	L	m
Polysilicon gate length	L	L_{gate}	m
Lateral diffusion/ gate-source overlap	LD	L_D	m
Transconductance parameter	KP	$\mu_n C_{ox}$	A/V^2
Threshold voltage / zero-bias threshold	VTO	V_{TOn}	V
Channel-length modulation parameter	$LAMBDA$	λ_n	V^{-1}
Bulk threshold/ backgate effect parameter	$GAMMA$	γ_n	$V^{1/2}$
Surface potential/ depletion drop in inversion	PHI	$-\phi_p$	V

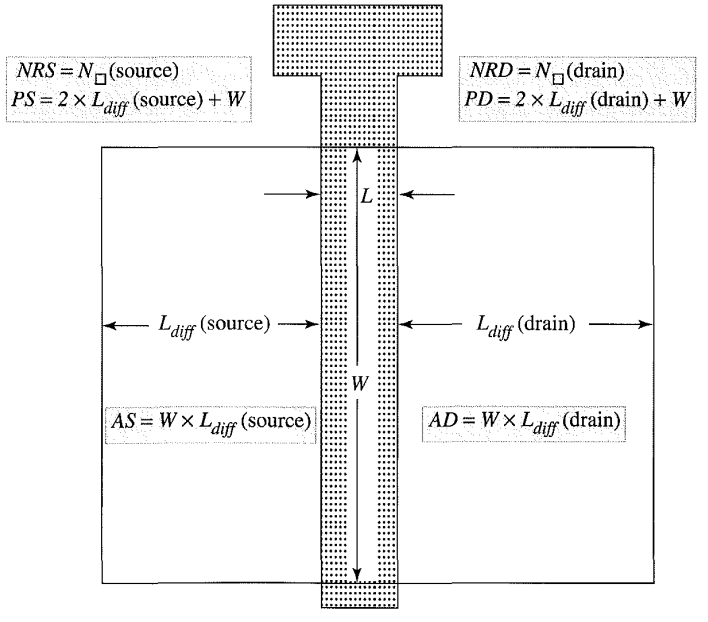


Figure 4.29 SPICE geometry parameters defined on the MOSFET layout.

Two-dimensional modeling (or curve-fitting to measurements) is needed to find the equivalent number of squares for the source and drain diffusions, which often have multiple contacts in order to minimize *NRD* and *NRS* (see the layout in Fig. 4.1, for example).

4.6.3 Level 1 Model Parameters

The MOSFET model statement in SPICE is

.MODEL *MODname* **NMOS/PMOS VTO=_ KP=_ GAMMA=_ PHI=_ LAMBDA=_ RD=_ RS=_ RSH=_ CBD=_ CBS=_CJ=_ MJ=_ CJSW=_ MJSW=_ PB=_ IS=_ CGDO=_ CGSO=_ CGBO=_ TOX=_ LD=_**

The MOSFET model must be specified as n-channel (**NMOS**) or p-channel (**PMOS**). The model parameters that have not been already defined in the DC drain current model Eqs. (4.93) and (4.94) are defined in this section.

The gate oxide thickness t_{ox} = TOX. When this parameter is specified, SPICE models the gate-source and gate-drain charge storage in cutoff, triode, and saturation. A cruder approximation for these capacitances is used if TOX is not specified. The capacitances *CGDO, CGSO*, and *CGBO* describe the overlap capacitances of the gate with the drain, source, and bulk. The units for all three parameters are F/m, with *CGDO* and *CGSO* scaling with the channel width *W*, while *CGBO* scales with the channel length *L*.

The resistance of the source and drain diffusions and their contacts can be described by directly specifying them or by giving their sheet resistance:

RD = Drain resistance (Ω)

RS = Source resistance (Ω)

RSH = Sheet resistance of the drain and source diffusions (*NRD* and *NRS* must be specified on the MOSFET statement) (Ω/□)

SPICE includes a model for the pn diode at both the drain and the source. *IS* is the saturation current of the diode. As shown in the large-signal circuit model in Fig. 4.28, the current into the drain terminal is the difference of the drain-source current I_{DS} and the diode current from the substrate. In normal operation, the diode is reverse biased and $I_D = I_{DS}$.

The final set of parameters concern the modeling of charge storage in the drain-bulk and source-bulk depletion regions. Again, SPICE allows alternative ways to specify the depletion capacitances. By specifying the parameters *CBD, CBS, PB*, and *MJ,* SPICE computes the voltage dependences of the drain-bulk and source-bulk capacitances:

$$C_{BD}(V_{BD}) = \frac{CBD}{(1 - V_{BD}/PB)^{MJ}} \tag{4.96}$$

$$C_{BS}(V_{BS}) = \frac{CBS}{(1 - V_{BS}/PB)^{MJ}} \tag{4.97}$$

CBD = Zero-bias drain-bulk junction capacitance (F)

CBS = Zero-bias source-bulk junction capacitance (F)

PB = Built-in potential for the bulk junction (V)
MJ = Bulk junction grading coefficient (dimensionless)

A more accurate simulation of these important capacitances incorporates both the planar junction capacitance and the perimeter, or sidewall, capacitance. If the parameters CJ, $CJSW$, MJ, $MJSW$, and PB are included, SPICE computes the drain-bulk and source-bulk capacitances as follows:

$$C_{BD}(V_{BD}) = \frac{CJ \cdot AD}{(1 - V_{BD}/PB)^{MJ}} + \frac{CJSW \cdot PD}{(1 - V_{BD}/PB)^{MJSW}} \tag{4.98}$$

$$C_{BS}(V_{BS}) = \frac{CJ \cdot AS}{(1 - V_{BS}/PB)^{MJ}} + \frac{CJSW \cdot PS}{(1 - V_{BS}/PB)^{MJSW}} \tag{4.99}$$

CJ = Zero-bias planar bulk junction capacitance (F/m^2)
$CJSW$ = Zero-bias sidewall bulk junction capacitance (F/m)
$MJSW$ = Sidewall junction grading coefficient (dimensionless)

The SPICE small-signal n-channel MOSFET model is identical to Fig. 4.24, except for the addition of the drain and source resistances RS and RD.

