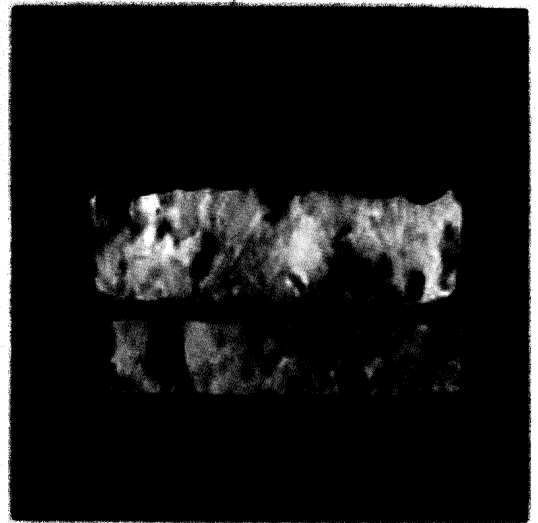


SEMICONDUCTOR DEVICES

**Physics
and
Technology**



2nd Edition

S.M. Sze

TSMC EX1019 p. 1

2ND EDITION

Semiconductor Devices

Physics and Technology

S. M. SZE

UMC Chair Professor

National Chiao Tung University

National Nano Device Laboratories

Hsinchu, Taiwan



JOHN WILEY & SONS, INC.

Acquisitions Editor *William Zobrist*
Marketing Manager *Katherine Hepburn*
Production Services Manager *Jeanine Furino*
Production Editor *Sandra Russell*
Designer *Harold Nolan*
Production Management Services *Argosy Publishing Services*

Cover Photography: A transmission-electron micrograph of a floating-gate nonvolatile semiconductor memory with a magnification of 100,000 times. (Photography courtesy of George T. T. Sheng.) For a discussion of the device, see Chapters 1, 6, and 14.

This book was typeset in *New Caledonia* by *Argosy Publishing* and printed and bound by *R. R. Donnelley and Sons, Inc. (Willard)*. The cover was printed by *The Lehigh Press*.

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

The book is printed on acid-free paper: 

Copyright © 1985, 2002 by John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (508) 750-8400, fax (508) 750-4470. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc. 605 Third Avenue, New York, NY 10158-0012, (212) 850-6008, E-mail: PERMREQ@WILEY.COM. To order books or for customer service call 1-800-CALL-WILEY (225-5945).

Library of Congress Cataloging in Publication Data:

Sze, S. M., 1936-

Semiconductor devices, physics and technology/S.M. Sze.—2nd ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-471-33372-7 (cloth: alk. paper)

1. Semiconductors. I. Title.

TK7871.85 .S9883 2001

621.3815'2—dc21

2001026003

ISBN 0-471-33372-7

Printed in the United States of America

10 9 8 7 6 5 4 3

Film Formation

- ▶ 11.1 THERMAL OXIDATION
- ▶ 11.2 DIELECTRIC DEPOSITION
- ▶ 11.3 POLYSILICON DEPOSITION
- ▶ 11.4 METALLIZATION
- ▶ SUMMARY

To fabricate discrete devices and integrated circuits, we use many different kinds of thin films. We can classify thin films into four groups: thermal oxides, dielectric layers, polycrystalline silicon, and metal films. Figure 1 shows a schematic view of a conventional silicon *n*-channel MOSFET (metal-oxide-semiconductor field-effect transistor) that uses all four groups of films. The first important thin film from the thermal oxide group is the gate-oxide layer under which a conducting channel can be formed between the source and the drain. A related layer is the field oxide, which provides isolation from other devices. Both gate and field oxides generally are grown by a thermal oxidation process because only thermal oxidation can provide the highest-quality oxides having the lowest interface trap densities.

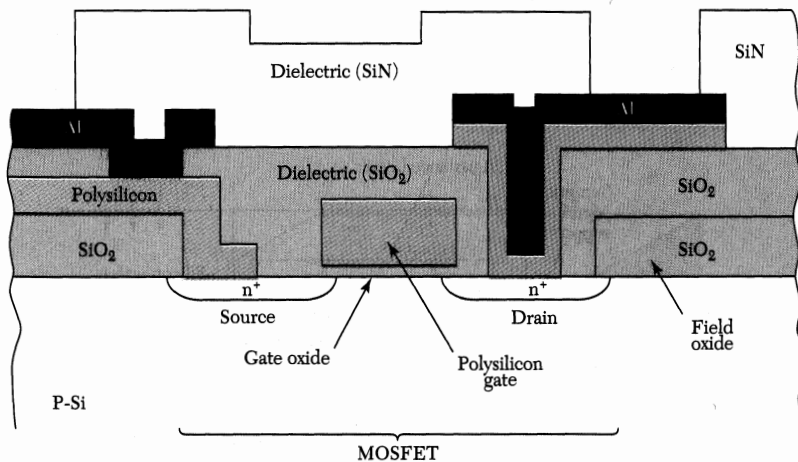


Fig. 1 Schematic cross section of a metal-oxide-semiconductor field-effect transistor (MOSFET).

SiO_2 SiN
Dielectric layers such as silicon dioxide and silicon nitride are used for insulation between conducting layers, for diffusion and ion implantation masks, for capping doped films to prevent the loss of dopants, and for passivation to protect devices from impurities, moisture, and scratches. Polycrystalline silicon, usually referred to as polysilicon, is used as a gate electrode material in MOS devices, a conductive material for multilevel metallization, and a contact material for devices with shallow junctions. Metal films such as aluminum and silicides are used to form low-resistance interconnections, ohmic contacts, and rectifying metal-semiconductor barriers.

Specifically, we cover the following topics:

- The thermal oxidation process to form silicon dioxide (SiO_2).
- Deposition techniques to form low-dielectric-constant and high-dielectric-constant films as well as polysilicon films.
- Deposition techniques to form aluminum and copper interconnections as well as the related global planarization process.
- Characteristics of these thin films and their compatibility with integrated-circuit processing.

► 11.1 THERMAL OXIDATION

Semiconductors can be oxidized by various ^{my process} methods. These include thermal oxidation, electrochemical anodization, and plasma reaction. Among these methods, thermal oxidation is by far the most important for silicon devices. It is the key process in modern silicon integrated-circuit technology. For gallium arsenide, however, thermal oxidation results in generally nonstoichiometric films. The oxides provide poor electrical insulation and semiconductor surface protection; hence, these oxides are rarely used in gallium arsenide technology. Consequently, in this section we concentrate on the thermal oxidation of silicon.

The basic thermal oxidation setup is shown¹ in Fig. 2. The reactor consists of a resistance-heated furnace, a cylindrical fused-quartz tube containing the silicon wafers held vertically in a slotted quartz boat, and a source of either pure dry oxygen or pure water vapor. The loading end of the furnace tube protrudes into a vertical flow hood where a filtered flow of air is maintained. Flow is directed as shown by the arrow in Fig. 2. The hood reduces dust and particulate matters in the air surrounding the wafers and mini-

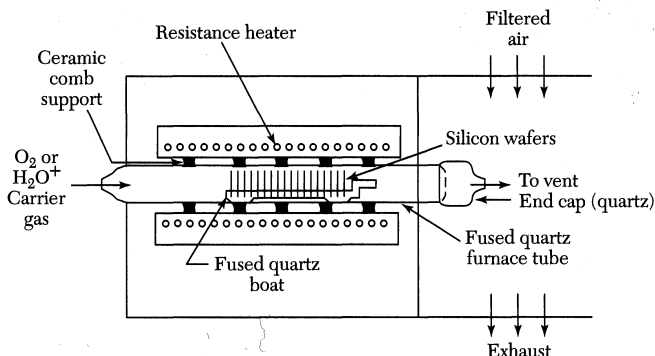
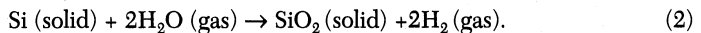


Fig. 2 Schematic cross section of a resistance-heated oxidation furnace.

mizes contamination during wafer loading. The oxidation temperature is generally in the range of 900°–1200°C and the typical gas flow rate is about 1 liter/min. The oxidation system uses microprocessors to regulate the gas flow sequence, to control the automatic insertion and removal of silicon wafers, to ramp the temperature up (i.e., to increase the furnace temperature linearly) from a low temperature to the oxidation temperature so that the wafers will not warp due to sudden temperature change, to maintain the oxidation temperature to within $\pm 1^\circ\text{C}$, and to ramp the temperature down when oxidation is completed.

11.1.1 Kinetics of Growth

The following chemical reactions describe the thermal oxidation of silicon in oxygen or water vapor:



The silicon-silicon dioxide interface moves into the silicon during the oxidation process. This creates a fresh interface region, with surface contamination on the original silicon ending up on the oxide surface. The densities and molecular weights of silicon and silicon dioxide are used in the following example to show that growing an oxide of thickness x consumes a layer of silicon $0.44x$ thick (Fig. 3).

► EXAMPLE 1

If a silicon oxide layer of thickness x is grown by thermal oxidation, what is the thickness of silicon being consumed? The molecular weight of Si is 28.9 g/mol, and the density of Si is 2.33 g/cm³. The corresponding values for SiO₂ are 60.08 g/mol and 2.21 g/cm³.

SOLUTION The volume of 1 mol of silicon is

$$\frac{\text{Molecular weight of Si}}{\text{Density of Si}} = \frac{28.9 \text{ g/mole}}{2.33 \text{ g/cm}^3} = 12.06 \text{ cm}^3/\text{mol}.$$

The volume of 1 mol of silicon dioxide is

$$\frac{\text{Molecular weight of SiO}_2}{\text{Density of SiO}_2} = \frac{60.08 \text{ g/mol}}{2.2 \text{ g/cm}^3} = 27.18 \text{ cm}^3/\text{mol}.$$

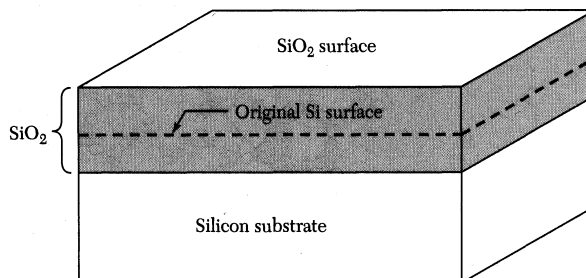


Fig. 3 Growth of silicon dioxide by thermal oxidation.

Since 1 mol of silicon is converted to 1 mol of silicon dioxide,

$$\frac{\text{Thickness of Si} \times \text{area}}{\text{Thickness of SiO}_2 \times \text{area}} = \frac{\text{volume of 1 mol of Si}}{\text{volume of 1 mol of SiO}_2},$$

$$\frac{\text{Thickness of Si}}{\text{Thickness of SiO}_2} = \frac{12.06}{27.18} = 0.44,$$

Thickness of silicon = 0.44 (thickness of SiO₂).

For example, to grow a silicon dioxide layer of 100 nm, a layer of 44 nm of silicon is consumed.

The basic structural unit of thermally grown silicon dioxide is a silicon atom surrounded tetrahedrally by four oxygen atoms, as illustrated¹ in Fig. 4a. The silicon-to-oxygen internuclear distance is 1.6 Å, and the oxygen-to-oxygen internuclear distance is 2.27 Å. These tetrahedra are joined together at their corners by oxygen bridges in a variety of ways to form the various phases or structures of silicon dioxide (also called silica). Silica has several crystalline structures (e.g., quartz) and an amorphous structure. When silicon is thermally oxidized, the silicon dioxide structure is amorphous. Typically amorphous silica has a density of 2.21 g/cm³, compared with 2.65 g/cm³ for quartz.