4.6.4 Example Level 1 SPICE Models for an n-well CMOS Technology

It is convenient to define "standard" SPICE models for the n-channel and p-channel MOSFETs in an n-well CMOS technology. These models can then be used throughout the text for simulating both digital and analog circuits. With the exception of the additional capacitance due to the sidewalls of the source and drain diffusions, the device parameters are similar to those derived for the n-channel MOSFET in Ex. 4.5.

For these models, we assume that the lateral diffusion $LD = L_D = 0$ μm, to avoid confusion in specifying the channel length. The Level 1 models are not so accurate that the omission of L_D will make a significant difference, in any case. The length of the source and drain diffusions is $L_{diff} = 6$ μm. Therefore, the drain and source areas and perimeters are given by

$$AD = AS = L_{diff} \cdot W = (6 \cdot W) \times 10^{-12}\ \text{m}^2 \text{ and} \tag{4.100}$$

$$PD = PS = 2L_{diff} + W = (12 + W) \times 10^{-6}\ \text{m}, \tag{4.101}$$

where W is in μm.

The gate-bulk overlap capacitance will be neglected by letting $CGBO = 0$ F/m (the default value in SPICE). Finally, the source and drain resistances are small and will be omitted by letting $RD = RS = 0\ \Omega$.

Table 4.2 describes the MOSFET parameters for the example n-well process.

■ **Table 4.2** n-well CMOS Level 1 SPICE Models.

Level 1 SPICE Parameter	n-channel MOSFET	p-channel MOSFET	Units
Gate oxide thickness *TOX*	150	150	Å
Transconductance parameter *KP*	50×10^{-6}	25×10^{-6}	A/V^2
Threshold voltage *VTO*	1.0	−1.0	V
Channel-length modulation parameter *LAMBDA*	$0.1/L$ *L* in μm	$0.1/L$ *L* in μm	V^{-1}
Bulk threshold parameter *GAMMA*	0.6	0.6	$V^{1/2}$
Surface potential *PHI*	0.8	0.8	V
Gate-drain overlap capacitance *CGDO*	5×10^{-10}	5×10^{-10}	F/m
Gate-source overlap capacitance *CGSO*	5×10^{-10}	5×10^{-10}	F/m
Zero-bias planar bulk depletion capacitance *CJ*	10^{-4}	3×10^{-4}	F/m^2
Zero-bias sidewall bulk depletion capacitance *CJSW*	5×10^{-10}	3.5×10^{-10}	F/m
Bulk junction potential *PB*	0.95	0.95	V
Planar bulk junction grading coefficient *MJ*	0.5	0.5	[ ]
Sidewall bulk junction grading coefficient *MJSW*	0.33	0.33	[ ]

For n-channel and p-channel MOSFETs with $L = 3$ μm, the parameters in Table 4.2 lead to the following model statements:

```
.MODEL MODN NMOS LEVEL = 1 VTO = 1 KP = 50U LAMBDA = .033 GAMMA = .6
+ PHI = 0.8 TOX = 1.5E-10 CGDO = 5E-10 CGSO = 5E-10 CJ = 1E-4 CJSW = 5E-10
+ MJ = 0.5 PB = 0.95
.MODEL MODP PMOS LEVEL = 1 VTO = -1 KP = 25U LAMBDA = .033 GAMMA = .6
+ PHI = 0.8 TOX = 1.5E-10 CGDO = 5E-10 CGSO = 5E-10 CJ = 3E-4 CJSW = 3.5E-10
+ MJ = 0.5 PB = 0.95
```

An n-channel MOSFET M_1 with $W = 150$ μm and $L = 3$ μm has source and drain areas and perimeters that are found from Eqs. (4.100) and (4.101)

$AD = AS = 9 \times 10^{-10}$ m^2 and $PD = PS = 1.62 \times 10^{-4}$ m

Transistor M_1 is specified in SPICE by

```
M1 4 3 2 1 MODN W = 150U L = 3U AD = 9E-10 AS = 9E-10 PD = 1.62E-4 PD = 1.62E-4
```

where the drain is node 4, the gate is node 3, the source is node 2, and the bulk is node 1.

➤ SUMMARY

This chapter has covered the device physics and device modeling of the MOS field-effect transistor, which is by far the most important microelectronic device. It is essential that you grasped the basic device operation and the large- and small-signal models, since the MOSFET will be used in digital and analog circuits throughout this text. Also, a good understanding of the circuit models for the MOSFET will be helpful in learning how to model the other important multi-terminal device—the bipolar junction transistor.

The most important topics in Chapter 4 are:

- The piecewise modeling of the n-channel and p-channel MOSFETs in the three regions of operation: cutoff, triode, and saturation.
- The relationship between drain current, channel charge, and drift velocity.
- The key assumption behind the gradual channel approximation.
- The channel electron charge variation from source to drain in the triode region, neglecting the body effect.
- The concept of saturation.
- The backgate effect on the threshold voltage of the MOSFET.
- MOSFET capacitances in the three regions of operation.
- Small-signal models for the n-channel and the p-channel MOSFETs.
- The Level 1 SPICE model for the MOSFET.

➤ FURTHER READING

MOSFET Device Structures

1. W. Maly, *Atlas of VLSI Technologies*, Addison-Wesley, 1986. Has layouts for simple MOS circuits in several technologies.

MOSFET Device Physics

1. R. F. Pierret, *Field Effect Devices*, Modular Series on Solid State Devices, Vol. IV, R. F. Pierret and G. W. Neudeck, eds., Addison-Wesley, 1983, Chapter 5 has a good discussion of the MOSFET at slightly more advanced level.
2. R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd ed., Wiley, 1986. Chapter 9 has a good description of basic MOSFET operation; derivations use energy bands rather than potential.

MOSFET SPICE Models

1. A. Vladimirescu, *The SPICE Book*, Wiley, 1994. Chapter 3 has a complete description of the Level 1 model; selected examples of the more advanced MOSFET models are detailed in an appendix.

➤ PROBLEMS

EXERCISES

Unless stated otherwise, use the following parameters in the exercises and problems.

n-channel MOSFET	p-channel MOSFET
$\mu_n C_{ox} = 50\ \mu A/V^2$	$\mu_p C_{ox} = 25\ \mu A/V^2$
$V_{TOn} = 1.0\ V$	$V_{TOp} = -1.0\ V$
$\gamma_n = 0.6\ V^{1/2}$	$\gamma_p = 0.6\ V^{1/2}$
$\lambda_n = (0.1/L)\ V^{-1}$, L in μm	$\lambda_p = (0.1/L)\ V^{-1}$, L in μm
$\phi_p = -0.42\ V$	$\phi_n = 0.42\ V$

E4.1 For the n-channel MOSFET with terminal voltages in Fig. E4.1, determine (a) the operating region (cutoff, triode, or saturation) and (b) the drain current I_D.

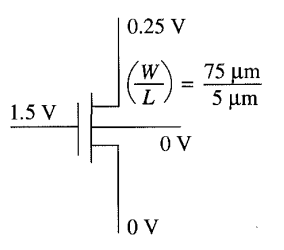
E4.2 Repeat E4.1 for the terminal voltages in Fig. E4.2.

E4.3 Repeat E4.1 for the terminal voltages in Fig. E4.3.

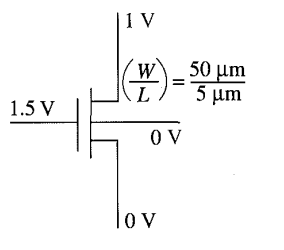
E4.4 For the n-channel MOSFET with terminal voltages in Fig. E4.4, determine (a) the operating region (cutoff, triode, or saturation) and (b) the drain current I_D. Note that the bulk-source voltage V_{BS} is *not* zero.

E4.5 Repeat E4.4 for the terminal voltages in Fig. E4.5.

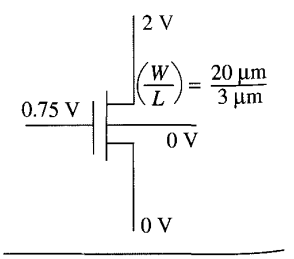
E4.6 Repeat E4.4 for the terminal voltages in Fig. E4.6.



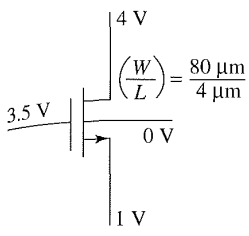
➤ **Figure E4.1**



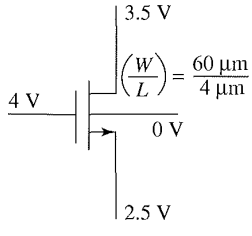
➤ **Figure E4.2**



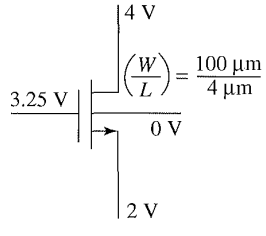
➤ **Figure E4.3**



➤ **Figure E4.4**



➤ **Figure E4.5**



➤ **Figure E4.6**

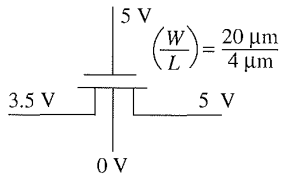
E4.7 For the n-channel MOSFET with terminal voltages in Fig. E4.7, (a) label the source and the drain terminals, (b) determine in which region the transistor is operating (cutoff, triode, or saturation), and (c) find the drain current I_D.

E4.8 Repeat E4.7 for the terminal voltages in Fig. E4.8.

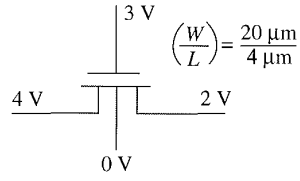
E4.9 For the p-channel MOSFET with terminal voltages in Fig. E4.9, determine (a) the operating region (cutoff, triode, or saturation) and (b) the drain current $-I_D$.

E4.10 Repeat E4.9 for the terminal voltages in Fig. E4.10.

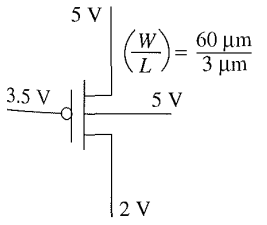
E4.11 Repeat E4.9 for the terminal voltages in Fig. E4.11.