The basic difference between the crystalline and amorphous structures is that the former is a periodic structure, extending over many molecules, whereas the latter has no periodic structure at all. Figure 4b is a two-dimensional schematic diagram of a quartz crystalline structure made up of rings with six silicon atoms. Figure 4c is a two-dimensional schematic diagram of an amorphous structure for comparison. In the amorphous

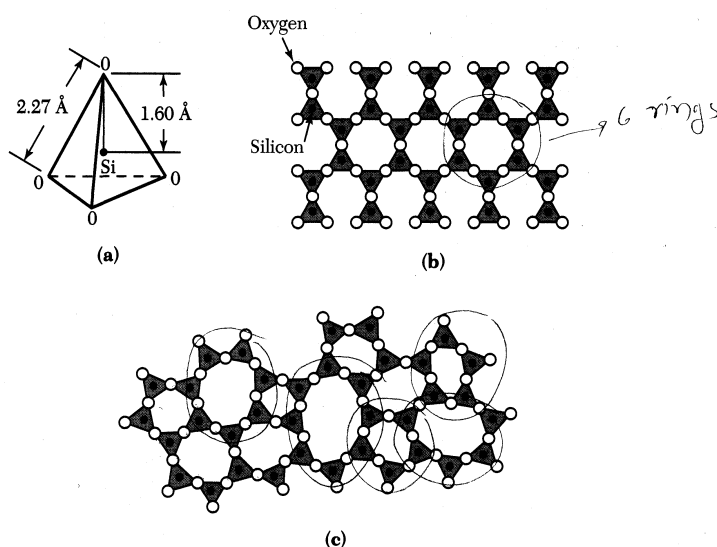


Fig. 4 (a) Basic structural unit of silicon dioxide. (b) Two dimensional representation of a quartz crystal lattice. (c) Two-dimensional representation of the amorphous structure of silicon dioxide.¹

structure there is still a tendency to form characteristic rings with six silicon atoms. Note that the amorphous structure in Fig. 4c is quite open because only 43% of the space is occupied by silicon dioxide molecules. The relatively open structure accounts for the lower density and allows a variety of impurities (such as sodium) to enter and diffuse readily through the silicon dioxide layer.

The kinetics of thermal oxidation of silicon can be studied based on a simple model illustrated² in Fig. 5. A silicon slice contacts the oxidizing species (oxygen or water vapor), resulting in a surface concentration of C_0 molecules/cm³ for these species. The magnitude of C_0 equals the equilibrium bulk concentration of the species at the oxidation temperature. The equilibrium concentration generally is proportional to the partial pressure of the oxidant adjacent to the oxide surface. At 1000°C and at a pressure of 1 atm, the concentration C_0 is 5.2×10^{16} molecules/cm³ for dry oxygen and 3×10^{19} molecules/cm³ for water vapor.

The oxidizing species diffuses through the silicon dioxide layer, resulting in a concentration C_s at the surface of silicon. The flux F_1 can be written as

$$F_1 = D \frac{dC}{dx} \equiv \frac{D(C_0 - C_s)}{x}, \quad (3)$$

where D is the diffusion coefficient of the oxidizing species and x is the thickness of the oxide layer already present.

At the silicon surface, the oxidizing species reacts chemically with silicon. Assuming the rate of reaction is proportional to the concentration of the species at the silicon surface, the flux F_2 is given by

$$F_2 = \kappa C_s, \quad (4)$$

where κ is the surface reaction rate constant for oxidation. At the steady state, $F_1 = F_2 = F$. Combining Eqs. 3 and 4 gives

$$F = \frac{DC_0}{x + (D/\kappa)}. \quad (5)$$

The reaction of the oxidizing species with silicon forms silicon dioxide. Let C_1 be the number of molecules of the oxidizing species in a unit volume of the oxide. There are 2.2×10^{22} silicon dioxide molecules/cm³ in the oxide, and we add one oxygen molecule

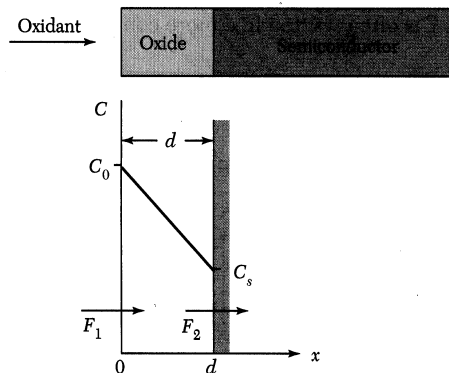


Fig. 5 Basic model for the thermal oxidation of silicon.²

(O₂) to each silicon dioxide molecule, whereas we add two water molecules (H₂O) to each silicon dioxide molecule. Therefore, C_1 for oxidation in dry oxygen is $2.2 \times 10^{22} \text{ cm}^{-3}$, and for oxidation in water vapor it is twice this number ($4.4 \times 10^{22} \text{ cm}^{-3}$). Thus, the growth rate of the oxide layer thickness is given by

$$\frac{dx}{dt} = \frac{F}{C_1} = \frac{DC_0/C_1}{x + (D/\kappa)}. \quad (6)$$

We can solve this differential equation subject to the initial condition, $x(0) = d_0$, where d_0 is the initial oxide thickness; d_0 can also be regarded as the thickness of oxide layer grown in an earlier oxidation step. Solving Eq. 6 yields the general relationship for the oxidation of silicon,

$$x^2 + \frac{2D}{\kappa}x = \frac{2DC_0}{C_1}(t + \tau), \quad (7)$$

where $\tau \equiv (d_0^2 + 2Dd_0/\kappa)C_1/2DC_0$, which represents a time coordinate shift to account for the initial oxide layer d_0 .

The oxide thickness after an oxidizing time t is given by

$$x = \frac{D}{\kappa} \left[\sqrt{1 + \frac{2C_0\kappa^2(t + \tau)}{DC_1}} - 1 \right]. \quad (8)$$

For small values of t , Eq. 8 reduces to

$$x \cong \frac{C_0\kappa}{C_1}(t + \tau), \quad (9)$$

and for larger values of t , it reduces to

$$x \cong \sqrt{\frac{2DC_0}{C_1}(t + \tau)}. \quad (10)$$

During the early stages of oxide growth, when surface reaction is the rate-limiting factor, the oxide thickness varies linearly with time. As the oxide layer becomes thicker, the oxidant must diffuse through the oxide layer to react at the silicon-silicon dioxide interface and the reaction becomes diffusion limited. The oxide growth then becomes proportional to the square root of the oxidizing time, which results in a parabolic growth rate.

Equation 7 is often written in a more compact form:

$$x^2 + Ax = B(t + \tau). \quad (11)$$

where $A = 2D/\kappa$, $B = 2DC_0/C_1$ and $B/A = \kappa C_0/C_1$. Using this form, Eqs. 9 and 10 can be written as

$$x = \frac{B}{A}(t + \tau) \quad (12)$$

for the linear region and as

$$x^2 = B(t + \tau). \quad (13)$$

for the parabolic region. For this reason, the term B/A is referred to as the linear rate constant and B is the parabolic rate constant. Experimentally measured results agree with the predictions of this model over a wide range of oxidation conditions. For wet oxidation, the initial oxide thickness d_0 is very small, or $\tau \approx 0$. However, for dry oxidation, the extrapolated value of d_0 at $t = 0$ is about 20 nm.

The temperature dependence of the linear rate constant B/A is shown in Fig. 6 for both dry and wet oxidation and for (111)- and (100)-oriented silicon wafers.² The linear rate constant varies as $\exp(-E_a/kT)$, where the activation energy E_a is about 2 eV for both dry and wet oxidation. This closely agrees with the energy required to break silicon-silicon bonds, 1.83 eV/molecule. Under a given oxidation condition, the linear rate constant depends on crystal orientation. This is because the rate constant is related to the rate of incorporation of oxygen atoms into the silicon. The rate depends on the surface bond structure of silicon atoms, making it orientation dependent. Because the density of available bonds on the (111)-plane is higher than that on the (100)-plane, the linear rate constant for (111)-silicon is larger.

Figure 7 shows the temperature dependence of the parabolic rate constant B , which can also be described by $\exp(-E_a/kT)$. The activation energy E_a is 1.24 eV for dry oxidation. The comparable activation energy for oxygen diffusion in fused silica is 1.18 eV. The corresponding value for wet oxidation, 0.71 eV, compares favorably with the value of 0.79 eV for the activation energy of diffusion of water in fused silica. The parabolic rate constant is independent of crystal orientation. This independence is expected because

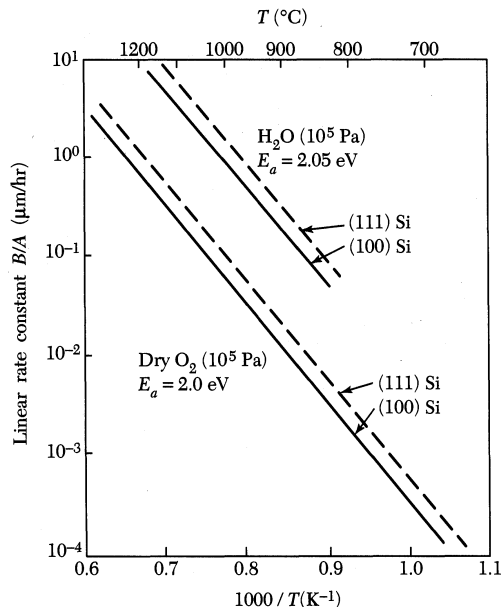


Fig. 6 Linear rate constant versus temperature.²

it is a measure of the diffusion process of the oxidizing species through a random network layer of amorphous silica.

Although oxides grown in dry oxygen have the best electrical properties, considerably more time is required to grow the same oxide thickness at a given temperature in dry oxygen than in water vapor. For relatively thin oxides such as the gate oxide in a MOS-FET (typically ≤ 20 nm), dry oxidation is used. However, for thicker oxides such as field oxides (≥ 20 nm) in MOS integrated circuits, and for bipolar devices, oxidation in water vapor (or steam) is used to provide both adequate isolation and passivation.

Figure 8 shows the experimental results of silicon dioxide thickness as a function of reaction time and temperature for two substrate orientations.³ Under a given oxidation condition, the oxide thickness grown on a (111)-substrate is larger than that grown on a (100)-substrate because of the larger linear rate constant of the (111)-orientation. Note that for a give temperature and time, the oxide film obtained using wet oxidation is about 5–10 times thicker than that using dry oxidation.

11.1.2 Thin Oxide Growth

Relative slow growth rates must be used to reproducibly grow thin oxide films of precise thickness. Various approaches to achieve such slower growth rates have been reported, including growth in dry O_2 at atmospheric pressure and lower temperatures ($800^\circ\text{--}900^\circ\text{C}$); growth at pressures lower than atmospheric pressure; growth in a reduced partial pressures of O_2 by using a diluent inert gas, such as N_2 , Ar, or He, together with the gas containing the oxidizing species; and the use of composite oxide films with the gate-oxide films consisting of a layer of thermally grown SiO_2 and an overlayer of chemical-vapor

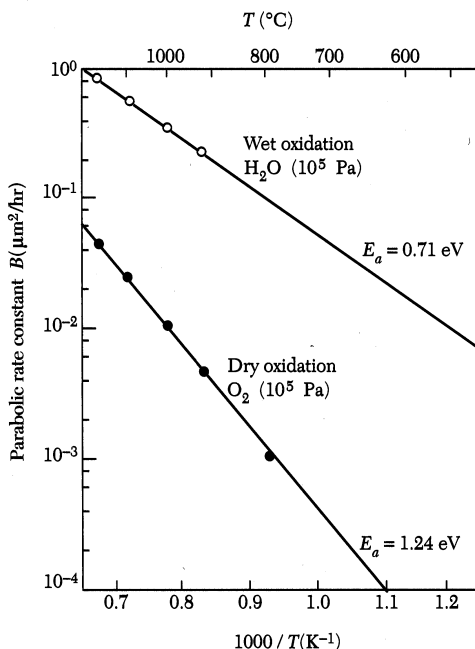


Fig. 7 Parabolic rate constant versus temperature.²

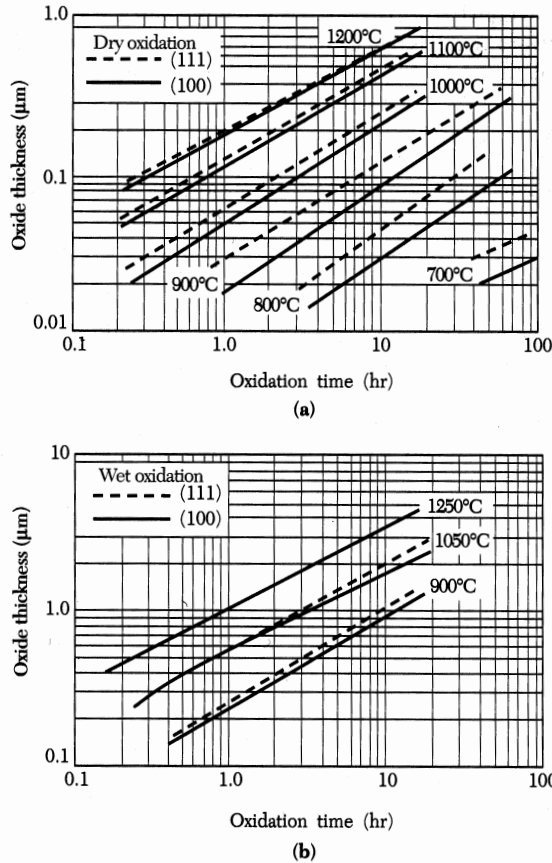


Fig. 8 Experimental results of silicon dioxide thickness as a function of reaction time and temperature for two substrate orientations. (a) Growth in dry oxygen. (b) Growth in steam.³

deposition (CVD) SiO_2 . However, the mainstream approach for gate oxides 10–15 nm thick is to grow the oxide film at atmospheric pressure and lower temperatures (800°–900°C). With this approach, processing using modern *vertical* oxidation furnaces can grow reproducible, high-quality 10 nm oxides to within 0.1 nm across the wafer.