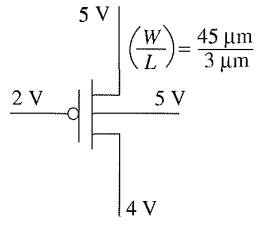
➤ **Figure E4.7**



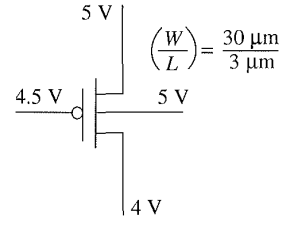
➤ **Figure E4.8**



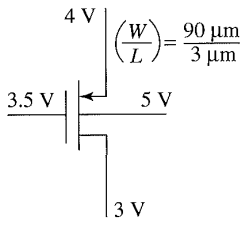
➤ **Figure E4.9**



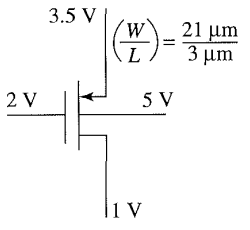
➤ **Figure E4.10**



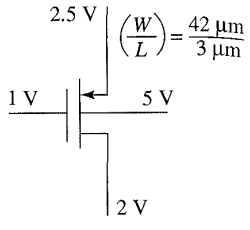
➤ **Figure E4.11**



➤ **Figure E4.12**



➤ **Figure E4.13**



➤ **Figure E4.14**

E4.12 For the p-channel MOSFET with terminal voltages in Fig. E4.12, determine (a) the operating region (cutoff, triode, or saturation) and (b) the drain current $-I_D$. Note that the source-bulk voltage V_{SB} is *not* zero.

E4.13 Repeat E4.12 for the terminal voltages in Fig. E4.13.

E4.14 Repeat E4.12 for the terminal voltages in Fig. E4.14.

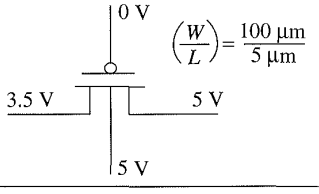
E4.15 For the p-channel MOSFET with terminal voltages in Fig. E4.15, (a) label the source and the drain terminals, (b) determine in which region the transistor is operating (cutoff, triode, or saturation), and (c) find the drain current I_D.

E4.16 Repeat E4.15 for the terminal voltages in Fig. E4.16.

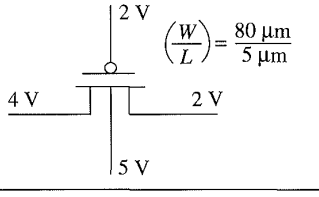
E4.17 Consider an n-channel MOSFET with $(W/L) = 25\ \mu m/2.5\ \mu m$.

(a) Plot the drain current I_D versus V_{GS} for $V_{DS} = 250$ mV and $V_{BS} = 0$. Note that the threshold voltage can be found from this plot of the MOSFET in the triode region.

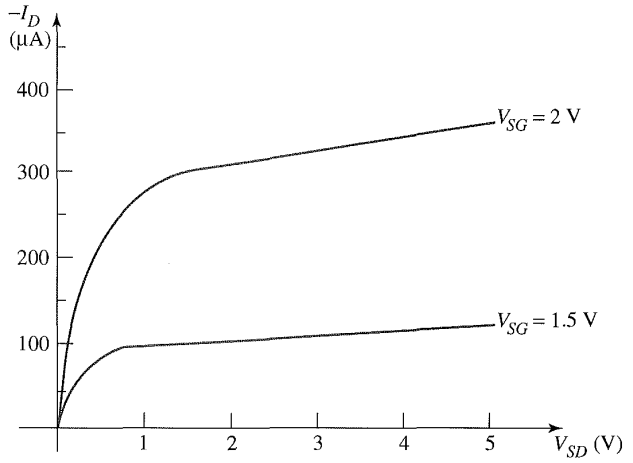
(b) Plot the square root of drain current $(I_D)^{1/2}$ versus V_{GS} for $V_{DS} = 5$ V and $V_{BS} = 0$. The threshold voltage can also be extracted from this plot of the MOSFET in the saturation region. Due to second-order effects, the measurement will differ somewhat from the threshold found in the triode region in part (a).



➤ **Figure E4.15**



➤ **Figure E4.16**



➤ **Figure E4.18**

E4.18 Figure E4.18 is a plot of two drain characteristics of a p-channel MOSFET, for $V_{BS} = 0$.

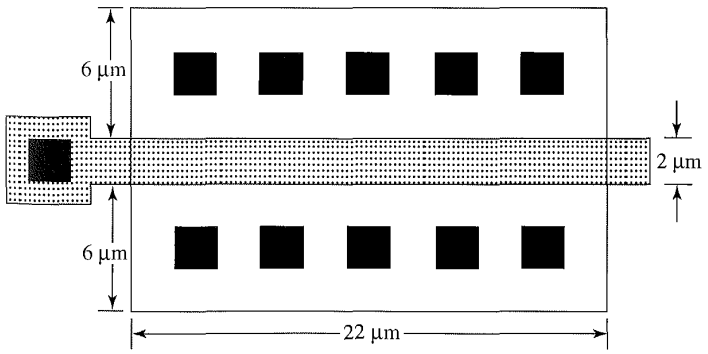
(a) From these measurements, what is the threshold voltage, V_{Tp}?

(b) Estimate the channel-length modulation parameter, λ_p.