We noted earlier that for dry oxidation, there is an apparently rapid oxidation that gives rise to an initial oxide thickness d_0 of about 20 nm. Therefore, the simple model presented in Section 11.1.1 is not valid for dry oxidation with an oxide thickness ≤ 20 nm. For ultralarge-scale integration (ULSI), the ability to grow thin (5 ~ 20 nm), uniform, high-quality reproducible gate oxides has become increasingly important. We briefly consider the growth mechanisms of such thin oxides.

In the early stage of growth in dry oxidation, there is a large compressive stress in the oxide layer that reduces the oxygen diffusion coefficient in the oxide. As the oxide becomes thicker, the stress will be reduced due to the viscous flow of silica and the diffusion coefficient will approach its stress-free value. Therefore, for thin oxides, the value of D/κ may be sufficiently small that we can neglect the term Ax in Eq. 11 and obtain

$$x^2 - d_0^2 = Bt, \quad (14)$$

where d_0 is equal to $\sqrt{2DC_0\tau/C_1}$, which is the initial oxide thickness when time is extrapolated to zero, and B is the parabolic rate constant defined previously. We therefore expect the initial growth in dry oxidation to follow a parabolic form.

► 11.2 DIELECTRIC DEPOSITION

Deposited dielectric films are used mainly for insulation and passivation of discrete devices and integrated circuits. There are three commonly used deposition methods: atmospheric-pressure CVD, low-pressure CVD (LPCVD), and plasma-enhanced chemical vapor deposition (PECVD, or plasma deposition). PECVD is an energy-enhanced CVD method, in which plasma energy is added to the thermal energy of a conventional CVD system. Considerations in selecting a deposition process are the substrate temperature, the deposition rate and film uniformity, the morphology, the electrical and mechanical properties, and the chemical composition of the dielectric films.

The reactor for atmospheric-pressure CVD is similar to the one shown in Fig. 2, except that different gases are used at the gas inlet. In a hot-wall, reduced-pressure reactor as shown in Fig. 9a, the quartz tube is heated by a three-zone furnace, and gas is introduced at one end and pumped out at the opposite end. The semiconductor wafers are held vertically in a slotted quartz boat.⁴ The quartz tube wall is hot because it is adjacent to the

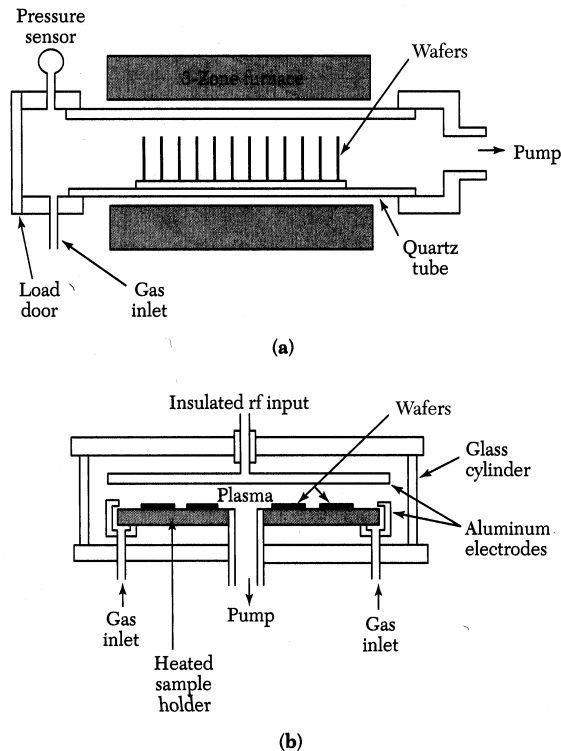


Fig. 9 Schematic diagrams of chemical-vapor deposition reactors. (a) Hot-wall, reduced-pressure reactor. (b) Parallel-plate plasma deposition reactor.⁴ rf, radio frequency.

furnace, in contrast to a cold-wall reactor such as the horizontal epitaxial reactor that uses radio frequency (rf) heating.

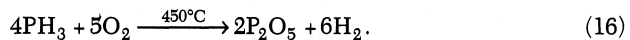
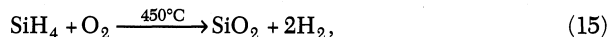
The parallel-plate, radial-flow, PECVD reactor shown in Fig. 9b consists of a cylindrical glass or aluminum chamber sealed with aluminum endplates. Inside are two parallel aluminum electrodes. An rf voltage is applied to the upper electrode, whereas the lower electrode is grounded. The rf voltage causes a plasma discharge between the electrodes. Wafers are placed on the lower electrode, which is heated between 100° and 400°C by resistance heaters. The reaction gases flow through the discharge from outlets located along the circumference of the lower electrode. The main advantage of this reactor is its low deposition temperature. However, its capacity is limited, especially for large-diameter wafers, and the wafers may become contaminated if loosely adhering deposits fall on them.

11.2.1 Silicon Dioxide

CVD silicon dioxide does not replace thermally grown oxides because the best electrical properties are obtained with thermally grown films. CVD oxides are used instead to complement the thermal oxides. A layer of undoped silicon dioxide is used to insulate multilevel metallization, to mask ion implantation and diffusion, and to increase the thickness of thermally grown field oxides. Phosphorus-doped silicon dioxide is used both as an insulator between metal layers and as a final passivation layer over devices. Oxides doped with phosphorus, arsenic, or boron are used occasionally as diffusion sources.

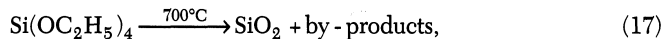
Deposition Methods

Silicon dioxide films can be deposited by several methods. For low-temperature deposition (300°–500°C), the films are formed by reacting silane, dopant, and oxygen. The chemical reactions for phosphorus-doped oxides are



The deposition process can be performed either at atmospheric pressure in a CVD reactor or at reduced pressure in an LPCVD reactor (Fig. 9a). The low deposition temperature of the silane-oxygen reaction makes it a suitable process when films must be deposited over a layer of aluminum.

For intermediate-temperature deposition (500°–800°C), silicon dioxide can be formed by decomposing tetraethylorthosilicate, $\text{Si}(\text{OC}_2\text{H}_5)_4$, in an LPCVD reactor. The compound, abbreviated TEOS, is vaporized from a liquid source. The TEOS compound decomposes as follows:



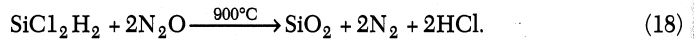
forming both SiO_2 and a mixture of organic and organosilicon by-products. Although the higher temperature required for the reaction prevents its use over aluminum, it is suitable for polysilicon gates requiring a uniform insulating layer with good step coverage. The good step coverage is a result of enhanced surface mobility at higher temperatures. The oxides can be doped by adding small amounts of the dopant hydrides (phosphines, arsine, or diborane), similar to the process in epitaxial growth.

The deposition rate as a function of temperature varies as $e^{-E_a/kT}$, where E_a is the activation energy. The E_a of the silane-oxygen reaction is quite low: about 0.6 eV for undoped oxides and almost zero for phosphorus doped oxide. In contrast, E_a for the TEOS reaction

is much higher: about 1.9 eV for undoped oxide and 1.4 eV when phosphorus doping compounds are present. The dependence of the deposition rate on TEOS partial pressure is proportional to $(1 - e^{-P/P_0})$, where P is the TEOS partial pressure and P_0 is about 30 Pa. At low TEOS partial pressures, the deposition rate is determined by the rate of the surface reaction. At high partial pressures, the surface becomes nearly saturated with adsorbed TEOS and the deposition rate becomes essentially independent of TEOS pressure.⁴

Recently, atmospheric-pressure and low-temperature CVD processes using TEOS and ozone (O_3) have been proposed,⁵ as shown in Fig. 10. This CVD technology produces oxide films with high conformality and low viscosity under low deposition temperature. In addition, the shrinkage of oxide film during annealing is also a function of ozone concentration, as shown in Fig. 11. Because of their porosity, O_3 -TEOS CVD oxides are often accompanied by plasma-assisted oxides to permit planarization in ULSI processing.

For high-temperature deposition (900°C), silicon dioxide is formed by reacting dichlorosilane, $SiCl_2H_2$, with nitrous oxide at reduced pressure:



This deposition gives excellent film uniformity and is sometimes used to deposit insulating layers over polysilicon.

Properties of Silicon Dioxide

Deposition methods and properties of silicon dioxide films are listed⁴ in Table 1. In general, there is a direct correlation between deposition temperature and film quality. At higher temperatures, deposited oxide films are structurally similar to silicon dioxide that has been thermally grown.

The lower densities occur in films deposited below 500°C . Heating deposited silicon dioxide at temperatures between 600° and 1000°C causes densification, during which the oxide thickness decreases, whereas the density increases to 2.2 g/cm^3 . The refractive index of silicon dioxide is 1.46 at a wavelength of $0.6328\text{ }\mu\text{m}$. Oxides with lower indices are porous, such as the oxide from the silane-oxygen deposition, which has a refractive index of 1.44. The porous nature of the oxide also is responsible for the lower dielectric strength, which is the applied electric field that will cause a high current to flow in the oxide film. The etch rates of oxides in a hydrofluoric acid solution depend on deposition

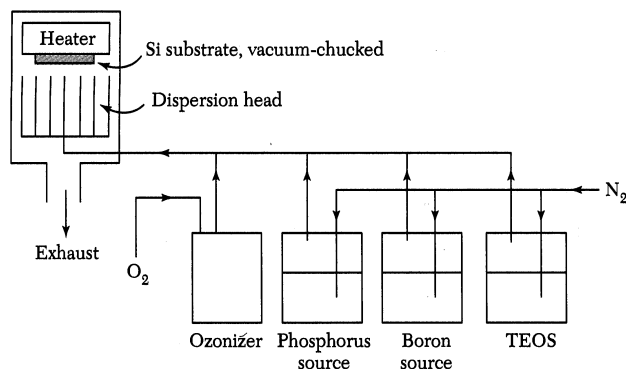


Fig. 10 Experimental apparatus for the O_3 -TEOS chemical-vapor deposition (CVD) system.

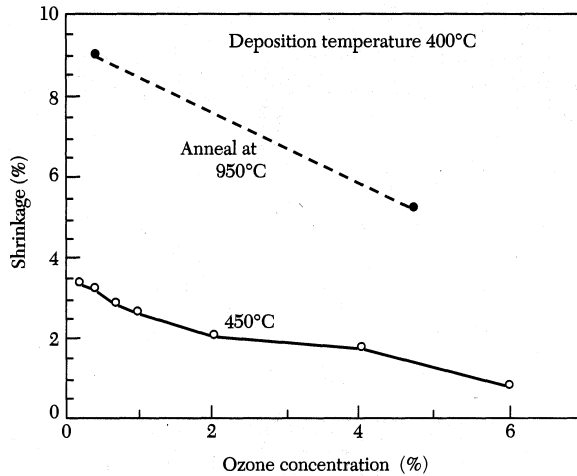


Fig. 11 Dependence of the shrinkage of the O_3 -TEOS CVD film on ozone concentration using annealing. (Courtesy of SAMCO Company, Japan.)

temperature, annealing history, and dopant concentration. Usually higher-quality oxides are etched at lower rates.

Step Coverage

Step coverage relates the surface topography of a deposited film to the various steps on the semiconductor substrate. In the illustration of ideal, or conformal, step coverage shown in Fig. 12a, film thickness is uniform along all surfaces of the step. The uniformity of the film thickness, regardless of topography, is due to the rapid migration of reactants after adsorption on the step surfaces.⁶

Figure 12b shows an example of nonconformal step coverage, which results when the reactants adsorb and react without significant surface migration. In this instance, the deposition rate is proportional to the arrival angle of the gas molecules. Reactants arriving along the top horizontal surface come from many different angles and ϕ_1 , the arrival

TABLE 1 Properties of SiO_2 Films

Property	Thermally grown at 1000°C	$SiH_4 + O_2$ at 450°C	TEOS at 700°C	$SiCl_2H_2 + N_2O$ at 900°C
Composition	SiO_2	$SiO_2 (H)$	SiO_2	$SiO_2(Cl)$
Density (g/cm^3)	2.2	2.1	2.2	2.2
Refractive index	1.46	1.44	1.46	1.46
Dielectric strength (10^6 V/cm)	>10	8	10	10
Etch rate ($\text{\AA}/\text{min}$) (100:1 $H_2O:HF$)	30	60	30	30
Etch rate ($\text{\AA}/\text{min}$) (buffered HF)	440	1200	450	450
Step coverage	—	Nonconformal	Conformal	Conformal

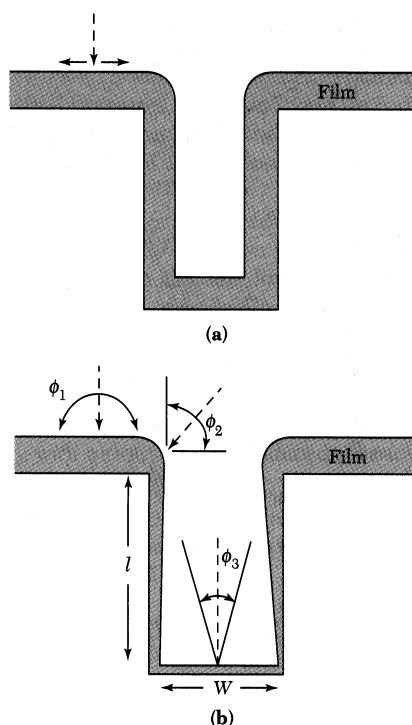


Fig. 12 Step coverage of deposited films. (a) Conformal step coverage. (b) Nonconformal step coverage.⁴

angle, varies in two dimensions, from 0 to 180°, whereas reactants arriving at the top of a vertical wall have an arrival angle ϕ_2 that varies from 0° to 90°. Thus, the film thickness on the top surface is double that of a wall surface. Further down the wall, ϕ_3 is related to the width of the opening, and the film thickness is proportional to

$$\phi_3 \cong \arctan \frac{W}{l}, \quad (19)$$

where l is the distance from the top surface and W is the width of the opening. This type of step coverage is thin along the vertical walls, with a possible crack at the bottom of step caused by self-shadowing.

Silicon dioxide formed by TEOS decomposition at reduced pressure gives a nearly conformal coverage due to rapid surface migration. Similarly, the high-temperature dichlorosilane-nitrous oxide reaction also results in conformal coverage. However, during silane-oxygen deposition, no surface migration takes place and the step coverage is determined by the arrival angle. Most evaporated or sputtered materials have a step coverage similar to that in Fig. 12b.

P-Glass Flow

A smooth topography is usually required for the deposited silicon dioxide used as an insulator between metal layers. If the oxide used to cover the lower metal layer is concave, circuit failure may result from an opening that may occur in the upper metal layer dur-

ing deposition. Because phosphorus-doped silicon dioxide (P-glass) deposited at low temperatures becomes soft and flows upon heating, it provides a smooth surface and is often used to insulate adjacent metal layers. This process is called P-glass flow.

Figure 13 shows four cross sections of scanning electron microscope photographs of P-glass covering a polysilicon step.⁶ All samples are heated in steam at 1100°C for 20 min. Figure 13a shows a sample of glass that contains a negligibly small amount of phosphorus and does not flow. Note the concavity of the film and that the corresponding angle θ is about 120°. Figures 13b, 13c, and 13d show samples of P-glass with progressively higher phosphorus contents up to 7.2 wt% (weight percent). In these samples the decreasing step angles of the P-glass layer indicate how flow increases with phosphorus concentration. P-glass flow depends on annealing time, temperature, phosphorus concentration, and the annealing ambient.⁶

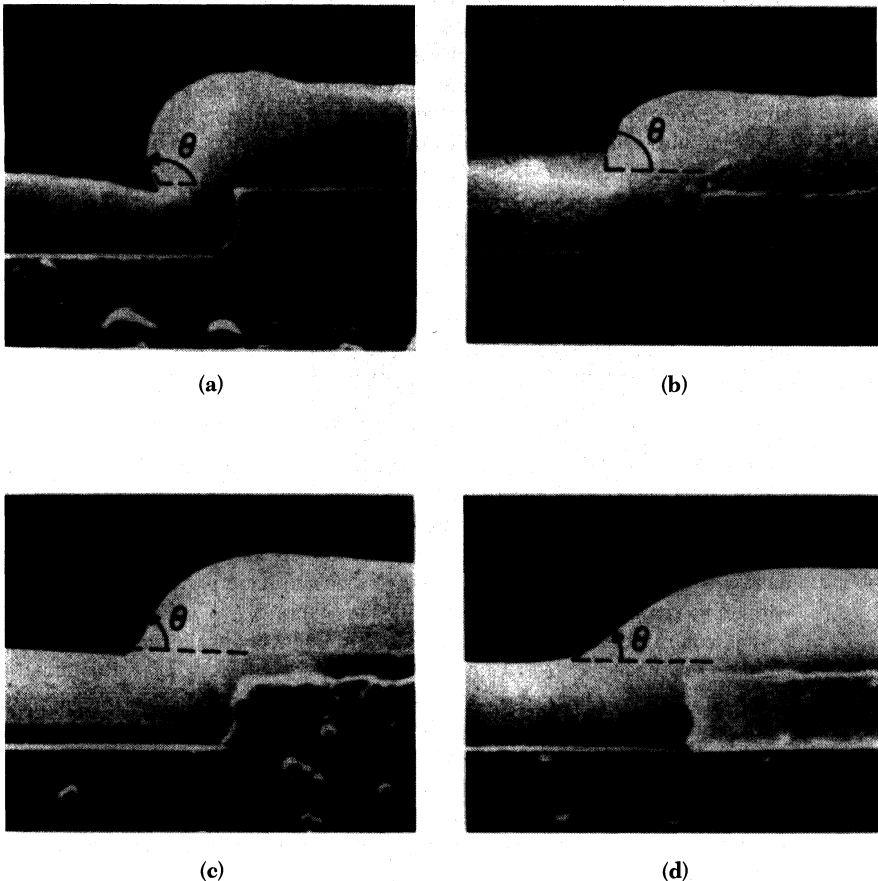


Fig. 13 Scanning-electron micrographs (10,000 $\times$) of samples annealed in steam at 1100°C for 20 minutes for the following weight percent of phosphorus.⁶ (a) 0 wt%; (b) 2.2 wt%; (c) 4.6 wt%; and (d) 7.2 wt%.

The angle θ as a function of weight percent of phosphorus as shown in Fig. 13 can be approximated by

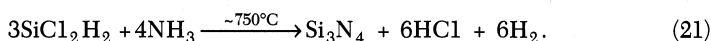
$$\theta \cong 120^\circ \left(\frac{10 - \text{wt}\%}{10} \right). \quad (20)$$

If we want an angle smaller than 45° we require a phosphorus concentration larger than 6 wt%. However, at concentrations above 8 wt%, the metal film (e.g., aluminum) may be corroded by the acid products formed during the reaction between the phosphorus in the oxide and atmospheric moisture. Therefore, the P-glass flow process uses phosphorus concentrations of 6–8 wt%.

11.2.2 Silicon Nitride

It is difficult to grow silicon nitride by thermal nitridation (e.g., with ammonia, NH_3) because of its low growth rate and high growth temperature. However, silicon nitride films can be deposited by an intermediate-temperature (750°C) LPCVD process or a low-temperature (300°C) plasma-assisted CVD process.^{7,8} The LPCVD films are of stoichiometric composition (Si_3N_4) with high density ($2.9\text{--}3.1\text{ g/cm}^3$). These films can be used to passivate devices because they serve as good barriers to the diffusion of water and sodium. The films also can be used as masks for the selective oxidation of silicon because silicon nitride oxidizes very slowly and prevents the underlying silicon from oxidizing. The films deposited by plasma-assisted CVD are not stoichiometric and have a lower density ($2.4\text{--}2.8\text{ g/cm}^3$). Because of the low deposition temperature, silicon nitride films can be deposited over fabricated devices and serve as their final passivation. The plasma-deposited nitride provides excellent scratch protection, serves as a moisture barrier, and prevents sodium diffusion.

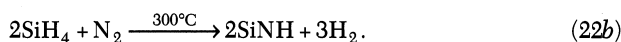
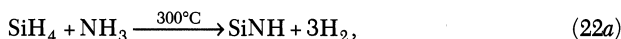
In the LPCVD process, dichlorosilane and ammonia react at reduced pressure to deposit silicon nitride at temperatures between 700° and 800°C . The reaction is



Good film uniformity and high wafer throughput (the number of wafers processed per hour) are advantages of the reduced-pressure process. As in the case of oxide deposition, silicon nitride deposition is controlled by temperature, pressure, and reactant concentration. The activation energy for deposition is about 1.8 eV. The deposition rate increases with increasing total pressure or dichlorosilane partial pressure and decreases with an increasing ammonia-to-dichlorosilane ratio.

Silicon nitride deposited by LPCVD is an amorphous dielectric containing up to 8 atomic percent (at %) hydrogen. The etch rate in buffered HF is less than 1 nm/min. The film has a very high tensile stress of approximately 10^{10} dynes/cm², which is nearly 10 times that of TEOS-deposited SiO_2 . Films thicker than 200 nm may crack because of the very high stress. The resistivity of silicon nitride at room temperature is about $10^{16}\ \Omega\text{-cm}$. Its dielectric constant is 7 and its dielectric strength is 10^7 V/cm .

In the plasma-assisted CVD process, silicon nitride is formed either by reacting silane and ammonia in an argon plasma or by reacting silane in a nitrogen discharge. The reactions are as follows:



The products depend strongly on deposition conditions. The radial-flow, parallel-plate reactor (Fig. 9b) is used to deposit the films. The deposition rate generally increases with increasing temperature, power input, and reactant gas pressure.

Large concentrations of hydrogen are contained in plasma-deposited films. The plasma nitride (also referred to as SiN) used in semiconductor processing generally contains 20–25 at % hydrogen. Films with low tensile stress ($\sim 2 \times 10^9$ dynes/cm²) can be prepared by plasma deposition. Film resistivities range from 10^5 to 10^{21} Ω -cm, depending on silicon-to-nitrogen ratio, whereas dielectric strengths are between 1×10^6 and 6×10^6 V/cm.

11.2.3 Low-Dielectric-Constant Materials

As devices continue to scale down to the deep submicron region, they require multilevel interconnection architecture to minimize the time delay due to parasitic resistance (R) and capacitance (C). The gain in device speed at the gate level will be offset by the propagation delay at the metal interconnects because of the increased RC time constant, as shown in Fig. 14. For example, in devices with gate length of 250 nm or less, up to 50% of the time delay is due to the RC delay of long interconnections.⁹ Therefore, the device interconnection network becomes a limiting factor in determining chip performance such as device speed, cross talk, and power consumption of ULSI circuits.