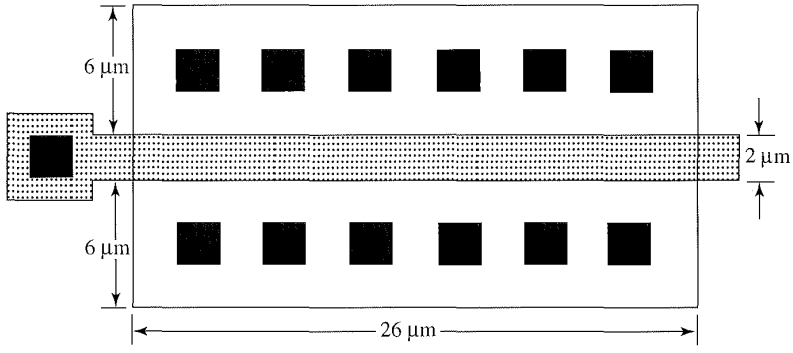
E4.19 The n-channel MOSFET shown in layout in Fig. E4.19 is biased at the operating point: $V_{GS} = 3.5$ V, $V_{DS} = 5$ V, $V_{BS} = -1$ V. For this problem, include $L_D = 0.1$ μm in finding the channel length L from the layout and let $C_{ox} = 1.42$ fF/μm^2.

(a) Find the small-signal parameters g_m, g_{mb}, and r_o at this operating point.

(b) Find the capacitances C_{gs}, C_{gd}, C_{db}, and C_{sb}. In calculating the overlap capacitances, only consider the under-diffusion of the drain and source



➤ **Figure E4.19**



➤ **Figure E4.20**

diffusions by L_D. Also, you can neglect the sidewall capacitance and use a substrate doping $N_a = 10^{17}$ cm^{-3}.

E4.20 The p-channel MOSFET shown in layout in Fig. E4.20 is biased at the operating point: $V_{SG} = 1.5$ V, $V_{SD} = 4$ V, $V_{SB} = -2$ V. For this problem, include $L_D = 0.1$ μm in finding the channel length L from the layout and let $C_{ox} = 1.42$ fF/μm^2.

(a) Find the small-signal parameters g_m, g_{mb}, and r_o at this operating point.

(b) Find the capacitances C_{gs}, C_{gd}, C_{db}, and C_{sb}. In calculating the overlap capacitances, only consider the under-diffusion of the drain and source diffusions by L_D. Also, you can neglect the sidewall capacitance and use a substrate doping $N_a = 10^{17}$ cm^{-3}.

E4.21 A p-channel MOSFET has $W/L = 500$ μm/2.5 μm. The source and drain diffusions have a length $L_{diff} = 6$ μm. The operating point is: $V_{SG} = 2.5$ V, $V_{SD} = 4$V, $V_{SB} = 0$ V. For this problem, let $L_D = 0.1$ μm and $C_{ox} = 1.42$ fF/μm^2.

(a) Find the small-signal parameters g_m, g_{mb}, and r_o at this operating point.

(b) Find the capacitances C_{gs}, C_{gd}, C_{db}, and C_{sb}. In calculating the overlap capacitances, only consider the under-diffusion of the drain and source diffusions by L_D. Also, you can neglect the sidewall capacitance and use a substrate doping $N_a = 10^{17}$ cm^{-3}.

E4.22 An n-channel MOSFET with $(W/L) = 20$ μm/3 μm is biased at $V_{GS} = 2$ V, $V_{DS} = 3.5$ V, and $V_{BS} = 0$ V. A small signal $v_{gs}(t) = 20$ mV cos(2π 1000 t) is added in series with V_{GS}.

(a) Find the DC drain current I_D.

(b) Find the small-signal drain current $i_d(t)$.

(c) Plot the total current $i_D(t)$ from your answers in (a) and (b).

E4.23 In a particular CMOS process, the gate oxide thickness is $t_{ox} = 150$ Å ±20 Å and the substrate doping is $N_a = 8 \times 10^{16}$ cm^{-3} ±10^{16} cm^{-3}. Using this process, we bias an n-channel MOSFET with dimensions $W = 100$ μm and $L = 2.5$ μm at the operating point: $V_{GS} = 3$ V, $V_{DS} = 4$ V, and $V_{BS} = -1.5$ V.

(a) What is the average threshold voltage and its uncertainty?

(b) What is the DC drain current I_D and its uncertainty at this operating point?

(c) What is the transconductance and its uncertainty at this operating point?

(d) What is the backgate transconductance and its uncertainty at this operating point?

PROBLEMS

P4.1 Figure P4.1 plots the drain current I_D of an n-channel MOSFFET as a function of the gate-source voltage V_{GS}, for a drain-source voltage $V_{DS} = 10$ mV and at three values of backgate bias V_{BS}. From the layout, you also know that $W/L = 10$.

(a) Extract the following device parameters from these measurements: the substrate doping concentration N_a, the oxide thickness t_{ox}, and the channel mobility μ_n.

(b) Estimate the uncertainty in your parameters, given that the current measurements have an uncertainty of ± 1 μA.

P4.2 In a particular CMOS technology, the substrate doping is $N_a = 4 \times 10^{17}$ cm^{-3}, the gate oxide is $t_{ox} = 150$ Å, and n$^+$ polysilicon is used for the gate.

(a) Calculate the surface potential ϕ_{so} in thermal equilibrium.

(b) Calculate the threshold voltage V_{TOn}.

(c) If we require $V_{TOn} = 0.7$ V, what oxide thickness is required?

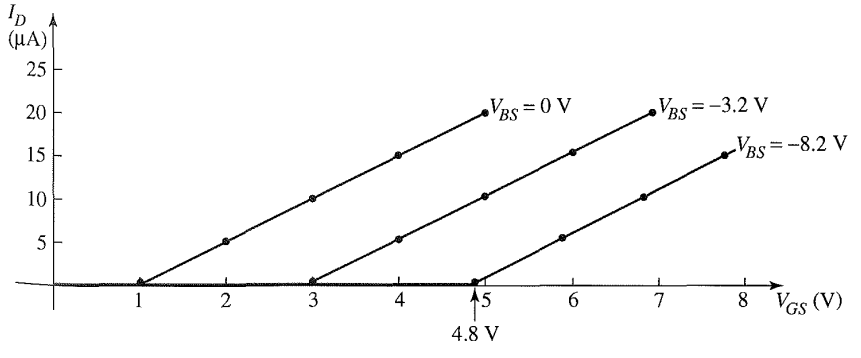
(d) If the oxide thickness couldn't be changed, what must the substrate doping concentration become in order to achieve $V_{TOn} = 0.7$ V.