In order to reduce the RC time constant of ULSI circuits, interconnection materials with low resistivity and interlayer films with low capacitance are required. On the low-capacitance topic ($C = \epsilon_f A/d$, where ϵ_f is the dielectric permittivity, A the area, and d the

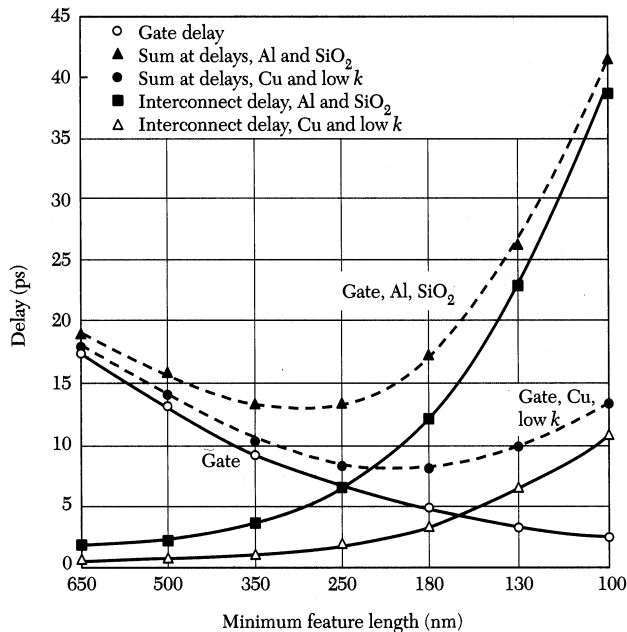


Fig. 14 Calculated gate and interconnect delay versus technology generation. The dielectric constant for the low- k material is 2.0. Both Al and Cu interconnects are $0.8 \mu\text{m}$ thick and $43 \mu\text{m}$ long.

missing page: P.386

TSMC EX1019 p. 21

11.2.4 High-Dielectric-Constant Materials

High- k materials are required for ULSI circuits, especially for dynamic random access memory (DRAM). The storage capacitor in a DRAM has to maintain a certain value of capacitance for proper operation (e.g., 40 fF). For a given capacitance ($\epsilon_f A/d$), usually a minimum d is selected to meet the conditions of the maximum allowed leakage current and the minimum required breakdown voltage. The area of the capacitor can be increased by using stacked or trench structures. These structures are considered in Chapter 14. However, for a planar structure, area A is reduced with increasing DRAM density. Therefore, the dielectric constant of the film must be increased.

Several high- k materials have been proposed, such as barium strontium titanate (BST) and lead zirconium titanate (PZT). They are shown in Table 3. In addition, there are titanates doped with one or more acceptors, such as alkaline earth metals, or doped with one or more donors, such as rare earth elements. The tantalum oxide (Ta_2O_5) has a dielectric constant in a range of 20–30. As a reference, the dielectric constant of Si_3N_4 is in a range of 6–7 and that for SiO_2 is 3.9. A Ta_2O_5 film can be deposited by a CVD process using gaseous TaCl_5 and O_5 as the starting materials.

▶ EXAMPLE 3

A DRAM capacitor has the following parameters: capacitance $C = 40$ fF, cell size $A = 1.28 \mu\text{m}^2$, and dielectric constant $k = 3.9$ for silicon dioxide. If we replace SiO_2 with Ta_2O_5 ($k = 25$) without changing thickness, what is the equivalent cell area of the capacitor?

SOLUTION

$$C = \frac{\epsilon_f A}{d},$$

$$\frac{3.9 \times 1.28}{d} = \frac{25 \times A}{d},$$

$$\therefore \text{Equivalent cell size } A = \frac{3.9}{25} \times 1.28 = 0.2 \mu\text{m}^2.$$

TABLE 3 High- k Materials

	Materials	Dielectric constant
Binary	Ta_2O_5	25
	TiO_2	40
	Y_2O_3	17
	Si_3N_4	7
Paraelectric perovskite	SrTiO_3 (STO)	140
	$(\text{Ba}_{1-x}\text{Sr}_x)\text{TiO}_3$ (BST)	300–500
	$\text{Ba}(\text{Ti}_{1-x}\text{Zr}_x)\text{O}_3$ (BZT)	300
	$(\text{Pb}_{1-x}\text{La}_x)(\text{Zr}_{1-y}\text{Ti}_y)\text{O}_3$ (PLZT)	800–1000
Ferroelectric perovskite	$\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3$ (PMN)	1000–2000
	$\text{Pb}(\text{Zr}_{0.47}\text{Ti}_{0.53})\text{O}_3$ (PZT)	>1000

► 11.3 POLYSILICON DEPOSITION

Using polysilicon as the gate electrode in MOS devices is a significant development in MOS technology. One important reason is that polysilicon surpasses aluminum for electrode reliability. Figure 15 shows the maximum time to breakdown for capacitors with both polysilicon and aluminum electrodes.¹⁰ The polysilicon is clearly superior, especially for thinner gate oxides. The inferior time to breakdown of aluminum electrode is due to the migration of aluminum atoms into the thin oxide under an electrical field. Polysilicon is also used as a diffusion source to create shallow junctions and to ensure ohmic contact to crystalline silicon. Additional uses include the manufacture of conductors and high-value resistors.

A low-pressure reactor (Fig. 9a) operated between 600° and 650°C is used to deposit polysilicon by pyrolyzing silane according to the following reaction.



Of the two most common low-pressure processes one operates at a pressure of 25–130 Pa using 100% silane, whereas the other process involves a diluted mixture of 20%–30% silane in nitrogen at the same total pressure. Both processes can deposit polysilicon on hundreds of wafers per run with good uniformity (i.e., thickness within 5%)

Figure 16 shows the deposition rate at four deposition temperatures. At low silane partial pressure, the deposition rate is proportional to the silane pressure.⁴ At higher silane

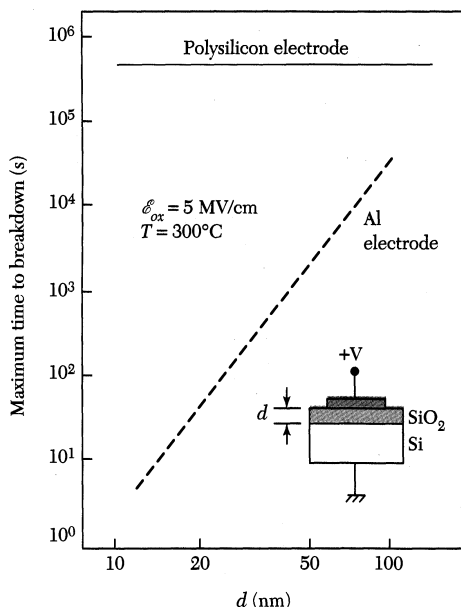


Fig. 15 Maximum time to breakdown versus oxide thickness for a polysilicon electrode and an aluminum electrode.¹⁰

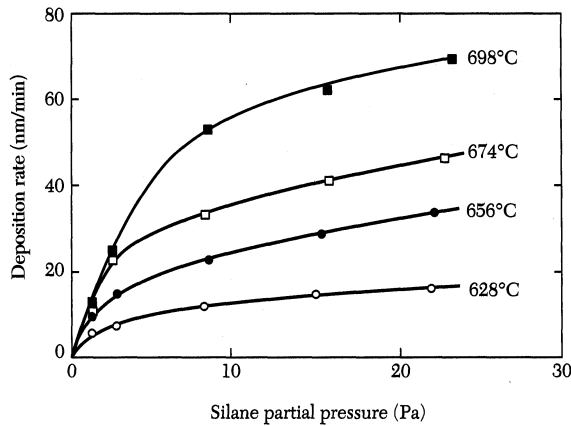


Fig. 16 Effect of silane concentration on the polysilicon deposition rate.⁴

concentrations, saturation of the deposition rate occurs. Deposition at reduced pressure is generally limited to temperatures between 600° and 650°C. In this temperature range, the deposition rate varies as $\exp(-E_a/kT)$, where the activation energy E_a is 1.7 eV, which is essentially independent of the total pressure in the reactor. At higher temperatures, gas-phase reactions that result in a rough, loosely adhering deposit become significant and silane depletion will occur, causing poor uniformity. At temperatures much lower than 600°C, the deposition rate is too slow to be practical.

Process parameters that affect the polysilicon structure are deposition temperature, dopants, and the heat cycle applied following the deposition step. A columnar structure results when polysilicon is deposited at a temperature of 600°–650°C. This structure is comprised of polycrystalline grains ranging in size from 0.03 to 0.3 μm at a preferred orientation of (110). When phosphorus is diffused at 950°C, the structure changes to crystallite and grain size increases to a size between 0.5 and 1.0 μm . When temperature is increased to 1050°C during oxidation, the grains reach a final size of 1–3 μm . Although the initially deposited film appears amorphous when deposition occurs below 600°C, growth characteristics similar to the polycrystalline-grain columnar structure are observed after doping and heating.

Polysilicon can be doped by diffusion, ion implantation, or the addition of dopant gases during deposition, referred to as in situ doping. The implantation method is most commonly used because of its lower processing temperatures. Figure 17 shows the sheet resistance of single crystal silicon and of 500 nm polysilicon doped with phosphorus and antimony using ion implantation.¹¹ The ion implantation process is considered in Chapter 13. Implant dose, annealing temperature, and annealing time all influence the sheet resistance of implanted polysilicon. Carrier traps at the grain boundaries cause a very high resistance in the lightly implanted polysilicon. As Fig. 17 illustrates, resistance drops rapidly, approaching that of implanted single crystal silicon, as the carrier traps become saturated with dopants.

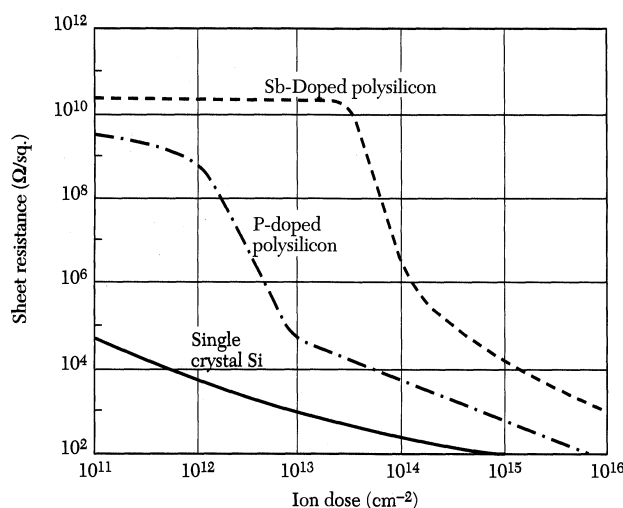


Fig. 17 Sheet resistance versus ion dose into 500 nm polysilicon at 30 keV.¹¹

► 11.4 METALLIZATION

11.4.1 Physical-Vapor Deposition

The most common methods of physical-vapor deposition (PVD) of metals are evaporation, e-beam evaporation, plasma spray deposition, and sputtering. Metals and metal compounds such as Ti, Al, Cu, TiN, and TaN can be deposited by PVD. Evaporation occurs when a source material is heated above its melting point in an evacuated chamber. The evaporated atoms then travel at high velocity in straight-line trajectories. The source can be molten by resistance heating, by rf heating, or with a focused electron beam. Evaporation and e-beam evaporation were used extensively in earlier generations of integrated circuits, but they have been replaced by sputtering for ULSI circuits.

In ion-beam sputtering, a source of ions is accelerated toward the target and impinged on its surface. Figure 18a shows the standard sputtering system. The sputtered material deposits on a wafer that is placed facing the target. The ion current and energy can be independently adjusted. Since the target and wafer are placed in a chamber that has lower pressure, more target material and less contamination are transferred to the wafer.

One method to increase the ion density and, hence, the sputter-deposition rate is to use a third electrode that provides more electrons for ionization. Another method is to use a magnetic field, such as electron cyclotron resonance (ECR), to capture and spiral electrons, increasing their ionizing efficiency in the vicinity of the sputtering target. This technique, referred to as magnetron sputtering, has found widespread applications for the deposition of aluminum and its alloys at a rate that can approach 1 $\mu\text{m}/\text{min}$.

Long-throw sputtering is another technique used to control the angular distribution. Figure 18b shows the long throw-sputtering system. In standard sputtering configurations there are two primary reasons for a wide angular distribution of incident flux at the surface: the use of a small target to substrate separation, d_{ts} , and scattering of the flux by the working gas as the flux travels from the target to the substrate. These two factors are linked because a small d_{ts} is needed to achieve good throughput, uniformity, and film properties when there is substantial gas scattering. A solution to this problem is to sputter at very low pressures, a capability that has been developed using a variety of sys-

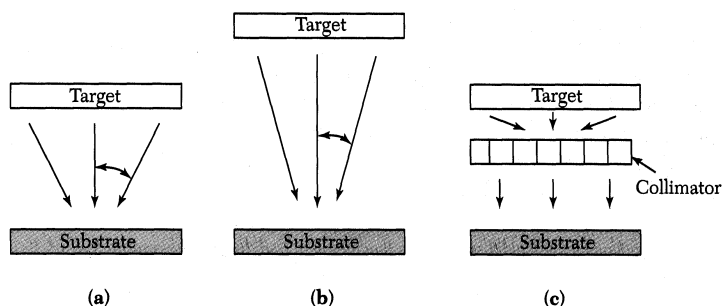


Fig. 18 (a) Standard sputtering, (b) long-throw sputtering, and (c) sputtering with a collimator.

tems, which can sustain the magnetron plasma under more rarefied conditions. These systems allow for sputtering at working pressures of less than 0.1 Pa. At these pressures, gas scattering is less important, and the target-to-substrate distance can be greatly increased. From a simple geometrical argument, this allows the angular distribution to be greatly narrowed, which permits more deposition at the bottom of high-aspect features such as contact holes.

Contact holes with large aspect ratio are difficult to fill with material, mainly because scattering events cause the top opening of the hole to seal before appreciable material has deposited on its floor. This problem can be overcome by collimating the sputtered atoms by placing an array of collimating tubes just above the wafer to restrict the depositing flux to normal $\pm 5^\circ$. Sputtering with a collimator is shown in Fig. 18c. Atoms whose trajectory is more than 5° from normal are deposited on the inner surface of the collimators.

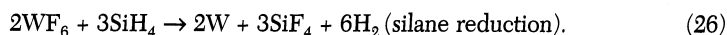
11.4.2 Chemical-Vapor Deposition

CVD is attractive for metallization because it offers coatings that are conformal, has good step coverage, and can coat a large number of wafers at a time. The basic CVD setup is the same as that used for deposition of dielectrics and polysilicon (see Fig. 9a). Low-pressure CVD (LPCVD) is capable of producing conformal step coverage over a wide range of topographical profiles, often with lower electrical resistivity than that from PVD.

One of the major new applications of CVD metal deposition for integrated circuit production is in the area of refractory-metal deposition. For example, tungsten's low electrical resistivity ($5.3 \mu\Omega\text{-cm}$) and its refractory nature make it a desirable metal for use in integrated circuit fabrication.

CVD-W

Tungsten is used both as a contact plug and as a first-level metal. Tungsten can be deposited by using WF_6 as the W source gas, since it is a liquid that boils at room temperature. WF_6 can be reduced by silicon, hydrogen, or silane. The basic chemistry for CVD-W is as follows:

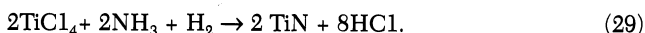
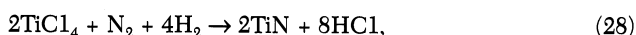


On a Si contact, the selective process starts from a silicon reduction process. This process provides a nucleation layer of W grown on Si but not on SiO₂. The hydrogen reduction process can deposit W rapidly on the nucleation layer, forming the plug. The hydrogen reduction process provides excellent conformal coverage of the topography. This process, however, does not have perfect selectivity, and the HF gas by-product of the reaction is responsible for the encroachment of the oxide, as well as for the rough surface of deposited W films.

The silane reduction process gives a high deposition rate and much smaller W grain size than that obtained with the hydrogen reduction process. In addition, the problems of encroachment and rough W surface are eliminated because there is no HF by-product generation. Usually, a silane reduction process is used as the first step in blanket W deposition to serve as a nucleation layer and to reduce junction damage. After the silane reduction, hydrogen reduction is used to grow the blanket W layer.

CVD TiN

TiN is widely used as a diffusion barrier-metal layer in metallization and can be deposited by sputtering from a compound target or by CVD. The CVD TiN can provide better step coverage than PVD methods in deep submicron technology. CVD TiN can be deposited,¹²⁻¹⁴ using TiCl₄ with NH₃, H₂/N₂, or NH₃/H₂:



The deposition temperature is about 400°–700°C for NH₃ reduction and is higher than 700°C for the N₂/H₂ reaction. The higher the deposition temperature, the better the TiN film and the less Cl incorporated in TiN (~5%).

11.4.3 Aluminum Metallization

Aluminum and its alloys are used extensively for metallization in integrated circuits. The Al film can be deposited by a PVD or CVD method. Because aluminum and its alloys have low resistivities, 2.7 μΩ-cm for Al and up to 3.5 μΩ-cm for its alloys, these metals satisfy the low-resistance requirements. Aluminum also adheres well to silicon dioxide. However, the use of aluminum in integrated circuits with shallow junctions often creates problems, such as spiking and electromigration. We consider the problems of aluminum metallization and their solutions in this section.

Junction Spiking

Figure 19 shows the phase diagram of the Al-Si system at 1 atm.¹⁵ The phase diagram relates these two components as a function of temperature. The Al-Si system exhibits eutectic characteristics; that is, the addition of either component lowers the system's melting point below that of either metal. Here, the minimum melting temperature, called eutectic temperature, is 577°C, corresponding to a 11.3% Si and 88.7% Al composition. The melting points of pure aluminum and pure silicon are 660°C and 1412°C, respectively. Because of the eutectic characteristics, during aluminum deposition the temperature on the silicon substrate must be limited to less than 577°C.

The inset of Fig. 19 also shows the solid solubility of silicon in aluminum. For example, the solubility of silicon in aluminum is 0.25 wt% at 400°C, 0.5 wt% at 450°C, and 0.8 wt% at 500°C. Therefore, wherever aluminum contacts silicon, the silicon will dis-

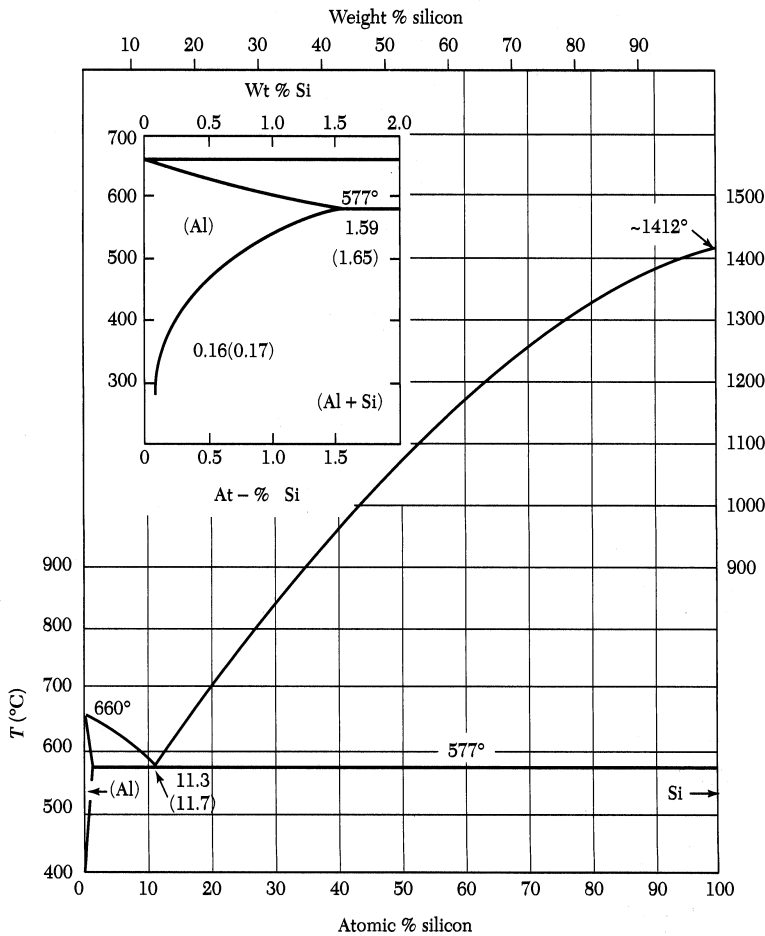


Fig. 19 Phase diagram of the aluminum-silicon system.¹⁵

solve into the aluminum during annealing. The amount of silicon dissolved will depend not only on the solubility at the annealing temperature but also on the volume of aluminum to be saturated with silicon. Consider a long aluminum metal line in contact with an area ZL of silicon as shown in Fig. 20. After an annealing time t , the silicon will diffuse a distance of approximately $\sqrt{Dt}$ along the aluminum line from the edge of the contact, where D is the diffusion coefficient given by $4 \times 10^{-2} \exp(-0.92/kT)$ for silicon diffusion in deposited aluminum films. Assuming that this length of aluminum is completely saturated with silicon, the volume of silicon consumed is then

$$\text{Vol} \cong 2\sqrt{Dt} (HZ) S \left(\frac{\rho_{\text{Al}}}{\rho_{\text{Si}}} \right), \quad (30)$$

where ρ_{Al} and ρ_{Si} are the densities of aluminum and silicon, respectively, and S is the solubility of silicon in aluminum at the annealing temperature.¹⁶ If the consumption takes

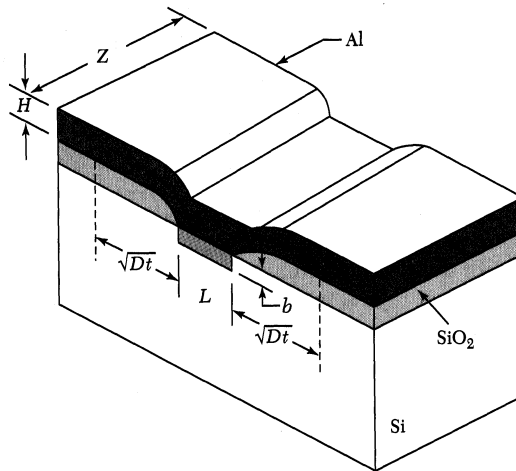


Fig. 20 Diffusion of silicon in aluminum metallization.¹⁶

place uniformly over the contact area A (where $A = ZL$ for uniform dissolution), the depth to which silicon would be consumed is

$$b \cong 2\sqrt{Dt} \left(\frac{HZ}{A} \right) S \left(\frac{\rho_{Al}}{\rho_{Si}} \right). \quad (31)$$

► EXAMPLE 4

For $T = 500^\circ\text{C}$, $t = 30$ min, $ZL = 16 \mu\text{m}^2$, $Z = 5 \mu\text{m}$, and $H = 1 \mu\text{m}$. Find the depth b , assuming uniform dissolution.

SOLUTION The diffusion coefficient of silicon in aluminum at 500°C is about $2 \times 10^{-8} \text{ cm}^2/\text{s}$; thus, $\sqrt{Dt}$ is $60 \mu\text{m}$. The density ratio is $2.7/2.33 = 1.16$.

At 500°C , S is $0.8 \text{ wt}\%$. From Eq. 31 we have

$$b = 2 \times 60 \left(\frac{1 \times 5}{16} \right) 0.8\% \times 1.16 = 0.35 \mu\text{m}.$$

Aluminum will fill a depth of $b = 0.35 \mu\text{m}$ from which silicon is consumed. If at the contact point there is a shallow junction whose depth is less than b , the diffusion of silicon into aluminum can short-circuit the junction.

In a practical situation, the dissolution of silicon does not take place uniformly but rather at only a few points. The effective area in Eq. 31 is less than the actual contact area; hence b is much larger. Figure 21 illustrates the actual situation in the p - n junction area of aluminum penetrating the silicon at only the few points where spikes are formed. One way to minimize aluminum spiking is to add silicon to the aluminum by coevaporation until the amount of silicon contained by the alloy satisfies the solubility requirement. Another method is to introduce a barrier metal layer between the aluminum and the silicon substrate (Fig. 22). This barrier metal layer must meet the following requirements: it forms low contact resistance with silicon, it will not react with aluminum, and

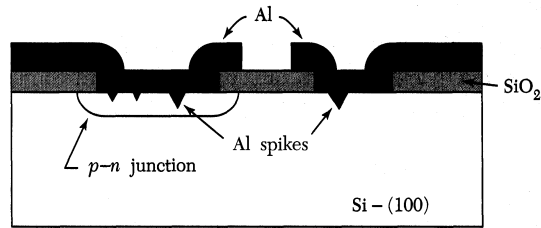


Fig. 21 Schematic view of aluminum films contacting silicon. Note the aluminum spiking in the silicon.