P4.3 In a particular CMOS technology, the n-well doping is $N_d = 2.5 \times 10^{17}$ cm^{-3}, the gate oxide is $t_{ox} = 200$ Å, and n$^+$ polysilicon is used for the gate.

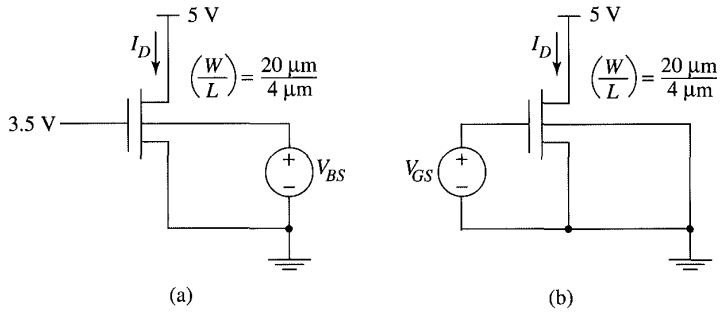
(a) Calculate the surface potential ϕ_{so} in thermal equilibrium.

(b) Calculate the threshold voltage V_{TOp}.

(c) If we require $V_{TOp} = -1.5$ V, what oxide thickness is required?



➤ **Figure P4.1**



➤ **Figure P4.6**

(d) If the oxide thickness couldn't be changed, what must the substrate doping concentration become in order to achieve $V_{TOp} = -1.5$ V.

P4.4 In a particular CMOS technology, the n-well doping is $N_d = 2.5 \times 10^{17}$ cm^{-3}, the gate oxide is $t_{ox} = 200$ Å, and p^+ polysilicon is used for the gate. The thermal-equilibrium potential of p^+ polysilicon is $\phi_{p+} = -550$ mV.

(a) Calculate the surface potential ϕ_{so} in thermal equilibrium.

(b) Calculate the threshold voltage V_{TOp}.

(c) If we require $V_{TOp} = -0.7$ V, what oxide thickness is required?

(d) If the oxide thickness couldn't be changed, what must the substrate doping concentration become in order to achieve $V_{TOp} = 0.7$ V.

P4.5 Silicon nitride (Si_3N_4, $\varepsilon_n = 6.64 \times 10^{13}$ F/cm) is a potential replacement for silicon dioxide as a gate insulator material. Assume the same operating point and transistor dimensions as used in Ex. 4.5.

(a) What thickness of Si_3N_4 is needed for $V_{TOn} = 0.6$ V, if the substrate doping is $N_a = 10^{17}$ cm^{-3}?

(b) Find the backgate effect parameter, γ_n for the silicon nitride gate insulator thickness from (a).

(c) Find the transconductance, backgate transconductance, and output resistance at the operating point and compare with the values for an SiO_2 from Ex. 4.5.

(d) Find the MOSFET capacitances and compare to those for the SiO_2 gate.

P4.6 Figure P4.6(a) shows an n-channel MOSFET with $V_{GS} = 3.5$ V and $V_{DS} = 5$ V with a variable bulk-source voltage V_{BS}. Figure P4.6(b) shows an n-channel MOSFET with $V_{BS} = 0$ V and $V_{DS} = 5$ V with a variable gate-source voltage V_{GS}.

(a) Plot the drain current as a function of V_{BS} for the circuit in Fig. P4.6(a), over the range $-5\ V < V_{BS} < 0$.

(b) Plot the drain current as a function of V_{GS} for the circuit in Fig. P4.6(b), over the range $0\ V < V_{GS} < 5$ V.

P4.7 This problem uses the results of P4.6 to extract small-signal parameters.

(a) From the $I_D(V_{BS})$ plot in P4.6(a), find graphically the backgate transconductance g_{mb} at the operating point $V_{GS} = 3.5$ V, $V_{DS} = 5$ V, $V_{BS} = -2$ V. Compare with the value you calculate from the analytical expression.

(b) From the plot of $I_D(V_{GS})$ in P4.6(b), find graphically the transconductance g_m at the operating point $V_{GS} = 3.5$ V, $V_{DS} = 5$ V, $V_{BS} = 0$ V. Again, compare your result with the analytical expression.

P4.8 A large-value, voltage-controlled resistor is useful in some analog ICs. Using the standard parameters, design a p-channel MOSFET with a resistance of $R = 1$ MΩ for $V_{GS} = 1.5$ V.

(a) What are the required channel width and length?

(b) Find the capacitance between the channel and the gate and between the channel and the substrate. Assume that the source and drain diffusion lengths are $L_{diff} = 6$ μm. Note that the device is operating in the triode region for use as a voltage-controlled resistor.

(c) Model the transistor as a low-pass filter and find the break (–3 dB) frequency.

P4.9 For the case of a p-channel MOSFET, apply the gradual channel approximation. Assume $C_{ox} = 1.42$ fF/μm^2, $V_{FB} = 1$ V, and $V_{TP} = -1$ V.

(a) Find the channel voltage $V_C(y)$ between the source and drain, assuming that $V_{SD} \leq V_{SG} + V_{Tp}$.

(b) Find the lateral electric field $E_y(y)$ between the source and drain for $V_{SD} \leq V_{SG} + V_{Tp}$.