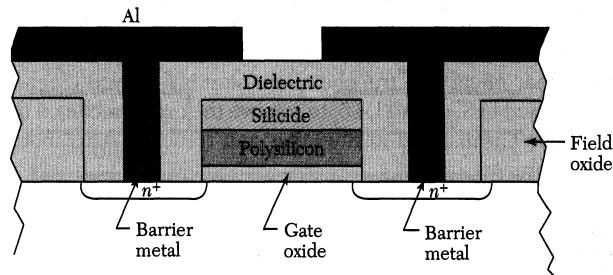


Fig. 22 Cross-sectional view of a MOSFET with a barrier metal between the aluminum and silicon and a composite gate electrode of silicide and polysilicon.

its deposition and formation are compatible with the overall process. Barrier metals such as titanium nitride (TiN) have been evaluated and found to be stable for contact annealing temperatures up to 550°C for 30 min.

Electromigration

In Chapter 6 we discussed scaled-down devices. As the device becomes smaller, the corresponding current density becomes larger. High current densities can cause device failure due to electromigration. The term electromigration refers to the transport of mass (i.e., atoms) in metals under the influence of current. It occurs by the transfer of momentum from the electrons to the positive metal ions. When a high current passes through thin metal conductors in integrated circuits, metal ions in some regions will pile up and voids will form in other regions. The pileup can short-circuit adjacent conductors, whereas the voids can result in an open circuit.

The mean time to failure (MTF) of a conductor due to electromigration can be related to the current density J and the activation energy E_a by

$$\text{MTF} \sim \frac{1}{J^2} \exp\left(\frac{E_a}{kT}\right). \quad (32)$$

Experimentally, a value of $E_a \approx 0.5$ eV is obtained for deposited aluminum. This indicates that low-temperature, grain-boundary diffusion is the primary vehicle of material transport, since $E_a \approx 1.4$ eV would characterize the self-diffusion of single-crystal aluminum. The electromigration resistance of aluminum conductors can be increased by using several techniques. These techniques include alloying with copper (e.g., Al with 0.5% Cu), encapsulating the conductor in a dielectric, or incorporating oxygen during film deposition.

11.4.3 Copper Metallization

It is well known that both high conductivity wiring and low-dielectric-constant insulators are required to lower the RC time delay of the interconnect network. Copper is the obvious choice for a new interconnection metallization because it has higher conductivity and higher electromigration resistance than aluminum. Copper can be deposited by PVD, CVD, and electrochemical methods. However, the use of Cu as an alternative material to Al in ULSI circuits has drawbacks, such as its tendency to corrode under standard chip manufacture conditions, its lack of a feasible dry-etching method or a stable self-passivating oxide similar to Al_2O_3 on Al, and its poor adhesion to dielectric materials, such as SiO_2 and low- k polymers. In this section, we discuss the copper metallization techniques.

Several different techniques for fabrication of multilevel Cu interconnects have been reported.^{17,18} The first method is a conventional method to pattern the metal lines followed by dielectric deposition. The second method is to pattern the dielectric layer first and fill copper metal into trenches. This step is followed by chemical mechanical polishing, discussed in Section 11.4.5, to remove the excess metal on the top surface of dielectric and leave Cu material in the holes and trenches. This method is also known as a damascene process.

Damascene technology

The approach for fabricating a copper/low- k dielectric interconnect structure is by the “damascene” or “dual damascene.” Figure 23 shows the dual damascene sequence for an advanced Cu interconnection structure. For a typical damascene structure, trenches for metal lines are defined and etched in the interlayer dielectric (ILD) and then followed by metal deposition of TaN/Cu. The TaN layer serves as a diffusion barrier layer and prevents copper from penetrating the low- k dielectric. The excess copper metal on the surface is removed to obtain a planar structure with metal inlays in the dielectric.

For the dual damascene process, the vias and trenches in the dielectric are defined using two lithography and reactive ion etching (RIE) steps before depositing the Cu metal (Fig. 23a–c). Then a Cu chemical-mechanical polishing process is used to remove the metal on the top surface, leaving the planarized wiring and via imbedded in the insulator.¹⁹ One special benefit of dual damascene is that the via plug is now of the same material as the metal line and the risk of via electromigration failure is reduced.

► EXAMPLE 5

If we replace Al with Cu wire associated with some low- k dielectric ($k = 2.6$) instead of SiO_2 layer, what percentage of reduction of RC time constant will be achieved? (The resistivity of Al is $2.7 \mu\Omega\text{-cm}$, and the resistivity of Cu is $1.7 \mu\Omega\text{-cm}$.)

SOLUTION

$$\frac{1.7}{2.7} \times \frac{2.6}{3.9} \times 100\% = 42\%.$$

11.4.5 Chemical-Mechanical Polishing

In recent years, the development of chemical-mechanical polishing (CMP) has become increasingly important for multilevel interconnection because it is the only technology that allows global planarization (i.e., to make a flat surface across the whole wafer). It

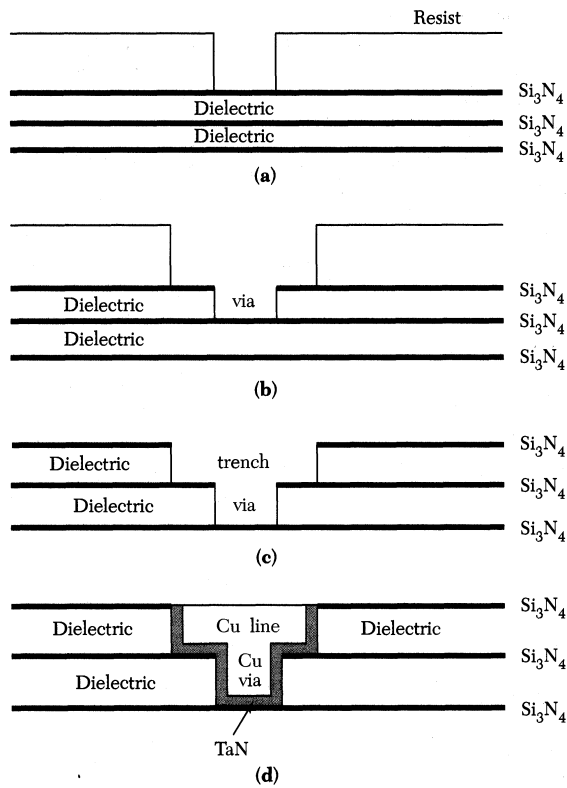


Fig. 23 Process sequence used to fabricate a Cu line-stud structure using dual damascene. (a) Resist stencil applied; (b) reactive ion etching dielectric and resist patterning; (c) trench and via definition; and (d) Cu depositions followed by chemical-mechanical polishing (CMP).

that allows global planarization (i.e., to make a flat surface across the whole wafer). It offers many advantages over other types of technologies—better global planarization over large or small structures, reduced defect density, and the avoidance of plasma damage. Three CMP approaches are summarized in Table 4.

The CMP process consists of moving the sample surface against a pad that carries slurry between the sample surface and the pad. Abrasive particles in the slurry cause mechanical damage on the sample surface, loosening the material for enhanced chemical attack or fracturing off the pieces of surface into a slurry where they dissolve or are swept away. The process is tailored to provide an enhanced material removal rate from

TABLE 4 Three Methods of Chemical-Mechanical Polishing (CMP)

Method	Wafer facing	Platen movement	Slurry feeding
Rotary CMP	Down	Rotary against rotating wafer carrier	Dripping to pad surface
Orbital CMP	Down	Orbital against rotating wafer carrier	Through the pad surface
Linear CMP	Down	Linear against rotating wafer carrier	Dripping to pad surface

high points on surfaces, thus affecting the planarization because most chemical actions are isotropic. Mechanical grinding alone, theoretically, may achieve the desired planarization but is not desirable because of extensive associated damage to the material surfaces. There are three main parts of the process: the surface to be polished, the pad—the key media enabling the transfer of mechanical action to the surface being polished, and the slurry—which provides both chemical and mechanical effects. Figure 24 shows the CMP setup.²⁰

► EXAMPLE 6

The oxide removal rate and the removal rate of a layer underneath the oxide (called a stop layer) are $1r$ and $0.1r$, respectively. To remove $1\text{ }\mu\text{m}$ oxide and $0.01\text{ }\mu\text{m}$ stop layer, the total removal time is 5.5 min. Find the oxide removal rate.

SOLUTION

$$\frac{1}{1r} + \frac{0.01}{0.1r} = 5.5$$

$$\frac{1.1}{1r} = 5.5, \quad r = 0.2\text{ }\mu\text{m} / \text{min.}$$

11.4.6 Silicide

Silicon forms many stable metallic and semiconducting compounds with metals. Several metal silicides show low resistivity and high thermal stability, making them suitable for ULSI application. Silicides such as TiSi_2 and CoSi_2 have reasonably low resistivities and are generally compatible with integrated-circuit processing. Silicides become important metallization materials as devices become smaller. One important application of silicide is for the MOSFET gate electrode either alone or with doped polysilicon (polycide) above the gate oxide. Table 5 shows a comparison of titanium silicide and cobalt silicide.

Metal silicides have been used to reduce the contact resistance of the source and drain, the gate electrodes, and the interconnections. The self-aligned metal silicide technology (salicide) has been proven to be a highly attractive technique for improving the performance of submicron devices and circuits. The self-aligned process uses the silicide gate electrode as the mask to form the source and drain electrodes of a MOSFET (e.g., by ion implantation, considered in Chapter 13). This process can minimize the overlaps of these electrodes and thus reduce the parasitic capacitances.

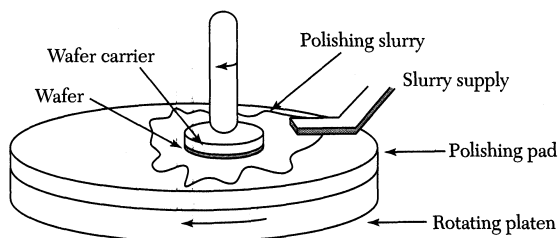


Fig. 24 A schematic of a CMP polisher.

TABLE 5 A Comparison of TiSi_2 and CoSi_2 Films

Properties	TiSi_2	CoSi_2
Resistivity	13–16	22–28
Silicide/metal ratio	2.37	3.56
Silicide/Si ratio	1.04	0.97
Reactive to native oxide	Yes	No
Silicidation temperature ($^{\circ}\text{C}$)	800–850	550–900
Film stress (dyne/cm^2)	1.5×10^{10}	1.2×10^{10}

Figure 25 shows the polycide and silicide process. A typical polycide formation sequence is shown in Fig. 25a. For sputter deposition, a high-temperature, high-purity compound target is used to ensure the quality of the silicide. The most commonly used silicides for the polycide process are WSi_2 , TaSi_2 , and MoSi_2 . They are all refractory, thermally stable, and resistant to processing chemicals. A self-aligned silicide process is illustrated in Fig. 25b. In the process, the polysilicon gate is patterned without any silicide, and a sidewall spacer (silicon oxide or silicon nitride) is formed to prevent shorting the gate to the source and drain during the silicidation process. A metal layer, either Ti or Co, is blanket-sputtered on the entire structure, followed by silicide sintering. Silicide is formed, in principle, only where the metal is in contact with Si. A wet chemical wash then rinses off the unreacted metal, leaving only the silicide. This technique eliminates

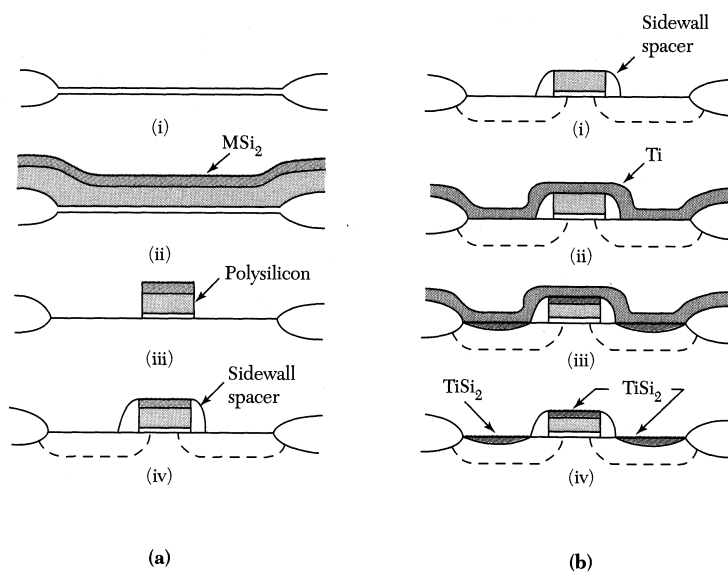


Fig. 25 Polycide and silicide processes. (a) Polycide structure: (i) gate oxide; (ii) polysilicon and silicide deposition, (iii) pattern polycide; and (iv) lightly doped drain (LDD) implant, sidewall formation, and S/D implant. (b) Salicide structure: (i) gate patterning (polysilicon only), LDD, sidewall, and S/D implant; (ii) metal (Ti,Co) deposition; (iii) anneal to form silicide; and (iv) selective (wet) etch to remove unreacted metal.

the need to pattern the composite polycide gate structure and adds silicide to the source/drain area to reduce the contact resistance.

The silicides are promising materials for ULSI circuits because of their low resistivity and excellent thermal stability. Cobalt silicide has been widely investigated recently because of its lowest resistivity and high-temperature thermal stability. However, cobalt is sensitive to native oxide as well as an oxygen-contained environment, and a large amount of silicon is consumed during silicidation.

► EXAMPLE 7

Calculate the thickness of the cobalt silicide. For a desired sheet resistance of $0.6 \Omega/\text{sq}$, the resistivity is $18 \mu\Omega\text{-cm}$.

SOLUTION The resistivity is equal to the product of the sheet resistance and the film thickness:

$$\rho = R_s \times t.$$

Then

$$t = \frac{\rho}{R_s} = \frac{18 \times 10^{-6}}{0.6} = 3 \times 10^{-5} \text{ cm} = 300 \text{ nm}.$$

► SUMMARY

Modern semiconductor device fabrication requires the use of thin films. Currently, there are four important types of films—thermal oxides, dielectric layers, polycrystalline silicon, and metal films. The major issues related to film formation are low-temperature processing, step coverage, selective deposition, uniformity, film quality, planarization, throughput, and large-wafer capacity.

Thermal oxidation offers the best quality for the Si-SiO₂ interface and has the lowest interface trap density. Therefore, it is used to form the gate oxide and the field oxide. LPCVD of dielectrics and polysilicon offer conformal step coverage. In contrast, PVD and atmospheric-pressure CVD generally result in nonconformal step coverage. CMP offers global planarization and reduces defect density. Conformal step coverage and planarization are also required for precise pattern transfer at the deep-submicron lithography level. Pattern transfer technology is discussed in the next chapter.

To minimize the *RC* time delay due to parasitic resistance and capacitance, the silicide process for ohmic contacts, copper metallization for interconnects, and low-dielectric-constant materials for interlayer films are extensively used to meet the requirements of the multilevel interconnect structures of ULSI circuits. In addition, we have investigated high-dielectric-constant materials to improve the gate insulator performance and to increase the capacitance per unit area for DRAM.

► REFERENCES

1. E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*, Wiley, New York, 1982.
2. B. E. Deal and A. S. Grove, "General Relationship for the Thermal Oxidation of Silicon," *J. Appl. Phys.*, **36**, 3770 (1965).
3. J. D. Meindl, et al., "Silicon Epitaxy and Oxidation," in F. Van de wile, W. L. Engl, and P. O. Jespers, Eds., *Process and Device Modeling for Integrated Circuit Design*, Noorhoff, Leyden, 1977.

4. For a discussion on film deposition, see, for example, A.C. Adams, "Dielectric and Polysilicon Film Deposition," in S. M. Sze, Ed., *VLSI Technology*, McGraw-Hill, New York, 1983.
5. K. Ejuno, et al., "Doped Silicon Oxide Deposition by Atmospheric Pressure and Low Temperature Chemical Vapor Deposition Using Tetraethoxysilane and Ozone," *J. Electrochem. Soc.*, **138**, 3019 (1991).
6. A. C. Adams and C. D. Capio, "Planarization of Phosphorus-Doped Silicon Dioxide," *J. Electrochem. Soc.*, **127**, 2222 (1980).
7. T. Yamamoto et al., "An Advanced 2.5nm Oxidized Nitride Gate Dielectric for Highly Reliable 0.25 μm MOSFETs," *Symp. on VLSI Technol. Dig. of Tech. Pap.*, 1997, p. 45.
8. K. Kumar, et al., "Optimization of Some 3 nm Gate Dielectrics Grown by Rapid Thermal Oxidation in a Nitric Oxide Ambient," *Appl. Phys. Lett.*, **70**, 384 (1997).
9. T. Homma, "Low Dielectric Constant Materials and Methods for Interlayer Dielectric Films in Ultralarge-Scale Integrated Circuit Multilevel Interconnects," *Mater. Sci. Eng.*, **23**, 243 (1998).
10. H. N. Yu, et al., "1 μm MOSFET VLSI Technology. Part I—An Overview," *IEEE Trans. Electron Devices*, **ED-26**, 318 (1979).
11. J. M. Andrews, "Electrical Conduction in Implanted Polycrystalline Silicon," *J. Electron. Mater.*, **8**, 3, 227 (1979).
12. M. J. Buiting, A. F. Otterloo, and A. H. Montree, "Kinetic Aspects of the LPCVD of Titanium Nitride from Titanium Tetrachloride and Ammonia," *J. Electrochem. Soc.*, **138**, 500 (1991).
13. R. Tobe, et al., "Plasma-Enhanced CVD of TiN and Ti Using Low-Pressure and High-Density Helicon Plasma," *Thin Solid Film*, **281-282**, 155 (1996).
14. J. Hu, et al., "Electrical Properties of Ti/TiN Films Prepared by Chemical Vapor Deposition and Their Applications in Submicron Structures as Contact and Barrier Materials," *Thin Solid Film*, **308**, 589 (1997).
15. M. Hansen and A. Anderko, *Constitution of Binary Alloys*, McGraw-Hill, New York, 1958.
16. D. Pramanik and A. N. Saxena, "VLSI Metallization Using Aluminum and Its Alloys," *Solid State Tech.*, **26**, No. 1, 127 (1983), **26**, No. 3, 131 (1983).
17. C. L. Hu, and J. M. E. Harper, "Copper Interconnections and Reliability," *Matter. Chem. Phys.*, **52**, 5 (1998).
18. P. C. Andricacos, et al., "Damascene Copper Electroplating for Chip Interconnects," *193rd Meet. Electrochem. Soc.*, 1998, p. 3.
19. J. M. Steigerwald, et al., "Chemical Mechanical Planarization of Microelectronic Materials," Wiley, New York, 1997.
20. L. M. Cook, et al., *Theoretical and Practical Aspects of Dielectric and Metal CMP*, Semicond. Int., p. 141 (1995).

► PROBLEMS (* DENOTES DIFFICULT PROBLEMS)

FOR SECTION 11.1 THERMAL OXIDATION

1. A p -type $\langle 100 \rangle$ -oriented, silicon wafer with a resistivity of $10 \Omega\text{-cm}$ is placed in a wet oxidation system to grow a field oxide of $0.45 \mu\text{m}$ at 1050°C . Determine the time required to grow the oxide.
- *2. After the first oxidation as given in Prob. 1, a window is opened in the oxide to grow a gate oxide at 1000°C for 20 minutes in dry oxidation. Find the thicknesses of the gate oxide and the total field oxide.
3. Show that Eq. 11 reduces to $x^2 = Bt$ for long times and to $x = B/A(t + \tau)$ for short times.
4. Determine the diffusion coefficient D for dry oxidation of $\langle 100 \rangle$ -oriented silicon samples at 980°C and 1 atm.

FOR SECTION 11.2 DIELECTRIC DEPOSITION

5. (a) In a plasma-deposited silicon nitride that contains 20 at% hydrogen and has a silicon-to-nitrogen ratio (Si/N) of 1.2, find x and y in the empirical formula of SiN_xH_y . (b) If the variation of film resistivity with Si/N ratio is given by $5 \times 10^{25} \exp(-33.3\gamma)$ for $2 > \gamma > 0.8$, where γ is the ratio, find the resistivity of the film in (a).
6. The dielectric constants of SiO_2 , Si_3N_4 , and Ta_2O_5 are about 3.9, 7.6, and 25, respectively. What is the capacitance ratio for the capacitors with the Ta_2O_5 and oxide/nitride/oxide dielectrics for the same dielectric thickness, provided the oxide/nitride/oxide has thickness ratio 1:1:1 for the oxide to the nitride?
7. In Prob. 6, if BST with a dielectric constant of 500 is chosen to replace Ta_2O_5 , calculate the area reduction ratio to maintain the same capacitance if the two films have the same thickness.
8. In Prob. 6, calculate the equivalent thickness of the Ta_2O_5 in terms of SiO_2 thickness if both have the same capacitance. Assume the actual thickness of Ta_2O_5 is $3t$.
9. In a silane-oxygen reaction to deposit undoped SiO_2 film, the deposition rate is 15 nm/min at 425°C . What temperature is required to double the deposition rate?
10. The P-glass flow process requires temperatures above 1000°C . As device dimensions become smaller in ULSI, we must use lower temperatures. Suggest methods to obtain a smooth topography at $< 900^\circ\text{C}$ for deposited silicon dioxide that can be used as an insulator between metal layers.

FOR SECTION 11.3 POLYSILICON DEPOSITION

11. Why is silane more often used for polysilicon deposition than silicon chloride?
12. Explain why the deposition temperature for polysilicon films is moderately low, usually between 600°C – 650°C .

FOR SECTION 11.4 METALLIZATION

13. An e-beam evaporation system is used to deposit aluminum to form MOS capacitors. If the flatband voltage of the capacitance is shifted by 0.5 V because of e-beam radiation, find the number of fixed oxide charges (the silicon dioxide thickness is 50 nm). How can these charges be removed?
14. A metal line ($L = 20 \mu\text{m}$, $W = 0.25 \mu\text{m}$) has a sheet resistance $5 \Omega/\text{sq}$. Calculate the resistance of the metal line.
15. Calculate the thickness of the TiSi_2 and CoSi_2 , where the initial Ti and Co film thickness are 30 nm.
16. Compare the advantages and disadvantages of TiSi_2 and CoSi_2 for salicide applications.
17. A dielectric material is placed between the two parallel metal lines. The length $L = 1 \text{ cm}$, width $W = 0.28 \mu\text{m}$, thickness $T = 0.3 \mu\text{m}$, and spacing $S = 0.36 \mu\text{m}$. (a) Calculate the RC time delay. The metal is Al with a resistivity of $2.67 \mu\Omega\text{-cm}$ and the dielectric is oxide with dielectric constant 3.9. (b) Calculate the RC time delay. The metal is Cu with a resistivity of $1.7 \mu\Omega\text{-cm}$ and the dielectric is organic polymer with dielectric constant 2.8. (c) Compare the results in (a) and (b). How much can we decrease the RC time delay?
18. Repeat Prob. 17 (a) and (b) if the fringing factor for the capacitors is 3. The fringing factor is due to the spreading of the electric-field lines beyond the length and the width of the metal lines.
- *19. To avoid electromigration problems, the maximum allowed current density in an aluminum runner is about $5 \times 10^5 \text{ A/cm}^2$. If the runner is 2 mm long, $1 \mu\text{m}$ wide, and nominally $1 \mu\text{m}$ thick, and if 20% of the runner length passes over steps and is only $0.5 \mu\text{m}$

thick there, find the total resistance of the runner if the resistivity is $3 \times 10^{-6} \Omega\text{-cm}$. Find the maximum voltage that can be applied across the runner.

- *20. To use Cu for wiring one must overcome several obstacles: the diffusion of Cu through SiO_2 , adhesion of Cu to SiO_2 , and corrosion of Cu. One way to overcome these obstacles is to use a cladding/adhesion layer (e.g. Ta or TiN) to protect the Cu wires. Consider a clad Cu wire with a square cross section of $0.5 \mu\text{m} \times 0.5 \mu\text{m}$ and compare it with a layered TiN/Al/TiN wire of the same size, with the top and bottom TiN layers 40 nm and 60 nm thick, respectively. What is the maximum thickness of the cladding layer if the resistance of the clad Cu wire and the TiN/Al/TiN wire is the same?