(c) Find the vertical electric field $E_x(y)$ at the silicon surface ($x = 0$) between the source and drain for $V_{SD} \leq V_{SG} + V_{Tp}$.

(d) For $V_{SD} = (3/4)(V_{SG} + V_{Tp})$ plot the ratio E_y/E_x between the source and the drain.

P4.10 In the gradual-channel approximation, we found Eq. (4.36) prior to eliminating the variation of the bulk charge.

(a) Starting with Eq. (4.36), separate variables and find a new triode region equation for the drain current as a function of V_{GS} and V_{DS}. You can assume that $V_{BS} = 0$.

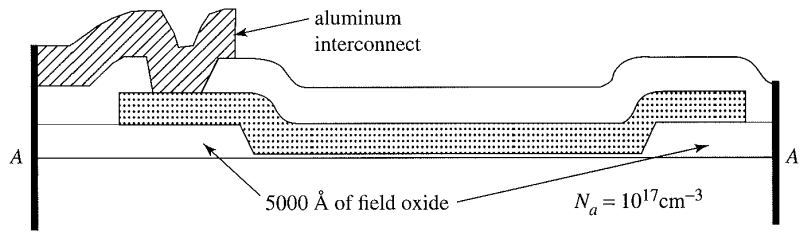
(b) What is the saturation drain current, from the triode equation in (a)?

(c) For $W/L = 25$ μm/2.5 μm, plot the drain characteristics for $V_{GS} = 1.5$ V, 2 V, 3 V, and 4 V over the range $0 < V_{DS} < 5$ V using your results from (a) and (b). In saturation, you can use the factor $(1 + \lambda_n V_{DS})$ in modeling channel-length modulation. On the same graph, plot Eqs. (4.59) and compare, using $N_a = 10^{17}$cm^{-3}, $C_{ox} = 1.42$ fF/μm^2, and standard Level 1 SPICE parameters.

P4.11 For this problem, use the approach used to derive Eq. (4.80) for the channel-charge in saturation.

(a) Find an expression for the channel charge in the triode region, $q_N(v_{GS}, v_{DS})$.

(b) From your result in (a), find expressions for the gate-source capacitance C_{gs} and the gate-drain capacitance C_{gd} in the triode region. Omit the overlap capacitance.



➤ **Figure P4.13**

(c) Plot the gate-source capacitance as a function of V_{DS} over the range 0 to $V_{GS} - V_{Tn}$. Show that the triode result converges with the saturation expression for C_{gs} at $V_{DS} = V_{GS} - V_{Tn}$.

P4.12 This problem concerns itself with the question of how small a "small signal" must be for the linearized model to be adequate. Consider an n-channel MOSFET with width-to-length ratio of 25 μm/5 μm biased at $V_{GS} = 2.5$ V, $V_{DS} = 4$ V, and $V_{BS} = 0$ V.

(a) Find the DC drain current I_D and the transconductance g_m.

(b) Expand the expression for the total drain current in a power series, retaining only the DC, linear, and quadratic terms.

(c) The small-signal model neglects the quadratic term. For a particular application, we can only tolerate a quadratic term that is 5% of the linear term. How large is the "small-signal" v_{gs} that meets this requirement, for the given bias condition?

P4.13 For the layout in Fig. E4.19, the polysilicon gate overlaps thick (field) oxide on both sides of the channel, as shown in cross section *AA* in Fig. P4.13. The field oxide thickness is $t_{fox} = 5000$ Å. The substrate is doped at $N_a = 10^{17}$ cm^{-3} in the field region. The n-channel MOSFET is biased at $V_{GS} = 2.5$ V, $V_{DS} = 4$ V, and $V_{BS} = 0$ V.

(a) What is the threshold voltage of the gate-bulk capacitor?

(b) Find the gate-bulk capacitor C_{gb} at the given operating point. Note that there is an overlap area at both ends of the gate. You will need to apply the MOS capacitor theory from Chapter 3.

(b) Sketch the gate-bulk capacitance as a function of the bulk-source bias V_{BS} over the range $-5 < V_{BS} < 0$.

P4.14 The field oxide thickness in a CMOS process (see Fig. 4.27) must be thick enough so that neither the p substrate nor the n-type well inverts under the extensions of the n^+ polysilicon gate. Assume that we have a case of symmetrical doping ($N_a = 10^{17}$ cm$^{-3} = N_d$) in the field regions for the substrate and the well.

(a) What field oxide thickness is needed to have the field threshold voltage equal to $V_{Tnf} = 4$ V for the substrate?

(b) What field oxide thickness is needed to have the field threshold voltage equal to $V_{Tpf} = -4$ V for the n-well?

P4.15 In a CMOS process it is sometimes desirable to match, as much as possible, the circuit models of the n-channel and the p-channel MOSFETs. Assume that both devices have channel lengths of $L_n = L_p = 2$ μm and that the n-channel MOSFET has a channel width $W_n = 20$ μm.

(a) In order to have identical saturation currents for the same gate-source drive ($V_{GS} = V_{SG}$), what is the width of the p-channel transistor?

(b) Given the dimensions from (a), lay out both devices assuming $L_{diff} = 6$ μm.

(c) Derive all the small-signal parameters (including capacitances) for the n-channel and the p-channel transistors.

P4.16 In some cases, it is helpful to have a small-signal model for the MOSFET in the triode region. Using Eq. (4.59) for the triode region.

(a) Find the transconductance g_m as a function of operating point in the triode region.

(b) Find the output resistance r_o as a function of operating point in the triode region.

(c) Plot the output conductance g_o as a function of V_{DS} and show that it approaches its saturation value at $V_{DS} = V_{GS} - V_{Tn}$.

P4.17 Figure P4.17 shows the layout of an n-channel MOSFET, in which the channel width tapers from $W_s = 50$ μm at the source to $W_d = 20$ μm at the drain. The channel length is $L = 10$ μm.

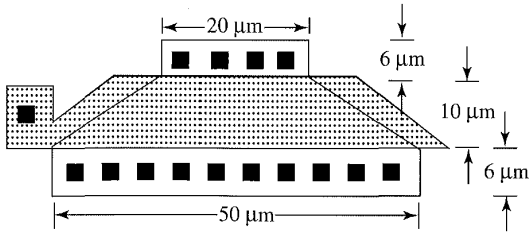
(a) Adapt the gradual channel approximation in Eq. (4.36) to this new geometry and find an expression for the drain current in the triode region.

(b) What is the effective channel width W_{eff} for this MOSFET?

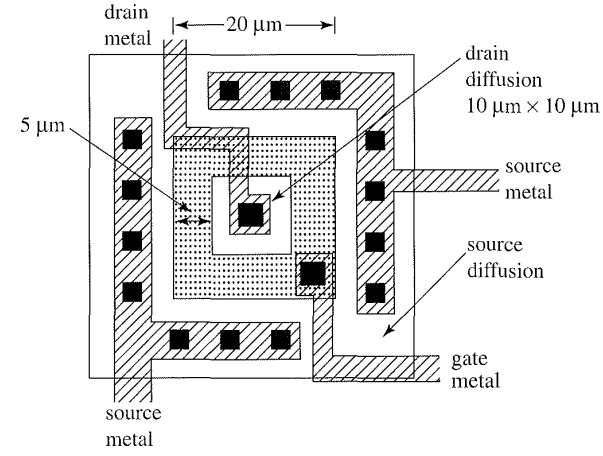
(c) Find the gate-drain and drain-bulk capacitances for this layout and compare to those for a standard MOSFET with $W = W_{eff}$. The overlap capacitance between gate and drain is $C_{gdo} = 0.5$ fF/μm and the operating point is $V_{GS} = 2$ V, $V_{DS} = 4$ V, and $V_{BS} = 0$ V.

P4.18 Figure P4.18 is a square MOSFET layout, which is an approach that is sometimes used to minimize the drain capacitance.

(a) Derive an approximate equation for the saturation drain current for this n-channel MOSFET. What is an estimate of the effective gate width W_{eff}? *Hint*: consider the role of the corners.



➤ **Figure P4.17**



➤ **Figure P4.18**

(b) Find the gate-drain and drain-bulk capacitances for this layout and compare to those for a standard MOSFET with $W = W_{eff}$. The overlap capacitance between gate and drain is $C_{gdo} = 0.5$ fF/μm and the operating point is $V_{GS} = 2$ V, $V_{DS} = 3$ V, and $V_{BS} = 0$ V. The CVD SiO_2 thickness is 5000 Å, which is needed to find C_{gd}.

Design Problems

D4.1 MOSFETs are useful as switches in power supplies. For this application, the *on resistance* of the transistor is a critical parameter. In this problem, the goal is to design an n-channel MOSFET to satisfy:

$R_{on} = 0.01\ \Omega$ for $V_{GS} = 5$ V

Given that the maximum oxide field is $E_{ox,\max} = 3 \times 10^6$ V/cm and that we limit the average lateral field in the channel to at most (1/200) of the vertical field in the channel. The electron mobility in the channel is a function of the vertical field in the silicon $E_x(x = 0)$:

$$\mu_n = \frac{\mu_{no}}{1 + \Theta E_x(0)} = \frac{215\ \text{cm}^2/\text{Vs}}{1 + (2 \times 10^{-6}\text{cm/V})\, E_x(0)}$$

The substrate doping is fixed at $N_a = 10^{16}$ cm^{-3}. The drain current should be linear with the drain-source voltage up to $V_{DS} = 500$ mV, which is why there is a restriction on the lateral electric field in the channel. Note that the peak drain current for linear operation is $V_{DS}/R_{on} = 500$ mV/100 mΩ = 5 A!

(a) Find a combination of oxide thickness t_{ox}, channel length L, and (W/L) ratio that meet the specifications. You should try to minimize L by using the maximum possible channel electric fields.

(b) This huge transistor is best laid out as 10 MOSFETs in parallel, each one of which is folded to make a more compact IC structure. Assuming that the source and drain diffusions have a length $L_{diff} = 6$ μm, what is the area required by each of the 10 MOSFETs? Draw the layout of one of

the 10 MOSFETs, using as many 2 μm × 2 μm contacts as you can fit in the source and drain in order to minimize the series resistance. Note that the MOSFET switch may consume most of the chip area in a "smart power" CMOS integrated circuit.

D4.2 In analog design, it is sometimes necessary to have a device with a very large output resistance. Given the standard p-channel MOSFET parameters, we would like to achieve

$$r_o = 650\ \text{k}\Omega \text{ with } I_D = -250\ \mu\text{A}$$

(a) What are the channel length and channel width to meet these specification?

(b) Draw the layout to scale, using $L_{diff} = 6\ \mu\text{m}$ and 2 μm × 2 μm contacts.

(c) Find the total capacitance between the drain diffusion and ground, assuming that $V_{SD} = V_{SB} = 2.5$ V. Assume that the gate can be considered at ground. You can assume that the substrate doping is $N_d = 10^{17}\ \text{cm}^{-3}$ and neglect the sidewall capacitance.

(d) What is the product of r_o and the capacitance from part (c)? The time constant that results will determine the break frequency beyond which the output impedance of the MOSFET begins to fall below r_o.